Experimental Test of the Influence of Evolutionary History on the Effect of Beneficial and Regulatory Mutations in *Escherichia coli*

by
Kelly Nicole Phillips

A dissertation submitted to the Department of Biology & Biochemistry,
College of Natural Sciences and Mathematics
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in Biology

Chair of Committee: Ricardo Azevedo

Committee Member: Tim F. Cooper

Committee Member: Rebecca Zufall

Committee Member: George Fox

Committee Member: Yousif Shamoo

University of Houston
December 2019

Acknowledgements

I would like to thank my committee members Dr. Tim F. Cooper, Dr. Ricardo Azevedo, Dr. Rebecca Zufall, Dr. George Fox, and Dr. Yousif Shamoo for their knowledge and advice, and for taking time out of their busy schedules so that I may pursue my PhD. A special thank you to Ricardo Azevedo for stepping in as my committee chair and helping me throughout the years. Most importantly I would like to thank my advisor, Dr. Tim F. Cooper, for his remarkable guidance and support. I am truly grateful for the opportunity to have worked with and learned from such an amazing researcher and mentor with patience and solicitude. I would also like to express my gratitude and appreciation to my undergraduate advisor at the University of Houston-Downtown, Dr. Lisa Morano, for giving me a chance by letting me work in her lab, which was pivotal in opening the door for where I am today.

I would like to thank my undergraduate helpers Jennifer Nguyen, Gerardo Castillo, Carolina Balmaceda, and Matthew Shirk for their hard work on my various projects. Thank you to Gábor Balázsi and Christian Johnstone Ray for their collaboration, insightful ideas, and advice with the *lac* operon regulation project. Thank you to all of the Cooper lab members for their informative discussions, help, and laughs over the years: Duy Dinh, Carolina Diaz Arenas, Yinhua Wang, Fen Peng, Trang Duong, Rebecca Satterwhite, Dustin Phipps, Zakaris Kwota,

Andrea Wünsche, Kristina Duan, Kedar Karkare, and Rachel Staples. Thank you to my fiancé, Scott Widmann, for helping me in the lab and for our beautiful daughter, Madeleine, who brings so much joy to our lives. Thank you to my family for their encouragement, support, and understanding of the time constraints of pursuing a PhD.

Abstract

Selection in fluctuating environments can lead to novel adaptations that may limit or facilitate evolution. I used *Escherichia coli* populations experimentally evolved in limited-glucose and -lactose environments to examine the effect of historical contingency and test for compensation of the cost normally produced by an evolved non-functional *lac* operon repressor (*lacI*-) in glucose. I surveyed 23 diverse strains for natural variation in regulatory function of the *E. coli lac* operon, which integrates transcriptional control of the *lac* operon with environmental signals from glucose and lactose availability. Although *lac* operon regulation has been extensively studied, a considerable amount of research is based on a small number of closely related strains. I found that populations initially evolved in a limited-glucose environment, and clones randomly selected from these populations, were more evolvable than the common ancestor during 1k generations of evolution in a lactose-limited environment, indicating specific genetic changes throughout glucose selection led to increased evolvability. Genome sequencing of the pre-evolved glucose clones revealed mutations in the gene *iclR* as candidates for increasing evolvability. Next, I tested for compensation of the glucose cost of *lacI*- in experimentally evolved populations after 8k generations of selection in minimal glucose and lactose fluctuating environments. Competitions measuring the *lacI*- fitness effect indicated that compensation rarely alleviated the cost in glucose, but epistasis commonly increased the benefit in lactose compared to the ancestor. A reporter

introduced into these evolved clones indicated *lac* operon expression changed but had different fitness effects. Finally, using the same reporter but placed into 28 distinct strains, I measured *E. coli lac* operon expression in inducers combined at different concentrations. The results encompassed regulatory functions of a vast range, and aspects that were associated with genetic relatedness were the most effective at predicting initial lactose growth. Hybrid reference strains containing l*acI* and the *lac* operon of five different natural isolates indicated regulatory elements that had more control over expression could be either global or local. My results demonstrate selection can lead to diverse adaptations that can depend on selective history, mutation interactions, or idiosyncrasies.

Table of contents

List of Tables

List of Figures

Chapter 1

General Introduction

1.1 Adaptation

Adaptation, the process through which a population becomes better fitted to its environment, is a cornerstone of evolution. When there are heritable traits that produce differential fitness between organisms, those organisms that are better adapted to their environment thrive and produce more offspring; a process known as natural selection. Fundamental questions regarding adaptation and natural selection include determining the underlying genetic changes that drive adaptation (Matange et al., 2019), mechanisms affecting the production of adaptive mutations (Velicer and Yu, 2003; Chubiz and Marx, 2017), how adaptive mutations interact with each other and their genetic background (Wang et al., 2012; Peng et al., 2017), how selective history affects future adaptation (Fragata et al., 2014; Travisano et al., 1995), and how changes in the environment affect adaptation (Bennett and Lenski, 2007; Melnyk et al., 2017).

Predicting what adaptations will arise in response to a given selection is difficult because it can depend on the unique selective history of a population (Gould, 1989; Buckling et al., 2003; McBride et al., 2008; Bedhomme et al., 2013; Phillips et al., 2016). Because past selective pressures select for different adaptations with different underlying genetic changes, independently evolving populations will have distinct combination of mutations. When new potential adaptive mutations depend on the broader genetic background for their effect, these distinct combinations will cause them to differ in the likelihood that they are selected.

Variation in the effect of available adaptive mutations is likely to depend on the specific genetic background, but how this background depends on past selection may not be predictable. The stochastic nature of mutations and random genetic drift cause differences in the genetic composition of populations that depend on chance, as well as on their selective environment. Gould believed that these chance differences would have a dominant role in determining subsequent evolutionary outcomes so that inevitable changes to the sequence of mutational events occurring during evolution, even of populations exposed to the same selective pressures, will result in different outcomes (Gould, 1989). In contrast, Simon Conway Morris suggested that evolution was largely repeatable (2003). He argued that an organism's history can be erased by natural selection, leading to the same adaptations (Morris, 2003). Morris believed that there are many

evolutionary paths but only few destinations which will constrain evolutionary outcomes (2003). One of the purposes of my study is to examine whether evolutionary outcomes are contingent on prior selective environments. In testing contingency, working with experimentally evolved populations confers the significant advantage that past selective environments are known and can be manipulated.

1.2 Benefits of Experimental Evolution

Experimental evolution allows scientists to study evolutionary processes in controlled environmental conditions which aids in isolating the cause of evolutionary outcomes (Lázaro et al., 2002; Spencer et al., 2008; Marks et al., 2010; Le Gac et al., 2012; Toprak et al., 2012; Hayden et al., 2013; Herron and Doebeli, 2013; Kvitek and Sherlock, 2013; Lang et al., 2013; Burke et al., 2014; Hong and Gresham, 2014; Plucain et al., 2014; Laan et al., 2015; Lind et al., 2015; Selmecki et al., 2015; Sunshine et al., 2015; Consuegra et al., 2017; Kram et al., 2017; Kutnjak et al., 2017; Sandberg et al., 2017). The selective history of experimentally evolved populations is known and, for many systems, at any point during evolution populations can be frozen in time. In natural populations, it is impossible to determine every selective pressure that occurred in its past or led to specific adaptations. Another advantage of working with experimentally evolved populations is

that evolved strains can be compared directly to their known common ancestor, which greatly simplifies identification of underlying genetic changes that led to specific adaptations. Comparisons between evolved populations can indicate adaptations that are specific to an environment and the effect of selective history (Le Gac et al., 2013). Experimental evolution studies can indicate what adaptations will likely evolve in a natural environment (Traverse et al. 2013).

1.3 Regulatory Functions

In this thesis, I examine the influence of selective history on the identity and effect of mutations that confer benefits mediated through changes in the expression of genes involved in nutrient acquisition. To understand these changes and the potential for them to depend on selective history, it is necessary to understand the basis of bacterial gene regulation and its potential to mediate adaptation. Expression of bacterial genes is regulated by *cis*- and *trans*- acting transcription elements whose influence and interaction is based on environmental signals. An important and well-studied adaptive mechanism by which gene regulation can evolve is through mutational changes that alter the function of transcription elements (Chan et al., 2010; Jones et al., 2012; Chang et al., 2013; Fraser, 2013; Halligan et al., 2013; Cleves et al., 2014; Guo et al., 2018; Payne

et al., 2018; Tangwancharoen et al., 2018; Lewis and Reed, 2019; Bleuven and Landry, 2016). When evolved in a glucose-rich environment, the bacterium *Shewanella oneidensis*, which does not normally catabolize glucose, gained the ability to metabolize glucose through a loss of function mutation in the transcriptional repressor *nagR* (Chubiz and Marx, 2017). However, the non-functional NagR that was necessary for the newly acquired metabolic ability diminished growth on the preferred resource lactate, creating a tradeoff between glucose and lactate metabolism (Chubiz and Marx, 2017). *Escherichia coli* grown in lactose-limited environments in chemostats has been shown to quickly result in constitutive expression of the *lac* operon, the genes necessary for lactose metabolism, caused by mutations in its repressor *lacI* (Dykhuizen and Davies, 1980; Novick and Horiuchi, 1961). Constitutive expression was critical in maintaining low frequencies of bacteria that utilize lactose in a limited mixed resource environment where 99% of the carbon source was maltose (Dykhuizen and Davies, 1980). Mutants constitutively expressing the *lac* operon did not fix when maltose and lactose were present at 90% and 10%, respectively, of the carbon source, and were undetectable in maltose single resource environments. These results indicate a trade-off between *lac* operon constitutive expression and maltose utilization (Dykhuizen and Davies, 1980). In both examples, the availability of a carbon source led to a selected change in regulatory functions that mediate trade-offs in fitness across environments that are likely to affect future adaptation depending on past selection for the trade-off.

1.4 Epistasis

When new mutations arise during evolution they may interact with other mutations in a way that is non-additive. Mutation interactions that produce phenotypic effects that are greater or less than their cumulative individual additive effects are called epistatic interactions. Epistatic interactions are context-dependent; meaning the mechanism of epistasis can not only depend on the interactions or function of the mutations and protein(s) involved, but also on the broader genetic background and environmental conditions (Flynn et al., 2013). There are different types of epistatic interactions. The phenotypic effect produced by epistatic interactions can be caused by mutations that interact positively – such that a combination of mutations causes a fitness higher than expected – or negatively – such that a combination of mutations causes a fitness lower than expected. Epistatic interactions between genes can cause widely different outcomes in the phenotypic effects of mutations (Vogwill et al., 2016; Press and Queitsch, 2017; Lagator et al., 2017; Domingo et al., 2018; Sackman and Rokyta, 2018; Pokusaeva et al., 2019).

Epistatic interactions between mutations can produce phenotypic effects that depend on the environment (de Vos et al., 2013; Flynn et al., 2013), where the sign and magnitude of an epistatic interaction is dependent on the environment, or cause fitness differences among individuals with

different genetic backgrounds (Wang *et al*., 2012). Epistasis can impose constraints on the fitness effect of an evolved allele that increases gene expression (Chou *et al*., 2009), and lead to a decelerating rate of adaptation based on mutations in the genetic background (Chou *et al*., 2011; Khan *et al*., 2011, Wünsche et al., 2017). Epistasis has important implications in adaptation (Wiser et al., 2013; Plucain et al., 2014; Kryazhimskiy et al., 2014; Peng et al., 2017; Dasmeh et al., 2017; Burch and Chao, 1999; Remold and Lenski, 2004; Sanjuán et al., 2005; Kryazhimskiy et al., 2011), developmental biology (Vijendravarma and Kawecki, 2013), pathogenesis of diseases (Mansoori et al., 2012), drug resistance in pathogens (Moura de Sousa et al., 2017; Schenk et al., 2013; Silva et al., 2011; Salverda et al., 2011; MacLean et al., 2010), and speciation (Schumer et al., 2015; Fishman and Willis, 2001; Brideau et al., 2006; Meiklejohn et al., 2013).

1.5 Consequences of Epistasis

Epistatic interactions affect phenotypes directly, but can also influence evolutionary processes to affect subsequent evolutionary outcomes (Chiotti et al., 2014; Draghi and Plotkin, 2013; Anderson et al., 2015; Good et al., 2017; Sailer and Harms, 2017; Wünsche et al., 2017; Matange et al., 2019; Zhao et al., 2019). Some related consequences of epistasis include evolution being heavily dependent on evolutionary history (Kvitek and Sherlock, 2011; Anderson et al., 2015; Gupta and

Adami, 2016; Steinberg and Ostermeier, 2016; Starr et al., 2017), differences in evolvability (Woods et al., 2011; Gifford et al., 2016; Jerison et al., 2017; Kutch and Fedorka, 2018), compensation of deleterious mutations (Szamecz et al., 2014; Filteau et al., 2015; Qi et al., 2016; Rojas Echenique et al., 2019), and differential fitness of horizontally transferred genes (Sota et al., 2010; Silva et al., 2011; San Millan et al., 2014; Vogwill and MacLean, 2015; Maddamsetti and Lenski, 2018).

1.5.1 Historical Contingency

Epistasis influences the effect of newly arising mutations such that the set of potentially beneficial mutations available to a population can depend on those that have already occurred, known as historical contingency (Lindsey et al., 2013). An example of historical contingency, which underlies Gould's belief that chance historical events will dominate future evolutionary outcomes, is the evolution of citrate utilization in experimentally evolved populations of *E. coli* (Gould, 1989; Blount et al., 2008). The capability to utilize citrate was contingent on specific genetic changes, and their epistatic interactions, that occurred during a population's history, and without these mutations or citrate in the environment, citrate utilization would not have been realized or even beneficial (Blount et al., 2008; Blount et al., 2012; Quandt et al., 2014; Quandt et al., 2015; Good

et al., 2017). Besides genetic changes, a lower level of competition caused by smaller fitness gains at later time points during evolution was also important in achieving citrate metabolism (Leon et al., 2018). Each step in the sequence of events that led to citrate metabolism were hypothesized to be necessary in preventing the ability to utilize citrate from being purged at earlier timepoints (Leon et al., 2018; Quandt et al., 2015).

Historical contingency has also affected evolution of the ancestral glucocorticoid receptor in bony vertebrates, changing the receptor from being promiscuous to cortisol-specific (Ortlund et al., 2007). Cortisol specificity was contingent on initial rare "permissive" mutations that allowed the receptor to remain functional in the presence of function-altering epistatic mutations that would otherwise destroy the protein's function (Harms and Thornton, 2014; Ortlund et al., 2007). Snakes resistant to high levels of tetrodotoxin (TTX) in prey evolved resistance over millions of years through sequential mutations in three different voltage-gated sodium channels (McGlothlin et al., 2016). The most recent sodium channel to evolve, which provides the highest level of resistance to TTX, was contingent on ancient evolved changes in sodium channel paralogs (McGlothlin et al., 2016).

Of course, just because chance events can influence future outcomes, it doesn't mean that, in some circumstances, selection will not have a dominant role. A well-studied example where a similar current selection has resulted in similar adaptation involves changes in pigmentation of Nebraska deer mice. Pigmentation changes have repeatedly been selected to match soil color even though many single nucleotide polymorphisms (SNPs) within the *Aguoti* locus for pigmentation and several other candidate regions were found to affect coloration (Pfeifer et al., 2018). In Nebraska deer mice the repeatable adaptive benefit of pigmentation, which is constrained to soil color of the habitat, is causing convergence in coloration of distinct locals based on the high fitness benefit it provides, which coincides with Morris' views (Pfeifer et al., 2018). Overall genetic diversity was low indicating recent divergence of populations but high at alleles affecting pigmentation (Pfeifer et al., 2018). Snake venom across extant species has also converged on similar adaptations due to constraints on the number of strategies that are effective against prey (Barua and Mikheyev, 2019). However, determining the exact past selective pressures that led to specific adaptations and which genetic changes led to those adaptations are not easily ascertained in natural populations.

1.5.2 Evolvability

Evolvability is a product of historical contingency in that it too is based on the presence of mutations that were obtained in a population's past. Evolvability is the capability of a population to adapt to an environment, usually operationally assessed by measuring the degree of fitness increase over time. Epistatic interactions between mutations can increase or decrease evolvability of an organism. *Saccharomyces cerevisiae* was shown to have differences in evolvability after 230 hybrid strains with unique genotypes were evolved for 500 generations in two different environments (Jerison et al., 2017). The hybrids were created by crossing two divergent *S. cerevisiae* strains. The rate of adaptation, or evolvability, was influenced by genotypic heterogeneity of founder strains, several quantitative trait loci (QTLs), and initial founder fitness (Jerison et al., 2017). Epistasis was widespread in their evolved populations, and epistatic interactions between a QTL and other mutations were found to affect evolvability in one of their evolution environments (Jerison et al., 2017). Woods and colleagues found that the source of differences in evolvability of *E. coli* populations was from negative epistasis among beneficial mutations, which led to significant shifts in predominant genotypes (2011). Mutations that led to quickly arising predominant genotypes restricted fitness gains at later timepoints, whereas, those with slower fitness gains obtained mutations that were more evolvable and eventually led to higher

fitness (Woods et al., 2011). *Pseudomonas* phage φ6 was shown to be less evolvable through mutations that restricted host range due to epistatic interactions between mutations necessary for host range expansion (Zhao et al., 2019). These studies indicate the importance of epistasis in the ability of organisms to adapt to current and future environmental conditions. I will present an experiment that consistently results in populations and clones that are more evolvable compared to the ancestor due to genetic changes that occurred in a previous environment.

1.5.3 Compensation

Another process that involves the effect of epistasis on an organism's ability to adapt to its environment is compensation. Compensatory mutations alleviate the negative effect of deleterious mutations while being neutral or deleterious in their absence. Compensatory mutations have been extensively studied in the context of antibiotic resistance (Levin et al., 2000; Björkman et al., 1998; Nagaev et al., 2001), compensation to specific perturbations (Björkman et al., 2000; Buckling et al., 2000), and in the determination of interacting components of gene networks (Jarvik and Botstein, 1975; Reinhart et al., 2000; Zhao et al., 1998; Gu et al., 1998; Johnson et al., 1990; Jenkins et al., 1986; Van Dyk et al., 1989; Takeda and Hirota, 1982). Multidrug-resistant *Mycobacterium tuberculosis* (MDR-TB) has been found to obtain a compensatory mutation that

reverses the deleterious effect of rifampicin resistant mutations (Meftahi et al., 2016). Not only did the secondary mutation increase doubling time to levels comparable to the wild type strain in an antibiotic free environment, it also increased the rifampicin minimum inhibitory concentration that MDR-TB could withstand (Meftahi et al., 2016). The mutations were hypothesized to both be necessary for MDR-TB to obtain the fitness benefit because of their locations in the RNA polymerase β subunit and the fact that the compensatory mutation always occurred with this specific resistance mutation (Meftahi et al., 2016). There are many examples of compensation associated with antibiotic resistance, however, few studies have focused on compensatory mutations during more general adaptation, especially in the presence of fluctuating environments. I aimed to determine whether compensation occurs to alleviate costs of an evolved pleiotropic mutation and if compensatory mutations have an effect on the benefit of the evolved pleiotropic mutation in its selected environment.

1.6 Lactose Utilization Network

Throughout the experiments in my dissertation, I take advantage of the utility of the lactose utilization (*lac*) network, because it is a well-characterized metabolic network that is regulated based on inducer concentrations in the environment (Dean, 1995; Dekel and Alon, 2005; Kalisky

et al., 2007). The *lac* operon structural genes encode: LacZ, a beta-galactosidase that cleaves lactose into glucose and galactose; LacY, a permease that imports lactose into the cell; and LacA, a transacetylase that triggers export of toxic derivative sugars that cannot be metabolized by the cell. The *lac* operon is negatively and positively regulated by transcription factors based on inducer concentrations in the environment (Figure 1.1). Negative regulation occurs in the absence of natural (lactose) or artificial (e.g., isopropyl β-D-thiogalactoside (IPTG)) inducers by the LacI repressor, which can bind three different operator binding sites to prevent *lac* operon expression (Lewis et al., 1996; Oehler et al., 1990). Positive regulation takes place in the absence of glucose, which increases intracellular cyclic AMP (cAMP) concentrations allowing the cAMP receptor protein-cAMP (CRP-cAMP) complex to form. The CRP-cAMP complex binds upstream of the promoter region and promotes transcription by recruiting RNA polymerase (Zubay et al., 1970).

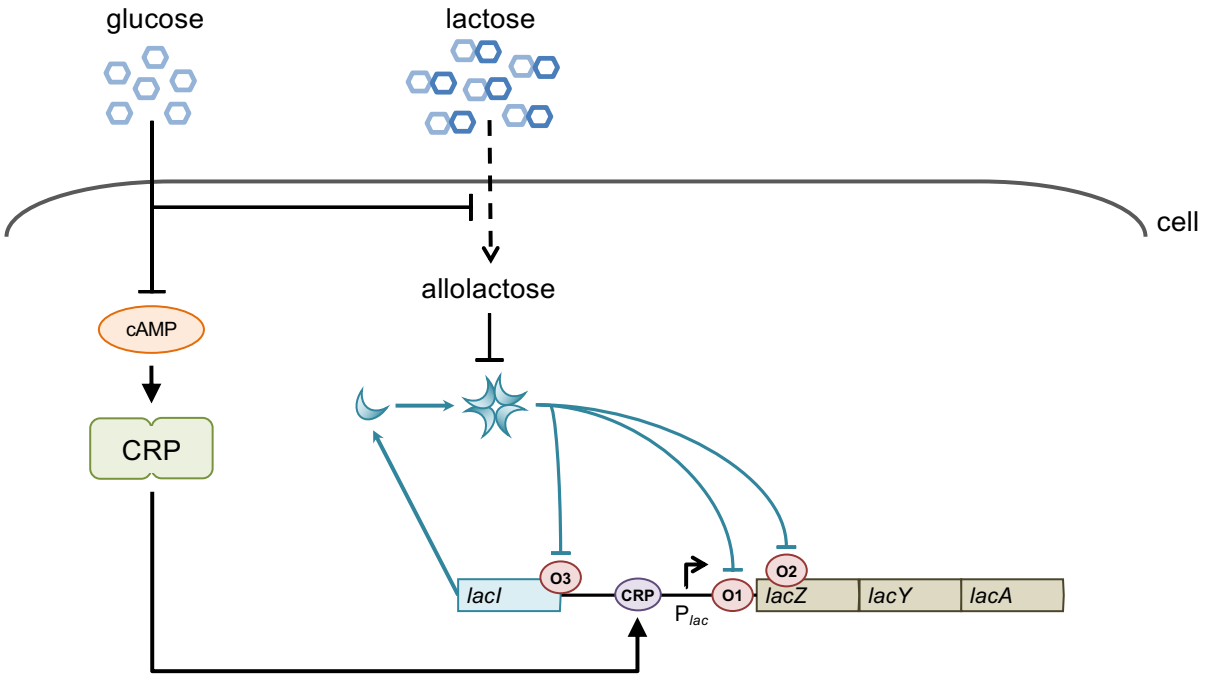


Figure 1.1 *lac* operon regulation schematic. The schematic shows positive and negative regulation of the *lac* operon with the approximate locations of the promoter region (P_{lac}), structural genes (*lacZ*, *lacY*, and *lacA*), and binding sites for cAMP-CRP (CRP) and *lacI* (O1, O2, and O3) (Lewis et al., 1996). Transcription of the *lac* operon is inhibited through binding of the LacI tetramer to its operator sites. In the presence of glucose, lactose is prevented from entering the cell, allowing the LacI tetramer to bind to its operator sites, which also occurs in the absence of lactose or artificial inducers. When glucose is depleted from the environment cAMP levels rise in the cell and available lactose enters the cell through the LacY permease. Lactose is either cleaved into glucose and galactose or converted to allolactose by LacZ. Allolactose forms a complex with LacI, which prevents LacI binding to its operator sites, allowing transcription of the *lac* operon. Transcription is further promoted through the formation of the cAMP-CRP complex that binds to the CRP binding site and attracts RNA polymerase to the promoter region.

The availability of genetic and chemical tools with which to manipulate and examine *lac* gene regulation and function has led to it becoming a model system in which to study the underlying cause of adaptive changes in the response of gene regulation to environmental changes. The *lac* network has been studied widely, for example as a model to: understand environment-response functions (Dean, 1995), predict and test evolutionary optimization (Dekel and Alon, 2005), and to examine the cost-benefit ratio for optimal protein expression (Kalisky *et al*., 2007). It has also been recognized as a focus of adaptive responses following selection of *E. coli* populations in environments containing lactose and artificial galactosides (Zhong et al., 2009; Quan et al., 2012; Satterwhite and Cooper, 2015). Mutations that change the regulation of the *lac* genes, usually to increase expression, increase fitness in lactose containing environments (Quan et al., 2012). These benefits depend on the focal genetic background, indicating that they may also be more or less likely to occur in different strains. Additionally, while differences in benefits have been examined, differences in costs of expression, the likely basis for selection of the initial tight gene regulation, have not. Finally, past work has focused almost exclusively on the functioning of the *lac* network in a few model lab strains. How the environment-response function has evolved in natural isolates and to contrasting environmental conditions in long-term experimental evolution populations is not well characterized.

1.7 Dissertation Overview

In this dissertation, I study the basis and consequences of adaptation to constant and fluctuating environments within the context of selective history and compensatory mutations, and examine within species variation in regulatory functions of a metabolic pathway relevant to my studies. Specifically, I examine: (1) if variation within a population led to increased evolvability in a new environment; (2) if the cost of an evolved pleiotropic mutation that produces a tradeoff is alleviated after evolution in fluctuating environments, and, if so, what is the mechanism of reducing that cost and does it affect the benefit; and (3) how control of *lac* operon expression varies between strains of *E. coli*, and if the level of *lac* operon expression is more dependent on global or local regulators.

In Chapter 2, I present experiments testing the influence of selective history on future adaptation. Specifically, I run a replay evolution experiment in a minimal lactose environment using populations previously evolved in a minimal glucose environment, clones selected from each of these populations, and the initial common ancestor. I find that after 500 generations of evolution in a lactose environment glucose evolved populations and clones had higher and faster fitness gains than the common ancestor that had never experienced selection in glucose. The previously evolved clones maintained a similar genomic mutation rate as the common ancestor. The results

demonstrate how adaptive genetic changes that occur during an individual's selective history increase evolvability during adaptation to a new environment, which may be a consequence of compensatory mutations.

In Chapter 3, I study compensation of an evolved pleiotropic mutation resulting from selection in fluctuating environments. The evolved pleiotropic mutation under study is *lacI*-, a nonfunctional regulator of the *lac* operon, which provides a benefit in lactose but a cost in glucose supplemented environments. After measuring the fitness effect of *lacI*- in evolved clones, I find that compensation for the cost in glucose was rare but possible and did not reduce the benefit in lactose. The mechanism for reduced cost was partly explained by a reduction in *lac* operon expression, but was also dependent on other mutations in the evolved background. The type of evolution environment, single resource and daily or long-term fluctuations, did not increase the likelihood of compensation to occur. Compensation in this system appears to be distinctive to or hindered by specific genetic adaptations.

In Chapter 4, I examine natural variation in regulatory control, described as logic functions, of the *lac* operon in a diverse set of 23 *E. coli* strains. After introducing, into each strain, a GFP reporter driven by the *lac* promoter and CRP binding site while controlled by the LacI repressor by

containing two of its operators (O1 and O3), I measured *lac* operon expression in a concentration gradient of inducer combinations. I found substantial variation in regulatory functions of natural isolates that was explained in part by genetic relatedness of strains, which best predicted initial growth in lactose, suggesting these regulatory characteristics were selected. I then examine the role of global (*trans*) and local (*cis*) regulators by transferring the *lac* operon and *lacI* gene of five natural isolates into a reference lab strain. I found that some of the variation in regulatory function is attributed to *trans*-regulators in the genetic background of the recipient strain. This work demonstrates that there can be variability in regulatory functions within a species that best matches an environment and is specific to a genetic background.

Chapter 2

Adaptation of *Escherichia coli* to glucose promotes evolvability in lactose

Phillips KN, Castillo G, Wünsche A, Cooper TF. 2016. Adaptation of *Escherichia coli* to glucose promotes evolvability in lactose. Evolution 70(2):465–470.[a]

[a.] This chapter is published in Evolution. I ran all experiments and statistical analyses with help from undergraduate students.

2.1 Introduction

Historical contingency—when the evolutionary history of a population affects its subsequent evolution—can profoundly influence evolutionary outcomes (Gould, 1989). For example, populations with different evolutionary histories might respond differently when they are selected in a common environment due to epistatic interactions that cause new mutations to have effects that depend on different genetic backgrounds (Barrick et al., 2010; Perfeito et al., 2013; Harms and Thornton, 2014; Kryazhimskiy et al., 2014). To the extent that historical contingency is

influential, it can mean that predicting evolutionary outcomes will require some knowledge of the past selection of a population.

In practice, it is difficult to distinguish between the action of contingent and chance effects in shaping evolutionary outcomes. To do so requires a comparison of the replicated evolutionary response of populations with known histories evolving in identical environments. Microbial evolution experiments represent one way to meet these conditions. Indeed, several such studies have found evidence for contingency. For example, replicate populations that diverged from one another during adaptation to glucose responded differently, largely erasing those differences, following subsequent selection in maltose (Travisano et al., 1995). A similar result was seen in a recent study that identified the current fitness of a population as the primary driver of subsequent evolutionary potential (Kryazhimskiy et al., 2014). Other examples include finding differences in the evolvability—usually measured as the extent of fitness improvement over a defined time interval—of bacterial and viral populations that were initially isogenic except for different deleterious or beneficial mutations (Burch and Chao, 2000; Moore et al., 2002; Cuevas et al., 2009; Barrick et al., 2010; Salverda et al., 2011; Woods et al., 2011; Perfeito et al., 2013; Plucain et al., 2014) or degree of specialization (Buckling et al., 2003), and cases where early 'potentiating'

mutational events are required for the evolution of subsequent phenotypic novelty (Blount et al., 2012; Meyer et al., 2012).

The studies cited above have clearly identified a role of history in influencing future evolutionary potential. In most cases, however, this history is either the result of: chance differences in mutation order leading to differences between genotypes despite selection in a common environment (Travisano et al., 1995; Blount et al., 2008; Salverda et al., 2011; Woods et al., 2011; Kryazhimskiy et al., 2014) or deliberately selected or engineered differences in genotypes (Moore et al., 2002; Barrick et al., 2010; Perfeito et al., 2013). Alternatively, genotypes with different selective histories have been compared, but without replication at the level of the past selective environment (Cuevas et al., 2009). It has rarely been possible to test for a systematic effect of past selection environments on future evolvability (Buckling et al., 2003; McBride et al., 2008; Bedhomme et al., 2013).

We previously presented an observation consistent with adaptation of *Escherichia coli* to a glucose-limited environment increasing the rate of subsequent adaptation to an otherwise identical lactose-limited environment (Satterwhite and Cooper, 2015). Here we carry out an experiment that repeats and extends the basis of this observation. We find that glucose-evolved populations had a

consistent advantage in subsequent adaptation to lactose, increasing fitness at a higher rate and to a higher final level, relative to populations started from the original ancestral strain. This result held even when populations were started from individual glucose-evolved clones, indicating that the advantage was a property of individual genotypes. We interpret these results as evidence that adaptation to one or more components of the glucose environment resulted in a repeatable increase in evolvability in the lactose environment.

2.2 Materials and Methods

2.2.1 Bacterial strains and selection experiments

The initial evolution experiment consisted of six *E. coli* populations evolved in each of seven selection treatments (Cooper and Lenski, 2010). These treatments differed only in the nature of the limiting carbon source(s) added to Davis Mingioli (DM) medium. All populations were founded by strain REL606 or strain REL607, an otherwise isogenic derivative of REL606 that is able to grow on arabinose (Ara+) (Lenski et al., 1991). Here, six populations—referred to as founder populations—previously evolved for 2,000 generations in glucose (175 µg/ml) supplemented medium were used to found a series of new replay populations that were selected in

lactose (210 μg/ml) supplemented medium. These replay populations comprised: (i) replication of each of the six founder populations into six new populations (36 total populations), (ii) isolation of a randomly chosen clone from each founder population that was used to found six new populations (36 total populations). In addition, eighteen control populations were founded from the ancestral strains REL606 or REL607. A schematic of this experimental design is presented in Fig 2.1. A derivative of REL606 that encodes GFP was used as a reference strain in competition experiments (Zhang et al., 2012).

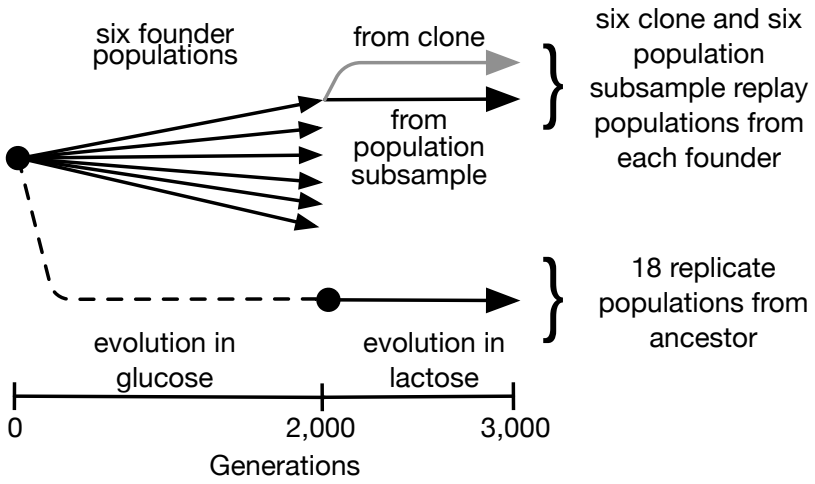


Figure 2.1 Schematic of experimental design. Six replicate populations were started from an ancestral strain and evolved for 2,000 generations in a glucose-limited environment. Each of these founder populations was used to found 12 new populations, six started from a random clone and six from a population sub-sample. These new populations and 18 populations started directly from the ancestor were evolved in lactose for 1,000 generations and their fitness compared.

A new evolution experiment involving the 90 populations described above was carried out using the same protocols and growth conditions used in the original evolution experiment (Cooper and Lenski, 2010). Briefly, populations were grown in 1 ml DM supplemented with 210 µg/ml lactose (DM210 lactose) for 1,000 generations by daily transfer of 10 µL samples. Propagation was in 96-deep well polypropylene plates. As far as possible, populations with different Ara marker states were arranged in a checker board pattern to facilitate checking for cross contamination of wells. Every 13 days cultures were stored at -80 ºC in 20% glycerol, and tested for cross contamination by inoculating 10 µL from each well onto DM+arabinose agar and plating onto tetrazolium arabinose (TA) indicator medium.

2.2.2 Fitness competitions

Relative fitness of evolved populations was measured at 0, 500 and 1,000 generations of evolution using competitive growth assays in the DM210 lactose selection environment. All assays were carried out as complete experimental blocks by using three 96-well deep well plates, one for each assayed evolutionary time point. Prior to each competition, experimental and reference fluorescent ancestor (one per competition) populations were acclimated to the selection environment over two daily transfer cycles. On the third day, paired reference and experimental populations were

combined at 1:1 and a 1:100 dilution of this mix was used to start competition cultures. An Accuri C6 (Becton Dickinson, NJ) flow cytometer was used to measure the ratio of competing strains. To do this, competitions were diluted 1:100 from the original 1:1 mix (t0) or from the end of one-day of competition (t1) into a 96-well polystyrene plate where each well contained a solution of 150 nM SYTO17 fluorescent dye and 10% DM made without $MgSO_4$ or thiamine. Cells were separated from background noise on the basis of SYTO17 fluorescence, which indicates the presence of nucleic acid. Reference cells were distinguished from their competitors on the basis of their GFP fluorescence signal. At least 5,000 events were collected from each sample time point. The percentage of each competitor at the beginning and end of each competition was used to estimate fitness as the ratio of Malthusian parameters (Satterwhite and Cooper, 2015). Using the percentage, rather than density, of competitors to calculate this ratio will underestimate fitness differences if evolved populations consistently reach lower stationary phase densities than ancestral populations. Although we cannot exclude this possibility, we were not able to find evidence for it in a previous study examining these populations (Satterwhite and Cooper, 2015).

2.2.3 Mutation rate estimates

Mutation rate to rifampicin resistance was measured for the ancestor and each of the six clones isolated from the founder glucose-evolved populations. To do this, freezer stocks were grown overnight at 37 °C in lysogeny broth (LB) and then diluted to give an inoculum of approximately ~10^3 cells that was added to each of 23 fresh LB cultures. After overnight growth, a sample from each replicate population was plated onto LB + rifampicin (100 μg/mL) to determine the number of Rif^r mutants. Samples from three populations were also plated onto LB to determine total cell density. Colonies were scored after 24 hours incubation at 37 °C. Analysis was carried out using the MSS-MLE method implemented in the FALCOR online calculator (http://www.keshavsingh.org/protocols/FALCOR.html) (Hall et al., 2009).

2.2.4 Genome sequencing

Illumina library preparation was carried out using the NexteraXT kit protocol, except that reaction volumes were reduced by a factor of four, or the TruSeq kit protocol. Sequencing was performed on HiSeq2000 and NextSeq500 machines. Mutational changes occurring in evolved clones were identified by comparison to the previously sequenced ancestral strain, REL606, using BRESEQ

(Barrick et al., 2009). A mutation in *recD* (V10A) was found in the three clones derived from REL607 and was subsequently identified to be present in our stock culture of that ancestral strain. Extensive competition experiments between the Ara- REL606 and our Ara+ derivative indicate that the mutation is neutral, so we omit it from further consideration (data not shown).

2.2.5 Statistical analyses

Statistical analyses were performed in R version 3.1.1 (R Core Team, 2013). Differences in evolvability were determined by using the package NLME4 to fit a linear mixed effects model (Bates et al., 2014). Replicate populations were nested within their initial founder population or genotype as appropriate and were fitted with random intercepts. Models were fit with and without the effect of interest, usually whether populations were evolved directly from the ancestor or from glucose-evolved founder genotypes, and their explanatory power compared using a χ^2 test with one degree of freedom. Replay populations started from the glucose-evolved founders started off at a lower fitness in lactose than did populations started from the ancestor, which can have a systematic effect on evolvability (Barrick et al., 2010; Kryazhimskiy et al., 2014). It was not possible to account for this effect by using initial fitness as a covariate because it is completely confounded with population history. For this reason, in most cases, we focus on comparing the

final fitness of replay populations, rather than overall change in fitness, because a higher final fitness of initially less fit populations clearly indicates a meaningful difference in evolvability due to genotype. Initial fitness was included as a covariate in analyses that test for differences between founder populations or clones.

2.3 Results

We previously reported the evolution of six replicate populations that were selected in a glucose-limited environment for 2,000 generations and then transferred to selection in an otherwise identical lactose-limited environment (Satterwhite and Cooper, 2015). Here, we follow up the observation that the six glucose-evolved populations improved in fitness more rapidly when switched to selection in lactose than did populations started directly from the same ancestral strain. Following transfer to lactose selection, glucose-evolved populations increased in fitness measured in lactose by an average of 32.1% after 1,000 generations compared to 23.1% for replicate populations derived directly from the ancestor. All fitness measurements are relative to the same ancestral strain. This observation is consistent with a period of selection in glucose systematically increasing the potential to adapt—which we here refer to as an increase in evolvability—to the lactose environment (ANOVA $F_{1,74}$=10.95, p = 0.001). Increased evolvability may be due to

mutations that commonly accumulate during selection in glucose acting to increase subsequent evolvability in lactose. Alternatively, higher evolvability might reflect a higher initial genetic variance in fitness within glucose-evolved populations relative to populations started from the ancestor. Higher variance in fitness is expected to increase the rate of adaptive response independently of any change in the adaptive potential of a specific genotype (Fisher, 1999).

To disentangle the effect of genotype specific evolvability from the effect of differences in initial population-level genetic variation on subsequent adaptation to lactose, we repeated the part of our previous evolution experiment involving adaptation to lactose including replication at the level of each of the six founder glucose-evolved populations. From each founder population 12 new 'replay' populations were started, six from population subsamples and six from a single randomly selected clone. These replay populations and 18 reference populations started directly from the original ancestral genotype were evolved for 1,000 generations in lactose.

After 500 generations of evolution in lactose, the fitness of the replay populations started from both clones and population subsamples were significantly greater than that of populations started directly from the ancestor (ANOVA; population subsamples vs. ancestor: $p < 0.001$; clones vs. ancestor: $p < 0.001$), but were not significantly different from each other (ANOVA; $p = 0.36$) (Fig.

2.2). These differences reflect that the glucose-evolved populations overcame an initially lower fitness to reach a higher fitness than the populations started directly from the ancestor. After 1,000 generations of evolution, the glucose-evolved populations still had higher total fitness increases than did populations started from the ancestor. At this time point, however, the final fitness of populations started from founder clones was marginally non-significantly higher than the fitness of populations started from the ancestor (final fitness: population subsamples vs. ancestor, $p = 0.003$; clones vs. ancestor, $p = 0.086$). Again, the fitness of replay populations started from clones and from population subsamples were not significantly different from one another (ANOVA; $p = 0.27$).

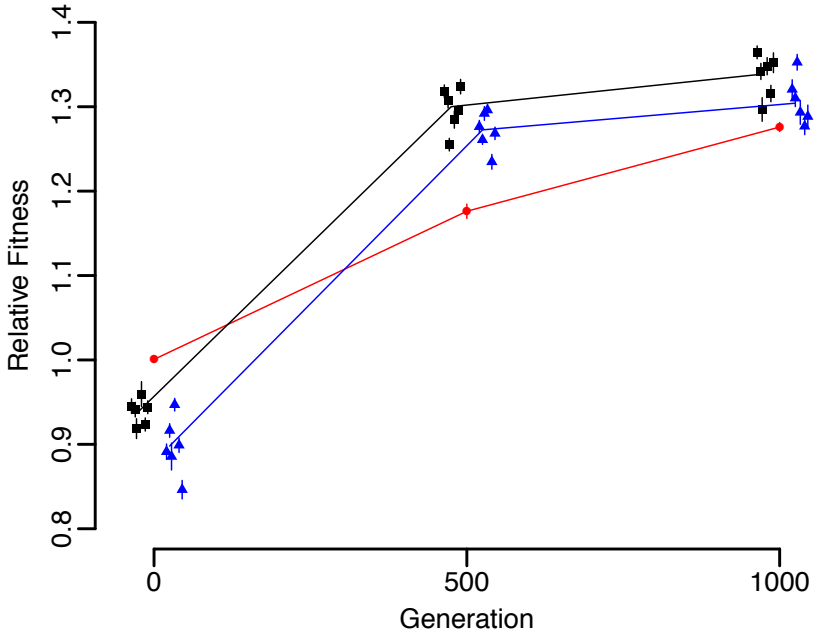


Figure 2.2 Fitness change in the lactose-limited environment of populations evolved in this study. Each symbol represents the mean of six replicate populations started from the same founding clone or population sub-sample (blue and black, respectively) or of 18 replicate populations started directly from the ancestor (red). For each treatment, solid lines connect the grand mean across assayed time points. Error bars indicate SEM. All fitness measurements are made relative to the ancestor. Points are offset slightly on the X-axis to increase legibility.

We focused on the six founder glucose-evolved clones to begin to examine a possible basis for their higher evolvability in lactose. A candidate explanation is that the glucose-evolved clones had a higher genomic mutation rate than did the ancestral strain. Higher genomic mutation rates have been associated with more rapid fitness improvement in several lab-evolution experiments (Chao and Cox, 1983; de Visser et al., 1997). To test this, we used fluctuation tests to measure the spontaneous mutation rate to rifampicin resistance of the glucose-evolved clones and the ancestral strain. We found no substantial difference in this rate, with all ancestor vs. glucose-evolved clone mutation rate estimates being within a factor of three of one another (Fig. 2.3).

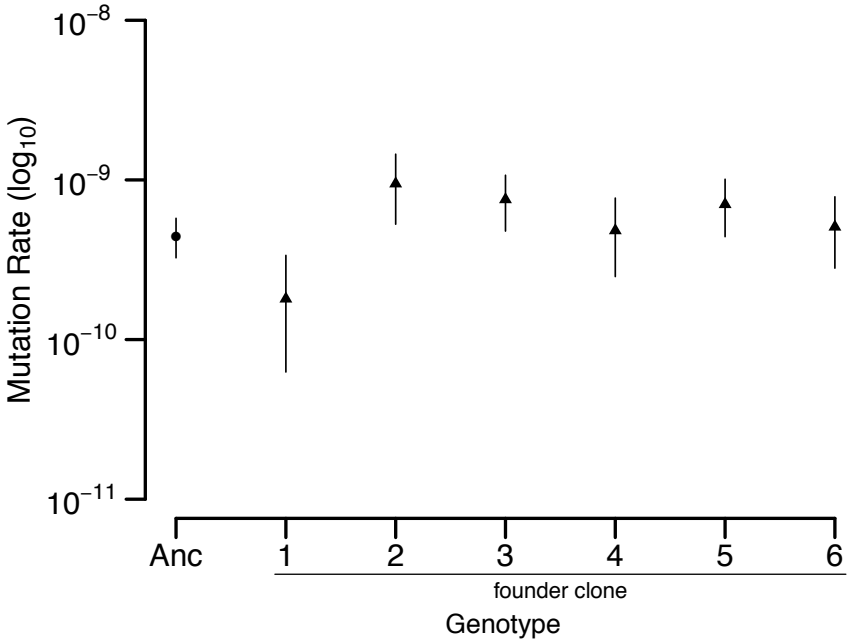


Figure 2.3 Mutation rate to rifampicin resistance of the ancestor and the six glucose-evolved founder clones. Symbols and error bars represent the estimated mutation rate to rifampicin resistance and 95% confidence intervals.

A similar mutation rate between founder clones and the ancestor indicates that selection in glucose led to accumulation of mutations that increased lactose evolvability directly. To identify candidate mutational changes that might contribute to evolvability, we sequenced the six founder clones. Because the increased lactose evolvability trait was common to all clones, judged by their similar fitness increases following selection in lactose, we reasoned that mutations that occurred in the same genes in multiple clones would be candidates for contributing to lactose evolvability. We identified 3-7 mutations per clone with four genes or gene regions being mutated in at least four clones; *rbs*, *spoT*, *iclR*, and *nadR* (Table 2.1). These genes are candidates for affecting evolvability, though a determination of their influence will require their removal from founder strains and measurement of the evolvability of those derived strains.

Table 2.1 Mutations in founder glucose-evolved clones

	Founder clone					
Mutated gene/region	1	2	3	4	5	6
bioA			G150S			
ECB510/nohB			-144/-245*			
fabG/acp	+156/-55*					
iclR		T60P	G219R		-58/-142	Δ48
nadR	::IS*150*	S178L	::IS*186*		+4	L289P
pykF				D336N		
rbs	Δ4,398	Δ4,649	Δ6,390	Δ6,213	Δ290	Δ3,252
spoT		G207D	Δ1	F409V	R701Q	R209L
wecF	N206D					
ybbN		::IS*150*				
ycjX						S294R
yijC		::IS*150*				
yjeP						Δ36
yoaD		G367G				

*For intergenic mutations, numbers indicate bases upstream/downstream of adjacent genes. For coding sequence mutations residue and change in one-letter amino acid code, insertion/deletion size, or identity of IS insertion are given.

2.4 Discussion

Several experiments have found that evolutionary history can influence subsequent evolutionary potential (Travisano et al., 1995; Burch and Chao, 2000; Buckling et al., 2003; Cuevas et al., 2009; Barrick et al., 2010; Woods et al., 2011; Perfeito et al., 2013; Kryazhimskiy et al., 2014). In general, this dependence reflects the pervasive effect of epistasis in making the influence of new beneficial mutations dependent on a particular genetic background. For example, to the extent that there is a generally antagonistic relationship between beneficial mutations, initially low fitness genotypes are expected to adapt more quickly than fitter genotypes (Chou et al., 2011; Khan et al., 2011; Wiser et al., 2013; Kryazhimskiy et al., 2014). Indeed, Kryazhmiskiy et al. (2014) found that differences in the effect of evolutionary history on adaptive potential could be explained mainly through the effect of differences in initial fitness rather than the different genotypes that are the specific basis of those fitness differences.

We found that despite ancestral and glucose-evolved populations having similar initial fitness in lactose, glucose-evolved populations increased in fitness more quickly and to higher levels than the ancestral genotype. Clones derived from the glucose-evolved populations were also more evolvable than the ancestor, indicating that increased evolvability was a consequence of specific

mutations, not just of initially higher within-population genetic variability. We can't distinguish between these mutations being selected to provide a benefit to the glucose component of the environment, or, perhaps to other components, for example, the particular growth regime (oxygenation and pH levels, etc.) prevailing during selection. This result is distinct from some previous studies that have found populations with different evolutionary histories, but with the same or unknown selective histories, to have different responses to subsequent selection (Burch and Chao, 2000; Woods et al., 2011). Here, we find that the difference in evolvability is associated with a difference in selective history.

The significant trend of independent glucose-evolved populations being more evolvable than the ancestor indicates that the genetic basis of higher lactose evolvability is a common outcome of evolution in glucose. To explain this, we suggest that increased evolvability could be due to the availability of compensatory mutations that reverse the effects of mutations accumulating during evolution in the glucose environment that are deleterious in the lactose environment. As part of the experiment reported here we also selected populations derived from the ancestral strain in a constant lactose environment (Cooper and Lenski, 2010; Satterwhite and Cooper, 2015). We found mutations in *rbs* and *spoT* in six of six, and mutations in *nadR* in five of six, of these populations, suggesting that mutations in these genes confer a benefit in lactose. Indeed a *rbs* deletion allele

and a *spoT* allele isolated previously (Khan et al., 2011), were found by themselves to confer significant fitness benefits in the lactose environment, as well as the glucose environment in which they were selected (*rbs*: glucose 3.2%, p <0.001, lactose 5.2%, p < 0.001; *spoT*: glucose 10.1%, p <0.001, lactose 5.6%, p = 0.002). Although it is possible that some evolved *rbs*, *spoT* and *nadR* alleles have glucose-specific benefits, we think a more likely explanation for our observation that the glucose-evolved founder populations had not increased in fitness in lactose is that they accumulated a mix of mutations with beneficial and deleterious effects in the lactose environment that effectively cancel out each others effects. A candidate for lactose deleterious mutations are those in *iclR*, which were found in 5 of 6 glucose but 0 of 6 lactose selected populations. During selection in lactose, any mutations that arose and compensated for the deleterious mutations would therefore 'reveal' the pre-existing beneficial mutations. Under this scenario, higher evolvability could result if compensatory mutations occurred at a higher rate than generally beneficial mutations of similar effect or if single mutations could confer very large benefits by compensating large-effect deleterious mutations (Moore et al., 2000; Barrick et al., 2010; Perfeito et al., 2013). In summary, while the genetic mechanism(s) of increased lactose evolvability remains unknown, what is clear is that adaptation to the glucose evolution environment moved genotypes to regions of genetic space where mutations that increased fitness in lactose were either more common or of higher benefit.

Selection in glucose consistently leads to changes in lactose evolvability through accumulation of specific mutations that change the effect of additional mutations. Our results demonstrate that even when adaptation to one environment confers no immediate benefit, or even a small cost, in a second environment, it can nevertheless confer a longer-term benefit. We emphasize that fitness in lactose had no direct benefit to the glucose-evolved populations prior to the switch to lactose as the limiting sugar in their selective environment at 2,000 generations. The increase in evolvability, while perhaps a consequence of mutations that increased direct fitness, was not itself selected. Of course, whether such a benefit will be realized will depend on the specific circumstances of environmental change.

Chapter 3

The cost of evolved constitutive *lac* expression is usually, but not always, maintained during evolution in environments with fluctuating expression demand

3.1 Introduction

Fluctuating environments pose several challenges to evolving populations. While some potential adaptations might confer benefits across all relevant selection regimes (Bennett and Lenski, 1999; Buckling et al., 2000; Buckling et al., 2007; Kassen and Bell, 1998; Satterwhite and Cooper, 2015), others will confer benefits in some and costs in others (Bailey and Kassen, 2012; Jasmin and Kassen, 2007; Lee and Marx, 2012; McGee et al., 2015; Roemhild et al., 2015). Indeed, even if unconditionally beneficial mutations are initially available, they are likely to become less common over time (Martin and Lenormand, 2015; Satterwhite and Cooper, 2015; Schick et al., 2015). When a mutation that confers a benefit in one environment, and a cost in another, fixes in a population, it creates selection for subsequent mutations that compensate for that cost (Fig. 3.1) (Maisnier-Patin et al., 2002; Moore et al., 2000; Moura de Sousa et al., 2017; Wood et al., 2013).

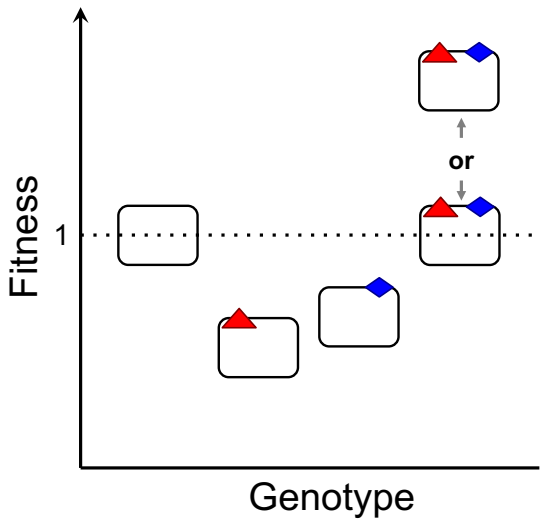


Figure 3.1 Schematic of the fitness effects of the individual and combined focal deleterious and compensatory mutation. The open rectangle on the far left represents a wild-type genotype. A compensatory mutation (blue diamond) alone is neutral or deleterious but will lessen the effect of an original deleterious mutation (red triangle). For example, addition of a secondary mutation (blue diamond) can return fitness to a level similar to that of the wild-type strain, or it can increase fitness to a level greater than the wild-type.

Compensatory mutations have long been used in molecular genetic studies as a tool to identify physical and genetic changes that can suppress the effects of a focal mutation, thereby identifying interacting components (Blank *et al*., 2014; Gu et al., 1998; Jarvik and Botstein, 1975; Kacar et al., 2017; Manson, 2000; Ponmani and Munavar, 2014; van Leeuwen et al., 2016). Increasingly, they are also recognized as being important in broadening the scope of evolutionary trajectories a population can follow (Szamecz et al., 2014; Zee et al., 2014), allowing adaptations to be selected that might otherwise prove to be evolutionary dead-ends (Harrison et al., 2015), and influencing the ability of populations to simultaneously adapt to multiple environments (Melnyk et al., 2017).

Mutation interactions arising after selection in fluctuating environments have been extensively studied in the context of antibiotic resistance evolution. Antibiotic resistance mutations often confer a cost to bacteria in antibiotic free environments (Moura de Sousa et al., 2017; Nilsson et al., 2003; Rozen et al., 2007), but subsequently generate selection for compensatory mutations that relieve the cost (Levin et al., 2000; Björkman et al., 1998; Björkman et al., 2000; Nagaev et al., 2001). These compensatory mutations are typically neutral or deleterious in the absence of the original resistance mutation (Brandis et al., 2012; Maisnier-Patin et al., 2002; Moura de Sousa et al., 2017). Compensatory mutations would not be selected by themselves but allow resistance mutations to be maintained when they would normally be selected against, influencing short-term

evolutionary outcomes and perhaps longer-term potential. Similar patterns of compensatory mutations depending on earlier resistance mutations for their benefit have been seen in studies of bacterial resistance to bacteriophage (Lenski 1988; Wielgloss 2016). Compensation during evolution has also been found to occur to overcome the loss of essential genes (Blank *et al*., 2014), the negative effects of synonymous (Knöppel et al., 2016) and gene deletion (Szamecz et al., 2014) mutations, and to restore a social trait (Zee et al., 2014).

In contrast to studies that have focused on mutations that compensate for a specific genetic perturbation, few studies have examined compensation during more general adaptation to an environment, especially when this adaptation involves repeating rounds of selection in contrasting environments. This distinction might be important. Compensation to a specific genetic perturbation, as to an antibiotic resistance mutation or a deletion of a focal gene, is thought to generally act locally to reverse the costly effect (Brandis et al., 2012; Filteau *et al*., 2015; Szamecz *et al*., 2014), although they can operate from pathways unrelated to the perturbation (Blank et al., 2014). At least in experimentally evolving populations, adaptation often involves mutations in regulatory genes that are likely to have highly pleiotropic consequences (Cooper *et al*., 2003; Kurlandzka et al., 1991; MacLean et al., 2004; Rosenzweig et al., 1994). Such mutations are likely to have widespread effects on cell physiology so that effective compensation may also be

pleiotropic (Velicer and Yu, 2003; Zee et al., 2014). Moreover, in many cases it remains unclear how compensatory mutations will affect fitness in the original environment in which a focal mutation was selected. It is easy to imagine that compensation causing a reduction in the cost of a focal mutation in a new environment might be associated with a reduction of the benefit it confers in the original environment. In that case, reversion to the original environment might select for reversal of the effects of the compensatory mutation. This could occur through its direct reversion or through a second compensatory mutation, creating potentially complex patterns of environmentally dependent epistatic interactions between selected mutations.

The particular nature of the environmental fluctuations a population is exposed to is expected to play a major role in the selection of compensatory mutations. In a rapidly changing environment, mutations that increase in frequency are likely to confer a net benefit across the different environments (Buckling et al., 2007; Melnyk et al., 2017; Turner and Elena, 2000). With this limitation, such mutations can only confer at most relatively small costs in any one environmental component so that the strength of selection for compensation may be small (Poon and Chao, 2005a; Poon and Chao, 2006). In a more slowly fluctuating environment, mutations selected in one component might fix before the population experiences a second component in which they might confer substantial costs (Bennett and Lenski, 2007; Cooper and Lenski, 2001; Kassen and Bell,

1998; Phillips et al., 2016). Such differential costs are consistent with the generally higher between-environment trade-offs seen in populations selected in slowly compared to quickly fluctuating environments (Bono et al., 2017; Satterwhite and Cooper, 2015; Schick et al., 2015).

I examine compensation to an adaptive mutation selected in a series of experimentally evolved populations selected in environments that contained either lactose or glucose alone or a combination of lactose and glucose fluctuating on daily or 2000 generation time-scales (Cooper and Lenski, 2010). Mutational inactivation of the LacI repressor was rapidly selected in many of the replicate bacterial populations that were selected in the presence of lactose (Quan et al., 2012). Loss of LacI causes the *lac* operon, a set of genes required for utilization of lactose, to be constitutively expressed (Markiewicz et al., 1994; Quan et al., 2012). Constitutive expression of *lac* genes provided a benefit of ~9% during growth in lactose by shortening the lag time before resumption of growth following transfer into fresh medium (Quan et al., 2012). It also conferred a cost of ~3% in an environment containing glucose as the sole resource, probably due to some combination of the energetic cost of expressing unnecessary genes and toxicity of the LacY permease (Dekel and Alon, 2005; Stoebel et al., 2008; Quan et al., 2012; Eames and Kortemme, 2012). In populations selected in environments containing both lactose and glucose, this trade-off

in the effect of *lacI*- mutations creates potential for subsequent mutations to provide a fitness benefit by compensating for the cost of the mutation in glucose.

I present a series of experiments to determine if evolved populations that fixed the *lacI*- mutation, and therefore constitutively express the *lac* operon, evolved mechanisms to alleviate the cost of this expression during growth in glucose and, if so, whether this compensation comes at a cost of the benefit conferred by *lacI*- in lactose. To do this, I isolated strains from populations evolved for 8,000 generations in lactose, glucose, and combinations of both fluctuating daily and every 2,000 generations. The *lacI*- mutation was reverted in those strains that had substituted it, and its effect on fitness determined. I found that the fitness cost of *lacI*- in glucose was variable, but did not differ consistently between populations evolved in lactose only, where compensation is not expected to be selected, and in environments containing glucose, where it is. Similarly, strains varied in their relationship between the fitness benefit conferred by the *lacI*- mutation in lactose and costs in glucose, but this variation did not depend on their selection environment. Together, these results demonstrate the potential for the action of compensatory mutations to influence costs of adaptation, but indicate that their effects may either be idiosyncratic or be overwhelmed by the effects of additionally selected mutations.

3.2 Materials and methods

3.2.1 Bacterial strains and strain construction

Bacterial strains were selected from replicate populations started with strains REL606 and REL607, which are isogenic except for a neutral arabinose marker and a mutation in *recD* that also appears to be neutral (Tenaillon et al., 2016). Populations were evolved in Davis-Mingioli (DM) minimal media supplemented with different presentations of glucose (175 μg/ml) and lactose (210 μg/ml) (Cooper and Lenski, 2010). The evolution environments included DM supplemented with lactose (Lac), daily fluctuations of glucose and lactose (G/L), or long-term switching from glucose to lactose every 2,000 generations (G_L) or from lactose to glucose every 2,000 generations (L_G). Each population was evolved for 8,000 generations, except for one G/L population, which was evolved for 7,000 generations (G/L2). Six replicate populations were evolved in each environment. Clones with a *lacI-* mutation were selected from populations based on their growing as a blue colony on indicator plates that contained X-gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside) and glucose (TGX plates) (Fig. 3.2; Quan et al. 2012). On TGX plates, the blue phenotype is indicative of a strain constitutively expressing the *lac* operon. A total of nine clones were selected, one from each of three lactose only populations, a G_L population, and five G/L

populations. In all clones the *lacI* gene was amplified and sequenced to verify that *lac* operon constitutive expression was caused by a mutant LacI repressor (Fig. 3.2). Amplification was carried out using the primers: 5'-GCGGAGCTGAATTACATTCC-3' (11-F) and 5'-GGGTGCCTAATGAGTGAGCT-3' (12-R).

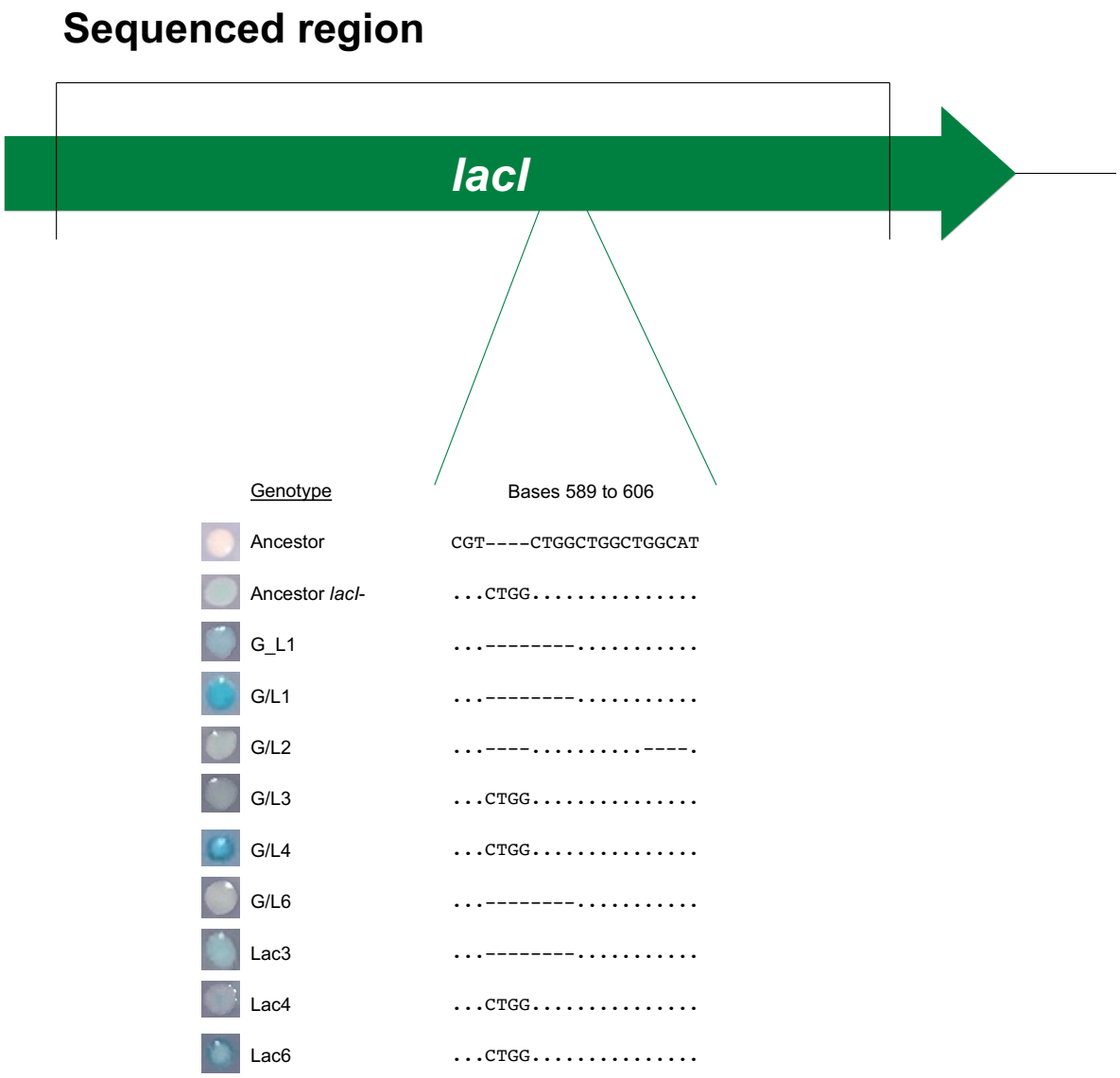


Figure 3.2 X-gal phenotype and *lacI* mutations of sequenced region. A representative colony for each clone on X-gal indicator plates is shown at left. All mutations across evolved clones were either a 4 bp insertion or deletion frameshift mutation in a mutational hotspot region of *lacI*. Only the insertion is shown for ancestor *lacI*-, but the deletion was also assessed and had the same fitness effect and X-gal phenotype.

To construct ancestral *lacI* allele derivatives of evolved strains that had substituted a *lacI*- mutation, I first PCR amplified the ancestral allele using primers 11-F and 12-R. The PCR product was ligated into pCR2.1 using a TA cloning kit (Invitrogen, California) and used to transform TOP10F' cells. After blue/white colony screening to identify transformants having a plasmid insert, pCR2.1::*lacI*+ was purified from TOP10F' cells (QIAGEN, Germany) and checked for the correct size insert by restriction fragment length polymorphism. To transfer the *lacI*+ allele to the suicide vector pDS132, pCR2.1::*lacI*+ and pDS132 were both digested using enzymes SacI and XbaI (New England Bio Labs, Massachusetts). The *lacI*+ encoding fragment from the pCR2.1::*lacI*+ digestion products were extracted from the gel and purified using a gel purification kit (QIAGEN, Germany). The pDS132 digestion products were purified using a PCR purification kit. The *lacI*+ gel extract inserts and purified pDS132 digestion products were ligated together and used to transform MFD*pir* cells (Ferrières *et al*., 2010). MFD*pir* (pDS132::*lacI*$^{+}$) was separately conjugated with each target evolved strain by mixing donor and recipient at a 1:2 ratio, respectively. Conjugation was carried out on LB agar plates supplemented with 2,6-diaminopimelic acid (DAP; 30 μM) for 3-4 hours at 37 °C. The conjugation mixture was resuspended in 200 μL DM medium and plated onto minimal glucose agar (MG) supplemented with chloramphenicol (Cm; 20 μg/mL) to select for strains with the pDS132 plasmid successfully integrated into the chromosome (Philippe *et al*., 2004). MFD*pir* were counter-selected by omitting DAP, which they require for

growth. After overnight incubation, six colonies were restreaked onto MG + Cm agar, again incubated overnight, and then restreaked a second time to the same medium. A colony descended from each of the original six was resuspended in 500 µL DM base liquid media and plated (50 µL) onto sucrose plates supplemented with X-gal to select for excision of the pDS132 plasmid (Philippe *et al*., 2004) and screen for clones that had retained the introduced *lacI*+ allele. One blue colony and one white colony were selected from each plate and each were restreaked twice onto sucrose + X-gal plates. White colonies indicate clones that successfully integrated the *lacI*+ allele and the blue colony (negative control) is a clone that went through the allele replacement process but did not retain the *lacI*+ allele. The blue negative control clone was used to test for the presence of additional mutations that occurred during the allele replacement process. All clones were tested for chloramphenicol sensitivity by spotting onto DM + glucose + Cm agar to ensure the plasmid was excised. Clones that were sensitive to chloramphenicol were grown up in 3 mL lysogeny broth (LB) and made into a freezer stock (600 µL LB culture mixed with 400 µL 80% glycerol) and stored at -80 °C. The *lacI*+ insertion was verified by using PCR to amplify a portion of *lacI* that contained the transferred mutation (using primers 11-F and 12-R) and sequencing this product using the 11-F primer.

To evaluate the possibility of secondary mutations being inadvertently added during construction of *lacI*+ strains, I measured the fitness of control strains that went through the conjugation process but that did not retain the introduced *lacI*+ allele in competition with an otherwise isogenic strain with a distinct neutral *ara* marker. The *ara* marker strains were constructed either by pairwise conjugations between the evolved strain and MFD*pir* (pDS132::*ara*-) (same method as above except X-gal was not used in the media), or by plating 100 μL of overnight LB cultures onto minimal arabinose agar (MA) and selecting for spontaneous *ara*+ mutants. Indistinguishable fitness between competing strains was interpreted as indicating the absence of a fitness-effecting secondary mutation.

3.2.2 Fitness assays

Fitness effect of the *lacI*- mutation in each isolated strain was measured in the relevant evolution environment and in glucose (175 μg/ml) and lactose (210 μg/ml) only environments. Cells were transferred daily into fresh media at a 1:100 dilution and incubated at 37 °C with shaking at 200 rpm. Strains were initially acclimated to the assay environment over two 24-hour transfer cycles with a 1:100 dilution occurring between each cycle. Pre-conditioned competitors were mixed at a 1:1 ratio by diluting each competing strain 1:200 directly into the assay environment. Competitions

were carried out over two or four transfer cycles as noted in figure legends. On the initial and final day of competitions, cells were plated onto TGX plates and incubated at 37 °C for 18-20 hours in order to distinguish competing genotypes. Relative fitness effect of the *lacI*- mutation was determined based on the change in density of blue (*lacI*-) and white (*lacI*+) colonies on TGX plates using the formula: $\ln(blue_2 \times 100^z/blue_0)/ \ln(white_2 \times 100^z/white_0)$, where subscripts indicate the time at which competitor density was estimated, and z accounts for transfer cycles during the competition (Lenski *et al*., 1991). Test competitions checking for additional mutations occurring during allele replacement procedures or selection of spontaneous *ara*+ mutants were performed as described above, except competitors were distinguished by plating onto tetrazolium arabinose agar (TA).

3.2.3 Expression assays

Expression of the *lac* operon was measured using a GFP reporter construct controlled by the P_{lac} promoter region including the O1 and O3 LacI operator sequences, and native CRP and ribosomal binding sites (Fig. 3.3) (Quan *et al*., 2012). This reporter was previously cloned into a mini-Tn*7* cassette in a suicide-vector (Quan *et al*., 2012) and was introduced into target strains by tri-parental conjugations between a target recipient evolved strain, a donor strain (MFD*pir*

(pUC18R6KT::P_{lac}-GFP, kan^r)), and a helper strain (MFD*pir* (pTSN2)) (Choi et al., 2005; Ferrières, et al., 2010; Quan *et al*., 2012). Strains were mixed at a 3:1:1 ratio (recipient : donor : helper) on LB + DAP (30 μM) agar and incubated at 37 °C for 3 hours. DAP is added to the medium because the donor and helper strains are both MFD*pir*, which is a diaminopimelic acid auxotroph (Ferrières *et al*., 2010). The conjugation mix was resuspended in DM medium and plated onto LB + kanamycin (Km; 60 μg/mL) + isopropyl β-D-1-thiogalactopyranoside (IPTG; 1 mM) agar plates. Kanamycin selects for clones that successfully obtained the P_{lac}-GFP reporter while IPTG induces expression of the reporter allowing identification of clones that stably incorporated the GFP reporter. After 24-36 hours of growth, six fluorescent colonies were restreaked onto LB + Km (60 μg/mL) plates. Restreaked colonies were streaked a second time on LB + Km + IPTG plates and then tested for ampicillin (Ap) sensitivity by spotting colonies onto LB + Ap (100 μg/mL) agar. Ampicillin sensitivity ensures that the pUC18R6KT vector was lost from the cell. Insertion of the P*lac*-GFP reporter into the *att*Tn7 site was confirmed by PCR (primers: 5'-TAACAGCCAGCACCACGCCG-3' (120-F) and 5'-CGCGAATCCGATCTGGCGCT-3' (121-R)). Transposition of the reporter into each recipient strain's *att*Tn*7* site allows consistent insertion of the reporter into the same region of the genome minimizing divergent effects on reporter fitness and expression costs (Choi et al., 2005).

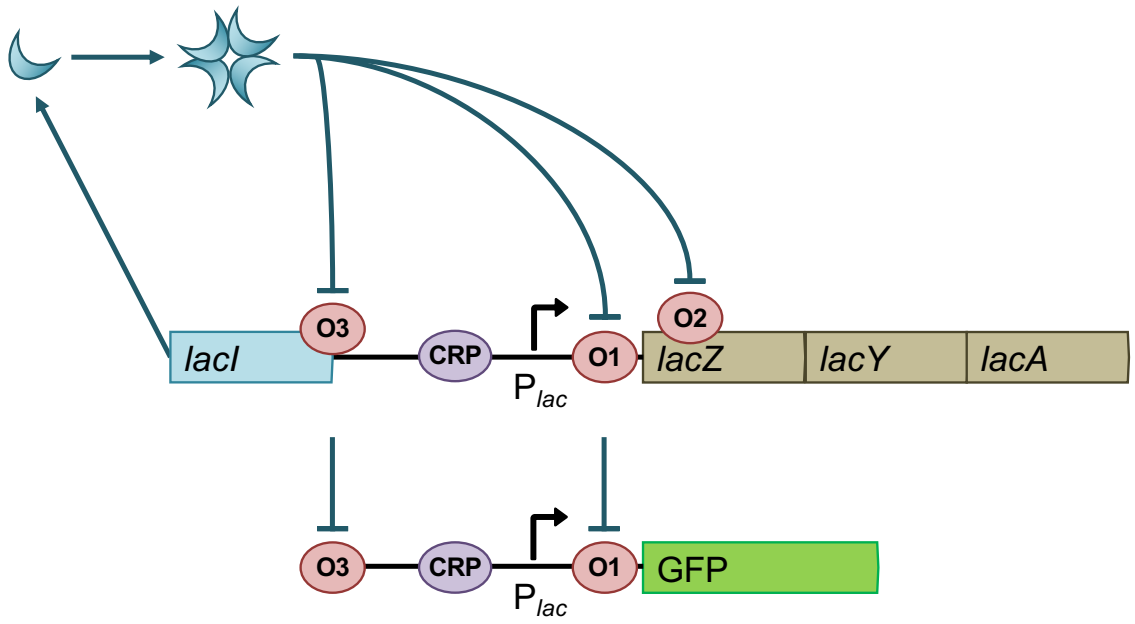


Figure 3.3 GFP reporter and *lac* operon schematic. The top schematic shows *lacI* and the *lac* operon with the approximate locations of the structural genes (*lacZ*, *lacY*, and *lacA*), binding sites of CRP and *lacI* (O1, O2, and O3), and the *lac* operon promoter region (Lewis et al., 1996). LacI forms a tetramer that binds to the operators. The bottom schematic shows the PlacGFP reporter, which includes the promoter region of the *lac* operon including the *lacI* operator binding sites O1 and O3, and CRP and ribosomal binding sites (Quan *et al*., 2012).

To measure expression, reporter strains were grown up in LB broth from freezer stocks. The overnight cultures were diluted 1:10,000 in the assay environment and allowed to grow for 24 hours. The following day cultures were diluted again 1:10,000 in the assay environment and grown in a VersaMax spectrophotometer (Molecular Dynamics, CA) until mid-log phase (OD_{450} ~0.1-0.15) to allow measurement of a steady-state level of P_{lac}-GFP reporter expression. P_{lac}-GFP expression was measured in an Accuri C6 flow cytometer (Becton Dickinson, NJ). Assay environments included DM+175 μg/mL glucose, DM+210 μg/mL lactose, and DM 4000 μg/mL glycerol + 100 μM TMG as noted in the text.

3.2.4 Statistical analyses

Data was analyzed using R (version 3.5.0) (R Core Team, 2018). Non-parametric tests were used to analyze data having a non-normal distribution or unequal variances, as determined using Shapiro-Wilk and Levene's tests, respectively. I used a Kruskal-Wallis rank sum test to determine if the evolution environment significantly influenced selection of compensatory mutations. I used a Fisher's exact test to determine if the presence of both glucose and lactose versus lactose alone influenced the presence of compensatory mutations.

3.3 Results

3.3.1 Evolution of compensatory mutations

In the ancestor to our evolution experiment, mutations that inactivate LacI confer a fitness cost of approximately 4% during growth in glucose (Quan et al., 2012). To determine whether this cost is compensated during evolution in environments containing both glucose and lactose resources, and, if so, whether this compensation differed depending on the presentation of the two resources, I reverted evolved *lacI*- mutations in a total of nine clones isolated from populations selected in lactose alone (three populations), long-term switching of glucose and lactose (one population), and daily switching of glucose and lactose (five populations). I ran fitness competitions in glucose to test the fitness effect of the *lacI*- mutation in the ancestor and in evolved backgrounds. Clones with compensatory mutations that alleviate the cost of *lacI*- in glucose will have a higher fitness than the *lacI*- ancestor (Fig. 3.1). I found that the cost of *lacI*- in glucose was significantly changed only in evolved clone G/L4, isolated from the daily fluctuating glucose and lactose environment (Fig. 3.4; Dunnett's Test, $P < 0.05$). In this clone, the *lacI*- mutation became beneficial.

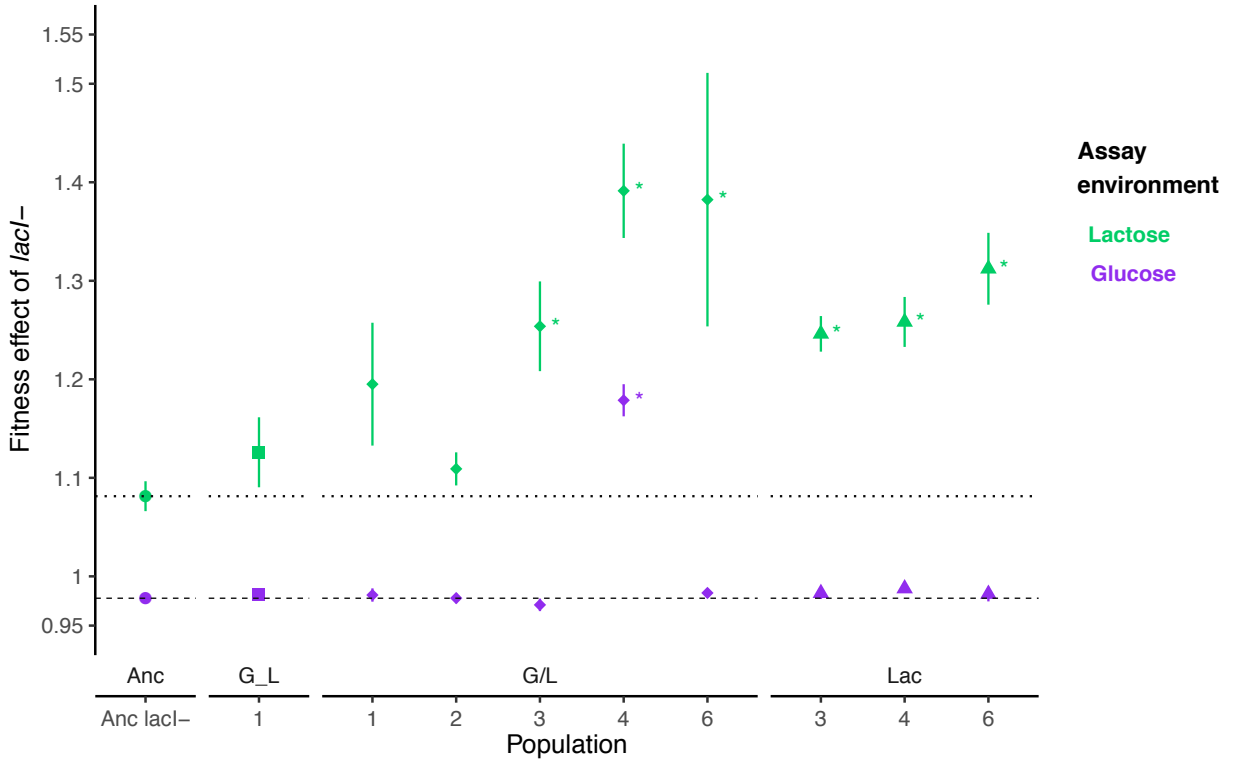


Figure 3.4 Pleiotropic effects of *lacI-*. This figure shows the fitness effect of *lacI-* in the ancestor and evolved genetic backgrounds measured in minimal glucose (purple) and lactose (green) environments. The asterisk indicates a significantly different *lacI-* fitness effect when compared to the effect in the ancestor (Dunnett's test, $P < 0.05$). Symbols and error bars indicate mean and standard error of at least 6 or 3 replicate fitness estimates for glucose and lactose competitions, respectively. Note that error bars are smaller than symbols for many of the glucose fitness estimates. The lines represent the fitness effect of *lacI-* in the ancestor in glucose (dashed) and lactose (dotted). Evolution environments are shown directly above clone names which denotes the population a clone was selected ("Anc" represents the ancestor). In the glucose environment, one clone had a significantly higher fitness than the ancestor indicating that compensatory mutations are alleviating the cost of *lacI-* in glucose.

3.3.2 Effect of environment on compensatory mutations

Although the cost of *lacI*- was significantly changed in only one clone when compared to the ancestor, there may be significant changes in costs seen when grouping clones based on their evolution environment. To test this, I used a Kruskal-Wallis rank sum test to compare the relative fitness of each group (strains evolved in Lac, G/L, or G_L). I expect clones evolved in the presence of fluctuations of glucose and lactose (i.e., G/L and G_L selection environments) to have a reduced fitness cost of the *lacI*- mutation in glucose because compensation for the original cost would provide an advantage. There was no significant difference among evolution environments (Kruskal-Wallis $\chi^2 = 2$, df = 2, $P = 0.37$), indicating that compensation was not more likely to occur in a specific evolution environment. Since a particular evolution environment doesn't seem to promote selection of compensatory mutations, I tested if clones exposed to both glucose and lactose at some time during their selective history were more likely to evolve compensatory mutations than clones that had only experienced selection in lactose. I was not able to detect a significant signal of compensation being dependent on exposure to both glucose and lactose (Fisher's exact test, $P = 0.67$).

3.3.3 Pleiotropic effect of *lacI*- compensation

Given that compensation for the cost of constitutive *lac* expression evidently can occur, one explanation for the low frequency at which it does occur is that it imposes a correlated cost in lactose. For example, it might be that compensation to constitutive expression of the *lac* operon involves a reduction in the maximum level of *lac* expression, perhaps reducing fitness in lactose and thereby causing compensation to be selectively disfavored. To test this possibility, I examined the fitness effect of *lacI*- mutations across glucose and lactose environments. To determine if there was any trend of a lower cost of *lacI*- in glucose corresponding to a lower benefit in lactose, I examined the relationship between the fitness effect of *lacI*- mutations in glucose and lactose. I found that there was, in fact, a marginally significant positive correlation between fitness in the two environments, indicating that a low cost of the *lacI*- mutation in glucose was, if anything, associated with an increased benefit in lactose (Fig. 3.5A and 3.5B; Spearman's rank correlation, rho = 0.72, $P = 0.04$). The positive relationship remained, although it was no longer significant, even when I considered only those clones that had experienced selection in both glucose and lactose (Fig. 3.5C and 3.5D; Spearman's rank correlation, rho = 0.6, $P = 0.24$). The significance of the correlation is dependent on the one evolved clone (G/L4) that had compensated for the *lacI*- mutations cost in glucose, which is an extreme outlier when compared to the other clones (fitness

in glucose: Bonferroni outlier test, P < 0.05 with Bonferroni correction for multiple comparisons). When G/L4 is removed the correlation is no longer significant, though was still positive, regardless of whether or not the lactose only evolved clones are included (with Lac clones: rho = 0.6, P = 0.13; without Lac clones: rho = 0.3, P = 0.68). Focusing on the G/L4 clone in which *lacI*- effects in glucose appear to have been compensated revealed that the *lacI*- mutation is beneficial in glucose and its effect in lactose is significantly higher than in three of the evolved clones tested, as well as the ancestor (G_L1, G/L1, and G/L2; Dunnett's test, $P < 0.05$). These results indicate that there is no universal constraint on the ability of populations to compensate for the fitness cost of *lacI*- mutation in glucose.

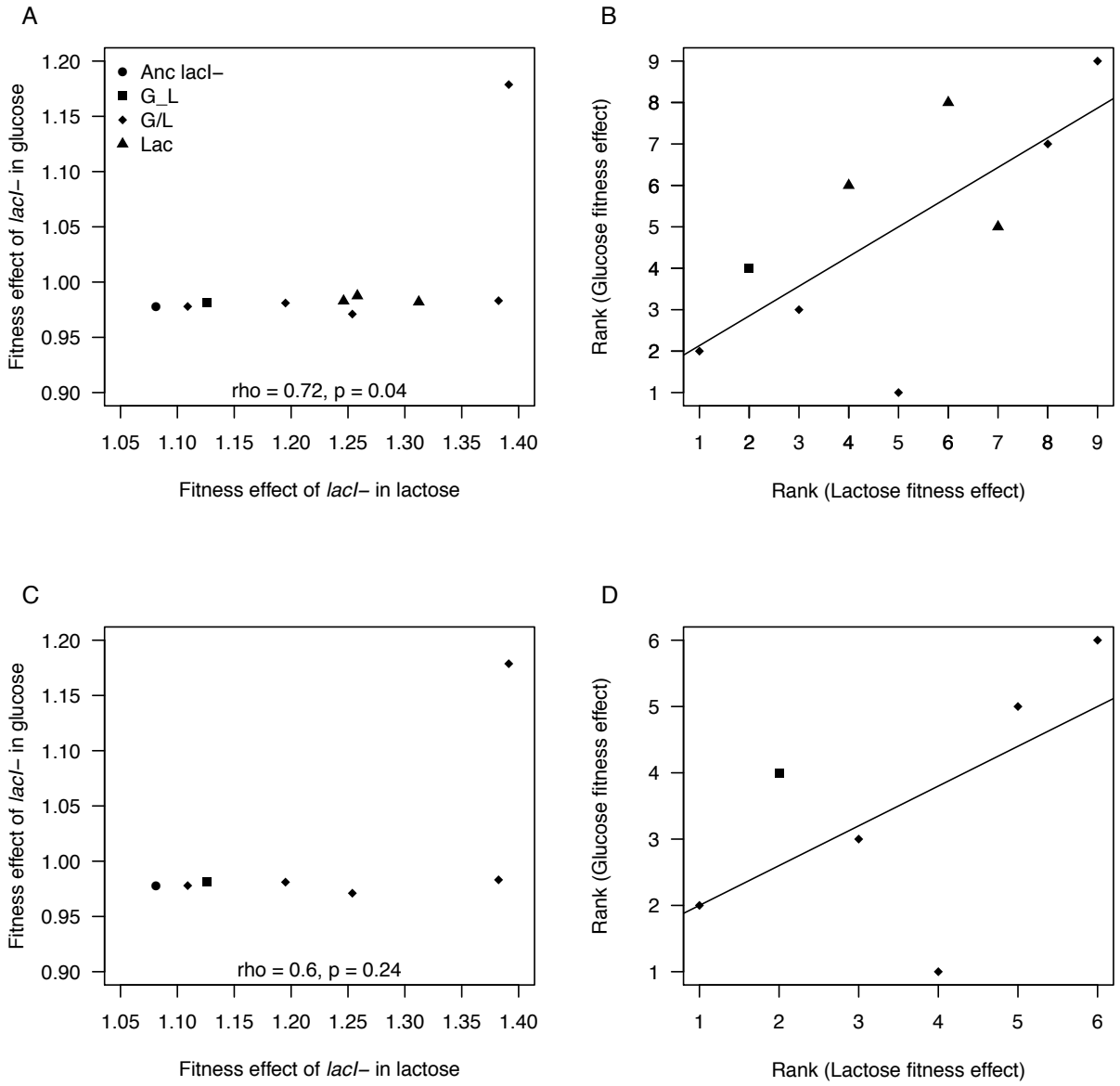


Figure 3.5 Reduced cost of *lacI*- in glucose does not reduce the benefit in lactose. Symbols represent the evolution environment or genotype of each clone that the fitness effect of *lacI*- was measured in a glucose or lactose environment. The closed circle, fitness effect of *lacI*- in the ancestor, shows what we expect if there were no epistatic interactions between *lacI*- and other mutations in the evolved background. To determine if a reduced fitness cost of *lacI*- in glucose was associated with a reduced benefit in lactose I ran a correlation between the fitness effect of *lacI*- in glucose and lactose. There was a significant positive correlation between fitness in glucose and lactose when including all clones (A and B), which became insignificant when excluding clones that had only experienced selection in lactose (C and D), indicating that alleviation of *lacI*- costs in glucose is not detrimental to the benefit in lactose. The ancestor was not included in analyses but is shown for reference.

3.3.4 Mechanisms of lowered costs

To determine if *lac* operon expression, an indicator of LacZ activity, is associated with changes in the fitness effects of the *lacI*- mutation, I measured *lac* operon expression in glucose using a reporter that is controlled by the P_{lac} promoter region (Fig. 3.6; expression in lactose and TMG were included as controls). Expression levels of the *lac* operon have been shown previously to contribute to the cost of constitutive expression, so I expected a negative relationship, such that clones that had higher *lac* operon expression would have lower fitness in the glucose environment (i.e., a higher cost) (Dekel and Alon, 2005; Stoebel et al., 2008). In fact, there was no correlation (Fig. 3.7A and 3.7B; Spearman's rank correlation, rho = 0.12, $P = 0.78$). This is especially surprising because half of the evolved strains had significantly higher *lac* expression in glucose than the ancestor *lacI*- (Dunnett's test, $P < 0.05$).

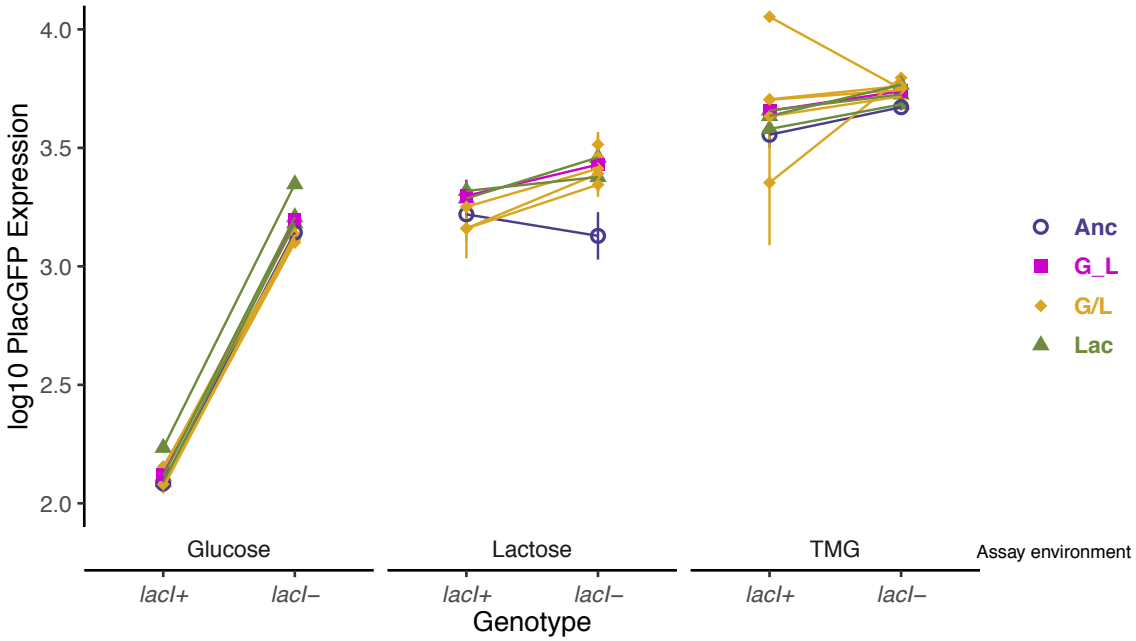


Figure 3.6 PlacGFP expression of *lacI*+ and *lacI*- clones in glucose, lactose, and TMG (maximum expression). The symbols and colors represent the genetic background and evolution environment of each clone. The x-axis shows the genotype and assay environment that PlacGFP expression was measured in. Each measurement is the average of the median expression values of independent replicates (at least 13, 2, and 4 replicates each for glucose, lactose and TMG, respectively). The error bars indicate the standard error.

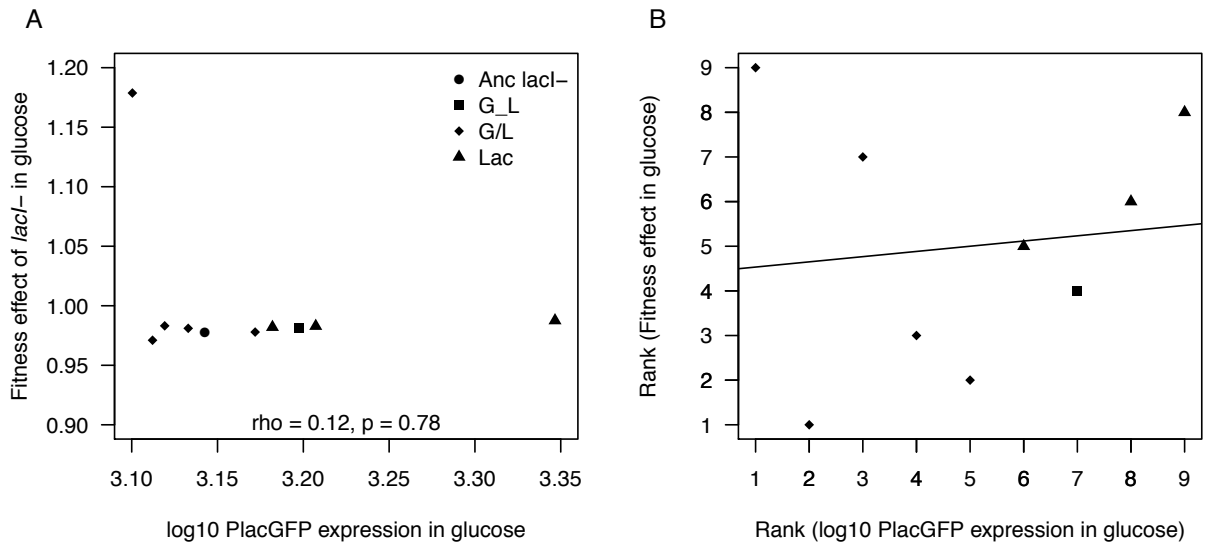


Figure 3.7 *lacI*- fitness effect is not caused by differences in *lac* operon expression. Symbols are the mean of replicates and represent the genotype or evolution environment of clones. To determine if the mechanisms of the fitness cost in glucose was due to higher *lac* operon expression I ran a correlation between the effect of *lacI*- on fitness and expression in glucose (A and B). There was not a significant correlation between *lac* operon expression and *lacI*- fitness effect in glucose, indicating that differences in fitness are not explained by the level of *lac* operon expression. The ancestor is shown for reference but was not included in analyses.

That increased *lac* expression was not associated with any fitness cost might indicate the action of compensation to some portion of the costs that would otherwise be associated with increased *lac* expression. Alternatively, there could be a limit to the costs associated with constitutive *lac* operon expression as has been found with *lacZ* expression (Eames and Kortemme, 2012), although the model most relevant to the situation prevailing in our experiments predicts exponentially increasing costs with increasing expression (Dekel and Alon, 2005). All clones except G/L4 had similar fitness costs when compared to the ancestor, but clones were divided when it came to differences in expression when compared to the *lacI*- ancestor. The clone that had the highest fitness in glucose (G/L4) also had significantly lower expression in glucose compared to all clones except two (Dunnett's test, $P < 0.05$; except for G/L3, $P = 0.87$; and G/L6, $P = 0.43$). Epistatic interactions between *lacI*- and the genetic background of evolved clones that affect expression were common in glucose. The effect of *lacI*- on expression (calculated as the change in expression caused by *lacI*-) was significantly lower for G/L2, G/L3 and G/L4 but higher for all Lac evolved clones when compared to the effect in the ancestor (Dunnett's test, $P < 0.05$). Together these results suggest that there was some differential selection of expression based on environment but the consequences are inconsistent.

3.4 Discussion

Few studies have focused on the influence of compensatory mutations during adaptation to a general environment rather than to a specific genetic perturbation. I found that compensation to the deleterious consequences of an adaptive *lacI* mutation is possible but rare among evolved clones in this experiment. Of the nine clones tested, only one clone (G/L4) evolved a mechanism to alleviate the cost of constitutive *lac* operon expression in glucose (Fig. 3.4). No differences in compensation were evident comparing groups of clones evolved in different environments, including some that were and some that were not expected to select for compensation, likely reflecting the small number of populations considered and rarity of compensation for small effect mutations.

A possible reason compensation did not occur more often is that beneficial mutations of large effect outcompete compensatory mutations that provide only small benefits. These small benefit compensatory mutations will continue to be outcompeted until mutations of large effect are exhausted. Previously, populations of the present study were shown to have higher fitness gains at early versus later time points (Satterwhite and Cooper, 2015), however, the latter gains may still be larger than the small ~2.2% cost of *lacI*- in glucose observed here. G_L populations were shown

to steadily increase in glucose fitness up to 6k generations (Satterwhite and Cooper, 2015) so small effect compensatory mutations may not be sufficiently competitive to increase in frequency. At 8k generations the G_L populations had just finished 2k generations of selection in a constant lactose environment so the need to compensate for the cost of *lacI*- in glucose was absent from 6k to 8k generations. However, G/L populations had an overall net decrease in fitness across glucose and lactose environments during 4-5k generations (Satterwhite and Cooper, 2015), which would have allowed small benefit mutations to get to higher frequencies in the population and may be why only a G/L clone was suggested to have compensatory mutations.

A caveat to the preceding interpretation for G/L populations is that fitness measurements for glucose and lactose environments were made independently (Satterwhite and Cooper, 2015), instead of the original evolution environment which is daily switching from glucose to lactose. Measuring fitness in only one environment, either glucose or lactose, could overlook a possible fitness advantage during the switch from glucose to lactose (Quan et al., 2012). To confirm declining fitness of G/L populations during 4-6k generations, fitness competitions would need to be remeasured in the original evolution environment of glucose and lactose daily fluctuations. Populations that have a decline in G/L fitness over an extended amount of time would indicate the presence of conditions that would permit small effect mutations to be compensated. If increases in

G/L fitness are observed over time, it would indicate that an advantage lies in a reduction in lag time when switching to a new carbon source. However, it may be possible that the phenomenon of compensation for small effect mutations is contingent on differential cell death during stationary phase or fitness during exponential growth in either glucose or lactose.

Another possible explanation for why compensation is rare is that multiple mutations are required to compensate the cost of *lacI*- in glucose without affecting its benefit in lactose (Poon and Chao, 2005b). Multiple compensatory mutations will take longer to fix and will be rare because successful compensation can depend on the order in which each required mutation occurs (Gong et al., 2013), the presence/absence of other mutations in the genetic background (Shah et al., 2015; Lunzer et al., 2010), and a clone's fitness at each step relative to others in the population which can subject an intermediate clone to being purged by purifying selection. Replacing *lacI*- with a functional LacI repressor in the G/L4 clone, in which compensation did occur, significantly reduces fitness in glucose and expression of the *lac* genes. In other words, this clone has somehow rewired the *lac* network to where *lac* operon expression is beneficial for growth in glucose. The basis of this rewiring is unknown, but might depend on multiple mutations as is the case, for example, of evolved citrate utilization selected in a population started from the same ancestor as the one used here (Blount et al., 2008, 2012). I note that both a clone isolated from a G/L selected

population, but not used in the work presented here, and clones from three glucose evolved populations, have also evolved in a way that caused the *lacI-* mutation to become beneficial in glucose (KP unpub. obs.; Satterwhite and Cooper, 2015). These observations argue that mechanisms underlying compensation are more likely to be constrained by selection than by the requirement for a specific mutational history.

We know that compensation is possible but the underlying mechanism of compensation is unknown. While unnecessary *lac* operon expression is likely to confer some cost (Dong et al.,1995; Stoebel et al., 2008; Scott et al., 2010), most of the cost has been associated with the mechanism of LacY (Eames and Kortemme, 2012). A possible explanation for the benefit of G/L4 constitutively expressing the *lac* operon in glucose is that at least one of the *lac* proteins has been repurposed for another task that provides a benefit in glucose. The most likely candidates involved in novel functions inside the cell are LacI and LacY, although it is possible that LacA has evolved to tag some unknown nonmetabolizable molecule for export. Perhaps the LacY permease is now importing glucose. Although the LacY permease is typically inhibited through inducer exclusion when glucose is present (Hariharan et al., 2015), it has been found that overactive LacY in *lacI-* constitutive cells can provide a fitness advantage in glucose (Dykhuizen and Dean, 1994). LacY has been proposed to differentiate between galactosides and glucosides by the position of the -OH

group on the C4 of the galactosyl moiety (Sahin-Tóth et al., 2001). When amino acids that recognize this area are mutated, they no longer have a specificity for galactosides, which can increase the affinity of non-galactoside sugars to LacY but sometimes decrease affinity to lactose (Gram and Brooker, 1992; King and Wilson, 1990). Although import of lactose is reduced, these studies indicate that LacY can potentially be mutated to interact with both lactose and glucose. A sugar similar to glucose that LacY has been shown to inefficiently import is galactose, an epimer of glucose, only differing by the conformation of the -OH group on C4 of the pyranose ring (Sahin-Tóth et al., 2000). LacY mutants have also been shown to import maltose, a disaccharide of two glucose monomers (Brooker and Wilson, 1985; King and Wilson, 1990). The ability to import maltose suggests that recognition of the galactosyl C4 -OH group in lactose is not an absolute requirement for some transport by LacY. Collectively, this indicates the potential for LacY to evolve to import glucose, however, it is unclear if it would provide a substantial benefit. Glucose transport is not likely a limiting factor at the concentration used in this experiment, because the K_m for the normal pathways in glucose uptake is well below the concentration used in this experiment (reviewed in Jahreis et al., 2008; Ferenci, 1996; and Postma et al., 1993). For this reason, the basis of the benefit of *lacI*- likely lies elsewhere but perhaps additional glucose transporters are advantageous when glucose is utilized and becomes increasingly low in the environment. A previous study found that glucose transport evolved during long-term evolution in

a minimal glucose environment through mutations in catalytic enzymes that indirectly affect glucose uptake through the phosphotransferase system (Lenski et al., 1998; Woods et al., 2006).

The 4 bp insertion frameshift mutation in *lacI*- of G/L4 results in a truncated non-functional LacI repressor that consists of 203 amino acids due to the replacement of Y204 with a premature stop codon. The ancestral LacI protein consists of 360 amino acids. The first 201 amino acid residues, which include the DNA binding domain of LacI (Platt et al., 1973), are identical to the ancestor, and since LacI has been shown to diffuse throughout the cell and non-specifically bind to DNA (Hammar et al., 2012; Garza de Leon et al., 2017), the LacI monomer could be interacting with another region in the genome. However, amino acids 202 and 203 are positively charged hydrophilic amino acids that are replaced with neutral hydrophobic residues (H202L, K203A). Genetic studies have shown that the specific amino acid replacements at these two positions still result in a functional LacI repressor when part of the whole protein (Kleina and Miller, 1990; Markiewicz et al., 1994), but when alanine is at the end of a truncated protein it results in an unstable protein that will likely be rapidly degraded (Parsell et al., 1990).

Another possible mechanism for a *lacI* mutation to provide an advantage in glucose is that an operator site for a different gene has evolved to provide a benefit in the absence of LacI+. The

LacI repressor has homology to other repressors in *E. coli* (Weickert and Adhya, 1992), and studies have shown that as few as two mutations are needed to increase the affinity of LacI to another operator (Lehming et al., 1987; Lehming et al., 1990; Salinas et al., 2005; Daber and Lewis, 2009). A potential candidate is a GalP operator binding site for the GalR repressor, which is a repressor with high similarity to LacI. The regions of the operators that are necessary for recognition by the GalR and LacI repressors are identical except at the fourth position. Mutations in GalR and its operator binding site have been previously found in only G/L clones of our evolved populations, including *galR* (28-29/1032) IS150 +2 bp in a different 8k clone selected from the same G/L4 population of the clone used in this experiment (TC unpub. obs.). GalR inhibits the GalP permease, a transporter with the ability to import glucose that is normally inhibited under low glucose concentrations (Hernández-Montalvo et al., 2003; Lu et al., 2011; Steinsiek and Bettenbrock, 2012). Although GalP has three independent repressors and multiple operator binding sites, mutations in its primary operator greatly reduce repression (El Qaidi et al., 2009). If the primary operator for the GalP transporter of G/L4 has mutations at only the fourth position in the operator half sites, it would resemble the O1 operator for LacI and may result in constitutive expression of GalP in the presence of *lacI*- but repression when *lacI*+ is added to the genetic background (Haber and Adhya, 1988). This is hypothesized because GalR has been shown to form a complex with a LacI operator variant that resembles the GalP operator due to a single point mutation (Lehming et

al., 1990). The benefit of GalP constitutive expression lies in the ability of the GalP transporter to import glucose (Hernández-Montalvo et al., 2003; Lu et al., 2011). Perhaps the glucose pathway in the G/L4 clone has evolved and alternative methods of glucose import are necessary (Steinsiek and Bettenbrock, 2012).

Finally, it is possible that the ancestral *lacI*+ allele, which was lost early in this population so that most mutations occurred in its absence, interacts epistatically with the evolved background to be deleterious so that the *lacI*- allele provides only a relative benefit (Bridgham et al., 2009; Shah et al., 2015; Starr et al., 2018). A previous study on genetic interactions throughout the genome of *E. coli* found that *lacI*+ interacts epistatically with 11 different genes that are not directly associated with lactose metabolism (Babu et al., 2014). Of these 11 genes, eight were associated with positive epistatic interactions (Babu et al., 2014), indicating the potential for negative interactions to arise if these genes evolved and no longer interacted positively with *lacI*+. To differentiate between novel functions of *lac* proteins including LacI- and the deleteriousness of LacI+, future research could perform deletion studies of each *lac* gene followed by competitions between knockout strains and G/L4 in glucose. A decrease in glucose fitness caused by knocking out any of the canonical *lac* genes suggests that expression of that particular gene is beneficial in glucose and it has a novel function. However, if knocking out each *lac* gene increases fitness relative to G/L4, it

indicates that expressing the *lac* genes is still costly but less than the effect of adding *lacI+* to this genetic background. If there is an indication of novel functions for any of the canonical *lac* genes, future research could determine differences in gene expression between G/L4 and each G/L4 *lac* knockout (lac^{KO}) that decreases fitness. A cluster analysis of gene expression data would indicate gene expression networks that change based on the presence or absence of the knockout gene. The networks of the genes that have changes in expression in the lac^{KO} strain compared to G/L4 would be candidates for processes that involve the novel protein function. Isolating which networks involve the novel protein function will facilitate identifying the novel function by reducing the number of potential interacting components.

Another more laborious method that could be used alone or in conjunction with DNA microarrays is testing for epistatic interactions between each evolved mutation (*ev*-) and evolved *lac* gene (*lac*-) in glucose. This could potentially narrow down interacting components of the novel protein function. Epistatic interactions between each *lac* gene (*lac*-) and each evolved mutation in glucose can be tested by replacing *lac*- and another evolved mutation (*ev*-) with the ancestral alleles (*lac*+ and *ev*+, respectively) in the evolved background and adding *lac*- and the same evolved mutation (*ev*-) to the ancestor background. Differences in the fitness effect of the double mutants in competitions between G/L4 vs. G/L4 *lac*+ *ev*+ and ancestor *lac*- *ev*- vs. ancestor in glucose would

indicate epistatic interactions between the two mutations and the potential for interactions that create novel functions.

Although compensation was rare for constitutive *lac* operon expression in glucose, alternative mechanisms to reduce costs of *lacI*- in glucose may be present in populations. The present study selected blue colonies on X-gal plates because it indicates that cells were likely to have constitutive *lac* expression and thus a cost in glucose that would need to be overcome. However, cells that have a *lacI*- mutation and produce white colonies on X-gal plates indicate an alternative mechanism to compensation for constitutive expression that reduces expression in the presence of glucose by reverting the phenotype. Future studies should consider analyzing a variety of phenotypes on indicator media to prevent excluding any possibilities and to determine if suppressor mutations are more likely to occur than compensation in this system, which could help explain why compensation was rare.

In summary, compensation was rare and did not occur based on specific fluctuations of glucose and lactose. In the clone that compensation of *lacI*- did occur *lac* operon expression was reduced but not more so than other strains that had a similar cost in glucose compared to the ancestor. This indicates that costs of constitutive expression were overcome by epistatic interactions with other

mutations in the evolved background, and that the reduction in cost was not due solely to a reduction in expression. Future studies should look into the potential long-term tradeoffs of compensation and what makes compensation rare.

Chapter 4

Diversity in *lac* operon regulation among diverse *Escherichia coli* isolates depends on the broader genetic background but is not explained by genetic relatedness

Phillips, K. N., Widmann, S., Lai, H.-Y., Nguyen, J., Ray, J. C. J., Balázsi, G., & Cooper, T. F. 2019. Diversity in *lac* operon regulation among diverse *Escherichia coli* isolates depends on the broader genetic background but is not explained by genetic relatedness. MBio 10(6):1–14.[a]

[a.] This chapter is published in mBio. I ran most experiments (with help from undergraduate students) and statistical analyses. I did not contribute to the experiments and analyses associated with Fig. A4.6 in the appendix.

4.1 Introduction

Gene regulatory networks allow bacteria to respond to changes in their environment by activating or repressing target genes (Lee et al., 2012). In this way, cells can exhibit phenotypes that balance the demands of expressing necessary genes while minimizing the diverse costs associated with

expression of genes that are not necessary (Kalisky et al., 2007; Dekel and Alon, 2005; Dean, 1989; Acar, 2008; Eames and Kortemme, 2012; Frumkin et al., 2017; Scott et al., 2010). Regulatory networks must respond to a diverse array of signals, for example, integrating information regarding availability of multiple resources that the organism uses with different preference (Aidelberg et al., 2014). For a particular group of co-regulated genes, the integration of these signals defines its regulatory input function. Knowledge of this function aids in the prediction of gene responses, understanding of the mechanistic basis of regulation, understanding the potential for regulation to evolve, and is likely to be helpful in the pursuit of engineering of specific responses in artificial circuits (Kashiwagi et al., 2006). Despite the importance of regulatory input functions, understanding of their variation within a species is limited. This variation is important as it reflects potential for evolutionary changes in regulatory function and might reveal differences in selection pressures affecting different sub-populations.

A good model system with which to study a regulatory input function is the *lac* operon (*lacZYA*) of *Escherichia coli* (Pardee et al., 1959; Jacob and Monod, 1961; Setty et al., 2003; Mayo et al., 2006; Kuhlman et al., 2007). This operon has been a focus of efforts to examine the effect on gene expression and regulation of transcription factor stochasticity (Choi et al., 2008), DNA topology (Kuhlman et al., 2007), transcriptional fidelity (Gordon et al., 2013), and hysteresis (Ozbudak et

al., 2004). It has also been examined to understand costs of protein expression (Dekel and Alon, 2005; Eames and Kortemme, 2012; Stoebel et al., 2008), the importance of coordinated gene expression (Ray et al., 2016), and is established as a target of selection during growth in defined environments (Dekel and Alon, 2005; Quan et al., 2012; Zhong et al., 2009). The wealth of information gained from empirical study of *lac* operon regulation has made it a focus of attempts to understand and model gene regulation, including attempts to learn how to manipulate the system to change regulatory outputs (Mayo et al., 2006; Ray et al., 2016; Perfeito et al., 2011; Savageau, 1998; Semsey, 2014).

The *lac* operon encodes three gene products. LacY is a permease that imports lactose into the cell where it is cleaved by LacZ, a β-galactosidase, into glucose and galactose. LacA is a transacetylase that is thought to facilitate export of toxic sugars that cannot be metabolized by the cell. These genes are beneficial to express in environments where lactose is the best available carbon source, being required for its import and initial catabolism, but their expression is also associated with a significant cost (Dekel and Alon, 2005; Eames and Kortemme, 2012; Quan et al., 2012). The *lac* operon is directly regulated by two environmental signals, positively by lactose and negatively by glucose, that modulate that activity of transcription factors that bind to *cis*-regulatory DNA regions. The LacI repressor, a *trans*-regulator, binds at three operator binding sites, the *cis*-regulators, in

the vicinity of the *lac* promoter and can interact to cause DNA looping, which promotes repressor binding and increases repression (Fulcrand et al., 2016; Vilar and Leibler, 2003). In the presence of allolactose (a derivative of lactose) or artificial inducers (e.g., isopropyl **β**-D-thiogalactoside (IPTG)) LacI is released from DNA, allowing transcription to occur (Jobe et al., 1972). The cAMP-CRP global regulator complex, another *trans*-acting factor, binds upstream of the *lac* operon promoter to its *cis*-regulatory region and enhances transcription by promoting the recruitment of RNA polymerase to the *lac* promoter (Zubay et al., 1970). Production of cAMP is decreased in the presence of glucose thereby decreasing availability of the cAMP-CRP complex.

The regulatory control of many genes can be described as logic functions. These functions integrate complex mechanistic details of regulatory control to describe how regulator activities combine at a *cis*-regulatory region to determine the expression of target genes (Setty et al., 2003; Mayo et al., 2006; Kaplan et al., 2008). A simple expectation is that *lac* genes will be controlled by AND type logic, whereby expression requires the presence of lactose and absence of glucose. In fact, experiments using the artificial IPTG inducer and exogenous cAMP to independently control LacI and CRP activity found that the underlying function is more complex, being intermediate between AND and OR functions (Setty et al., 2003; Mayo et al., 2006). That work, however, focused on the gene input function of a single K-12 *E. coli* strain, MG1655, and close

derivatives, which may not be representative of other strains. Though often considered a wild-type strain, MG1655 was isolated in 1922 and during subsequent propagation and storage may have been subject to inadvertent selection that affected the *lac* gene input function (Bachmann, 1996; Jensen, 1993). Even if the *lac* regulatory function has not changed, it remains unknown if different natural isolate strains demonstrate different functions.

Two factors suggest the potential for variation in a given regulatory function within a species. In the case of the *lac* operon, model and experiments have revealed many different regulatory functions can evolve through single mutations (Mayo et al., 2006; Quan et al., 2012; de Vos et al., 2013). Second, *lac* regulation can be affected by changes occurring outside its immediate regulatory network. Indeed, a previous study of *E. coli* populations evolved in environments containing lactose or combinations of lactose and glucose, evolved changes in *lac* expression that were common and, at least in part, due to mutations occurring outside the canonical regulatory network (Quan et al., 2012). Moreover, that work found that the nature of *lac* regulatory changes reflected the selection environment. For example, most populations evolving in an environment that fluctuated daily between glucose and lactose evolved to constitutively express the *lac* genes, whereas populations evolved in the simultaneous presence of glucose and lactose evolved a graded response function, allowing a continuous expression response. Similar findings of selection

dependent changes in gene regulation have been found in populations adapted to chemostat environments (Zhong et al., 2009; Dykhuizen et al., 1987) and during evolution of a stress response network (González et al., 2015), and inferred from selective benefits of naturally occurring variants controlling biosynthesis of arginine (Suiter et al., 2003). Although studies have not compared detailed *lac* logic functions of different *E. coli* strains, *lac* structural gene enzyme activity and fitness effect can vary between isolated *lac* operons (Dean, 1989; Dykhuizen et al., 1987).

To the extent that there is variation in gene regulatory functions, a key question is the relative contribution of *cis*-regulatory changes that affect expression of a specific transcriptional unit (i.e., an operon) and *trans*-regulatory changes that have the potential to affect expression of a regulon potentially containing hundreds of genes (Cooper et al., 2008; Treviño et al., 2012). This distinction is important because a few *trans*-regulatory changes may allow a large number of key expression changes to evolve relatively quickly, whereas the same expression change occurring through *cis*-regulatory change would take much longer, though perhaps with fewer pleiotropic side-effects. The distinction between *cis* and *trans* control of gene regulation is also relevant to the consequences of horizontal gene transfer. If adaptive changes in gene regulation are *cis*-regulatory they are likely to have fewer antagonistic pleiotropic consequences following transfer to alternative genetic backgrounds, allowing transfer to more genetically diverged recipients.

To examine natural variation in the *lac* regulatory input function we introduced a GFP reporter driven by the *lac* promoter and containing the primary (O1) and upstream (O3) LacI repressor, and CRP, binding sites into 21 divergent natural isolate strains and into two reference lab strains, MG1655 and REL606. We found substantial variation in regulatory functions, which we quantified by fitting a simple regulatory model to the observed expression data. Some aspects of this variation were explained by the genetic relatedness of strains, assessed using phylogenies constructed from core and accessory genes, and from only the *lac* genes. Other parameters varied, but without any phylogenetic signal, consistent with them changing on a relatively short time-scale. Transfer of a subset of *lac* operons into a common reference strain indicated that at least some of the variation is determined by *trans*-regulators encoded by the recipient strain, not the *cis*-regulatory sequences local to the *lac* genes. To the extent that regulatory functions are influenced by *trans*-regulators that have pleiotropic activity that varies between strains, adaptive changes in gene regulatory functions may be less likely to remain beneficial following horizontal transfer to new strains.

4.2 Materials and Methods

4.2.1 Bacterial strains and strain construction

Natural isolate strains used as recipients of a *lac* reporter construct were chosen from a collection of 96 strains collected and sequenced as part of a Broad institute project and obtained from the Michigan State University STEC Center, and from strains described in (Moore and Woods, 2006) (Fig. 4.1, Table 4.1). Genome sequences of strains were downloaded from the Broad Institute (http://www.broadinstitute.org/annotation/genome/escherichia_antibiotic_resistance/MultiDownloads.html) or obtained by *de novo* Illumina sequencing as described in (Wang et al., 2016). One strain, B156, was included in this work despite being classified as *E. albertii*. This species lacks a functional LacI repressor and LacY permease and is unable to grow on lactose. Throughout, we include this strain in individual strain descriptions but omit it from summary data.

Figure 4.1 Phylogeny based on the core genome shared between 96 diverse natural isolates of *E. coli*. Strains whose *lac* regulatory function was determined and whose *lacI-ZYA region* was transferred to the reference strain, REL606, are indicated by the red symbols in columns labelled 'expression' and 'transfer', respectively. The former group of strains represents a random sample of the complete phylogeny (Fig. A4.1). The *lac* regulatory function was also measured for three strains for which we do not have genome sequence and, therefore, are not included here: B156, B1167 and TA263 (Table 4.1). The *lac* operon of TA263 was also transferred to REL606. Phylogeny construction is described in Materials and Methods.

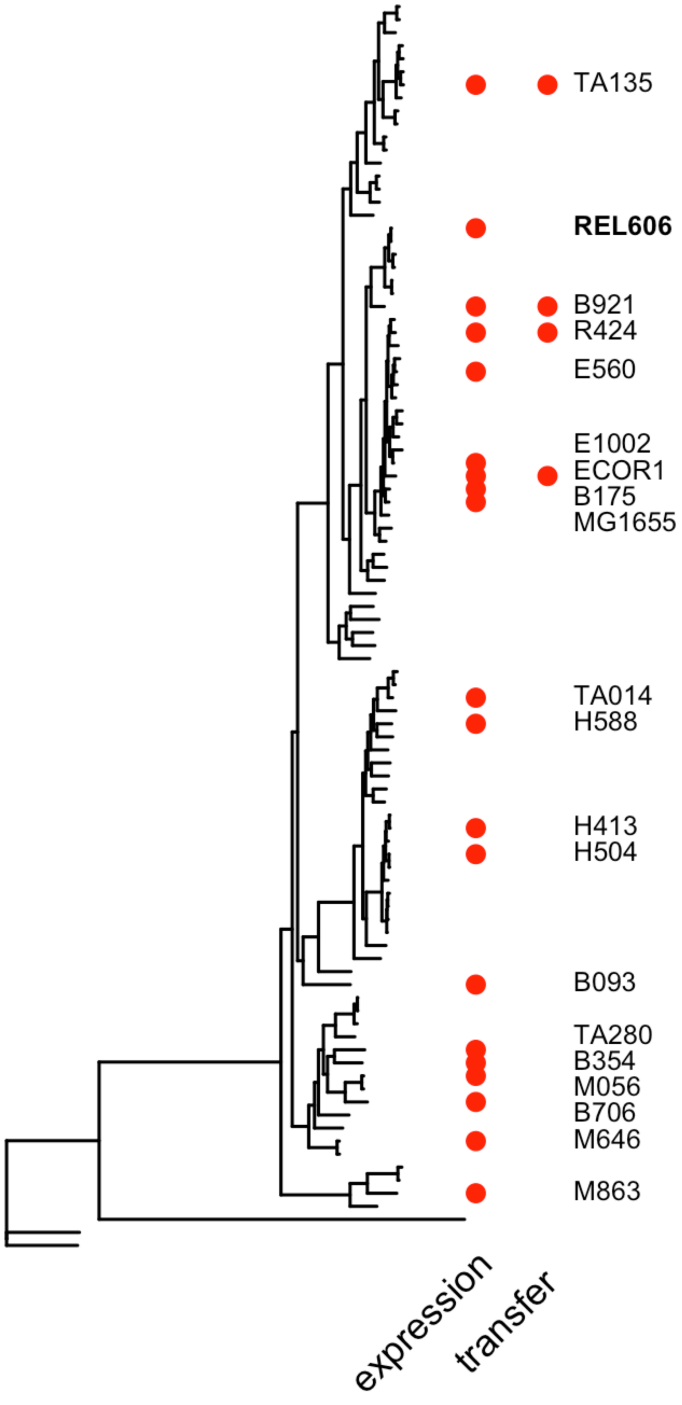

TA135
REL606
B921
R424
E560
E1002
ECOR1
B175
MG1655
TA014
H588
H413
H504
B093
TA280
B354
M056
B706
M646
M863
expression
transfer

Table 4.1 Strains used in this study.

Strain ID[a]	Alternative ID	Obtained from	Genome sequence
B093	TW15931	Broad Institute via MSU STEC center	Broad[b]
B156	TW15934	Broad Institute via MSU STEC center	
B175	TW15935	Broad Institute via MSU STEC center	Broad
B354	TW15938	Broad Institute via MSU STEC center	Broad
B706	TW15943	Broad Institute via MSU STEC center	Broad
B921	TW15945	Broad Institute via MSU STEC center	Broad
B1167	TW15933	Broad Institute via MSU STEC center	
E560	TW15955	Broad Institute via MSU STEC center	Broad
E1002	TW15946	Broad Institute via MSU STEC center	Broad
TA135	FBGM4	F.B.-G. Moore	Broad
TA263	FBGM10	F.B.-G. Moore	
ECOR1	FBGM17	F.B.-G. Moore	Wang et al. 2016
H413	TW15974		Broad
H504	TW15981	Broad Institute via MSU STEC center	Broad
H588	TW15982	Broad Institute via MSU STEC center	Broad
M056	TW15990	Broad Institute via MSU STEC center	Broad
M646	TW15993	Broad Institute via MSU STEC center	Broad
M863	TW15995	Broad Institute via MSU STEC center	Broad
MG1655		CGSC	NCBI: NZ_CP027060
R424	TW15997	Broad Institute via MSU STEC center	Broad
REL606		R. E. Lenski	Jeong et al. 2009
TA014	TW16005	Broad Institute via MSU STEC center	Broad
TA280	TW16018	Broad Institute via MSU STEC center	Broad

[a] ID from Moore and Woods (2006) or MSU STEC center documentation.

[b] Broad genome sequences downloaded from: http://www.broadinstitute.org/annotation/genome/escherichia_antibiotic_resistance/MultiDownloads.html

The lab strain REL606 was used as the recipient for transfer of *lacI-ZYA* genes from five natural isolate strains. First, we deleted the corresponding genes in REL606 and replaced them with a chloramphenicol resistance (Cm^r) gene cassette. To do this we amplified the chloramphenicol cassette from pKD3 (Datsenko and Wanner, 2000) using primers containing 5' extensions complementary to REL606 sequence on either side of the *lac*I-ZYA genes (forward primer (overlaps with region immediately downstream of *lacA* and pKD3): 5'-GCTGAACTTGTAGGCCTGATAAGCGCAGCGTATCAGGCAATTTTTATAATTGTGTAG GCTGGAGCTGCTTC; reverse primer (overlaps with region immediately downstream of *lacI* and pKD3): 5'-GCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGTGAAT CATATGAATATCCTCCTTAG). This product was used to transform REL606 containing the red recombineering plasmid pSIM5 (Datta et al., 2006) and Cm^r transformants were selected. These strains had the *lacI-ZYA* gene region replaced by the Cm^r gene. In the second step, this replacement strain containing pSIM5 was transformed with the *lacI-ZYA* region amplified from donor natural isolate strains using Phusion Hot Start polymerase (New England Biolabs, MA) (forward primer (overlaps at 39 bases downstream of *lacA*): 5'-AGGCCTGATAAGCGCAGCGT; reverse primer (overlaps at 44 bases upstream of *lacI*): 5'-TGGCATGATAGCGCCCGGAA). Transformants were selected for incorporation of the incoming DNA by plating on Davis-Mignoli (DM) minimal medium supplemented with thiamine and containing lactose as the sole carbon source. The

transformed cells contain *lacI-ZYA* and 44 bases upstream of the *lacI* gene from the donor, while maintaining the -35 promoter site of *lacI* from REL606. Sequencing of junctions between recipient and incoming DNA was performed to confirm the successful incorporation of incoming DNA into the target chromosomal site.

Expression of the *lac* operon was measured using a reporter construct controlled by the P*lac* promoter region including the O1 and O3 LacI, and the primary CRP, binding sites (Quan et al., 2012). This reporter was cloned into a mini-Tn*7* cassette in a suicide-vector that was introduced into target strains by conjugation (Choi et al., 2005). Transposition into the recipient strain *att*Tn*7* site was confirmed by PCR. Although the reporter encodes its own *cis*-regulatory sites and is present at a chromosomal location separate from the native *lac* operon, it does reflect expression of the native operon because it responds to inducer levels in the cell as a whole, which are determined by expression of the LacY permease encoded by the native operon. Previous work has shown that reporter driven GFP expression is correlated to native *lac* operon expression as judged by direct enzymatic assays (Setty et al., 2003; Quan et al., 2012).

4.2.2 Expression assays

Regulatory input functions were characterized by measuring the expression of a P*lac*-GFP reporter at different combinations of cAMP and IPTG in DM supplemented with 2000 μg/ml glucose. This environment was used because glucose inhibits production of cAMP, allowing measurement of the regulatory input function from as close to the basal level of P*lac*-GFP expression as possible. Strains containing the P*lac*-GFP reporter were pre-conditioned in DM medium supplemented with 2000 μg/mL (DM2000) glucose for 24 hours then transferred at a 1:1000 dilution to the test environments containing combinations of DM2000 supplemented with cAMP and IPTG. cAMP was added at eight concentrations (0, 0.625, 1.25, 2.5, 5, 10, 20, and 40 mM) and IPTG was added at 10 or 6 concentrations (0, 0.78, 1.56, 3.125, 6.25, 12.5, 25, 50, 100, and 200 μM; 0, 6.25, 12.5, 25, 50, and 100 μM) as noted in the text. Strains were grown in these environments for ~16 hours to an OD_{450} of ~0.1-0.2, which corresponded to mid-log growth phase as determined by tracking changes in population OD using a VersaMax spectrophotometer (Molecular Dynamics, CA). An OD of 0.1 reflects approximately six population doublings from the initial inoculum such that we assume GFP expression is at steady state and at a level dependent on promoter activity. GFP expression was measured using an Accuri C6 (Becton Dickinson, NJ) flow cytometer. The analysis pipeline was implemented in R. Expression estimates are presented as arbitrary fluorescence units

following subtraction of the fluorescence value of the corresponding strain that did not contain the P*lac*-GFP reporter. In comparisons of gene regulatory functions involving the reference strain (REL606), a natural isolate strain, and a hybrid with the natural isolate *lacI-ZYA* region replacing that of REL606, all compared strains were measured in the same experimental block.

4.2.3 Phylogeny construction

Core (shared across all recipient strains) and accessory (shared amongst a subset of strains) gene regions were identified comparing DNA sequence windows as implemented in PANSEQ (Laing et al., 2010). Core regions were defined as regions of 250 bp present in an arbitrary reference strain that were present at a match of >80% identity in all other strains. A phylogeny was built from the core genome by concatenating core regions for each strain and performing a multiple sequence alignment. Variable sites in this alignment were extracted as a SNP file. We also generated alignments based on the *lacI-ZYA* region alone. The gene-region alignment, and core and accessory genomes, were used to build phylogenies with which to test for a phylogenetic signal in regulatory parameters estimated from the different test strains. In all cases, PhyML was used to build maximum likelihood trees. For the accessory genome, a binary input file indicating the

presence/absence of each accessory gene in each strain was analyzed using default parameters of PARS in PHYLIP (Felsenstein, 1989).

4.2.4 Growth rate estimation

Strains were inoculated into LB and grown overnight at 37 °C with shaking. A 2 μl aliquot of each culture was transferred to each of three wells in a microtiter plate containing 200 μl DM200 glucose medium. Following 24 hours incubation at 37 °C with shaking, a 1:100 dilution was made into another microtiter plate containing the same medium. After a second 24-hour incubation, another 1:100 dilution was made into a microtiter plate containing DM1500 lactose and the new plate was incubated in a Versamax plate spectrophotometer. OD_{450} readings and three second shaking periods were carried out every three minutes for 24 hours. A custom script was used to fit a modified Gompertz growth function to the resulting growth data (Zwietering et al., 1990). Growth parameters for each strain were estimated as the average of estimates for individual replicates weighted by the quality of each fit. In the modified Gompertz function, the parameter best interpreted as lag time, λ, corresponds to the time taken for a population to reach its maximum growth rate.

4.2.5 Model and statistical analyses

All analyses were carried out using R (version 3.4.3) (R Core Team, 2017). Regulatory input functions were analyzed in two stages. First, the optim function was used to estimate parameters of a simple model incorporating key features of *lac* regulation that best fit observed GFP expression at each combination of cAMP-IPTG concentrations (Eqn. 1 of (Mayo et al., 2006)). A detailed outline of this model is presented in the accompanying Supplementary Information. Briefly, it includes terms that describe: CRP activity (fraction bound to cAMP) (*A*); cAMP-CRP binding cooperativity (*n*); LacI activity (fraction not bound to IPTG)(*R*); LacI-IPTG binding cooperativity (*m*); affinity to binding sites of RNA polymerase in the absence of cAMP-CRP (*a*), cAMP-CRP (*d*), and LacI (*d*); the effect of cAMP-CRP binding on RNA polymerase binding stability (η); and maximum (α) and basal (γ) expression rates. The model omits some molecular details, such as DNA looping stabilized by bound LacI tetramers, that are known to influence *lac* expression (Kuhlman et al., 2007). Nevertheless, for all strains, the fitted models captured a substantial portion of the overall expression variation (RMSE of fitted models was low relative to overall variation in expression (mean = 0.106, standard deviation = 0.055)).

Estimates of each model parameter were used to predict an idealized regulatory function that characterized the individual and combined effect of IPTG and cAMP on *lac* expression (Mayo et al., 2006). Following previous work, we used an artificial inducer, IPTG, to manipulate LacI activity. IPTG is not metabolized, allowing concentrations to be maintained through cell growth and reducing potential feedback between inducer concentration and cell growth rates (Ray et al., 2016). The resulting expression profiles will probably differ from those that would be seen if the natural lactose inducer was used. One reason for this is that IPTG can passively diffuse into cells, allowing a baseline intracellular concentration independent of the LacY permease and reducing the influence of inducer exclusion, a post-translational regulation mechanism through which glucose indirectly reduces activity of the LacY permease (Kuhlman et al., 2007; Ozbudak et al., 2004; Fernández-Castané et al., 2012). Reduced inducer exclusion has the effect of allowing LacI-mediated negative regulation and cAMP-CRP-mediated positive regulation to be controlled independently so that all combinations of their activity can be measured even when some combinations may not be accessible during growth in environments containing only natural inducers. We note that many of the analyses we present focus on *lac* expression occurring at saturating inducer concentrations, where LacY-independent uptake of IPTG is not expected to have any additional regulatory effect. Supporting this, we observed good correspondence between *lac* expression estimates using high levels of IPTG and TMG, an inducer that depends on LacY

for import (Fig. A4.6). We also find a significant correlation between expression levels during growth of strains in: (i) glycerol, an environment supporting production of high levels of endogenous cAMP, and in glucose supplemented with exogenous cAMP, and (ii) lactose and glucose supplemented with cAMP and IPTG (Fig. A4.6).

Regulatory parameter estimates were tested for an association with genetic variation in the *lacI-ZYA* region of 18 of the strains for which expression and regulatory parameter information was available (sequence of the entire *lacI-ZYA* region was not available for strains B156, B1167, TA135, TA263 and H413). Alignment of this region included 6298 bases of which 322 sites were polymorphic. The function BUS in the BUS package was used to determine the association between estimated *lac* expression parameters and polymorphism. The mutual information between these variables was determined and significance was estimated using a permutation approach to correct for testing over multiple sites (using option: method=2).

Tests for phylogenetic signal were performed using the function phylosig in the Phytools package. The functions pd.calc and pd.bootstrap in the package Caper were used to test whether the strains we used were representative of the diversity present in our larger collection of 96 strains (Fig. A4.1). To do this, we compared the distance separating the strains used here to a distribution of

distances between 1000 randomly chosen sets of the same number of strains from the 96 sequenced strains contained in our overall phylogeny. Comparisons between strain expression descriptions (model parameterization, logic phenotypes and the complete expression landscape) were performed using non-parametric Mantel tests as implemented in the Ecodist package.

4.3 Results

4.3.1 *lac* gene input function of natural isolate *E. coli* strains

We introduced a P*lac*-GFP reporter into 21 natural isolates and two lab strains of *E. coli* (Fig. 4.1 and Table 4.1). Fluorescence from this reporter was measured in combinations of IPTG and cAMP to determine the *lac* operon expression profile of each strain. These profiles exhibit substantial variation across strains (Figs. 4.2 and A4.2). We follow two approaches to quantify this variation. First, we fit a simple model to estimate regulatory parameters that explain each strain's expression profile. This model includes terms corresponding to the interaction of regulatory molecules (IPTG and cAMP) and the transcription factors they bind to (LacI and CRP, respectively), the activity of those transcription factors, and their interaction with RNA polymerase binding (see Methods and Supplementary Information for details; (Mayo et al., 2006)). Second, we use the fitted model to

infer the regulatory logic function of each response–a measure of the individual and combined effect of cAMP and IPTG inducers on expression (Fig. 4.3). For example, a requirement of both cAMP and IPTG for *lac* expression represents an AND function, whereas either individual inducer being sufficient for high expression represents an OR function.

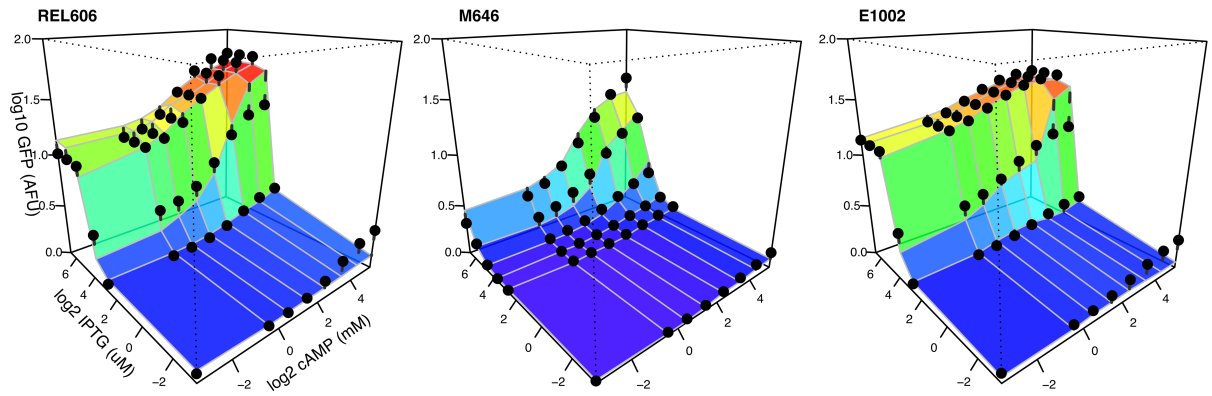


Figure 4.2 Empirical and modelled gene regulatory profiles. Expression of a *lac* reporter was determined during growth in glucose supplemented with combinations of the inducers cAMP (mM) and IPTG (μM). Expression was measured from a chromosomally integrated reporter at mid-log phase and is reported in arbitrary fluorescence units. Solid symbols indicate expression predicted at each measured inducer combination using a simple regulatory model fitted to the observed data (SI, (14)). Dashed lines connect model estimates and empirically determined expression values. The three profiles shown here are for a lab strain (REL606) and two natural isolate strains (M646 and E1002) have profiles that differ in the sufficiency of IPTG to induce *lac* expression to a high level. Additional profiles are shown in Fig. A4.2.

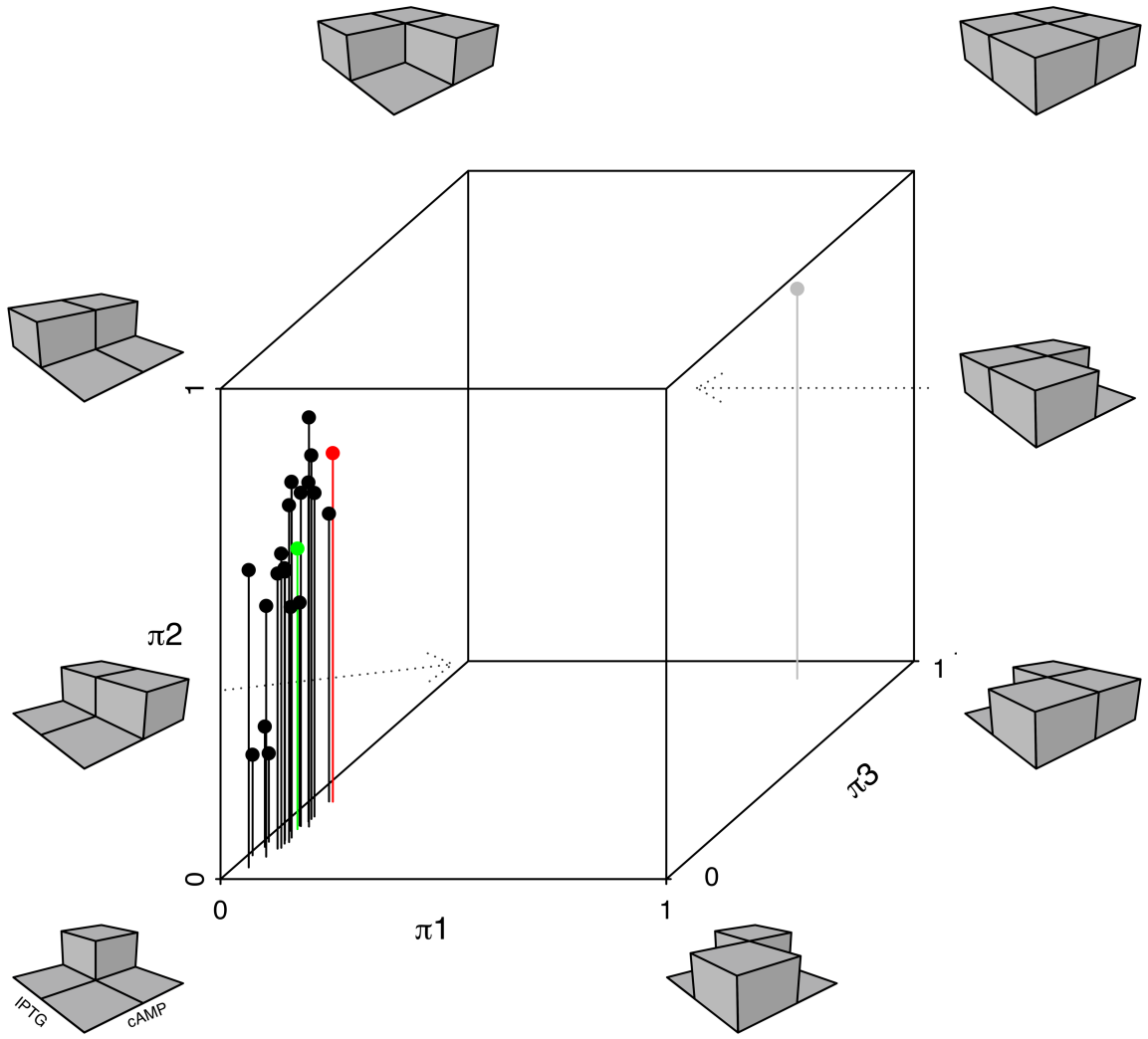


Figure 4.3 *lac* regulatory logic of *E. coli* strains. Models describing the *lac* expression phenotype were fitted for each of the tested *E. coli* stains (Fig. A4.1). Model parameters were used to determine the ratio of log expression at low IPTG–low cAMP, high IPTG–low cAMP, and low IPTG–high cAMP combinations to the high IPTG–high cAMP combination giving parameters, $\pi 1$, $\pi 2$ and $\pi 3$, respectively (Mayo et al., 2006). Combinations of these parameters describe a particular regulatory logic input function. For example, low values of $\pi 1$, $\pi 2$, and $\pi 3$, indicate high *lac* expression only when both IPTG and cAMP levels are present, reflecting an AND type logic function. Black symbols indicate parameter estimates of natural isolate strains. Green and Red points indicate estimates of the lab strains REL606 and MG1655, respectively. The grey point indicates an *E. albertii* strain, B156, that does not encode several components of the canonical lactose utilization system, including a LacI repressor, and therefore expresses the reporter at high levels regardless of IPTG (see also Fig. A4.2).

We find considerable variation in both the fitted model parameters and in the logic function characterizing *lac* operon regulation in the different strains. Considering first the regulatory logic phenotype, we find that, by itself, IPTG causes between 18% and 83% (mean 57 ± 17% (SD)) of maximum *lac* expression. By comparison, by itself, cAMP causes between 5% and 36% (mean 21% ±8% (SD)) of maximum *lac* expression. Synergy between cAMP and IPTG was estimated as the difference in maximum expression observed when both are present to that expected based on the product of their individual effects. By this measure, strains depended on the combination of inducers for between -1 and 70% of maximum expression (mean 28% ± 21% (SD)). Together these results indicate a range of regulatory logic phenotypes where some strains depend strongly on both inducers (AND type logic) and others depends largely on the activity of the LacI repressor. We note that while logic phenotypes omit potentially important aspects of the overall expression phenotype, they nevertheless capture similar relationships among strains as do the overall expression profiles that are described below (one-tailed Mantel test: $r = 0.41$, $P = 0.005$). Comparing logic functions and the parameterization of the regulatory model fitted to the underlying expression profiles, we find differences in how these descriptions of expression cluster strains (Table A4.1). Logic and model characterizations were only moderately well correlated, consistent with a mapping whereby the same logic function can be realized by different underlying regulatory parameters (one-tailed Mantel test: $r = 0.22$, $P = 0.11$).

4.3.2 Comparison of gene input functions to evolutionary distance

It is of interest to examine whether differences in *lac* regulation have been selected or whether they represent effectively neutral variation. The ideal test would be to examine differences in fitness consequences of different *lac* regulation phenotypes in ecologically relevant environments. In practice, however, what constitutes such an environment is not known. Moreover, the effect of *lac* regulation on fitness will be confounded by comparisons across different genetic backgrounds. We therefore follow two complementary approaches to assess the potential for regulatory parameters to have been selected. First, we test the expectation that, if regulatory variation is neutral, differences in estimated parameters will correspond to the underlying strain phylogeny (Fay and Wittkopp, 2008). Selected differences may correspond to the phylogeny, but are more likely to be driven by different ecological pressures relevant to each strain (Le Gall et al., 2005; Whitehead and Crawford, 2006). We have previously found that ecological performance of a subset of strains considered here was not correlated with their phylogenetic relationships, indicating that underlying ecological selection is likely to vary independent of phylogenetic relationships (Wang et al., 2016). Second, we examine the effect of *lac* operon regulatory parameters on growth following their transfer to an environment where *lac* expression is likely to be influential in determining dynamics.

To test for phylogenetic signal present in regulatory logic and model parameters, we assessed variation in those parameters in the context of phylogenies generated based on the core genome common to all strains, the accessory genome comprising genes present in some but not all strains, and a phylogeny based on the *lacI-ZYA* genes. Phylogenetic signal was assessed using Pagel's λ, which tests for signal against the null hypothesis of a trait evolving independently of an underlying phylogeny, as would be the case if it varied either neutrally or due to selection pressures that were not correlated to genetic relatedness (Pagel, 1999). In most cases, the pattern of regulatory parameter variation was not consistent with any of the tested phylogenies. There were two exceptions to this trend: the η parameter—corresponding to the effect of cAMP-CRP on binding of RNA polymerase to the *lac* promoter—exhibited phylogenetic signal over all phylogenies considered, and the *m* parameter—corresponding to the extent of cooperativity of IPTG affecting LacI activity—followed the *lacI-ZYA* phylogeny (Table 4.2 and Fig. A4.3). Consistent with an overall lack of phylogenetic signal in *lac* regulatory parameters, Mantel tests examining the relationship between pair-wise strain distance based on genetic relatedness and expression landscapes, did not find significant associations (core: r = -0.107, P = 0.37; accessory: r = -0.12, P = 0.34; *lacI-ZYA*: r = -0.15, P = 0.27).

Table 4.2 Pagel's λ test of phylogenetic signal in parameter variation

	Phylogeny		
Parameter	Core	Accessory	*lacI-ZYA*
a			0.162
η	0.024[a]	0.012	0.004
c			
d			
α			
γ			
n			
m			<0.001
Km_{cAMP}			
Km_{IPTG}		0.055	
$\pi 1$		0.733	
$\pi 2$			
$\pi 3$			

[a]P-values <0.05 are interpreted as a significant deviation from the null model that parameter values are distributed randomly over the phylogeny. Note: For clarity, P-values estimated as 1.0 are omitted.

4.3.3 Relationship between regulatory parameters and growth

The ideal experiment to test for ecologically meaningful effects of among-strain regulatory differences would be to compare strains that are otherwise identical and determine the fitness consequences of focal regulation phenotypes in a lab, or even natural, environment. The strains we examined are, however, evolutionarily and ecologically divergent and are likely to have growth differences independent of *lac* regulation. Nevertheless, in environments where the effect of *lac* regulation differences are substantial relative to effects of broader background differences, we might see a relationship between these *lac* regulation parameters and a growth phenotype. We chose to focus on the phenotype of lag phase following a transition from growth in glucose to lactose because this transition is likely to depend on the regulatory induction of the *lac* genes, which is a process dependent on the parameters we have measured (Fig. 4.4A).

To assess the relationship between *lac* regulatory parameters and lag time we used partial least squares regression, an approach suited to the analysis of relationships involving a large number of correlated parameters with relatively few data points. This approach was applied to the 23 strains described above as well as five hybrid strains having *lac* genes moved from natural isolate strains into REL606 (Materials and Methods). We found that the largest component of the regression

explained 45% of the variation in lag time. The largest contributors to this component were the activity of RNA polymerase in the absence of cAMP-CRP (a) and its relative stabilization in the presence of cAMP-CRP (η), which together accounted for 59% of the component (Fig. 4.4B). The proportion of lag time variation explained in this analysis was meaningful by comparison to a set of 1000 permutations in which assignment of estimated lag times to strains was randomized (Fig. A4.4). Moreover, both the a and η parameters were individually significantly correlated with lag times (Pearson correlation, a: r = -0.59, P = 0.002; η: r = 0.54, P = 0.005). Expression landscapes illustrating the regulatory influence of observed variation in these parameters (and α, which determines the maximum expression) are shown in Fig. 4.4C & D.

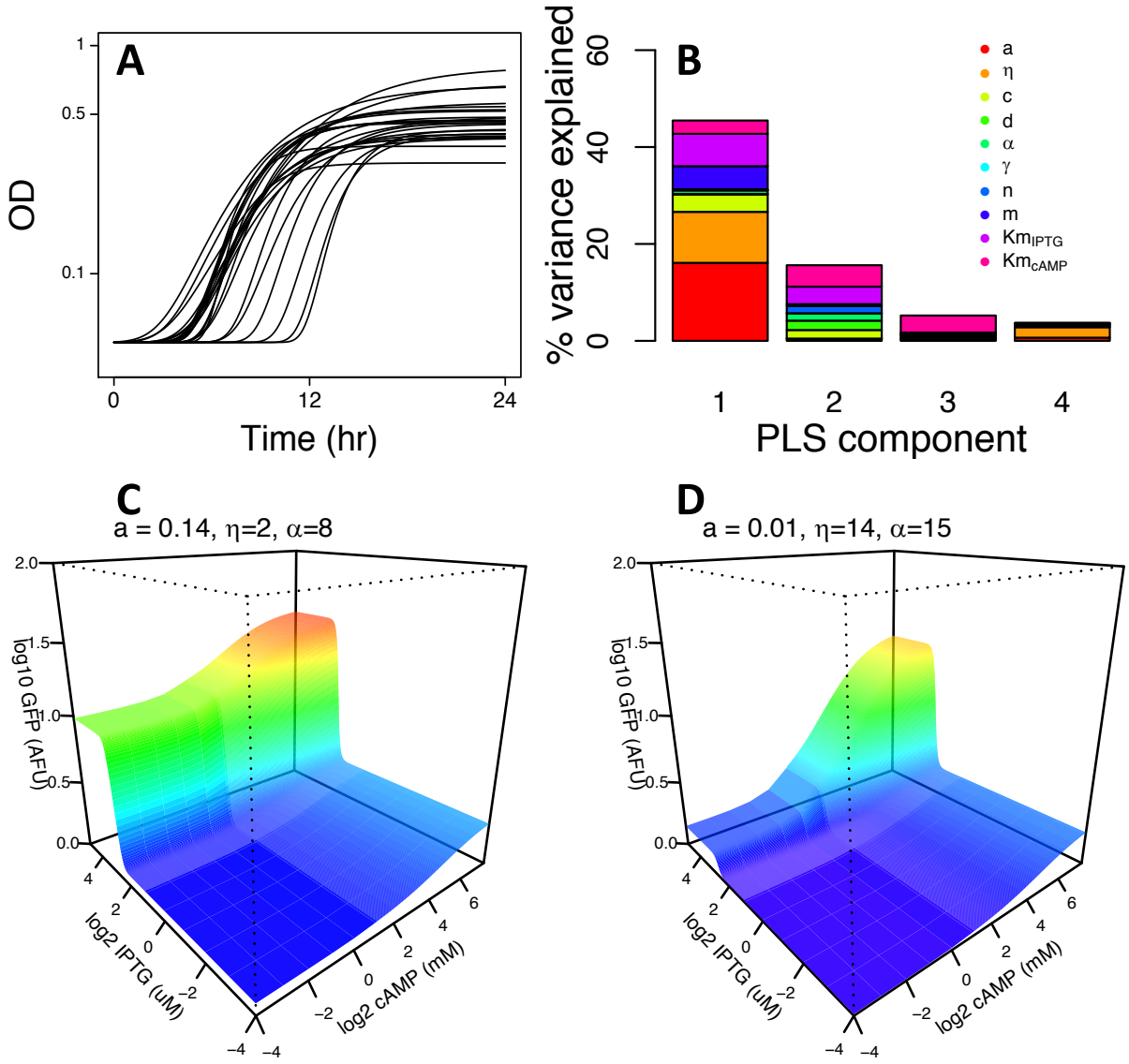


Figure 4.4 Relationship between regulatory parameters and lag time following transition from a glucose to lactose supplemented growth environment. A. Gompertz fits to growth data of natural isolate strains and hybrid strains containing the *lacI-ZYA* region from a natural isolate strain replacing the same region in the REL606 genetic background. Growth is in lactose following a transition from a day of growth in glucose. **B.** Partial least square regression indicating contribution of regulatory model parameters to the largest four components explaining variation in lag time. **C** and **D.** Changes in expression landscapes dependent on changing the two parameters, *a* and η, that explain the most lag time variation. Parameters are changed between the extremes of their estimated ranges and preserving their negative correlation. Other parameters are as for REL606, except that α is increased in panel D so that maximum expression level is comparable. C and D correspond to landscapes associated with short and long lag times, respectively.

4.3.4 Mutual information between regulatory function and genetic polymorphism

We next sought to identify variable sites in known regulatory regions that are correlated with variation in estimated regulatory parameters. We used a measure of mutual information to assess the association between 322 variable sites throughout the *lacI-ZYA* region with variation in estimated regulatory parameters (Fig. 4.5). This analysis identified a large number of sites associated with regulatory parameters, though significance levels were both generally low and similar across sites for a given parameter, a signature of linkage between genetic variants that are and are not driving regulatory variation. Together, these results suggest that regulatory variation is driven by some combination of a complex genotype-phenotype mapping (e.g., multiple genetic variants may cause similar phenotypic effects or phenotypes are due to the combined effect of multiple variable sites) and by variable regions outside the one considered here.

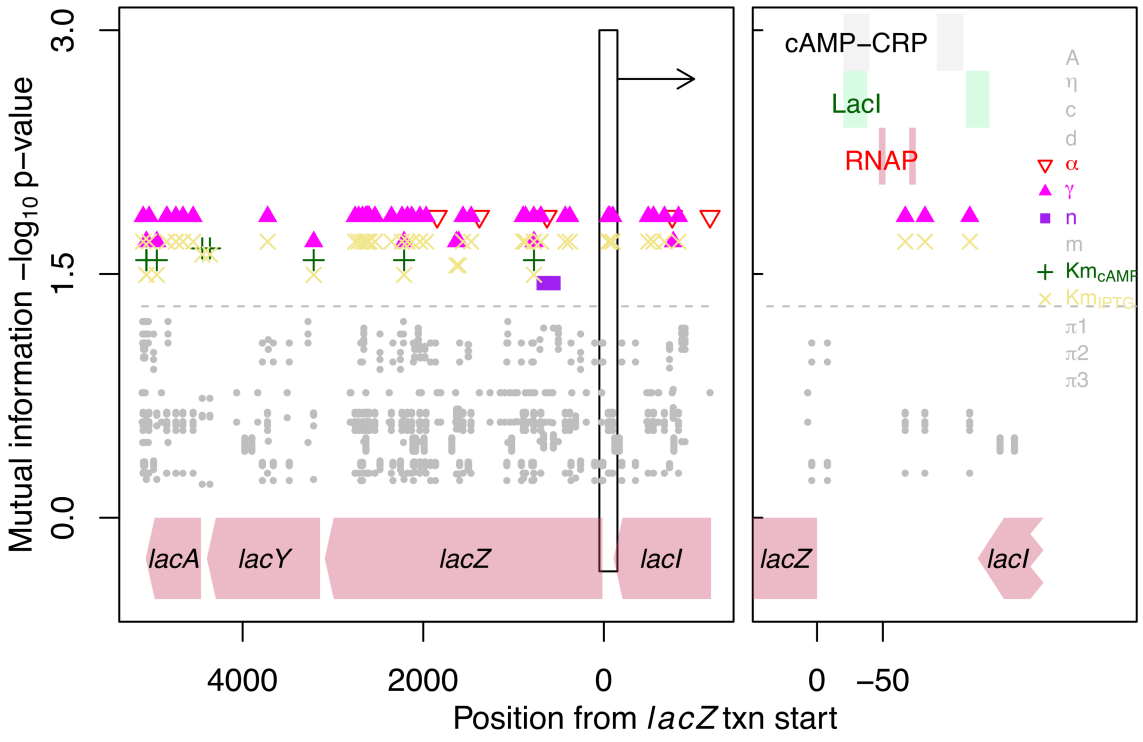


Figure 4.5 Association between polymorphism in the *lacI-ZYA* region and variation in estimated regulatory parameters. For each parameter, mutual information was estimated between estimates and genetic variation at each site in the genetic region (see M&M for details). All polymorphic sites are plotted. The dashed line indicates a significance cutoff at P = 0.05. Grey symbols indicate parameter-polymorphism associations below this cutoff, colored symbols indicate associations above this cutoff. Only parameters with at least one significant association are colored in the legend. These significant associations primarily affect basal *lac* expression (γ) and the dissociation constant of IPTG from LacI (Km_{IPTG}). The left panel presents the entire region considered. The right panel provides higher resolution around the key regulatory area between *lacI-Z* indicated by the box in the left panel. Transcription factor binding sites and the promoter region are indicated in the right panel (binding site information from Regulon DB).

4.3.5 Dependence on genetic background of gene input function

To characterize the dependence of *lac* operon regulation on its broader genetic background, we assessed regulation of different *lac* operons in their native and in a common genetic background. We replaced the *lacI-ZYA* region of REL606 with the corresponding region of five natural isolate strains and determined *lac* expression profiles (Figs. 4.6 & A4.5). In general, there was relatively little divergence in profiles, but there were examples of the hybrid strain having a *lac* expression more similar to the strain comprising the broader genetic background (i.e., REL606)—such as in the cross between REL606 and B921—indicating that regulatory elements outside the immediate *lacI-ZYA* region are important in determining its regulation. We also saw examples of the *lac* regulation in the hybrid being more similar to that of the donor strain—e.g., the cross between REL606 and FBGM17—indicating dominance of local *cis*-regulatory sequences.

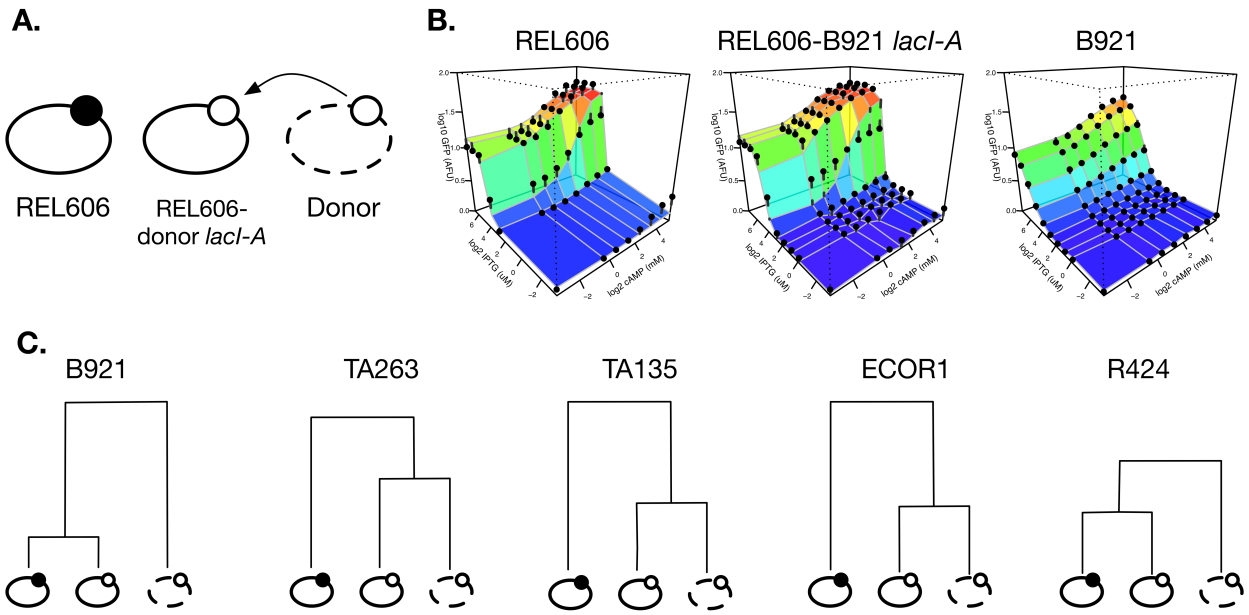


Figure 4.6 Effect of genetic background on *lac* expression. A. Schematic of expression comparisons. The *lac* expression profile was obtained from a common reference strain (REL606), different donor natural isolate strains, and a hybrid constructed by swapping the donor strain's *lacI-ZYA* genes into REL606 (details in Material and Methods). **B.** Example expression profiles of one comparison set. In this case, the hybrid strain has an expression profile more similar to that of the recipient background strain (REL606) than of the donor (B921). **C.** Dendrograms clustering for each of five donors the set of three strains based on Euclidean distances among modelled expression landscapes. The height of dendrograms is scaled to the distance between strains. Expression profiles for each strain are presented in Fig. A4.5.

4.4 Discussion

We characterized and compared *lac* regulation of 23 diverse *E. coli* strains (Fig. 4.1). We found substantial variation between strains, especially in the degree to which IPTG was individually able to regulate expression to a maximum level (Fig. 4.3). This variation is consistent with findings of regulatory models that predict that small genetic changes can have large regulatory consequences, revealing that a substantial portion of this potential is realized among natural isolate strains (Mayo et al., 2006). Such regulatory changes can evolve quickly and have ecological consequences (Behringer et al., 2018). Regulatory variation was not well explained by the genetic relatedness of strains, consistent with it being selected, rather than evolving neutrally. We also found that a significant part of regulatory variation is likely to depend on factors determined by the genetic background in which the *lac* genes are expressed, as well as on the identity of those genes themselves.

The most variable of the regulatory parameters we examined was the ratio of expression induced by IPTG alone to maximum *lac* expression induced by the presence of both cAMP and IPTG ($\pi 2$ in Fig. 4.3). This parameter describes the extent to which *lac* expression depends on the LacI repressor, with less influence by cAMP. One consequence is the possibility that strains where *lac*

expression depends less on cAMP, and by extension, the absence of glucose in the environment, might weaken the hierarchy of resource that is determined by the concentrations of preferred resources below which cells switch to catabolism of other alternative resources. Concentration of cAMP has been shown to be critical for determining these concentration crossover points (Aidelberg et al., 2014). Resources catabolized by genes that were less dependent on cAMP for expression were used preferentially to resources that depended on higher cAMP for their utilization. A previous study of cell evolved in a mix of glucose and lactose sugars found evolved changes in *lac* operon regulation that caused cells to become more sensitive to inducer, consistent with relaxation of the resource use hierarchy (Quan et al., 2012).

Diversity of *lac* regulation indicates the likelihood of a diversity of regulatory responses to different natural environmental conditions, consistent with previous work finding that different *lac* structural genes can confer different growth responses (Dean, 1989). It is clearly of interest to determine exactly what ecological consequences the different regulatory profiles might have, especially given that profiles were determined using artificial conditions. In practice, however, this is difficult to do because the strains we examine differ in ways other than in their regulation of the *lac* operon, so it is not possible to isolate the influence of *lac* regulatory differences to strain fitness across particular environments. This issue is controlled for among strains we constructed in which

different *lac* genes were transferred to a common background, but here regulation often differed from that in the donor strains, so that differences in effects cannot be easily interpreted with respect to their donor context. Despite the confounding effect of different backgrounds, we still found a significant relationship between some *lac* regulatory parameters and the transition of our strains from growth in glucose to growth in lactose. This result underlines the potential ecological relevance of the regulatory differences we see.

Several studies have identified natural genetic variation underlying ecologically relevant differences in regulation of focal genes (Dean, 1989; Suiter et al., 2003; Osborne et al., 2009). We found limited indication of an association between genetic polymorphism in the *lacI-ZYA* region and variation in regulatory parameters. Although it is not possible to determine, which, if any, of the SNPs we considered might be driving regulatory variation, we note that there were clusters of significant associations between polymorphisms at the end of the *lacZ* and *lacA* genes with the basal level of *lac* reporter expression (determined by γ). There are several possible sources of regulatory variation within these regions. In the end of the *lacZ* gene, there are sites that are responsible for substrate binding, and variation in these sites can affect LacZ catalytic activity and allolactose production (Huber et al., 2003; Juers et al., 2003; Sutendra et al., 2007). In *lacA*, associations occur in the stem-loop transcription terminator and in the preceding AT rich region,

suggesting they might affect transcriptional termination and thereby influence levels of *lacZYA* transcripts. We note as well that the *lac* expression parameters we identify using IPTG and exogenous cAMP may not be realizable in natural environments, for example because inducer exclusion causes lactose uptake to be more dependent than IPTG uptake on the absence of glucose. To the extent this is true, some features of the underlying regulatory network are not expected to be directly accessible for selection. In general, however, we interpret the lack of clear association between polymorphisms and regulatory variation as indicating that most regulatory variation is complex, having a different genetic basis in different strains as well as likely involving the action of several sites, including genes outside the canonical regulatory network.

A substantial portion of the regulatory variation we considered was not explained by patterns of relatedness determined on the basis of core or accessory genomes, or of the genes involved in *lac* utilization. Discordant patterns of phenotypic and genetic evolution are consistent either with regulatory parameters varying neutrally at high rates or being selected in a pattern distinct from that determined by the genetic relatedness of strains. A previous study found that ecological performance of strains from the same collection used here was not correlated with core or accessory phylogenies, consistent with the possibility that selection might be important (Wang et

al., 2016). This possibility is supported by our finding that regulatory parameters correlated with growth dynamics in at least one environment.

Our finding that the broader genetic background can have substantial influence on the regulation of transferred *lac* operons highlights the importance of non-canonical regulation in determining expression of *lac* genes. An example of such regulation is the influence of DNA supercoiling on accessibility of regulatory proteins to the *lac* promoter (Fulcrand et al., 2016). We note that an influence of the broader background on gene regulation complicates goals of rational design of regulatory networks, potentially putting a premium on strategies that increase robustness. A strong dependence on genetic background might also lead to greater variation in regulation between strains, increasing the chance that an effective regulatory strategy can be found in changing environments, but also making it less likely that regulation will be successful following horizontal transfer of the *lac* genes to other recipient strains.

In summary, we found that diverse strains of *E. coli* have different *lac* regulatory profiles, most of which was realized as difference in the form of the regulatory function and of the relative influence of the regulators, cAMP and IPTG, on expression. This variation reveals a wealth of raw material on which selection can act to optimize gene regulation to new environmental challenges. It also

poses a challenge to relevant models to be able to explain this diversity of regulation, with some of it coming from outside the canonical regulatory network.

Appendices

A. Chapter 4 Supplementary Information

Model

Promoter activity at the *lac* operon was characterized based on the steady-state binding of the LacI and CRP regulators using the model and approach presented by (Setty et al., 2003; Mayo et al., 2006). Promoter activity was calculated as:

$$P = (\alpha - \gamma)\frac{a(1 + \eta.d.A)}{1 + a + (a.\eta + 1)d.A + c.R} + \gamma$$

where activity of the CRP activator, assessed as the fraction of CRP bound to cAMP, is A = $X^n/(1+X^n)$, in which X = [cAMP]/K_{cAMP} (i.e., cAMP concentration in units of its dissociation constant for CRP). Cooperativity of cAMP binding to CRP is given by the Hill coefficient, n. Similarly, the fraction of active LacI repressor not bound to the IPTG inducer is R = $1/(1+Y^m)$, in which Y = [IPTG]/K_{IPTG} (i.e., IPTG concentration in units of its dissociation constant for LacI).

Cooperativity of IPTG binding with LacI is given by the Hill coefficient, *m*. The binding affinity of regulators to DNA binding sites is given by parameters *a*, *c*, and *d*. $a = [\text{RNAp}]/K_p$ is the concentration of RNAp (RNA polymerase) in units of its dissociation constant to the free promoter (i.e., CRP is not bound). $c = [\text{LacI}]/K_R$ is the concentration of LacI repressor in units of its dissociation constant with its binding sites. $d = [\text{CRP}]/K_A$ is the concentration of CRP is units of its dissociation constant with its binding site in the *lac* operon *cis*-regulatory region. Stabilization of RNAp by the presence of CRP is given by the ratio of its dissociation constants in the presence and absence of CRP: $\eta = Kp/Kcp$. The terms α and γ give maximum and basal transcript rates, respectively.

B. Chapter 4 Supplementary Table

Table A4.1 Model parameterization for natural isolate and lab strains used in this study.

Strain	*a*	η	*c*	*d*	α	γ	*n*	*m*	K_{IPTG}	K_{cAMP}	$\pi 1$	$\pi 2$	$\pi 3$
B1167	0.12	1.69	14.24	3.29	9.02	0.00	2.45	9.28	6.17	9.98	0.05	0.69	0.26
B156	0.00	16.79	7.26	11.36	12.82	1.11	4.58	3.98	6.77	1.92	0.78	0.79	0.92
B175	0.14	1.81	13.53	3.46	8.10	0.00	11.21	7.06	11.09	11.35	0.05	0.66	0.29
B354	0.27	1.61	20.17	4.32	5.65	0.01	2.89	59.23	17.80	12.08	0.05	0.74	0.28
B706	0.12	2.03	13.28	5.03	9.16	0.00	3.57	2.44	7.91	4.45	0.04	0.59	0.36
B921	0.07	3.88	40.74	0.35	14.12	0.03	1.37	1.65	4.50	4.38	0.03	0.51	0.05
E1002	0.15	1.72	14.44	1.78	8.61	0.00	1.31	14.21	6.55	10.81	0.05	0.73	0.19
E560	0.11	1.55	36.34	9.53	12.12	0.00	5.71	7.31	7.97	8.55	0.02	0.70	0.25
TA263	0.11	3.33	10.66	0.70	9.81	0.00	0.94	1.86	1.08	1.44	0.05	0.56	0.16
ECOR1	0.10	3.70	11.18	0.53	9.61	0.00	3.58	1.57	0.69	1.12	0.05	0.56	0.14
TA135	0.08	3.17	10.94	1.96	9.79	0.00	1.26	1.46	1.51	1.60	0.04	0.46	0.26
M863	0.04	8.26	13.89	1.02	10.73	0.00	0.97	2.71	17.93	14.82	0.02	0.25	0.15
MG1655	0.12	1.57	13.51	5.29	9.34	0.00	2.69	4.16	11.60	6.62	0.05	0.71	0.36
R424	0.11	2.85	12.47	1.21	9.67	0.00	1.24	2.50	5.86	3.02	0.04	0.55	0.18
REL606	0.14	2.41	13.33	2.10	8.13	0.00	1.70	9.16	10.09	9.89	0.05	0.57	0.23
TA280	0.14	1.32	13.68	2.69	10.24	0.00	2.97	7.95	6.86	8.57	0.06	0.83	0.24
B093	0.09	15.52	11.21	0.08	9.91	0.00	1.05	1.73	4.15	5.00	0.05	0.70	0.25
H413	0.02	10.07	13.12	1.36	10.64	0.00	1.23	3.64	2.67	10.88	0.01	0.18	0.17
H504	0.13	1.74	15.92	3.25	9.28	0.00	1.75	15.07	11.73	10.98	0.05	0.68	0.24
H588	0.14	2.09	12.09	1.03	8.85	0.00	1.56	3.94	5.10	8.04	0.06	0.69	0.17
M056	0.12	3.46	10.81	0.49	9.33	0.00	0.96	2.05	3.65	4.06	0.06	0.60	0.14
M646	0.03	10.72	17.67	0.86	12.79	0.00	1.07	2.72	14.09	17.61	0.01	0.20	0.10
TA014	0.08	3.10	13.44	2.17	9.49	0.00	1.69	3.25	6.35	10.13	0.03	0.46	0.2

Parameters determining activity of the *lac* promoter were estimated by fitting the regulatory model described in SI (from Mayo et al., 2006). The phenotypic parameters $\pi 1$, $\pi 2$ and $\pi 3$ give the predicted ratio of expression at no induction, high-IPTG_low-cAMP, and low-IPTG_high-cAMP to expression at maximum cAMP and IPTG levels, respectively.

C. Chapter 4 Supplementary Figures

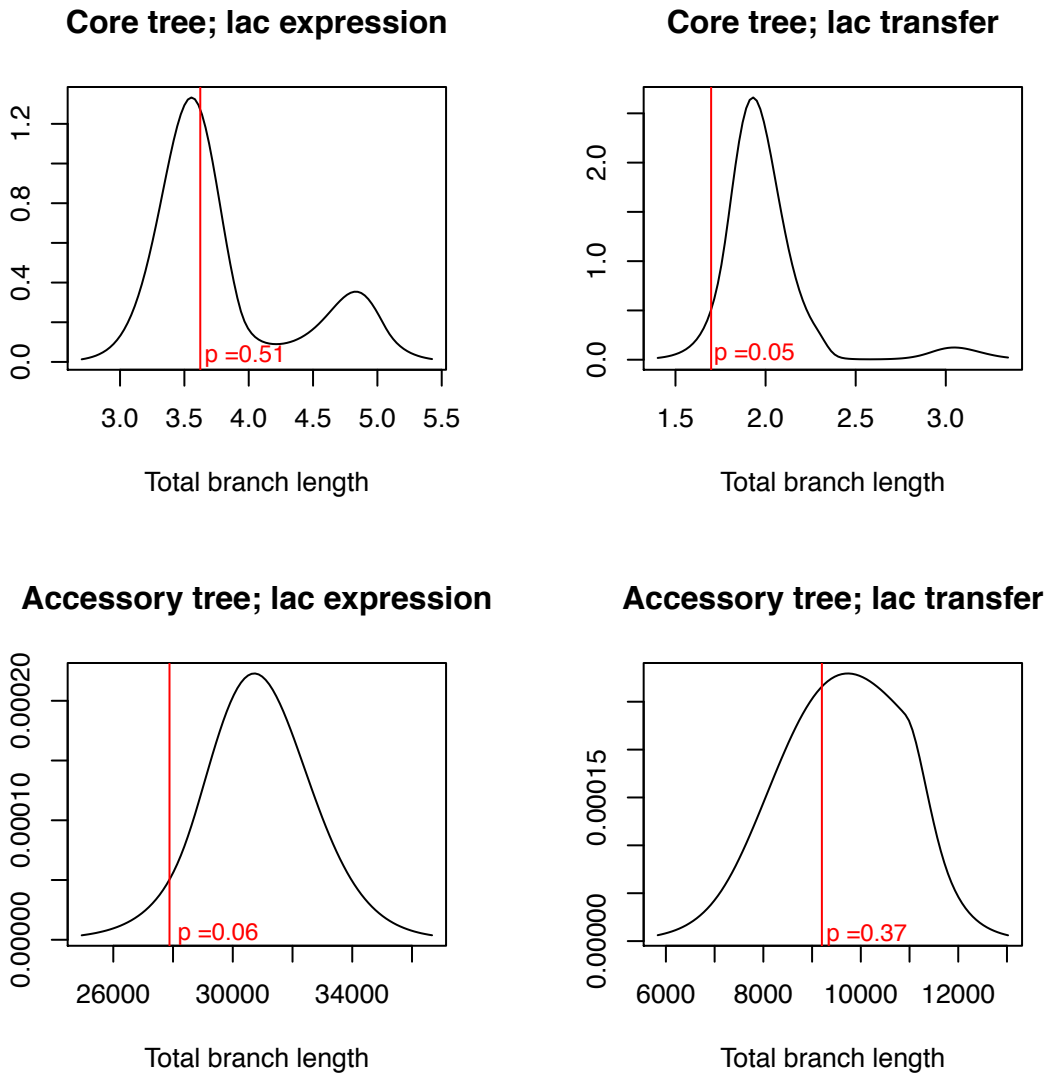
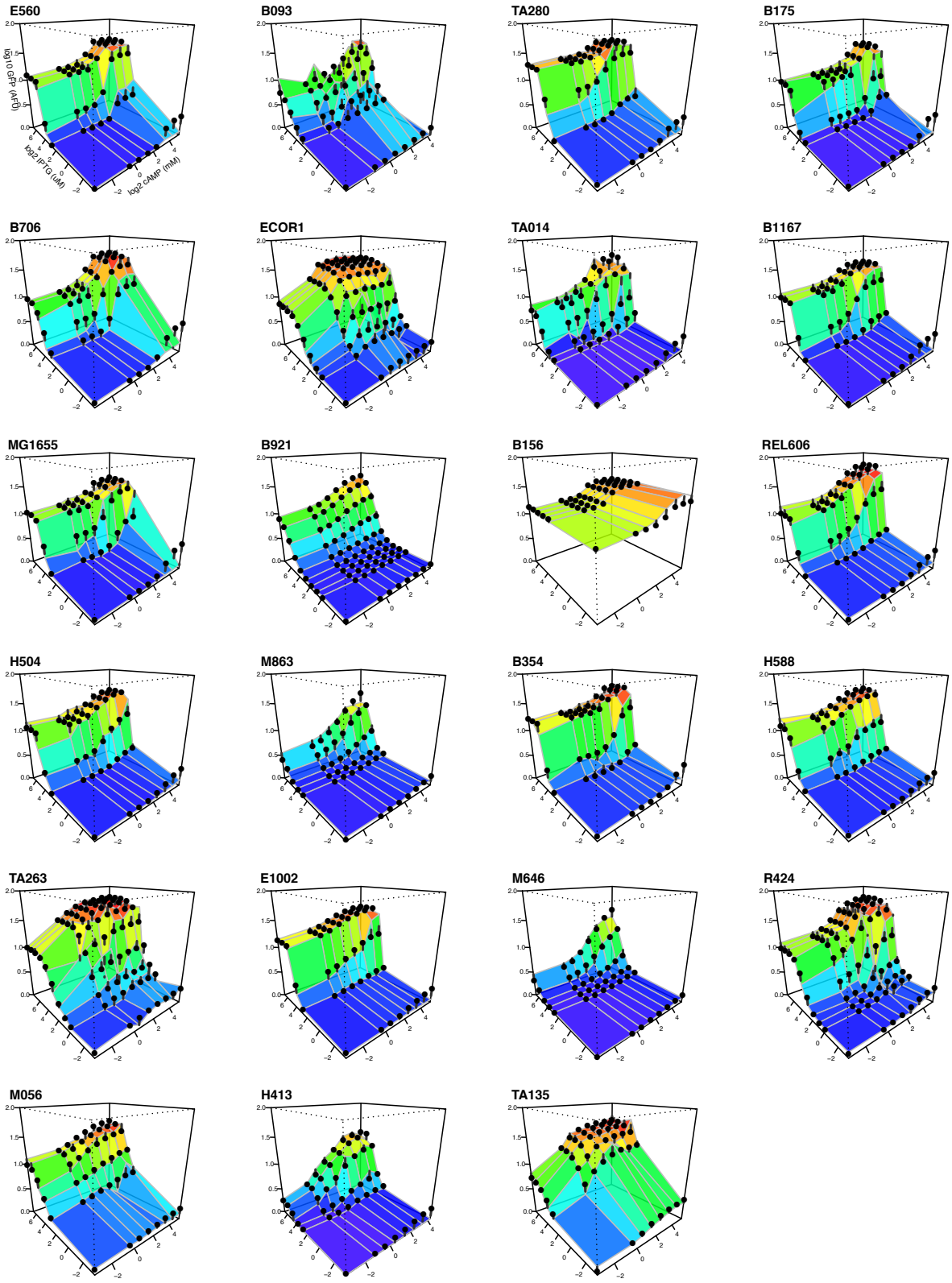


Figure A4.1 Sampled strains are generally representative of the broader strain collection. The solid black line indicates the distribution of shared branch lengths following 1000 draws from the broader strain collection (Fig. 4.1) of the same number of strains used in expression and transfer experiments. The red line indicates the shared branch length of the actual recipient strains. P-values are calculated as the fraction of bootstrapped samples with a lower branch length than the actual sample. Phylogeny (core or accessory)-mutation combinations are indicated in the title of each panel.

Figure A4.2 Observed and modelled regulatory functions of 23 strains. Colored landscapes indicate experimentally determined *lac* expression profiles using at least 48 different cAMP-IPTG concentration combinations. Solid symbols and drop lines indicate predicted expression at each measured point derived from a regulatory model fitted to the observed expression data (Materials and Methods for details). Note that B156 is an *E. albertii* strain that does not encode several components of the canonical lactose utilization system, including the LacI repressor, and therefore expresses the *lac* reporter at high levels regardless of the presence of IPTG.


E560
B093
TA280
B175
B706
ECOR1
TA014
B1167
MG1655
B921
B156
REL606
H504
M863
B354
H588
TA263
E1002
M646
R424
M056
H413
TA135
log10 GFP (AFU)
log2 IPTG (uM)
log2 cAMP (mM)

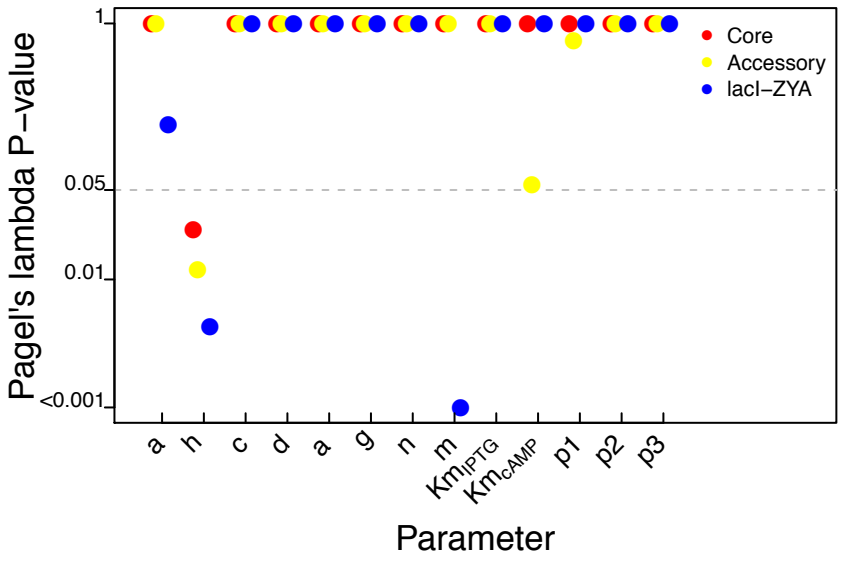


Figure A4.3 P-values of Pagel's λ tests of phylogenetic signal for regulatory model and regulatory logic parameters. The null hypothesis is the absence of phylogenetic signal, P-values below 0.05 are consistent with a parameter evolving neutrally over a particular phylogeny. Tests for phylogenetic signal are shown against phylogenies based on the core genome, the accessory genome, and sequencing encompassing the *lacI-ZYA* region.

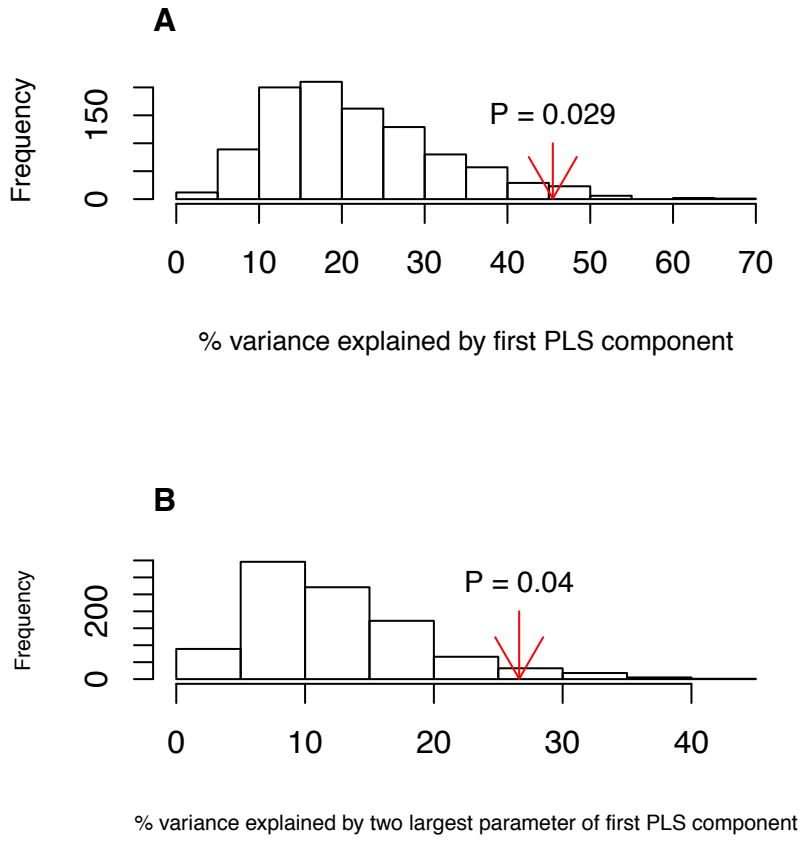
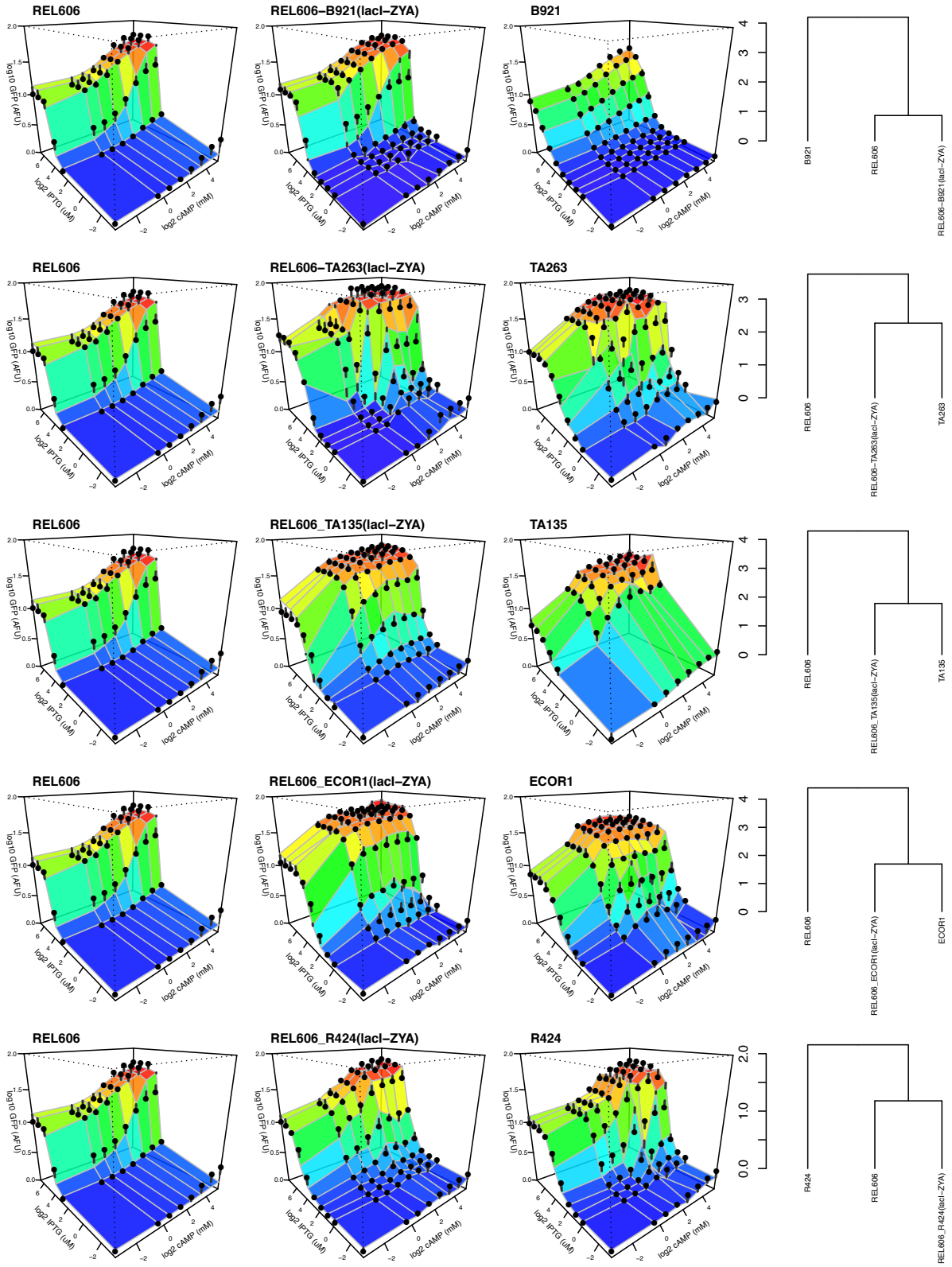


Figure A4.4 Permutation tests to assess the significance of variation in lag time explained by partial least squares regression (Fig. 4.4). Histograms show the percent of variation in lag time explained by the first PLS component (A) and by the two largest contributors to this component (B) based on 1000 regressions in which the association between strain names and measured lag times was randomized while the association between names and regulatory parameters was unchanged. Red arrows indicate corresponding estimates of the actual regression. P-values are based on rank ordering of actual estimates against randomized regression outputs.

Figure A4.5 Effect of genetic background on *lac* expression. Expression profile of the lab strain REL606 is presented in the left column (mean of 5 separate profiles). The column second from left presents expression profiles of hybrid strains constructed by adding the *lacI-ZYA* region from a donor natural isolate strain into REL606 (details in Material and Methods). The column second from right presents the expression profile of the donor natural isolate strain. Dendrograms indicate clustering of each set of three strains (i.e., REL606, REL606 with introduced *lac* region, and natural isolate strain) based on Euclidean distance of empirically determined expression profiles in arbitrary fluorescence units.


REL606
REL606−B921(lacI−ZYA)
B921
REL606
REL606−TA263(lacI−ZYA)
TA263
REL606
REL606_TA135(lacI−ZYA)
TA135
REL606
REL606_ECOR1(lacI−ZYA)
ECOR1
REL606
REL606_R424(lacI−ZYA)
R424
log10 GFP (AFU)
log2 IPTG (uM)
log2 cAMP (mM)

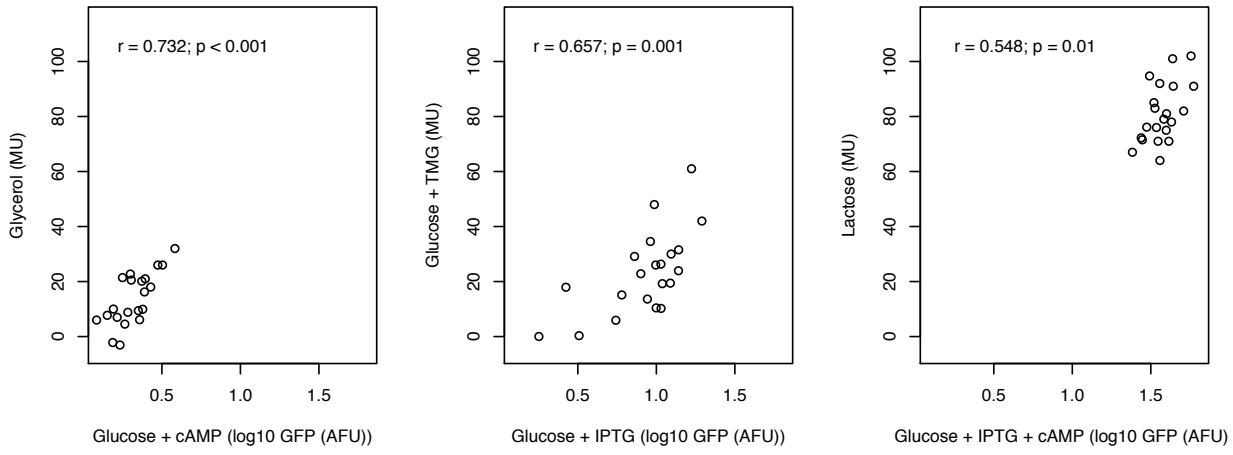


Figure A4.6 Comparison of assay and inducer types on inferred *lac* expression. Expression values indicated on the horizontal axes are in arbitrary fluorescence units (AFU) as reported in the main text and are inferred from fluorescence derived from a P_{lac}-*gfp* reporter construct. Strains were grown in glucose supplemented with the artificial inducer, IPTG (200 μM) and/or exogenously supplied cAMP (40 mM) as noted on axis titles. Expression profiles in Miller units (MU) reported on the vertical axes are inferred from Miller assays (Zhang and Bremer, 1995) in environments using the natural inducers glycerol (which supports high intracellular cAMP) and lactose, and the artificial inducer TMG (200 μM), which depends on LacY for uptake. Combinations of these inducers are chosen to manipulate LacI and CRP regulation in a way that recapitulates as closely as possible effects using IPTG and exogenous cAMP (glucose+cAMP and glycerol environments, high cAMP-low IPTG; glucose+IPTG and glucose+TMG, low cAMP-high IPTG; glucose+cAMP+IPTG and lactose, high cAMP-high IPTG). Pearson correlation analyses are reported on each plot. Miller assays measure LacZ enzyme activity directly, providing a control for expression inferred from the fluorescent reporter construct used in our main set of experiments.

Bibliography

Acar M, Mettetal JT, van Oudenaarden A. 2008. Stochastic switching as a survival strategy in fluctuating environments. Nat Genet 40(4):471–475.

Aidelberg G, Towbin BD, Rothschild D, Dekel E, Bren A, Alon U. 2014. Hierarchy of non-glucose sugars in *Escherichia coli*. BMC Syst Biol 8:1–12.

Anderson DW, McKeown AN, Thornton JW. 2015. Intermolecular epistasis shaped the function and evolution of an ancient transcription factor and its DNA binding sites. Elife 4:1–26.

Andersson DI, Hughes D. 2010. Antibiotic resistance and its cost: is it possible to reverse resistance? Nat Rev Microbiology 8:260–271.

Babu M, Arnold R, Bundalovic-Torma C, Gagarinova A, Wong KS, Kumar A, et al. 2014. Quantitative genome-wide genetic interaction screens reveal global epistatic relationships of protein complexes in *Escherichia coli*. PLoS Genet 10(2):1–15.

Bachmann BJ. 1996. Derivations and genotypes of some mutant derivatives of *Escherichia coli* K-12. In: *Escherichia coli* and *Salmonella typhimurium*: Cellular and molecular biology. 2nd ed. Washington, D. C.: ASM Press. p 2460-2488.

Bailey SF, Kassen R. 2012. Spatial Structure of Ecological Opportunity Drives Adaptation in a Bacterium. Am Nat 180(2):270–283.

Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, et al. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature 461:1243–1247.

Barrick JE, Kauth MR, Strelioff CC, Lenski RE. 2010. *Escherichia coli rpoB* mutants have increased evolvability in proportion to their fitness defects. Mol Biol Evol 27(6):1338–1347.

Barua A, Mikheyev AS. 2019. Many options, few solutions: over 60 million years snakes converged on a few optimal venom formulations. Mol Biol Evol 36(9):1964–1974.

Bates D, Maechler M, Bolker B, Walker S. 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. URL: http://CRAN.R-project.org/package=lme4.

Bedhomme S, Lafforgue G, Elena SF. 2013. Genotypic but not phenotypic historical contingency revealed by viral experimental evolution. BMC Evol Biol 13:1–13.

Behringer MG, Choi BI, Miller SF, Doak TG, Karty JA, Guo W, et al. 2018. *Escherichia coli* cultures maintain stable subpopulation structure during long-term evolution. Proc Natl Acad Sci U S A 115(20):E4642–E4650.

Bennett AF, Lenski RE. 1999. Experimental evolution and its role in evolutionary physiology. Integr Comp Biol 39(2):346–362.

Bennett AF, Lenski RE. 2007. An experimental test of evolutionary trade-offs during temperature adaptation. Proc Natl Acad Sci U S A 104(Supplement 1):8649–8654.

Björkman J, Hughes D, Andersson DI. 1998. Virulence of antibiotic-resistant *Salmonella typhimurium*. Proc Natl Acad Sci U S A 95(7):3949–3953.

Björkman J, Nagaev I, Berg OG, Hughes D, Andersson DI. 2000. Effects of environment on compensatory mutations to ameliorate costs of antibiotic resistance. Science 287(5457):1479–1482.

Blank D, Wolf L, Ackermann M, Silander OK. 2014. The predictability of molecular evolution during functional innovation. Proc Natl Acad Sci U S A 111(8):3044–3049.

Bleuven C, Landry CR. 2016. Molecular and cellular bases of adaptation to a changing environment in microorganisms. Proc Biol Sci 283(1841):1–9.

Blount ZD, Borland CZ, Lenski RE. 2008. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. Proc Natl Acad Sci U S A 105(23):7899–7906.

Blount ZD, Barrick JE, Davidson CJ, Lenski RE. 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. Nature 489:513–518.

Bono LM, Smith LB, Pfennig DW, Burch CL. 2017. The emergence of performance trade-offs during local adaptation: insights from experimental evolution. Mol Ecol 26(7):1720–1733.

Brandis G, Wrande M, Liljas L, Hughes D. 2012. Fitness-compensatory mutations in rifampicin-resistant RNA polymerase. Mol Microbiol 85(1):142–151.

Brideau NJ, Flores HA, Wang J, Maheshwari S, Wang X, Barbash DA. 2006. Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. Science 314(5803):1292–1295.

Bridgham JT, Ortlund EA, Thornton JW. 2009. An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. Nature 461(7263):515–519.

Brooker RJ, Wilson TH. 1985. Isolation and nucleotide sequencing of lactose carrier mutants that transport maltose. Proc Natl Acad Sci U S A 82:3959–3963.

Buckling A, Kassen R, Bell G, Rainey PB. 2000. Disturbance and diversity in experimental microcosms. Nature 408:961–964.

Buckling A, Wills MA, Colegrave N. 2003. Adaptation limits diversification of experimental bacterial populations. Science 302(5653):2107–2109.

Buckling A, Brockhurst MA, Travisano M, Rainey PB. 2007. Experimental adaptation to high and low quality environments under different scales of temporal variation. J Evol Biol 20(1):296–300.

Burch CL, Chao L. 1999. Evolution by small steps and rugged landscapes in the RNA virus Φ6. Genetics 151(3):921–927.

Burch CL, Chao L. 2000. Evolvability of an RNA virus is determined by its mutational neighbourhood. Nature 406:625–628.

Burke MK, Liti G, Long AD. 2014. Standing genetic variation drives repeatable experimental evolution in outcrossing populations of *Saccharomyces cerevisiae*. Mol Biol Evol 31(12):3228–3239.

Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, et al. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. Science 327(5963):302–305.

Chang J, Zhou Y, Hu X, Lam L, Henry C, Green EM, et al. 2013. The molecular mechanism of a *cis*-regulatory adaptation in yeast. PLoS Genet 9(9):1–8.

Chao L, Cox E. 1983. Competition between high and low mutating strains of *Escherichia coli*. Evolution 37(1):125–134.

Chiotti K, Kvitek DJ, Schmidt K, Koniges G, Schwartz K, Donckels EA, et al. 2014. The valley of death: reciprocal sign epistasis constrains adaptive trajectories in a constant, nutrient limiting environment. Genomics 104(6 0 0):431–437.

Choi K-H, Gaynor JB, White KG, Lopez C, Bosio CM, Karkhoff-Schweizer RR, Schweizer HP. 2005. A Tn7-based broad-range bacterial cloning and expression system. Nat Methods 2(6):443–448.

Choi PJ, Cai L, Frieda K, Xie XS. 2008. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. Science 322(5900):442–446.

Chou H-H, Berthet J, Marx CJ. 2009. Fast growth increases the selective advantage of a mutation arising recurrently during evolution under metal limitation. PLoS Genet 5(9):1–12.

Chou H-H, Chiu H-C, Delaney NF, Segrè D, Marx CJ. 2011. Diminishing returns epistasis among beneficial mutations decelerates adaptation. Science 332(6034):1190–1192.

Chubiz LM, Marx CJ. 2017. Growth trade-offs accompany the emergence of glycolytic metabolism in *Shewanella oneidensis* MR-1. J Bacteriol 199(11):1–15.

Cleves PA, Ellis NA, Jimenez MT, Nunez SM, Schluter D, Kingsley DM, Miller CT. 2014. Evolved tooth gain in sticklebacks is associated with a *cis*-regulatory allele of *Bmp6*. Proc Natl Acad Sci U S A 111(38):13912–13917.

Consuegra J, Plucain J, Gaffé J, Hindré T, Schneider D. 2017. Genetic basis of exploiting ecological opportunity during the long-term diversification of a bacterial population. J Mol Evol 85(1-2):26–36.

Cooper TF, Rozen DE, Lenski RE. 2003. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. Proc Natl Acad Sci U S A 100(3):1072–1077.

Cooper TF, Remold SK, Lenski RE, Schneider D. 2008. Expression profiles reveal parallel evolution of epistatic interactions involving the CRP regulon in *Escherichia coli*. PLoS Genet 4(2):0001-0010.

Cooper TF, Lenski RE. 2010. Experimental evolution with *E. coli* in diverse resource environments. I. Fluctuating environments promote divergence of replicate populations. BMC Evol Biol 10:1–10.

Cuevas JM, Moya A, Sanjuan R. 2009. A genetic background with low mutational robustness is associated with increased adaptability to a novel host in an RNA virus. J Evol Biol 22(10):2041–2048.

Daber R, Lewis M. 2009. A novel molecular switch. J Mol Biol 391(4):661–670.

Dasmeh P, Girard É, Serohijos AWR. 2017. Highly expressed genes evolve under strong epistasis from a proteome-wide scan in *E. coli*. Sci Rep 7:1–6.

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

Datta S, Costantino N, Court DL. 2006. A set of recombineering plasmids for Gram-negative bacteria. Gene 379:109–115.

Dean AM. 1989. Selection and neutrality in lactose operons of *Escherichia coli*. Genetics 123(3):441–454.

Dean AM. 1995. A molecular investigation of genotype by environment interactions. Genetics 139(1):19–33.

Dekel E, Alon U. 2005. Optimality and evolutionary tuning of the expression level of a protein. Nature 436(7050):588–592.

Domingo J, Diss G, Lehner B. 2018. Pairwise and higher-order genetic interactions during the evolution of a tRNA. Nature 558(7708):117–121.

Dong H, Nilsson L, Kurland CG. 1995. Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. J Bacteriol 177(6):1497–1504.

de Visser JAGM, Zeyl CW, Gerrish PJ, Blanchard JL, Lenski RE. 1999. Diminishing returns from mutation supply rate in asexual populations. Science 283:404–406.

de Vos MGJ, Poelwijk FJ, Battich N, Ndika JD, Tans SJ. 2013. Environmental dependence of genetic constraint. PLoS Genet 9(6):1–8.

Draghi JA, Plotkin JB. 2013. Selection biases the prevalence and type of epistasis along adaptive trajectories. Evolution 67(11):3120–3131.

Dykhuizen D, Davies M. 1980. An experimental model: bacterial specialists and generalists competing in chemostats. Ecology 61(5):1213–1227.

Dykhuizen D, Dean A, Hartl DL. 1987. Metabolic flux and fitness. Genetics 115(1):25–31.

Dykhuizen DE, Dean AM. 1994. Predicted fitness changes along an environmental gradient. Evol Ecol 8:524–541.

Eames M, Kortemme T. 2012. Cost-benefit tradeoffs in engineered *lac* operons. Science 336(6083):911–915.

El Qaidi S, Allemand F, Oberto J, Plumbridge J. 2009. Repression of *galP*, the galactose transporter in *Escherichia coli*, requires the specific regulator of N-acetylglucosamine metabolism. Mol Microbiol 71(1):146–157.

Fay JC, Wittkopp PJ. 2008. Evaluating the role of natural selection in the evolution of gene regulation. Heredity (Edinb) 100:191–199.

Felsenstein J. 1989. PHYLIP - phylogeny inference package (Version 3.2). Cladistics 5:164–166.

Ferenci T. 1996. Adaptation to life at micromolar nutrient levels: the regulation of *Escherichia coli* glucose transport by endoinduction and cAMP. FEMS Microbiol Rev 18(4):301–317.

Fernández-Castané A, Vine CE, Caminal G, López-Santín J. 2012. Evidencing the role of lactose permease in IPTG uptake by *Escherichia coli* in fed-batch high cell density cultures. J. Biotechnol 157(3):391–398.

Ferrières L, Hémery G, Nham T, Guérout A-M, Mazel D, Beloin C, Ghigo JM. 2010. Silent mischief: bacteriophage Mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. J Bacteriol 192(24):6418–6427.

Filteau M, Hamel V, Pouliot M-C, Gagnon-Arsenault I, Dubé AK, Landry CR. 2015. Evolutionary rescue by compensatory mutations is constrained by genomic and environmental backgrounds. Mol Syst Biol 11(10):1–10.

Fisher RA. 1999. The genetical theory of natural selection: a complete variorum edition. Oxford Univ. Press, Oxford, U.K.

Fishman L, Willis JH. 2001. Evidence for Dobzhansky-Muller incompatibilities contributing to the sterility of hybrids between *Mimulus guttatus* and *M. nasutus*. Evolution 55(10):1932–1942.

Flynn KM, Cooper TF, Moore FB-G, Cooper VS. 2013. The environment affects epistatic interactions to alter the topology of an empirical fitness landscape. PLoS Genet 9(4):1–9.

Fragata I, Simões P, Lopes-Cunha M, Lima M, Kellen B, Bárbaro M, et al. 2014. Laboratory selection quickly erases historical differentiation. PLoS One 9(5):1–9.

Fraser HB. 2013. Gene expression drives local adaptation in humans. Genome Res 23(7):1089–1096.

Frumkin I, Schirman D, Rotman A, Li F, Zahavi L, Mordret E, et al. 2017. Gene architectures that minimize cost of gene expression. Mol Cell 65(1):142–153.

Fulcrand G, Dages S, Zhi X, Chapagain P, Gerstman BS, Dunlap D, Leng F. 2016. DNA supercoiling, a critical signal regulating the basal expression of the *lac* operon in *Escherichia coli*. Sci Rep 6:1–12.

Garza de Leon F, Sellars L, Stracy M, Busby SJW, Kapanidis AN. 2017. Tracking low-copy transcription factors in living bacteria: the case of the *lac* repressor. Biophys J 112(7):1316–1327.

Gifford DR, Toll-Riera M, MacLean RC. 2016. Epistatic interactions between ancestral genotype and beneficial mutations shape evolvability in *Pseudomonas aeruginosa*. Evolution 70(7):1659–1666.

Gong LI, Suchard MA, Bloom JD. 2013. Stability-mediated epistasis constrains the evolution of an influenza protein. Elife 2:1–19.

González C, Ray JCJ, Manhart M, Adams RM, Nevozhay D, Morozov AV, Balázsi G. 2015. Stress-response balance drives the evolution of a network module and its host genome. Mol Syst Biol 11(8):1–18.

Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM. 2017. The dynamics of molecular evolution over 60,000 generations. Nature 551(7678):45–50.

Gordon AJE, Satory D, Halliday JA, Herman C. 2013. Heritable change caused by transient transcription errors. Casadesús J, editor. PLoS Genet 9(6):1–12.

Gould SJ. 1989. Wonderful life: the burgess shale and the nature of history. Norton, New York.

Gram CD, Brooker RJ. 1992. An analysis of the side chain requirement at position 177 within the lactose permease which confers the ability to recognize maltose. J Biol Chem 267(6):3841–3846.

Gu T, Orita S, Han M. 1998. *Caenorhabditis elegans* SUR-5, a novel but conserved protein, negatively regulates LET-60 Ras activity during vulval induction. Mol Cell Biol 18(8):4556–4564.

Guo L, Wang X, Zhao M, Huang C, Li C, Li D, et al. 2018. Stepwise *cis*-regulatory changes in *ZCN8* contribute to maize flowering-time adaptation. Curr Biol 28(18):3005–3015.

Gupta A, Adami C. 2016. Strong selection significantly increases epistatic interactions in the long-term evolution of a protein. PLoS Genet 12(3):1–23.

Haber R, Adhya S. 1988. Interaction of spatially separated protein-DNA complexes for control of gene expression: operator conversions. Proc Natl Acad Sci U S A 85(24):9683–9687.

Hall BM, Ma C-X, Liang P, Singh KK. 2009. Fluctuation AnaLysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. Bioinformatics 25(12):1564–1565.

Halligan DL, Kousathanas A, Ness RW, Harr B, Eöry L, Keane TM, et al. 2013. Contributions of protein-coding and regulatory change to adaptive molecular evolution in murid rodents. PLoS Genet 9(12):1–14 .

Hammar P, Leroy P, Mahmutovic A, Marklund EG, Berg OG, Elf J. 2012. The *lac* repressor displays facilitated diffusion in living cells. Science 336(6088):1595–1598.

Hariharan P, Balasubramaniam D, Peterkofsky A, Kaback HR, Guan L. 2015. Thermodynamic mechanism for inhibition of lactose permease by the phosphotransferase protein IIAGlc. Proc Natl Acad Sci U S A 112(8):2407–2412.

Harms MJ, Thornton JW. 2014. Historical contingency and its biophysical basis in glucocorticoid receptor evolution. Nature 512(7513):203–207.

Harrison E, Guymer D, Spiers AJ, Paterson S, Brockhurst MA. 2015. Parallel compensatory evolution stabilizes plasmids across the parasitism-mutualism continuum. Curr Biol 25(15):2034–2039.

Hayden EJ, Bratulic S, Koenig I, Ferrada E, Wagner A. 2013. The effects of stabilizing and directional selection on phenotypic and genotypic variation in a population of RNA enzymes. J Mol Evol 78(2):101–108.

Hernández-Montalvo V, Martínez A, Hernández-Chavez G, Bolívar F, Valle F, Gosset G. 2003. Expression of *galP* and *glk* in a *Escherichia coli* PTS mutant restores glucose transport and increases glycolytic flux to fermentation products. Biotechnol Bioeng 83(6):687–694.

Herron MD, Doebeli M. 2013. Parallel evolutionary dynamics of adaptive diversification in *Escherichia coli*. PLoS Biol 11(2):1–11.

Hong J, Gresham D. 2014. Molecular specificity, convergence and constraint shape adaptive evolution in nutrient-poor environments. PLoS Genet 10(1):1–16.

Huber RE, Hakda S, Cheng C, Cupples CG, Edwards RA. 2003. Trp-999 of β-galactosidase (*Escherichia coli*) is a key residue for binding, catalysis, and synthesis of allolactose, the natural *lac* operon inducer. Biochemistry 42(6):1796–803.

Jacob F, Monod J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 3(3):318–356.

Jahreis K, Pimentel-Schmitt EF, Brückner R, Titgemeyer F. 2008. Ins and outs of glucose transport systems in eubacteria. FEMS Microbiol Rev 32(6):891–907.

Jarvik J, Botstein D. 1975. Conditional-lethal mutations that suppress genetic defects in morphogenesis by altering structural proteins. Proc Natl Acad Sci U S A 72(7):2738–2742.

Jasmin J-N, Kassen R. 2007. Evolution of a single niche specialist in variable environments. Proc Biol Sci 274(1626):2761–2767.

Jenkins AJ, March JB, Oliver IR, Masters M. 1986. A DNA fragment containing the *groE* genes can suppress mutations in the *Escherichia coli dnaA* gene. Mol Genet Genomics 202(3):446–454.

Jensen KF. 1993. The *Escherichia coli* K-12 “wild types” W3110 and MG1655 have an *rph* frameshift mutation that leads to pyrimidine starvation due to low *pyrE* expression levels. J Bacteriol 175(11):3401–3407.

Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi S-H, et al. 2009. Genome sequences of *Escherichia coli* B strains REL606 and BL21 (DE3). J Mol Biol 394(4):644–652.

Jerison ER, Kryazhimskiy S, Mitchell JK, Bloom JS, Kruglyak L, Desai MM. 2017. Genetic variation in adaptability and pleiotropy in budding yeast. Elife 6:1–27.

Jobe A, Riggs AD, Bourgeois S. 1972. *lac* repressor-operator interaction: V. characterization of super- and pseudo-wild-type repressors. J Mol Biol 64(1):181–199.

Johnson LM, Kayne PS, Kahn ES, Grunstein M. 1990. Genetic evidence for an interaction between SIR3 and histone H4 in the repression of the silent mating loci in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A 87(16):6286–6290.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484(7392):55–61.

Juers DH, Hakda S, Matthews BW, Huber RE. 2003. Structural basis for the altered activity of Gly794 variants of *Escherichia coli* β-galactosidase. Biochemistry 42(46):13505–13511.

Kacar B, Ge X, Sanyal S, Gaucher EA. 2017. Experimental evolution of *Escherichia coli* harboring an ancient translation protein. J Mol Evol 84(2-3):69–84.

Kalisky T, Dekel E, Alon U. 2007. Cost-benefit theory and optimal design of gene regulation functions. Phys Biol 4(4):229–245.

Kaplan S, Bren A, Zaslaver A, Dekel E, Alon U. 2008. Diverse two-dimensional input functions control bacterial sugar genes. Mol Cell 29(6):786–792.

Kashiwagi A, Urabe I, Kaneko K, Yomo T. 2006. Adaptive response of a gene network to environmental changes by fitness-induced attractor selection. PLoS One 1(1):1–10.

Kassen R, Bell G. 1998. Experimental evolution in *Chlamydomonas*. IV. Selection in environments that vary through time at different scales. Heredity 80(6):732–741.

Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF. 2011. Negative epistasis between beneficial mutations in an evolving bacterial population. Science 332(6034):1193–1196.

Kimura M. 1985. The role of compensatory neutral mutations in molecular evolution. J Genet 64(1):7–19.

King SC, Wilson TH. 1990. Identification of valine 177 as a mutation altering specificity for transport of sugars by the *Escherichia coli* lactose carrier. Enhanced specificity for sucrose and maltose. J Biol Chem 265(17):9638–9644.

Kleina LG, Miller JH. 1990. Genetic studies of the *lac* repressor. XIII. Extensive amino acid replacements generated by the use of natural and synthetic nonsense suppressors. J Mol Biol 212(2):295–318.

Knöppel A, Näsvall J, Andersson DI. 2016. Compensating the fitness costs of synonymous mutations. Mol Biol Evol 33(6):1461–1477.

Kuhlman T, Zhang Z, Saier MH Jr, Hwa T. 2007. Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. Proc Natl Acad Sci U S A 104(14):6043–6048.

Kurlandzka A, Rosenzweig RF, Adams J. 1991. Identification of adaptive changes in an evolving population of *Escherichia coli*: the role of changes with regulatory and highly pleiotropic effects. Mol Biol Evol 8(3):261–281.

Kram KE, Geiger C, Ismail WM, Lee H, Tang H, Foster PL, et al. 2017. Adaptation of *Escherichia coli* to long-term serial passage in complex medium: evidence of parallel evolution. mSystems 2(2):1–12.

Kryazhimskiy S, Dushoff J, Bazykin GA, Plotkin JB. 2011. Prevalence of epistasis in the evolution of influenza A surface proteins. PLoS Genet 7(2):1–11.

Kryazhimskiy S, Rice DP, Jerison ER, Desai MM. 2014. Global epistasis makes adaptation predictable despite sequence-level stochasticity. Science 344(6191):1519–1522.

Kutch IC, Fedorka KM. 2018. Y-chromosomes can constrain adaptive evolution via epistatic interactions with other chromosomes. BMC Evol Biol 18:1–8.

Kutnjak D, Elena SF, Ravnikar M, Simon AE. 2017. Time-sampled population sequencing reveals the interplay of selection and genetic drift in experimental evolution of potato virus y. J Virol 91(16):1–17.

Kvitek DJ, Sherlock G. 2011. Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. PLoS Genet 7(4):1–11.

Kvitek DJ, Sherlock G. 2013. Whole genome, whole population sequencing reveals that loss of signaling networks is the major adaptive strategy in a constant environment. PLoS Genet 9(11):1–11.

Laan L, Koschwanez JH, Murray AW. 2015. Evolutionary adaptation after crippling cell polarization follows reproducible trajectories. Elife 4:1–18.

Lagator M, Sarikas S, Acar H, Bollback JP, Guet CC. 2017. Regulatory network structure determines patterns of intermolecular epistasis. Elife 6:1–22.

Laing C, Buchanan C, Taboada EN, Zhang Y, Kropinski A, Villegas A, et al. 2010. Pan-genome sequence analysis using Panseq: an online tool for the rapid analysis of core and accessory genomic regions. BMC Bioinformatics 11:1–14.

Lang GI, Rice DP, Hickman MJ, Sodergren E, Weinstock GM, Botstein D, Desai MM. 2013. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. Nature 500(7464):571–574.

Lázaro E, Escarmís C, Domingo E, Manrubia SC. 2002. Modeling viral genome fitness evolution associated with serial bottleneck events: evidence of stationary states of fitness. J Virol 76(17):8675–8681.

Lee DJ, Minchin SD, Busby SJW. 2012. Activating transcription in bacteria. Annu Rev Microbiol 66(1):125–152.

Lee M-C, Marx CJ. 2012. Repeated, selection-driven genome reduction of accessory genes in experimental populations. PLoS Genet 8(5):1–8.

Le Gac M, Plucain J, Hindré T, Lenski RE, Schneider D. 2012. Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with *Escherichia coli*. Proc Natl Acad Sci U S A 109(24):9487–9492.

Le Gac M, Cooper TF, Cruveiller S, Médigue C, Schneider D. 2013. Evolutionary history and genetic parallelism affect correlated responses to evolution. Mol Ecol 22(12):3292–3303.

Le Gall T, Darlu P, Escobar-Páramo P, Picard B, Denamur E. 2005. Selection-driven transcriptome polymorphism in *Escherichia coli*/*Shigella* species. Genome Res 15(2):260–268.

Lehming N, Sartorius J, Niemöller M, Genenger G, v.Wilcken-Bergmann B, Müller-Hill B. 1987. The interaction of the recognition helix of *lac* repressor with *lac* operator. EMBO J 6(10):3145–3153.

Lehming N, Sartorius J, Kisters-Woike B, von Wilcken-Bergmann B, Müller-Hill B. 1990. Mutant *lac* repressors with new specificities hint at rules for protein--DNA recognition. EMBO J 9(3):615–621.

Lenski RE. 1988. Experimental studies of pleiotropy and epistasis in *Escherichia coli*. II. Compensation for maladaptive effects associated with resistance to virus T4. Evolution 42(3):433–440.

Lenski RE, Rose MR, Simpson SC, Tadler SC. 1991. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. Am Nat 138:1315–1341.

Lenski RE, Mongold JA, Sniegowski PD, Travisano M, Vasi F, Gerrish PJ, Schmidt TM. 1998. Evolution of competitive fitness in experimental populations of *E. coli*: what makes one genotype a better competitor than another? Antonie van Leeuwenhoek 73:35–47.

Leon D, D'Alton S, Quandt EM, Barrick JE. 2018. Innovation in an *E. coli* evolution experiment is contingent on maintaining adaptive potential until competition subsides. PLoS Genet 14(4):1–22.

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

Lewis JJ, Reed RD. 2019. Genome-wide regulatory adaptation shapes population-level genomic landscapes in *Heliconius*. Mol Biol Evol 36(1):159–173.

Lewis M, Chang G, Horton NC, Kercher MA, Pace HC, Schumacher MA, et al. 1996. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. Science 271(5253):1247–1254.

Lind PA, Farr AD, Rainey PB. 2015. Experimental evolution reveals hidden diversity in evolutionary pathways. Elife 4:1–17.

Lindsey HA, Gallie J, Taylor S, Kerr B. 2013. Evolutionary rescue from extinction is contingent on a lower rate of environmental change. Nature 494(7438):463–467.

Liu D, Zuo T, Hora B, Song H, Kong W, Yu X, et al. 2014. Preexisting compensatory amino acids compromise fitness costs of a HIV-1 T cell escape mutation. Retrovirology 11:1–11.

Lu J, Tang J, Liu Y, Zhu X, Zhang T, Zhang X. 2011. Combinatorial modulation of *galP* and *glk* gene expression for improved alternative glucose utilization. Appl Microbiol Biotechnol 93(6):2455–2462.

Lunzer M, Golding GB, Dean AM. 2010. Pervasive cryptic epistasis in molecular evolution. PLoS Genet 6(10):1–10.

MacLean RC, Bell G, Rainey PB. 2004. The evolution of a pleiotropic fitness tradeoff in *Pseudomonas fluorescens*. Proc Natl Acad Sci U S A 101(21):8072–8077.

MacLean RC, Perron GG, Gardner A. 2010. Diminishing returns from beneficial mutations and pervasive epistasis shape the fitness landscape for rifampicin resistance in *Pseudomonas aeruginosa*. Genetics 186(4):1345–1354.

Maddamsetti R, Lenski RE. 2018. Analysis of bacterial genomes from an evolution experiment with horizontal gene transfer shows that recombination can sometimes overwhelm selection. PLoS Genet 14(1):1–30.

Maisnier-Patin S, Berg OG, Liljas L, Andersson DI. 2002. Compensatory adaptation to the deleterious effect of antibiotic resistance in *Salmonella typhimurium*. Mol Microbiol 46(2):355–366.

Manson MD. 2000. Allele-specific suppression as a tool to study protein–protein interactions in bacteria. Methods 20(1):18–34.

Mansoori N, Tripathi M, Luthra K, Alam R, Lakshmy R, Sharma S, et al. 2012. MTHFR (677 and 1298) and IL-6-174 G/C genes in pathogenesis of Alzheimer's and vascular dementia and their epistatic interaction. Neurobiol Aging 33(5):1003.e1–1003.e8.

Markiewicz P, Kleina LG, Cruz C, Ehret S, Miller JH. 1994. Genetic studies of the *lac* repressor. XIV. Analysis of 4000 altered *Escherichia coli lac* repressors reveals essential and non-essential residues, as well as "spacers" which do not require a specific sequence. J Mol Biol, 240(5):421–433.

Marks ME, Castro-Rojas CM, Teiling C, Du L, Kapatral V, Walunas TL, Crosson S. 2010. The genetic basis of laboratory adaptation in *Caulobacter crescentus*. J Bacteriol 192(14):3678–3688.

Martin G, Lenormand T. 2015. The fitness effect of mutations across environments: Fisher's geometrical model with multiple optima. Evolution 69(6):1433–1447.

Matange N, Hegde S, Bodkhe S. 2019. Adaptation through lifestyle switching sculpts the fitness landscape of evolving populations: implications for the selection of drug-resistant bacteria at low drug pressures. Genetics, 211(3):1029–1044.

Mayo AE, Setty Y, Shavit S, Zaslaver A, Alon U. 2006. Plasticity of the *cis*-regulatory input function of a gene. PLoS Biol 4(4):0555–0561.

McBride RC, Ogbunugafor CB, Turner PE. 2008. Robustness promotes evolvability of thermotolerance in an RNA virus. BMC Evol Biol 8:1–14.

McGee LW, Sackman AM, Morrison AJ, Pierce J, Anisman J, Rokyta DR. 2016. Synergistic pleiotropy overrides the costs of complexity in viral adaptation. Genetics 202(1):285–295.

McGlothlin JW, Kobiela ME, Feldman CR, Castoe TA, Geffeney SL, Hanifin CT, et al. 2016. Historical contingency in a multigene family facilitates adaptive evolution of toxin resistance. Curr Biol 26(12):1616–1621.

Meftahi N, Namouchi A, Mhenni B, Brandis G, Hughes D, Mardassi H. 2016. Evidence for the critical role of a secondary site *rpoB* mutation in the compensatory evolution and successful transmission of an MDR tuberculosis outbreak strain. J Antimicrob Chemother 71(2):324–332.

Meiklejohn CD, Holmbeck MA, Siddiq MA, Abt DN, Rand DM, Montooth KL. 2013. An incompatibility between a mitochondrial tRNA and its nuclear-encoded tRNA synthetase compromises development and fitness in *Drosophila*. PLoS Genet 9(1):1–12.

Melnyk AH, McCloskey N, Hinz AJ, Dettman J, Kassen R. 2017. Evolution of cost-free resistance under fluctuating drug selection in *Pseudomonas aeruginosa*. mSphere 2(4):1–14.

Meyer JR, Dobias DT, Weitz JS, Barrick JE, Quick RT, Lenski RE. 2012. Repeatability and contingency in the evolution of a key innovation in phage lambda. Science 335:428–432.

Moore FB-G, Rozen DE, Lenski RE. 2000. Pervasive compensatory adaptation in *Escherichia coli*. Proc Biol Sci 267(1442):515–522.

Moore FB-G, Woods R. 2006. Tempo and constraint of adaptive evolution in *Escherichia coli* (Enterobacteriaceae, Enterobacteriales). Biol J Linn Soc Lond 88(3):403–411.

Morris SC. 2003. Life's solution. Cambridge, UK: Cambridge University Press.

Moura de Sousa J, Balbontín R, Durão P, Gordo I. 2017. Multidrug-resistant bacteria compensate for the epistasis between resistances. PLoS Biol 15(4):1–24.

Nagaev I, Björkman J, Andersson DI, Hughes D. 2001. Biological cost and compensatory evolution in fusidic acid-resistant *Staphylococcus aureus*. Mol Microbiol 40(2):433–439.

Nilsson AI, Berg OG, Aspevall O, Kahlmeter G, Andersson DI. 2003. Biological costs and mechanisms of fosfomycin resistance in *Escherichia coli*. Antimicrob Agents Chemother 47(9):2850–2858.

Novick A, Horiuchi T. 1961. Hyper-production of β-galactosidase by *Escherichia coli* bacteria. Cold Spring Harb Symp Quant Biol 26:239–245.

Oehler S, Eismann ER, Krämer H, Müller-Hill B. 1990. The three operators of the *lac* operon cooperate in repression. EMBO J 9(4):973–979.

Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW. 2007. Crystal structure of an ancient protein: evolution by conformational epistasis. Science 317(5844):1544–1548.

Osborne SE, Walthers D, Tomljenovic AM, Mulder DT, Silphaduang U, Duong N, et al. 2009. Pathogenic adaptation of intracellular bacteria by rewiring a *cis*-regulatory input function. Proc Natl Acad Sci U S A 106(10):3982–3987.

Ozbudak EM, Thattai M, Lim HN, Shraiman BI, van Oudenaarden A. 2004. Multistability in the lactose utilization network of *Escherichia coli*. Nature 427:737–740.

Pagel M. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.

Pardee AB, Jacob F, Monod J. 1959. The genetic control and cytoplasmic expression of "inducibility" in the synthesis of β-galactosidase by *E. coli*. J Mol Biol 1(2):165–178.

Parsell DA, Silber KR, Sauer RT. 1990. Carboxy-terminal determinants of intracellular protein degradation. Genes Dev 4(2):277–286.

Payne JL, Khalid F, Wagner A. 2018. RNA-mediated gene regulation is less evolvable than transcriptional regulation. Proc Natl Acad Sci U S A 115(15):E3481–E3490.

Peng F, Widmann S, Wünsche A, Duan K, Donovan KA, Dobson RCJ, et al. 2017. Effects of beneficial mutations in *pykF* gene vary over time and across replicate populations in a long-term experiment with bacteria. Mol Biol Evol 35(1):202-210.

Perfeito L, Ghozzi S, Berg J, Schnetz K, Lässig M. 2011. Nonlinear fitness landscape of a molecular pathway. PLoS Genet 7(7):1–10.

Perfeito L, Sousa A, Bataillon T, Gordo I. 2013. Rates of fitness decline and rebound suggest pervasive epistasis. Evolution 68:150–162.

Pfeifer SP, Laurent S, Sousa VC, Linnen CR, Foll M, Excoffier L, et al. 2018. The evolutionary history of nebraska deer mice: local adaptation in the face of strong gene flow. Mol Biol Evol 35(4):792–806.

Philippe N, Alcaraz J-P, Coursange E, Geiselmann J, Schneider D. 2004. Improvement of pCVD442, a suicide plasmid for gene allele exchange in bacteria. Plasmid 51(3):246–255.

Phillips KN, Castillo G, Wünsche A, Cooper TF. 2016. Adaptation of *Escherichia coli* to glucose promotes evolvability in lactose. Evolution 70(2):465–470.

Phillips KN, Widmann S, Lai H-Y, Nguyen J, Ray JCJ, Balázsi G, Cooper TF. 2019. Diversity in *lac* operon regulation among diverse *Escherichia coli* isolates depends on the broader genetic background but is not explained by genetic relatedness. MBio 10(6):1–14.

Platt T, Files JG, Weber K. 1973. Lac repressor. Specific proteolytic destruction of the NH_2-terminal region and loss of the deoxyribonucleic acid-binding activity. J Biol Chem 248(1):110–121.

Plucain J, Hindré T, Le Gac M, Tenaillon O, Cruveiller S, Médigue C, et al. 2014. Epistasis and allele specificity in the emergence of a stable polymorphism in *Escherichia coli*. Science 343(6177):1366–1369.

Pokusaeva VO, Usmanova DR, Putintseva EV, Espinar L, Sarkisyan KS, Mishin AS, et al. 2019. An experimental assay of the interactions of amino acids from orthologous sequences shaping a complex fitness landscape. PLoS Genet 15(4):1–30.

Ponmani T, Munavar MH. 2014. G_{673} could be a novel mutational hot spot for intragenic suppressors of *pheS5* lesion in *Escherichia coli*. Microbiologyopen 3(3):369–382.

Poon A, Chao L. 2005a. The rate of compensatory mutation in the DNA bacteriophage ΦX174. Genetics 170(3):989–999.

Poon A, Davis BH, Chao L. 2005b. The coupon collector and the suppressor mutation: estimating the number of compensatory mutations by maximum likelihood. Genetics 170(3):1323–1332.

Poon AFY, Chao L. 2006. Functional origins of fitness effect-sizes of compensatory mutations in the DNA bacteriophage ΦX174. Evolution 60(10):2032–2043.

Postma PW, Lengeler JW, Jacobson GR. 1993. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. Microbiol Rev 57(3):543–594.

Press MO, Queitsch C. 2017. Variability in a short tandem repeat mediates complex epistatic interactions in *Arabidopsis thaliana*. Genetics 205(1):455–464.

Qi Q, Toll-Riera M, Heilbron K, Preston GM, MacLean RC. 2016. The genomic basis of adaptation to the fitness cost of rifampicin resistance in *Pseudomonas aeruginosa*. Proc Biol Sci 283(1822):1–9.

Quan S, Ray JCJ, Kwota Z, Duong T, Balázsi G, Cooper TF, Monds RD. 2012. Adaptive evolution of the lactose utilization network in experimentally evolved populations of *Escherichia coli*. PLoS Genet 8(1):1–18.

Quandt EM, Deatherage DE, Ellington AD, Georgiou G, Barrick JE. 2014. Recursive genomewide recombination and sequencing reveals a key refinement step in the evolution of a metabolic innovation in *Escherichia coli*. Proc Natl Acad Sci U S A 111(6):2217–2222.

Quandt EM, Gollihar J, Blount ZD, Ellington AD, Georgiou G, Barrick JE. 2015. Fine-tuning citrate synthase flux potentiates and refines metabolic innovation in the Lenski evolution experiment. Elife 4:1–22.

Ray JCJ, Wickersheim ML, Jalihal AP, Adeshina YO, Cooper TF, Balazsi G. 2016. Cellular growth arrest and persistence from enzyme saturation. Rao CV, editor. PLoS Comput Biol 12(3):1–21.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL http://www.R-project.org/.

R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL https://www.R-project.org/.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL https://www.R-project.org/.

Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, et al. 2000. The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. Nature 403(6772):901–906.

Remold SK, Lenski RE. 2004. Pervasive joint influence of epistasis and plasticity on mutational effects in *Escherichia coli*. Nat Genet 36(4):423–426.

Roemhild R, Barbosa C, Beardmore RE, Jansen G, Schulenburg H. 2015. Temporal variation in antibiotic environments slows down resistance evolution in pathogenic *Pseudomonas aeruginosa*. Evol Appl 8(10):945–955.

Rojas Echenique JI, Kryazhimskiy S, Nguyen Ba AN, Desai MM. 2019. Modular epistasis and the compensatory evolution of gene deletion mutants. PLoS Genet 15(2):1–23.

Rosenzweig RF, Sharp RR, Treves DS, Adams J. 1994. Microbial evolution in a simple unstructured environment: genetic differentiation in *Escherichia coli*. Genetics 137(4):903–917.

Rozen DE, McGee L, Levin BR, Klugman KP. 2007. Fitness costs of fluoroquinolone resistance in *Streptococcus pneumoniae*. Antimicrob Agents Chemother 51(2):412–416.

Sackman AM, Rokyta DR. 2018. Additive phenotypes underlie epistasis of fitness effects. Genetics 208(1):339–348.

Sahin-Tóth M, Akhoon KM, Runner J, Kaback HR. 2000. Ligand recognition by the lactose permease of *Escherichia coli*: specificity and affinity are defined by distinct structural elements of galactopyranosides. Biochemistry 39(17):5097–5103.

Sahin-Tóth M, Lawrence MC, Nishio T, Kaback HR. 2001. The C-4 hydroxyl group of galactopyranosides is the major determinant for ligand recognition by the lactose permease of *Escherichia coli*. Biochemistry 40(43):13015–13019.

Sailer ZR, Harms MJ. 2017. High-order epistasis shapes evolutionary trajectories. PLoS Comput Biol 13(5):1–16.

Salinas RK, Folkers GE, Bonvin AMJJ, Das D, Boelens R, Kaptein R. 2005. Altered specificity in DNA binding by the *lac* repressor: a mutant *lac* headpiece that mimics the *gal* repressor. Chembiochem 6(9):1628–1637.

Salverda MLM, Dellus E, Gorter FA, Debets AJM, van der Oost J, Hoekstra RF, et al. 2011. Initial mutations direct alternative pathways of protein evolution. PLoS Genet 7(3):1–11.

Sandberg TE, Lloyd CJ, Palsson BO, Feist AM, Kivisaar M. 2017. Laboratory evolution to alternating substrate environments yields distinct phenotypic and genetic adaptive strategies. Appl Environ Microbiol 83(13):1–15.

Sanjuán R, Cuevas JM, Moya A, Elena SF. 2005. Epistasis and the adaptability of an RNA virus. Genetics 170(3):1001–1008.

San Millan A, Heilbron K, MacLean RC. 2014. Positive epistasis between co-infecting plasmids promotes plasmid survival in bacterial populations. ISME J 8(3):601–612.

Satterwhite RS, Cooper TF. 2015. Constraints on adaptation of *Escherichia coli* to mixed-resource environments increase over time. Evolution 69(8):2067–2078.

Savageau MA. 1998. Demand theory of gene regulation. II. Quantitative application to the lactose and maltose operons of *Escherichia coli*. Genetics 149(4):1677–1691.

Schenk MF, Szendro IG, Salverda MLM, Krug J, de Visser JAGM. 2013. Patterns of epistasis between beneficial mutations in an antibiotic resistance gene. Mol Biol Evol 30(8):1779–1787.

Schick A, Bailey SF, Kassen R. 2015. Evolution of fitness trade-offs in locally adapted populations of *Pseudomonas fluorescens*. Am Nat 186(S1):S48–S59.

Schumer M, Cui R, Rosenthal GG, Andolfatto P. 2015. Reproductive isolation of hybrid populations driven by genetic incompatibilities. PLoS Genet 11(3):1–21.

Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T. 2010. Interdependence of cell growth and gene expression: origins and consequences. Science 330(6007):1099–1102.

Selmecki AM, Maruvka YE, Richmond PA, Guillet M, Shoresh N, Sorenson AL, et al. 2015. Polyploidy can drive rapid adaptation in yeast. Nature 519(7543):349–352.

Semsey S. 2014. Mutations in transcriptional regulators allow selective engineering of signal integration logic. MBio 5(3):1–7.

Setty Y, Mayo AE, Surette MG, Alon U. 2003. Detailed map of a cis-regulatory input function. Proc Natl Acad Sci U S A 100(13):7702–7707.

Shah P, McCandlish DM, Plotkin JB. 2015. Contingency and entrenchment in protein evolution under purifying selection. Proc Natl Acad Sci U S A 112(25):E3226–E3235.

Silva RF, Mendonça SCM, Carvalho LM, Reis AM, Gordo I, Trindade S, Dionisio F. 2011. Pervasive sign epistasis between conjugative plasmids and drug-resistance chromosomal mutations. PLoS Genet 7(7):1–10.

Sota M, Yano H, Hughes JM, Daughdrill GW, Abdo Z, Forney LJ, Top EM. 2010. Shifts in the host range of a promiscuous plasmid through parallel evolution of its replication initiation protein. ISME J 4(12):1568–1580.

Spencer CC, Tyerman J, Bertrand M, Doebeli M. 2008. Adaptation increases the likelihood of diversification in an experimental bacterial lineage. Proc Natl Acad Sci U S A 105(5):1585–1589.

Starr TN, Picton LK, Thornton JW. 2017. Alternative evolutionary histories in the sequence space of an ancient protein. Nature 549(7672):409–413.

Starr TN, Flynn JM, Mishra P, Bolon DNA, Thornton JW. 2018. Pervasive contingency and entrenchment in a billion years of Hsp90 evolution. Proc Natl Acad Sci U S A 115(17):4453–4458.

Steinberg B, Ostermeier M. 2016. Environmental changes bridge evolutionary valleys. Sci Adv 2(1):1–9.

Steinsiek S, Bettenbrock K. 2012. Glucose transport in *Escherichia coli* mutant strains with defects in sugar transport systems. J Bacteriol 194(21):5897–5908.

Stoebel DM, Dean AM, Dykhuizen DE. 2008. The cost of expression of *Escherichia coli lac* operon proteins is in the process, not in the products. Genetics 178(3):1653–1660.

Suiter AM, Bänziger O, Dean AM. 2003. Fitness consequences of a regulatory polymorphism in a seasonal environment. Proc Natl Acad Sci U S A 100(22):12782–12786.

Sunshine AB, Payen C, Ong GT, Liachko I, Tan KM, Dunham MJ. 2015. The fitness consequences of aneuploidy are driven by condition-dependent gene effects. PLoS Biol 13(5):1–34.

Sutendra G, Wong S, Fraser ME, Huber RE. 2007. β-Galactosidase (*Escherichia coli*) has a second catalytically important Mg^{2+} site. Biochem Biophys Res Commun 352(2):566–570.

Szamecz B, Boross G, Kalapis D, Kovács K, Fekete G, Farkas Z, et al. 2014. The genomic landscape of compensatory evolution. PLoS Biol 12(8):1–16.

Takeda Y, Hirota Y. 1982. Suppressor genes of a *dnaA* temperature sensitive mutation in *Escherichia coli*. Mol Genet Genomics 187(1):67–71.

Tangwancharoen S, Moy GW, Burton RS. 2018. Multiple modes of adaptation: regulatory and structural evolution in a small heat shock protein gene. Mol Biol Evol 35(9):2110–2119.

Tenaillon O, Barrick JE, Ribeck N, Deatherage DE, Blanchard JL, Dasgupta A, et al. 2016. Tempo and mode of genome evolution in a 50,000-generation experiment. Nature 536(7615):165–170.

Toprak E, Veres A, Michel J-B, Chait R, Hartl DL, Kishony R. 2012. Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. Nat Genet 44(1):101–105.

Traverse CC, Mayo-Smith LM, Poltak SR, Cooper VS. 2013. Tangled bank of experimentally evolved *Burkholderia* biofilms reflects selection during chronic infections. Proc Natl Acad Sci U S A 110(3):E250–E259.

Travisano M, Mongold JA, Bennett AF, Lenski RE. 1995. Experimental tests of the roles of adaptation, chance, and history in evolution. Science 267(5194):87–90.

Treviño S, Sun Y, Cooper TF, Bassler KE. 2012. Robust detection of hierarchical communities from *Escherichia coli* gene expression data. Miyano S, editor. PLoS Comput Biol 8(2):1–15.

Turner PE, Elena SF. 2000. Cost of host radiation in an RNA virus. Genetics 156(4):1465–1470.

Van Dyk TK, Gatenby AA, LaRossa RA. 1989. Demonstration by genetic suppression of interaction of GroE products with many proteins. Nature 342(6248):451–453.

van Leeuwen J, Pons C, Mellor JC, Yamaguchi TN, Friesen H, Koschwanez J, et al. 2016. Exploring genetic suppression interactions on a global scale. Science 354(6312):1–11.

Velicer GJ, Yu Y-TN. 2003. Evolution of novel cooperative swarming in the bacterium *Myxococcus xanthus*. Nature 425(6953):75–78.

Vijendravarma RK, Kawecki TJ. 2013. Epistasis and maternal effects in experimental adaptation to chronic nutritional stress in *Drosophila*. J Evol Biol 26(12):2566–2580.

Vilar JMG, Leibler S. 2003. DNA Looping and physical constraints on transcription regulation. J Mol Biol 331(5):981–989.

Vogwill T, MacLean RC. 2015. The genetic basis of the fitness costs of antimicrobial resistance: a meta-analysis approach. Evol Appl 8(3):284–295.

Vogwill T, Kojadinovic M, MacLean RC. 2016. Epistasis between antibiotic resistance mutations and genetic background shape the fitness effect of resistance across species of *Pseudomonas*. Proc Biol Sci 283(1830):1–8.

Wang Y, Arenas CD, Stoebel DM, Cooper TF. 2012. Genetic background affects epistatic interactions between two beneficial mutations. Biol Lett 9:1–4.

Wang Y, Díaz Arenas C, Stoebel DM, Flynn K, Knapp E, Dillon MM, et al. 2016. Benefit of transferred mutations is better predicted by the fitness of recipients than by their ecological or genetic relatedness. Proc Natl Acad Sci U S A 113(18):5047–5052.

Weickert MJ, Adhya S. 1992. A family of bacterial regulators homologous to Gal and Lac repressors. J Biol Chem 267(22):15869–15874.

Whitehead A, Crawford DL. 2006. Neutral and adaptive variation in gene expression. Proc Natl Acad Sci U S A 103(14):5425–5430.

Wielgoss S, Bergmiller T, Bischofberger AM, Hall AR. 2016. Adaptation to parasites and costs of parasite resistance in mutator and nonmutator bacteria. Mol Biol Evol 33(3):770–782.

Wiser MJ, Ribeck N, Lenski RE. 2013. Long-term dynamics of adaptation in asexual populations. Science 342(6164):1364–1367.

Wood MP, Cole AL, Ruchala P, Waring AJ, Rohan LC, Marx P, et al. 2013. A compensatory mutation provides resistance to disparate HIV fusion inhibitor peptides and enhances membrane fusion. PloS One 8(2):1–9.

Woods R, Schneider D, Winkworth CL, Riley MA, Lenski RE. 2006. Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. Proc Natl Acad Sci U S A 103(24):9107–9112.

Woods RJ, Barrick JE, Cooper TF, Shrestha U, Kauth MR, Lenski RE. 2011. Second-order selection for evolvability in a large *Escherichia coli* population. Science 331(6023):1433–1436.

Wünsche A, Dinh DM, Satterwhite RS, Arenas CD, Stoebel DM, Cooper TF. 2017. Diminishing-returns epistasis decreases adaptability along an evolutionary trajectory. Nat Ecol Evol 1:1–6.

Zee PC, Mendes-Soares H, Yu Y-TN, Kraemer SA, Keller H, Ossowski S, et al. 2014. A shift from magnitude to sign epistasis during adaptive evolution of a bacterial social trait. Evolution 68(9):2701–2708.

Zhao X, Muller EGD, Rothstein R. 1998. A suppressor of two essential checkpoint genes identifies a novel protein that negatively affects dNTP pools. Mol Cell 2(3):329–340.

Zhao L, Seth-Pasricha M, Stemate D, Crespo-Bellido A, Gagnon J, Draghi J, et al. 2019. Existing host range mutations constrain further emergence of RNA viruses. J Virol 93(4):1-16.

Zhang X, Bremer H. 1995. Control of the *Escherichia coli rrnB* P1 promoter strength by ppGpp. J Biol Chem 270:11181–11189.

Zhang W, Sehgal V, Dinh DM, Azevedo RBR, Cooper TF, Azencott R. 2012. Estimation of the rate and effect of new beneficial mutations in asexual populations. Theor Pop Biol 81:168–178.

Zhong S, Miller SP, Dykhuizen DE, Dean AM. 2009. Transcription, translation, and the evolution of specialists and generalists. Mol Biol Evol 26(12):2661–2678.

Zubay G, Schwartz D, Beckwith J. 1970. Mechanism of activation of catabolite-sensitive genes: a positive control system. Proc Natl Acad Sci U S A 66:104–110.

Zwietering MH, Jongenburger I, Rombouts FM, van 't Riet K. 1990. Modeling of the bacterial growth curve. Appl Environ Microbiol 56(6):1875–1881.