DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

# Adaptation and Specialization in the Evolution of Bacterial Metabolism

## Citation

Leiby, Nicholas. 2014. Adaptation and Specialization in the Evolution of Bacterial Metabolism. Doctoral dissertation, Harvard University.

## Permanent link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:12274296

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

# Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

# Adaptation and Specialization in the Evolution of Bacterial Metabolism

A dissertation presented

by

Nicholas Leiby

to

The Committee on Higher Degrees in Systems Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Systems Biology

Harvard University

Cambridge, Massachusetts

April 2014

© 2014 Nicholas Leiby
All rights reserved.

**Adaptation and Specialization in the Evolution of Bacterial Metabolism**

## Abstract

Specialization is a balance of evolutionary adaptation and its accompanying costs. Here we focus on the Lenski Long-Term Evolution Experiment, which has maintained cultures of *Escherichia coli* in the same, defined seasonal environment for 50,000 generations. This dissertation explores the extent and means by which metabolic specialization occurs over an extended period in the same environment.

Chapter 1 provides an overview and introduction of the problem.

In Chapter 2 we investigated the acquired dependence on citrate for optimal growth of some of the adapting populations. Earlier work uncovered that one of the adapting populations gained the ability to utilize citrate as a sole carbon source. We showed that in addition to this population, three other lineages evolving in parallel began to rely on citrate as a chelator of iron for optimal growth on glucose. This specialization seemed to have occurred through loss of function, most consistent with the accumulation of mutations in iron transport genes that were obviated by abundant citrate.

In Chapter 3 we examined changes in fitness of the evolving populations on carbon sources other than the glucose on which they adapted. We demonstrated that declines in performance were much less widespread than suggested by previous results, and surprisingly accompanied by improvements on a variety of substrates. Strains with higher mutation rate exhibited significantly more declines, and these were ameliorated by growth at lower temperature. These findings suggested that specialization does not

mainly result as a consequence of adaptive tradeoffs, but rather due to the gradual accumulation of disabling mutations in unused portions of the genome.

In Chapter 4 we tested the ability of Flux Balance Analysis models to predict evolved changes in central metabolism. We measured metabolic fluxes for evolved populations from the Lenski experiment, and compared them along with datasets from two other experiments to flux predictions. We found that improved growth largely derived from increased rate of substrate use. Flux predictions were more accurate for two experiments initiated with relatively sub-optimal ancestors, whereas ancestors near the optimum tended to move away from predictions over experimental evolution.

# Table of Contents

**CHAPTER 1: UNDERSTANDING THE MECHANISMS UNDERLYING SPECIALIZATION** **1**

REFERENCES 6

**CHAPTER 2: MULTIPLE LONG-TERM, EXPERIMENTALLY-EVOLVED POPULATIONS OF *ESCHERICHIA COLI* ACQUIRE DEPENDENCE UPON CITRATE AS AN IRON CHELATOR FOR OPTIMAL GROWTH ON GLUCOSE** **7**

INTRODUCTION 9

RESULTS 15

DISCUSSION 26

CONCLUSIONS 29

MATERIALS AND METHODS 30

REFERENCES 34

**CHAPTER 3: METABOLIC EROSION PRIMARILY THROUGH MUTATION ACCUMULATION, AND NOT TRADEOFFS, DRIVES LIMITED EVOLUTION OF SUBSTRATE SPECIFICITY IN *ESCHERICHIA COLI*** **39**

INTRODUCTION 40

RESULTS 43

DISCUSSION 58

MATERIALS AND METHODS 68

REFERENCES 73

**CHAPTER 4: THE ABILITY OF FLUX BALANCE ANALYSIS TO PREDICT EVOLUTION OF CENTRAL METABOLISM SCALES WITH THE INITIAL DISTANCE TO THE OPTIMUM** **79**

INTRODUCTION 81

RESULTS 91

DISCUSSION 104

MATERIALS AND METHODS 110

REFERENCES 120

**SUPPLEMENTARY MATERIAL** **125**

MATERIAL FOR CHAPTER 3 125

MATERIAL FOR CHAPTER 4 150

# Chapter 1

# Understanding the mechanisms underlying specialization

Nicholas Leiby[1,2]

1 Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA, 2 Systems Biology Graduate Program, Harvard University, Cambridge, MA, USA

Microbes evolve quickly. To successfully control them in medicine and engineer them in industry, we need a predictive understanding of microbial evolution. Ideally, we could anticipate the adaptive genetic and phenotypic changes that occur in response to a given environment, on what time scale they occur, and with what degree of repeatability.

While the same mutations responsible for adaptive improvement in one environment may provide benefit in alternative environments, we do not expect this to generally be the case. If there were no cost for improvements, there would be a single optimum genotype regardless of environment. Instead we expect adaptive improvements to one condition lead to costs in others. These costs may result from antagonistic pleiotropy, where the same mutations causing adaptive improvements in one environment are harmful in another. Antagonistic pleiotropy is also referred to as a tradeoff, and represents a direct cost. Adaptive costs may also be a result of mutation accumulation, where mutations with

neutral selective effect in the evolutionary environment have a detrimental effect in an alternative environment. Mutation accumulation is an opportunity cost- an absence of purifying selection resulting in the degradation of unused functions.

These adaptive costs force organisms into one of two general groups- specialists and generalists. Specialists have a high fitness in a narrow range of niches, but not necessarily in others. Generalists maintain a moderate level of fitness across a broad range of environments. Tradeoffs and mutation accumulation prevent the rise of 'Darwinian Demons' [1] or super generalists- a single genotype that is optimally fit across all environments. These processes are fundamental to specialization and the rise and maintenance of global biodiversity. However, the relative role and mechanisms of tradeoffs and mutation accumulation are poorly understood.

Our lack of understanding of specialization exists in part because it is challenging to study. The relative contributions of natural selection, random events, and constraints to adaptation and divergence are not easily separable, and the divergence of evolving populations may merely reflect their adaptation to different environments. Experiments testing hypotheses about specialization are also hampered by the glacial pace of evolution. An ideal model of specialization would provide a stable environment to which adaptation can occur, would be sufficiently long-running to allow adaptation, and would run in parallel to allow the differentiation between chance events and adaptive signal.

Fortunately, a model exists that allows us to test fundamental assumptions about specialization. In Richard Lenski's Long-Term Evolution Experiment (LTEE) [2], 12 parallel populations of *Escherichia coli* have been passaged daily in minimal glucose batch cultures for 26 years and ~58,000 generations of growth. Samples were frozen periodically, and they can be revived to study the populations at a given time or to rewind and replay evolution (e.g., [3-5]). Fitness of the evolved strains in the adaptive environment increased on average by approximately 80% in 50,000 generations [6], and the LTEE strains have been evolving in the same consistent environment for such a long time that we would expect specialization to that environment to have occurred.

One environmental condition that drives microbial evolution and specialization is carbon source availability. In the LTEE, strains were propagated in glucose minimal media. One hypothesis is that evolution in this environment would yield adaptive improvements to glucose growth resulting in tradeoffs in the ability to use other carbon sources. In fact, this has been observed in specific cases, where adaptive improvement occurred at the cost of transport of non-glucose sugars such as maltose [7], or where portions of the ribose operon were deleted [8].

A further hypothesis is that the parallel populations will have adapted in parallel ways. This has certainly been seen in the LTEE - the populations all improved their fitness in the evolutionary environment to similar extents [6]. Many known mutations driving the improvements were different between populations, but many occurred in more than one population, and at least three occurred in all

twelve parallel populations [9], including the above examples of maltose and ribose [7,10].

A previous study attempted to assess the relative roles of mutation accumulation and tradeoffs behind the costs of adaptation in the Lenski LTEE [11]. This study used an assay of respiration as a proxy for fitness on a wide range of carbon sources, and tested performance of the evolved strains at different time points up to 20,000 generations. It found that the strains significantly reduced function across a wide range of substrates, that these reductions occurred in parallel across many or all strains, and that these reductions occurred early on in the course of evolution- in the timeframe associated with rapid increases in fitness in the evolutionary environment. The authors concluded that these findings were most consistent with antagonistic pleiotropy. A further feature of the LTEE allowed the authors of this study to assess the extent of costs associated with mutation accumulation. A number of the parallel populations had acquired a 'mutator' phenotype caused by mutations to DNA repair genes, causing an increase in their mutation rate of approximately 100-fold [11,12]. If mutation accumulation played an important role in adaptive costs, it was expected that these mutator strains would have reduced catabolic function on more substrates than the non-mutators. However, after 20,000 generations, no significant differences in functional decreases were observed between the mutators and non-mutators, and it was concluded that the direct costs of antagonistic pleiotropy were the main drivers of adaptive costs [11].

Much has changed in the 13 years since the publication of this previous study. One of the LTEE populations acquired the novel ability to catabolize citrate as a sole carbon source [13], a finding that made the experiment widely famous. Another population was found to harbor a stable phenotypic polymorphism maintained in a negative frequency-dependent fashion. One of these subpopulations optimized its growth on the glucose in the media, while the other became a cross-feeder that cannibalizes resources and lysed cells in stationary phase [14-16]. Two more populations acquired the mutator phenotype [3,9], and those that were already mutators by the time of the previous study have had another 30,000 generations to accumulate mutations.

This dissertation is a collection of 3 published journal articles that revisit the story of specialization in the LTEE. In chapter 3 I examine the roles of mutation accumulation and tradeoffs in specialization to the glucose carbon source niche. I find that, contrary to previous thinking, the costs of adaptation are driven largely by mutation accumulation. In chapter 2 I examine a different form of specialization- the narrowing of niche breadth based on an acquired dependence to a supplied environmental resource. In chapter 4 I discuss some of the underlying physiological changes involved in metabolic adaptation and specialization, and test models that attempt to predict these changes.

## References


1. Law R (1979) Optimal life histories under age-specific predation. Am Nat 114: 399–417.

2. 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. doi:10.1086/285289.

3. 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. doi:10.1038/nature11514.

4. Bennett AF, Lenski RE (2007) An experimental test of evolutionary trade-offs during temperature adaptation. Proc Natl Acad Sci USA 104 Suppl: 8649–8654. doi:10.1073/pnas.0702117104.

5. Herron MD, Doebeli M (2013) Parallel Evolutionary Dynamics of Adaptive Diversification in Escherichia coli. PLoS Biol 11: e1001490. doi:10.1371/journal.pbio.1001490.

6. Wiser MJ, Ribeck N, Lenski RE (2013) Long-Term Dynamics of Adaptation in Asexual Populations. Science 342: 1364–1367. doi:10.1126/science.1243357.

7. Travisano M, Lenski RE (1996) Long-term experimental evolution in *Escherichia coli*. IV. Targets of selection and the specificity of adaptation. Genetics 143: 15–26.

8. Cooper VS, Schneider D, Blot M, Lenski RE (2001) Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. J Bacteriol 183: 2834–2841. doi:10.1128/JB.183.9.2834.

9. Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature 461: 1243–1247. doi:10.1038/nature08480 -sup.

10. Pelosi L, Kühn L, Guetta D, Garin J, Geiselmann J, et al. (2006) Parallel changes in global protein profiles during long-term experimental evolution in Escherichia coli. Genetics 173: 1851–1869. doi:10.1534/genetics.105.049619.

11. Cooper VS, Lenski RE (2000) The population genetics of ecological specialization in evolving *Escherichia coli* populations. Nature 407: 736–739. doi:10.1038/35037572.

12. Sniegowski PD, Gerrish PJ, Lenski RE (1997) Evolution of high mutation rates in experimental populations of *E. coli*. Nature 387: 703–705.

13. 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 USA 105: 7899–7906. doi:10.1073/pnas.0803151105.

14. Rozen DE, Philippe N, Arjan de Visser J, Lenski RE, Schneider D (2009) Death and cannibalism in a seasonal environment facilitate bacterial coexistence. Ecol Lett 12: 34–44. doi:10.1111/j.1461-0248.2008.01257.x.

15. Rozen DE, Schneider D, Lenski RE (2005) Long-term experimental evolution in *Escherichia coli*. XIII. Phylogenetic history of a balanced polymorphism. J Mol Evol 61: 171–180. doi:10.1007/s00239-004-0322-2.

16. Rozen DE, Lenski RE (2000) Long-term experimental evolution in *Escherichia coli*. VIII. Dynamics of a balanced polymorphism. Am Nat 155: 24–35. doi:10.1086/303299.

## Chapter 2

# Multiple long-term, experimentally-evolved populations of *Escherichia coli* acquire dependence upon citrate as an iron chelator for optimal growth on glucose

Nicholas Leiby[1,2], William R. Harcombe[2], and Christopher J. Marx[2,3,*]

[1]Systems Biology Graduate Program, Harvard University, Cambridge, MA, USA, [2]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA, [3]Faculty of Arts and Sciences Center for Systems Biology, Harvard University, Cambridge, MA, USA.

*Corresponding Author

*This chapter was originally published in BMC Evolutionary Biology. I performed all of the experiments and analysis for this work, with the exception of those for Figure 2.3, which were performed in part by William Harcombe. I was also the primary author of the manuscript.*

Specialization for ecological niches is a balance of evolutionary adaptation and its accompanying tradeoffs. Here we focus on the Lenski Long-Term Evolution Experiment, which has maintained cultures of *Escherichia coli* in the same, defined seasonal environment for 50,000 generations. Over this time, much adaptation and specialization to the environment has occurred. The presence of citrate in the growth media selected one lineage to gain the novel ability to utilize citrate as a carbon source after 31,000 generations. Here we test whether other

strains have specialized to rely on citrate after 50,000 generations. We show that in addition to the citrate-catabolizing strain, three other lineages evolving in parallel have acquired a dependence on citrate for optimal growth on glucose. None of these strains were stimulated indirectly by the sodium present in disodium citrate, nor exhibited even partial utilization of citrate as a carbon source. Instead, all three of these citrate-stimulated populations appear to rely on it as a chelator of iron. The strains we examine here have evolved specialization to their environment through apparent loss of function. Our results are most consistent with the accumulation of mutations in iron transport genes that were obviated by abundant citrate. The results present another example where a subtle decision in the design of an evolution experiment led to unexpected evolutionary outcomes.

## Introduction

Evolutionary adaptation to a new environment leads to changes that improve function and increase fitness, but it may also result in the deterioration of functions not needed in that selective environment. These declines in function that accompany adaptation – tradeoffs – are often considered inevitable costs or constraints of adaptation [1]. Tradeoffs can either result from antagonistic pleiotropy, when adaptive mutations in the selective environment are deleterious elsewhere, or from the accumulation of mutations that are neutral in the selective environment. A consequence of these tradeoffs is that they prevent 'Darwinian demons': single super-genotypes that are optimally fit across a spectrum of

environments [2]. Instead, there are usually 'generalist' organisms that perform adequately in a variety of environments that can coexist with 'specialist' organisms that occupy smaller niches but with greater effectiveness.

Critical to the emergence of specialization is the constancy of an environment [3]. By occupying the same environment for an extended period of time, the advantage of maintaining fitness in alternate environments is not realized. Organisms may thus be expected to become increasingly specialized to their current, static environment. The most obvious form of specialization is a decline in the ability to utilize resources or exist in conditions that were not experienced during adaptation. Many such examples have been observed during laboratory evolution whereby organisms evolved to narrow or shift their range of preferred temperatures [4], carbon sources [5,6], host organisms [7], or even laboratory water supply [8]. Alternatively, organisms can lose the ability to grow well in the absence of a resource that is currently available. Although we are unaware of an experimental demonstration of a novel change to dependence upon a supplied resource, this has been dramatically and repeatedly observed for microbes infecting hosts. These microbes commonly lose the ability to synthesize a substantial number of components essential for growth because these are available in the host environment, such as the cytoplasm [9] or mouse gut [10].

One of the most prominent examples of prolonged adaptation to a single environment is the Lenski Long-Term Evolution Experiment (LTEE) [11]. In this experiment, 12 populations of *Escherichia coli* were founded with either the

arabinose-negative strain REL606 (populations A-1 to A-6) or the otherwise isogenic arabinose-positive derivative, REL607 (A+1 to A+6). These have evolved since 1988 in Davis-Mingioli (DM) minimal media [12] batch cultures containing glucose as a growth substrate. Over 50,000 generations, the fitness of the evolved strains in the evolutionary environment has increased substantially, and in fact continues to do so [13]. In line with the expectation of a generalist-specialist tradeoff, evolving strains have also specialized for aspects of their static environment. Evolved isolates have lower fitness in some alternative environments [4, 5, 14, 15], despite the fact that many individual mutations were generally beneficial across environments [16].

Perhaps the most surprising adaptive change to have occurred during the LTEE was the huge increase in fitness of one population due to evolving the ability to metabolize citrate, the "inert" metal chelator present in DM minimal medium. Disodium citrate (which we will hereafter refer to as citrate) was included in the evolutionary growth media as a historical artifact of the media's original formulation for penicillin enrichment of auxotrophs [12, 17]. The common use of citrate in minimal media formulations owes to its ability to serve as a chelator of Fe (III). Indeed, no direct addition of iron is made to DM minimal media. Given the quite low level of glucose used in the LTEE (25 mg/L = 0.14 mM), substantially more citrate was present (1.7 mM) than glucose. A diagnostic trait of most *E. coli* strains is that they can only grow on citrate anaerobically, whereas *Salmonella*, for example, can grow on it aerobically. This inability of the ancestral

strain to grow on citrate during the aerobic conditions of the LTEE therefore initially rendered this organic acid an unavailable secondary resource. Incredibly, after 31,000 generations – and by just one of the 12 replicate populations (A-3) – the ability emerged to a utilize citrate as a sole carbon source during aerobic growth [18]. Interestingly, this was not even the first time *E. coli* has acquired the ability to aerobically utilize citrate. $Cit^+$ *E. coli* K12 had previously been observed to arise spontaneously [19] or via the high expression of the anaerobic citrate transporter on multi-copy plasmids in *E. coli* B [20].

Besides the remarkable story of novel aerobic utilization of citrate, it is unclear whether any of the LTEE lines may have changed in terms of use of citrate as an iron chelator. Bacteria commonly have multiple acquisition systems for various metals, particularly iron, which is required for aerobic respiration due to the hemes found in cytochromes. The acquisition of iron is particularly challenging for microbes because of the low solubility of the Fe (III) species that is available at neutral pH in oxygenated environments. Consequently, metals are often growth limiting in nature in environments ranging from the open ocean [21], to host infections [22]. In the context of infections, iron is sequestered by high-affinity eukaryotic proteins, and the pathogenesis of many infectious diseases relies on the ability of bacteria to strip iron from their host.

Consistent with the fundamental role that iron plays in microbial growth, and its relative scarcity, microbes have developed an arsenal of techniques to procure it. In *E. coli*, transcriptional regulators sensitive to the intracellular concentrations

of iron down-regulate iron uptake genes when supplies are adequate. Under conditions of iron deprivation, these same regulators simultaneously up-regulate iron acquisition systems while down-regulating proteins requiring iron [23, 24]. One mechanism to obtain iron is to secrete and reabsorb small molecules called siderophores that chelate extracellular Fe (III), as well as the transporters to utilize them. *E. coli* also has the ability to take up Fe (II), and a transport system to directly capture Fe (III) from host proteins like transferrin or lactoferrin, or from heme [25].

There is precedent for selection acting upon metal acquisition during experimental evolution. Growing *Methylobacterium* at low levels of cobalt repeatedly selected for transposition events upstream of a single cobalt transporter, leading to increased uptake rates [26]. Interestingly, the selective effect of these mutations was dependent upon both the carbon source and the growth rate of cells. Additionally, there have been mutations in metal acquisition as a social trait. Genotypes of *Pseudomonas aeruginosa* that fail to produce siderophores have emerged during adaption in laboratory conditions, as well as over the course of infection in the lungs of cystic fibrosis patients [27, 28]. Because excreted siderophores become public goods, in well-mixed environments, non-producing cheaters can have a selective advantage over producers [29].

Here we investigated whether other strains from generation 50,000 of the Lenski LTEE have evolved to become dependent on citrate for their performance growing on glucose. There are four non-exclusive hypotheses for the dependence of glucose growth of each population upon citrate. *H0:* The null hypothesis is that there is no significant stimulation (or perhaps even inhibition), *H1:* As disodium citrate is the sole source of sodium in the LTEE media, there may be stimulation by sodium ions in a manner independent of citrate itself, *H2:* Evolved strains use – at least partially – citrate as a growth substrate (e.g., the $Cit^+$ A-3 population), and *H3:* Evolved strains have come to rely upon citrate to chelate iron present in the media.

We found that the ancestors and most of the evolved populations were largely insensitive to the presence of citrate (H0). In contrast, however, three lineages in addition to the $Cit^+$ A-3 population have evolved increased growth rate and yield in the presence of citrate. In no case was this was due to sodium ions (H1). Unlike the $Cit^+$ A-3 (H2), however, these lineages neither grew on citrate directly nor incorporated isotopic label from citrate during growth on 100% U-$^{13}$C-glucose. In contrast, these three lineages were at least partially rescued by the direct addition of Fe (II). These data are consistent with H3: these three populations have evolved to rely on citrate for its original intent, as a chelator of iron. As such, these three populations have evolved a novel dependence upon citrate due to

reduction in citrate-independent glucose growth rather than the gain of growth on citrate as a carbon and energy source.

## Results

*Evolved strains acquired dependence on citrate*

In order to test for acquired dependence upon citrate (i.e., rule out H0), we compared growth rate and yield on glucose either with or without citrate for the ancestors and 12 clonal isolates from the 50,000 generation LTEE populations (Figure 2.1). Growth rate is known to be strongly correlated with fitness in batch cultures in general, and in the LTEE in particular [30]. Yield, while not a component of fitness during batch culture, conveys additional information about cell physiology and was therefore also analyzed. Differences seen in apparent final yield was not explainable by the changed osmolality due to the addition of citrate to the media; citrate caused no detectable change in optical density ($p=0.95$, two-tailed paired T test). Both 'small' and 'large' clones from the A-2 population were included [31] in the analysis.

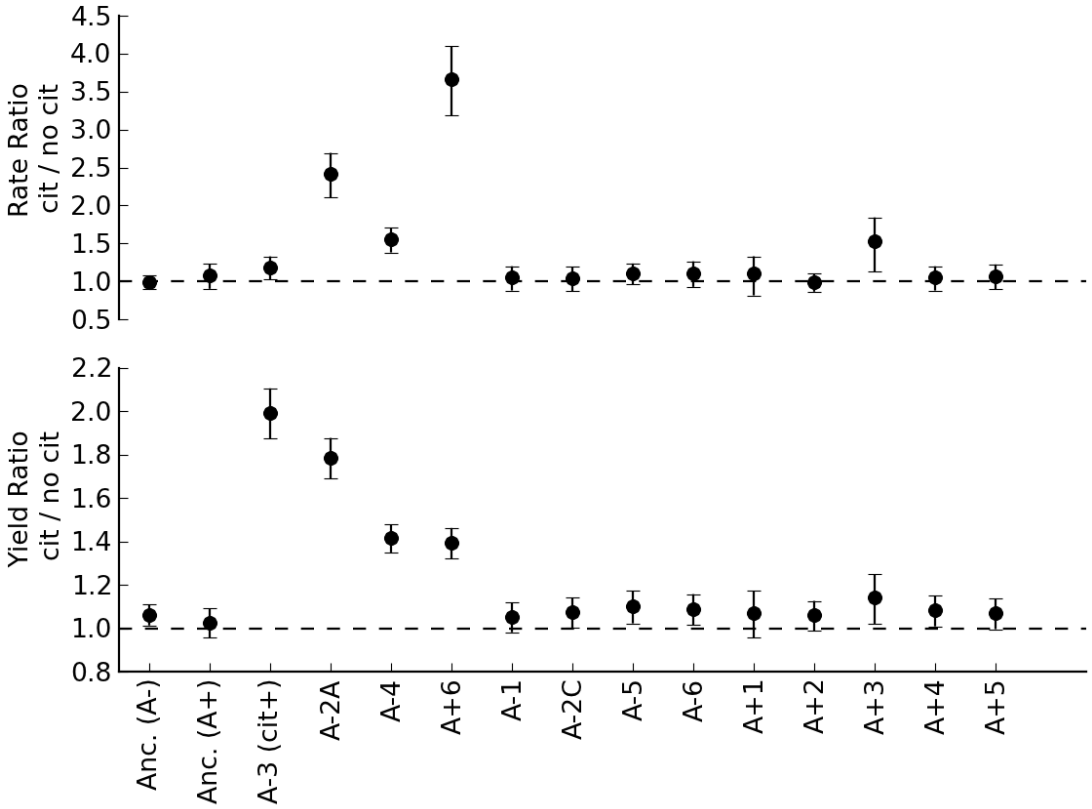



**Figure 2.1: Dependence of growth on glucose upon the presence of citrate for the ancestors and evolved strains.** The ratio and error for growth parameters was calculated by fitting a log-linear model with the strain:media interaction terms and block as the dependent variables. Unlike the ancestors (first two values), strains from three of the evolved populations were strongly stimulated by citrate in terms of both (A.) growth rate and (B.) yield. The remaining nine Cit$^-$ lineages, like their ancestors, exhibit little to no dependence upon citrate. We tested both 'large' (A-2A) and 'small' (A-2C) clones from the A-2 population, as it is known to have a stable, long-term polymorphism. Values represent the mean and 95% confidence intervals for the ratio of growth on glucose with citrate to growth without citrate for a given strain.

The Cit$^+$ A-3 population was seen to grow faster and to a much higher final yield in the presence of citrate, as previously reported [18]. The growth rate of the ancestor REL606 was neither significantly decreased nor enhanced by the addition of citrate (P=0.71, two-tailed Welch's T test unless stated otherwise), but there was a very slight, but significant increase in yield (P=0.0001). This difference in yield was not significant for the REL607 ancestor, for which we ran fewer replicates. Nine of the 13 evolved strains had similarly small increases in rate or yield. While in some cases there were statistically significant increases, the effect was small. For the remaining three strains, however, the increase was much greater for both rate and yield. In the presence of citrate, A-2A, A-4 and A+6 all experienced significant increases in both rate and yield (Respectively, rate: $P=1.6x10^{-7}$, $P=5.3x10^{-13}$, and $3.1x10^{-12}$, yield $1.6x10^{-14}$, $1.7x10^{-7}$, and $1.8x10^{-9}$). Indeed, these effects were quite substantial: the growth rate of A+6 with citrate was 3.75 times faster than growth without citrate, and the final yield of A-2A increased by 80% by the addition of citrate. In comparison, at the 1mM concentration of glucose used, the Cit$^+$ A-3 strain grew 1.2 times faster and increased in final yield by 100% with the addition of citrate. To explore the nature of the citrate dependence, we focused on these three strains for which the effect was greatest, and omitted strains for which the effect was marginal (as well as the Cit$^+$ A-3).

*Disodium citrate dependence is not driven by sodium*

One possible explanation for the effect of citrate could be stimulation by the addition of sodium (3.4 mM $Na^+$ cations), as disodium citrate was the only source of the cation in the medium (H1). In order to test for this possible effect, we compared the stimulatory effect of disodium citrate to citric acid of the same concentration (which was pH-normalized with KOH). Both growth rates and yields with citric acid for the three citrate-stimulated evolved strains were nearly identical to those from disodium citrate (Figure 2.2). The only differences that were statistically distinguishable were those for yield for A-2A (P=0.016) and A-4 (P=0.004), which were in fact slightly higher on citric acid than on disodium citrate. This suggests that the growth phenotype we see in these evolved strains is related to citrate rather than a side-effect of sodium.

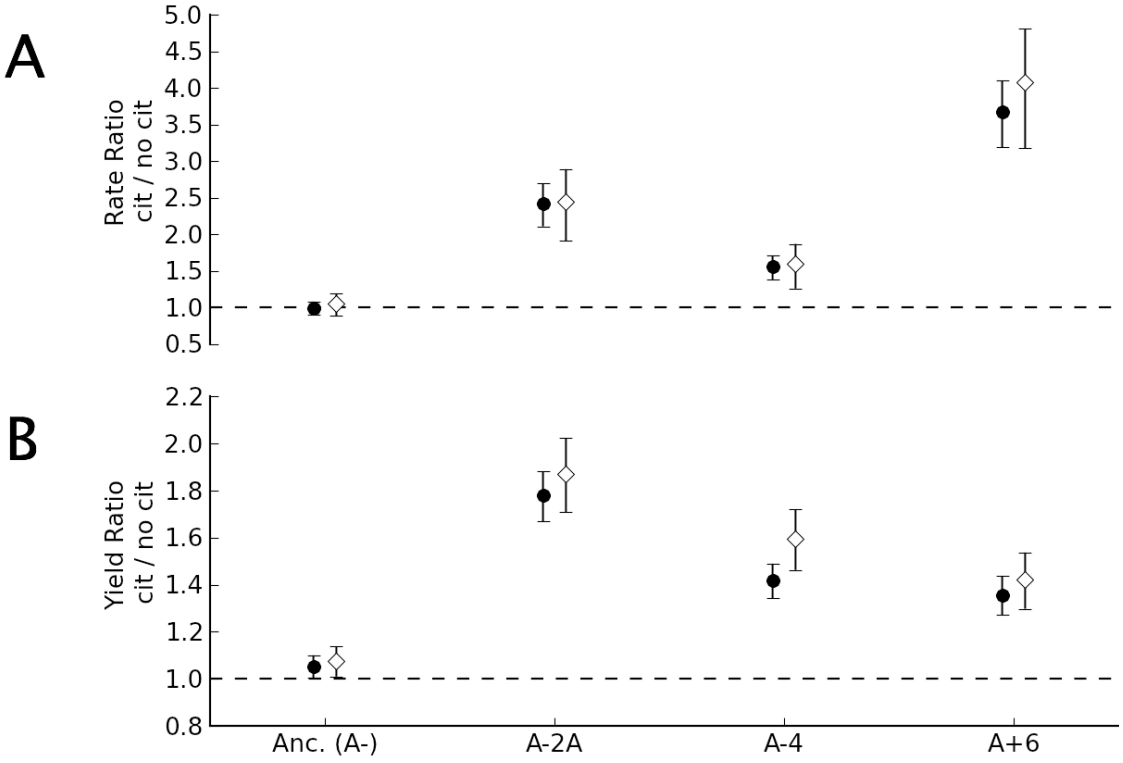



**Figure 2.2: Dependence of growth on glucose is due to citrate itself rather than the coincidental addition of sodium.** The ratio and error for growth parameters was calculated by fitting a log-linear model with the strain:media interaction terms and block as the dependent variables. Both the relative (A.) growth rate and (B.) yield on glucose for the three lineages that were strongly stimulated by the addition of sodium citrate (black circles) can be achieved by simply adding citric acid (white diamonds). Values represent the mean and 95% confidence intervals for the ratio of growth of that strain on glucose with the supplement to growth without.

*Evolved strains are unable to utilize citrate as a carbon source*

Given that citrate stimulation of glucose growth was not due to sodium, we next examined whether or not strains could directly metabolize citrate (H2). First, we asked whether any of these three citrate-stimulated strains grow with citrate as a sole carbon source. When cultured in DM containing citrate but no glucose, there was no measurable increase in OD after 24 hours (A-2A P=0.45, A-4 P=0.40, A+6 P=0.72). Thus, unlike the $Cit^+$ A-3 population, none of these isolates appear to be capable of growth on citrate alone.

Although the three citrate-stimulated evolved isolates were found to still be $Cit^-$, this result did not rule out possible co-metabolic use during growth on glucose. To explore this possibility, we determined whether carbon from the citrate was incorporated into biomass during growth on glucose. We grew cells in 100% U-$^{13}C$ labeled glucose at 1 mM, both with and without unlabeled (i.e., ~99% $^{12}C$) citrate, and looked for differential incorporation of $^{13}C$ into the amino acids of the growing cells under these two conditions. Using gas chromatography-mass spectrometry (GC-MS), we compared the mass distribution vectors for fragments of derivatized amino acids from the cultures (Figure 2.3). Since $CO_2$-utilizing reactions represent a source of naturally occurring carbon, it is expected to observe less than 100% $^{13}C$ in amino acid fragments. If citrate were co-utilized it would enter directly into the citric acid cycle and be incorporated into biomass,

which would dilute the incorporation of $^{13}C$ from glucose into amino acids. We compared the peaks corresponding to fully-labeled fragments using a Wilcoxon signed-rank test. For the $Cit^+$ A-3 population, there was a significant difference in incorporated label in the presence of unlabeled citrate (P=0.0006), but for the other citrate-dependent strains there was no significant difference (A-2A: P=0.33, A-4: P=0.73, A+6: P=0.47) (Figure 2.3B-C). These data confirm that these three strains are not utilizing citrate as a carbon source to any measurable extent.

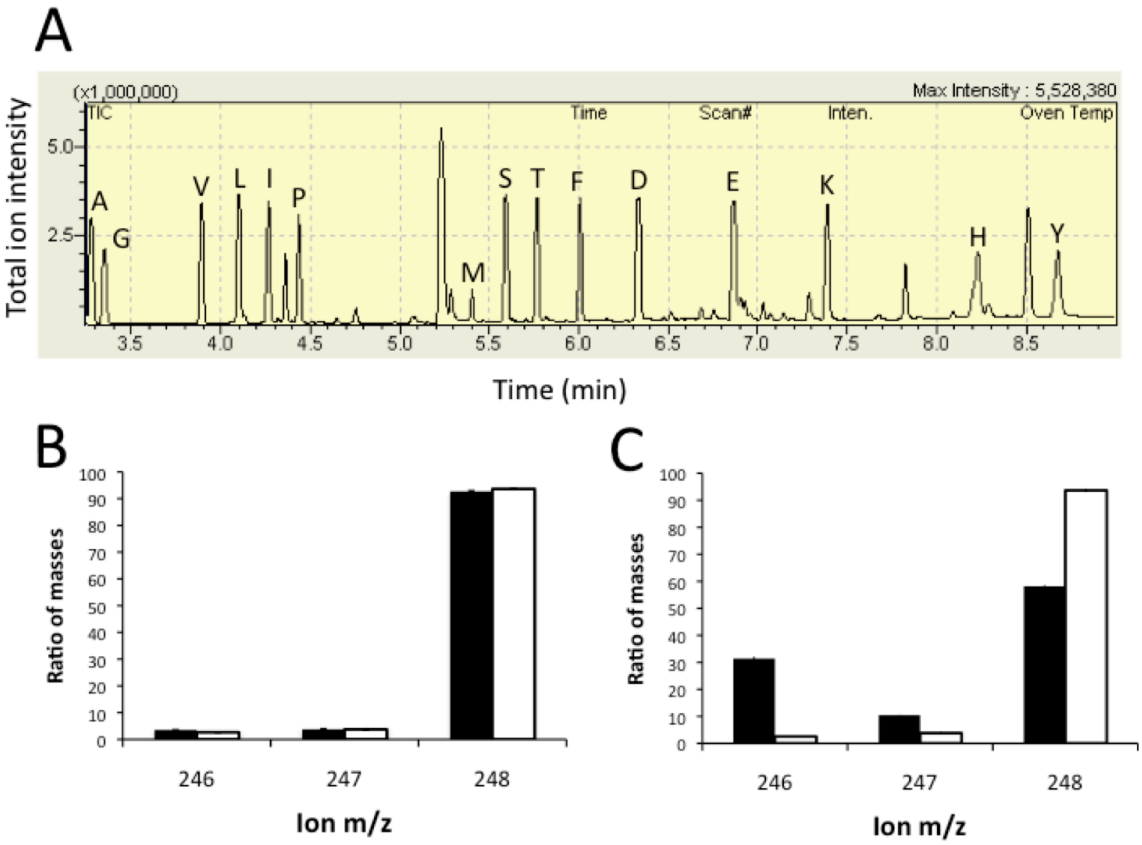



**Figure 2.3: GC-MS technique to determine if citrate was incorporated into the amino acids of biomass during glucose growth.** Amino acids from cells grown on 100% U-$^{13}$C-glucose with or without citrate were analyzed via GC-MS. (A.) A representative GC-MS chromatogram of total ion intensity indicating the presence and separation of derivatized amino acids. (B) Mass spectra for an example fragment (from glycine) when A+6 was grown on glucose with citrate (black) or without (white), and (C) when the citrate-consuming A-3 was grown with and without citrate. For each amino acid, one or more characteristic mass fragments are observed whose mass distribution spectra

**Figure 2.3, continued:** shows the relative amount of fully $^{13}C$-labeled fragment (peak at 248 for the M-57 fragment of glycine) compared to total of all ions for that fragment (sum of peaks 246 to 248). Across all amino acids, there was not significant decrease in $^{13}C$-labeling from glucose except in the REL-3 line where unlabelled citrate was clearly incorporated into biomass. Error bars represent standard deviations of 3 biological replicates.

*Citrate dependence is related to its role as an iron chelator*

Having rejected hypotheses H0-H2 for the three citrate-dependent strains, we tested the remaining hypothesis that the effect of citrate is related to its role as a chelator in iron acquisition (H3). *E. coli* can take up ferrous iron with the Feo and Efe transporters without the involvement of citrate [25]. We first tested whether providing sufficient iron (as ferrous iron to 10 μM) in the absence of citrate would enhance growth on glucose of the three citrate-dependent evolved isolates (Figure 2.4). For one of the citrate-dependent strains, A-4, the iron greatly increased the growth rate and yield to values indistinguishable from that achieved via citrate (P=0.06 and P=0.11, respectively). A similar effect was seen for A-2A, where iron increased growth rate and yield to ~90% that achieved by citrate, a small, but significant difference (P=0.008 and $P=5.6x10^{-5}$, respectively). For the third strain, A+6, the direct addition of iron significantly increased rate ($P=5.6x10^{-8}$) and yield (P=0.0008), but this increase amounted to only ~40% of the yield increase and ~13% of the rate increase seen with the addition of citrate.

One possible explanation for the evolved strains' acquired dependence upon citrate would be a defect in either production or uptake of iron chelating siderophores. To test whether the citrate-dependent strains had become deficient in their ability to secrete siderophores, we grew the three citrate-dependent evolved isolates in media without citrate that had been conditioned by the growth of the ancestral REL606 and then re-supplemented with glucose. Unlike what has been observed previously for siderophore-defective strains [32,33], none of the citrate-dependent strains grew faster or to a higher final yield on the spent media than on DM without citrate (at an 0.05 level, Welch's 1-tail T test) (Figure 2.4). This argues against defects in siderophore production as the sole cause of citrate dependency, but does not rule out the possibility that it is related to the ability to transport iron-siderophore complexes.

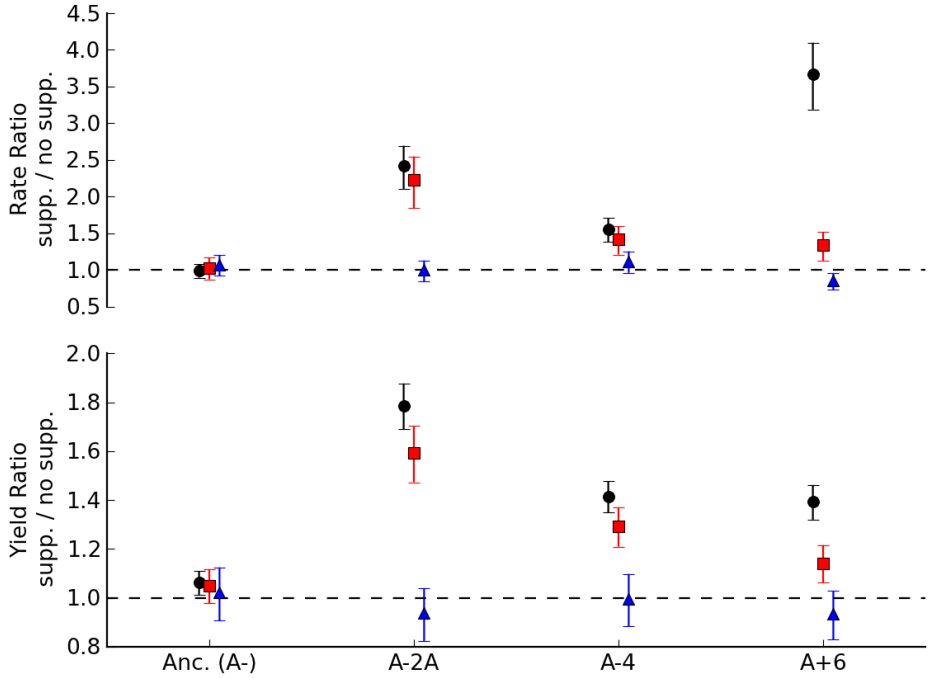



**Figure 2.4: Direct addition of iron, but not spent medium, stimulates growth on glucose for citrate-dependent evolved lineages.** The ratio and error for growth parameters was calculated by fitting a log-linear model with the strain:media interaction terms and block as the dependent variables. Both the relative (A.) growth rate and (B.) yield of the three strains from populations stimulated by sodium citrate (black circles) are at least partially enhanced by direct addition of 10 μM Fe (II) (as ammonium iron sulfate, red squares). In contrast, growth in spent media re-supplemented with glucose (blue triangles) did not stimulate growth. Values represent the mean and 95% confidence intervals for the ratio of growth of that strain on glucose with the supplement to growth without.

## Discussion

Here we report that by 50,000 generations, three strains from the 11 LTEE populations not previously shown to be dependent on citrate now grow on glucose in a manner that is strongly dependent on the presence of citrate. Thus, eight of the populations have retained the ancestral phenotype of marginal - if any - stimulation by citrate (H0 above) and one has famously become $Cit^+$ and can readily grow on it as a sole carbon and energy source (H2). For the remaining three populations whose glucose-growth was substantially enhanced by citrate, our experiments allowed us to reject both the hypothesis that this stimulation was due to a side-effect of adding sodium (H1) as well as that it was due to co-metabolism during glucose growth (H2). The remaining hypothesis – citrate as an iron chelator (H3)– appears to best explain this phenotype, as direct addition of iron in the absence of citrate stimulated glucose growth.

What physiological mechanism may have caused this evolved dependence on citrate? The ability of ferrous iron to largely restore the phenotype of rapid growth on glucose is the strongest evidence for a loss of function in metal acquisition. The primary iron acquisition system for *E. coli* under aerobic conditions is via siderophores [25]. The fact that citrate-dependent strains do not benefit from growth on media conditioned by their citrate-independent ancestor, however, suggests that their dependence is unlikely to be due to an inability to produce or secrete siderophores. The most plausible explanation for their citrate-dependence would therefore be mutations involved in siderophore uptake. Since

strains A-2A and A-4 were almost completely rescued by ferrous iron, they appear to have no defects in their systems for direct iron uptake. On the other hand, A+6 was only mildly stimulated by ferrous iron, which would suggest this strain either has additional mutations in genes related to direct ferrous iron transport, defects in global regulators that stimulate up-regulation of iron-acquisition genes under iron starvation [24], and/or another non-iron-related dependency on the presence of citrate. Given the large number of genes involved in iron transport (the ferric uptake regulon alone consists of ~30 iron-transport-related genes [34]), the most useful first step for answering these questions will come from analysis of candidate causal mutations from the genome sequences from these lineages, which are in progress (personal communication, R.E. Lenski & D. Schneider).

At this point, it is not possible to conclude whether the new citrate dependence during growth on glucose is the result of adaptation or the accumulation of mutations. Previous experiments in the LTEE have suggested that the predominant mechanism for tradeoffs in the evolved strains has been antagonistic pleiotropy, and not the accumulation of mutations [5]. However, since this initial study, 30,000 more generations have passed, and many more mutations have occurred. In the genome sequence published for the A-1 lineage, for example, there were only 29 nucleotide substitutions after the first 20,000 generations, but 598 more accumulated in the subsequent 20,000 generations. This striking increase in mutation rate was due to acquiring a mutator phenotype

during the latter period [13]. In general, the respective impacts of mutation accumulation versus antagonistic pleiotropy should be revisited given the increased mutation rate in many of the strains, and the amount of time that has passed.

Though definitive statements cannot be made, mutation accumulation seems like a plausible cause of the observed acquisition of citrate dependence. Iron transport genes were presumably superfluous in the constant presence of high levels of citrate for 50,000 generations. Additionally, the large number of genes involved in iron uptake would provide a reasonably large mutational target. Finally, all three of the strains that show strong citrate dependence during growth on glucose are amongst those that have also acquired a mutator phenotype; mutations to *uvrD*, *mutL*, and *mutS* have increased their mutation rates 100-1000 fold [5, 35]. This result in itself is not statistically significant, as 6 of the parallel populations have acquired this phenotype by 50,000 generations, so the probability by chance that all three citrate-dependent strains are mutators is 0.1. The fact that they are all mutators, however, is consistent with the hypothesis that mutation accumulation drove the observed phenotypic changes. As with the above speculation about the identity of the mutational targets leading to citrate-dependency, the selective mechanism leading to their incorporation will require identification and introduction into the appropriate backgrounds to test the fitness of the dependency-inducing mutations.

## Conclusions

We have uncovered a second form of interaction with citrate that occurred during the 50,000 generations of the LTEE: three of the 12 $Cit^-$ strains we examine now require its presence as an iron chelator for maximal growth on glucose. This interaction is due to an acquired dependence on citrate to carry out a function that populations have lost, which contrasts with the gain of novel function observed in the $Cit^+$ A-3 population. Unlike the strong selection for the ability of A-3 to catabolize citrate, selection was likely either weak or absent in the loss of citrate-independent glucose performance. In this regard, the constant environment of the LTEE was perhaps not unlike that encountered along the evolutionary path to intracellular symbionts, whereby specialization emerges not so much from gain of new traits as from the persistent loss of traits rendered unnecessary.

A final subtext for considering the emergence of citrate-dependence during glucose growth is that subtle decisions in the design of evolution experiments can have unpredictable impacts. By addressing one complication in experimental design, one can unknowingly create new adaptive opportunities. Citrate was present in DM media from the start of the experiment despite a quite small, marginal difference between growth of the ancestor in its presence or absence. This seemingly insignificant detail – combined with tens of thousands of generations of adaptation – has led to unexpected phenomena that range from

the incredible acquisition of aerobic citrate metabolism, to the dependence on citrate as an iron chelator.

## Materials and Methods

*Strains and LTEE conditions*

*Escherichia coli* B isolates were obtained from the Lenski LTEE [11] after 50,000 generations. Briefly, 12 populations of *E. coli* were founded with either the arabinose-negative strain REL606 (populations A-1 to A-6) or the arabinose-positive derivative, REL607 (A+1 to A+6). These have been evolved since 1988 in 50 mL flasks containing 10 mL of modified Davis-Mingioli (DM) minimal media [36] (which we refer to as DM media) with 139 μM glucose (25 mg/L) as a growth substrate. One liter of DM consists of 7 g potassium phosphate (dibasic trihydrate), 2 g potassium phosphate (monobasic anhydrous), 1 g ammonium sulfate, 0.5 g disodium citrate, 1mL 10% magnesium sulfate, 1mL 0.2% thiamine (vitamin B1), and a carbon source- here glucose. These populations have been cultured at 37 °C while shaking at 120 rpm, and have been transferred daily via 1:100 dilutions (~6.64 net doublings per day).

The isolates analyzed in this experiment consisted of the ancestral lines REL606 and REL607, as well as the 'A' clone frozen at 50,000 generations from 11 of the 12 populations (A-1A: REL11330, A-2A: REL11333, A-3: REL11364, A-4A:

REL11336, A-5A: REL11339, A-6A: REL11389, A+1A: REL11392, A+2A: REL11342, A+3A: REL 11345, A+4A: REL11348, A+5A: REL11367, A+6A: REL11370). The A-2A clone is from the 'large' lineage that has coexisted with a cross-feeding 'small' lineage for tens of thousands of generations [31]. We therefore also examined an A-2C 'small' clone (REL11335).

Before growth and GC-MS experiments, acclimation cultures were inoculated from freezer stocks and grown for 24 hours in DM 1 mM glucose without citrate.

*Growth Rate and Yield Experiments*

To measure growth rate and yield, cultures were grown in 50 mL flasks containing 10 mL of DM minimal media (the same media as the evolution experiment) with 1 mM glucose. In order to avoid extremely low optical densities, growth assays were conducted at 1 mM, which is higher than the 25 mg/L = 0.14 mM that was used in the "$DM_{25}$" medium of the evolution experiment. It is, however, closer to the concentration of glucose in the evolutionary environment than either the 250mg/L or 1000mg/L that has previously been used for physiological assays on these strains [13, 30, 37]. DM+citrate media contained disodium citrate at a concentration of 1.7 mM, as normal. DM+iron was supplemented with ammonium iron sulfate, mixed fresh the day of the experiment, to 10 μM of Fe (II). Ammonium is already present in DM (as ammonium sulfate) at a concentration of 7.6mM, so 10 μM addition in

iron sulfate does not meaningfully change its concentration. DM+citric acid was DM supplemented with 1.7 mM citric acid and pH balanced with KOH. Spent media was produced by growing the ancestral strain REL606 in DM 1 mM glucose without citrate for 24 hours, centrifuging to remove the majority of the cells, and then filtering with a 0.2 μm filter to sterilize. Citrate was omitted in order to ensure that it did not suppress siderophore production. This spent media was supplemented to 1 mM glucose with filter-sterilized 10% glucose solution, and compared to DM without citrate identically supplemented with glucose.

For growth rate experiments, $OD_{600}$ was measured on a BioRad SmartSpec Plus (Philadelphia, PA) spectrophotometer every hour until there was a detectable increase in OD, then approximately every 30 minutes until stationary phase was reached. Yield was measured as $OD_{600}$ after 24 hours. Growth rate was calculated with Growth Curve Fitter, an in-house software package that fits exponential curves using points in the log-linear range of observations (N.F. Delaney & CJM, unpublished). To determine growth rates, we subtracted the blank (optical density at t=0), and fitted all points after a change in OD was measurable until the point before stationary phase was reached (4-6 points). The exception was for the Cit+ A-3 strain, for which the growth rate was fitted for growth on only the glucose portion of the diauxic growth curve. Between 4 biological replicates (for citrate-insensitive strains) and 19 (for the ancestor) were run for each condition. The statistical difference between growth parameters for different conditions was

calculated directly from the data using Welch's T test. In order to plot the relationship between the data in an intuitive way as a ratio, the ratio and error for growth parameters was calculated by fitting a log-linear model with the strain:media interaction terms and block as the dependent variables using R (R-Project software).

*GC-MS analysis of incorporation of carbon from glucose versus citrate*

Incorporation of $^{13}C$ was measured as previously described [38]. Strains were grown with or without citrate in 100 mL of Davis minimal media with 1 mM U-$^{13}C$ glucose (Cambridge Isotope Laboratories, Andover, MA). At stationary phase, all cells were pelleted from the media, hydrolyzed overnight in 6 M HCl, and dried. The dry cell material was then derivatized for an hour at 85 °C with 40 μL each of dimethylformamide (DMF) and N-tert-butyldimethylsilyl-N-methyltrifluoroacetamide with 1% tert-butyldimethyl-chlorosilane (TBDMSTFA). The derivatized cell material was injected into a Shimadzu QP2010 GC-MS (Columbia, MD). The injection source was 230 °C. The oven was held at 160 °C for 1 min, ramped to 310 °C at 20 °C $min^{-1}$, and finally held at 310 °C for 0.5 min. The flow rate was 1 mL $min^{-1}$ and the split was 10. The column was a 30 m Rx1-1ms (Restek, Bellefonte, PA). Three biological replicates were run for each isolate (except A-2A in which one run failed).

In all samples the percent of each amino acid that was uniformly labeled with $^{13}$-C molecules was determined using FiatFlux qualitycheck [38]. For each amino acid

fragment that was detected in all biological replicates, the difference in average percent of uniformly labeled fragment was determined for samples grown with and without citrate. Finally we used a non-parametric Wilcoxon signed-rank test to determine whether there was a significant difference in labeling between treatments.

*Authors' Contributions*

NL, WRH and CJM conceived and designed the study. NL and WRH conducted the experiments and analyzed data. NL, WRH, and CMJ wrote the manuscript. All authors read and approved the final manuscript.

*Acknowledgements*

We are very grateful to Richard Lenski and Neerja Hajela for providing us with the 50,000 generation isolates of the long-term evolved *E. coli*, and thank members of the Marx laboratory for their helpful comments on the manuscript. WRH was supported by NIH grant F32GM090760 and NL by NIH grant GM080177.

## References

1. Futuyma D (1988) The evolution of ecological specialization. Annual Review of Ecology and Systematics 19: 207-233.

2. Law R (1979) Optimal life histories under age-specific predation. Am Nat 114: 399-417.

3. Kassen R (2002) The experimental evolution of specialists, generalists, and the maintenance of diversity. J Evolution Biol 15: 173-190.

4. Bennett AF, Lenski RE (2007) An experimental test of evolutionary trade-offs during temperature adaptation. Proc Natl Acad Sci USA 104 Suppl: 8649-54.

5. Cooper VS, Lenski RE (2000) The population genetics of ecological specialization in evolving *Escherichia coli* populations. Nature 407: 736-9.

6. Lee M-C, Chou H-H, Marx CJ (2009) Asymmetric, bimodal trade-offs during adaptation of *Methylobacterium* to distinct growth substrates. Evolution 63: 2816-30.

7. Crill WD, Wichman HA, Bull JJ (2000) Evolutionary reversals during viral adaptation to alternating hosts. Genetics 154: 27-37.

8. O'Keefe KJ, Morales NM, Ernstberger H, Benoit G, Turner PE (2006) Laboratory-Dependent Bacterial Ecology: a Cautionary Tale. Appl Environ Microbiol 72: 3032-3035.

9. Moran NA (2002) Microbial minimalism: genome reduction in bacterial pathogens. Cell 108: 583-6.

10. Giraud A, Matic I, Tenaillon O, Clara A, Radman M, Fons M, & Taddei F (2001) Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. Science 291: 2606-8.

11. 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: 1315-41.

12. Davis B (1949) The isolation of biochemically deficient mutants of bacteria by means of penicillin. Proc Natl Acad Sci USA 35: 1-10.

13. Barrick J, Yu D, Yoon S, Jeong H, Oh T (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature 461: 1243-7.

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

15. Cooper VS, Schneider D, Blot M, Lenski RE (2001) Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. J Bacteriol 183: 2834-41.

16. Ostrowski E a, Rozen DE, Lenski RE (2005) Pleiotropic effects of beneficial mutations in *Escherichia coli*. Evolution 59: 2343-52.

17. Davis B, Mingioli E (1950) Mutants of *Escherichia coli* requiring methionine or vitamin B12. J Bacteriol 60: 17-28.

18. 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 USA 105: 7899-906.

19. Hall BG (1982) Chromosomal mutation for citrate utilization by *Escherichia coli* K-12. Journal of Bacteriology 151: 269-73.

20. Pos KM, Dimroth P, Bott M (1998) The *Escherichia coli* citrate carrier CitT: a member of a novel eubacterial transporter family related to the 2-oxoglutarate/malate translocator from spinach chloroplasts. J Bacteriol 180: 4160-5.

21. Hopkinson BM, Mitchell BG, Reynolds RA, Wang H, Selph KE, Measures CI, Hewes CD, Holm-Hansen O, Barbeau KA (2007) Iron limitation across chlorophyll gradients in the southern Drake Passage: Phytoplankton responses to iron addition and photosynthetic indicators of iron stress. Limnol Oceanogr 52: 2540-54.

22. West SA, Buckling A (2003) Cooperation, virulence and siderophore production in bacterial parasites. Proc R Soc Lond Biol 270: 37-44.

23. Bagg A, Neilands JB (1987) Molecular mechanism of regulation of siderophore-mediated iron assimilation. Microbiol Rev 51: 509-18.

24. McHugh JP, Rodríguez-Quinoñes F, Abdul-Tehrani H, Svistunenko DA, Poole RK, Cooper CE, Andrews SC (2003) Global iron-dependent gene regulation in *Escherichia coli*. A new mechanism for iron homeostasis. *The* J Biol Chem 278: 29478-86.

25. Braun V, Killmann H (1999) Bacterial solutions to the iron-supply problem. TIBS 24: 104-9.

26. 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 Genetics 5: e1000652.

27. Harrison F, Buckling A (2007) High relatedness selects against hypermutability in bacterial metapopulations. Proc R Soc Lond Biol 274: 1341-7.

28. De Vos D, De Chial M, Cochez C, Jansen S, Tümmler B, Meyer JM, Cornelis P (2001) Study of pyoverdine type and production by *Pseudomonas aeruginosa* isolated from cystic fibrosis patients: prevalence of type II pyoverdine isolates and accumulation of pyoverdine-negative mutations. Arch Microbiol 175: 384-8.

29. Griffin AS, West SA, Buckling A (2004) Cooperation and competition in pathogenic bacteria. Nature 430: 1024-7.

30. Vasi F, Travisano M, Lenski RE (1994) Long-Term Experimental Evolution in *Escherichia coli*. II. Changes in Life-History Traits During Adaptation to a Seasonal Environment. Am Nat: 432–56.

31. Rozen D, Lenski R (2000) Long-Term Experimental Evolution in *Escherichia coli*. VIII. Dynamics of a Balanced Polymorphism. Am Nat 155: 24-35.

32. Jadhav RS, Desai A (1994) Role of siderophore in iron uptake in cowpea *Rhizobium* GN1 (peanut isolate): Possible involvement of iron repressible outer membrane proteins. FEMS Microbiol Lett 115: 185-90.

33. Dale S, Doherty-Kirby A, Lajoie, G, Heinrichs, DE (2004) Role of Siderophore Biosynthesis in Virulence of *Staphylococcus aureus*: Identification and Characterization of Genes Involved in Production of a Siderophore. Infect Immun 72: 29-37.

34. Karp PD, Riley M, Saier M, Paulsen IT, Collado-Vides J, Paley SM, Pellegrini-Toole A, Bonavides C, Gama-Castro S (2002) The EcoCyc Database. Nucleic Acids Res 30: 56-8.

35. Sniegowski PD, Gerrish PJ, Lenski RE (1997) Evolution of high mutation rates in experimental populations of *E. coli*. Nature 387: 703–5.

36. Carlton B, Brown B (1981) Gene Mutation. In *Manual of Methods for General Bacteriology* edited by Gerhardt P. Washington DC: American Society for Microbiology: 222-242.

37. Rozen DE, Philippe N, de Visser JA, Lenski RE, Schneider D (2009) Death and cannibalism in a seasonal environment facilitate bacterial coexistence. Ecol Lett 12: 34-44.

38. Zamboni N, Fendt S-M, Rühl M, Sauer U (2009) (13)C-based metabolic flux analysis. Nat Protoc 4: 878-92.

# Chapter 3

# Metabolic erosion primarily through mutation accumulation, and not tradeoffs, drives limited evolution of substrate specificity in *Escherichia coli*

Nicholas Leiby[1,2] and Christopher J. Marx[1,3,*]

1 Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA, 2 Systems Biology Graduate Program, Harvard University, Cambridge, MA, USA, 3 Faculty of Arts and Sciences Center for Systems Biology, Harvard University, Cambridge, MA, USA.

* Corresponding author

This chapter was originally published in PLoS Biology.

*I performed all experiments and analysis for this manuscript, and was its primary author.*

Evolutionary adaptation to a constant environment is often accompanied by specialization and a reduction of fitness in other environments. We assayed the ability of the Lenski *Escherichia coli* populations to grow on a range of carbon sources after 50,000 generations of adaptation on glucose. Using direct measurements of growth rates, we demonstrated that declines in performance were much less widespread than suggested by previous results from Biolog assays of cellular respiration. Surprisingly, there were many performance increases on a variety of substrates. In addition to the now famous example of citrate, we

observed several other novel gains of function for organic acids that the ancestral strain only marginally utilized. Quantitative growth data also showed that strains with higher mutation rate exhibited significantly more declines, suggesting that most metabolic erosion was driven by mutation accumulation and not by physiological tradeoffs. These reductions in growth by mutator strains were ameliorated by growth at lower temperature, consistent with the hypothesis that this metabolic erosion is largely caused by destabilizing mutations to the associated enzymes. We further hypothesized that reductions in growth rate would be greatest for substrates used most differently from glucose, and we used flux balance analysis to formulate this question quantitatively. To our surprise, we found no significant relationship between decreases in growth and dissimilarity to glucose metabolism. Taken as a whole, these data suggest that in a single resource environment, specialization does not mainly result as an inevitable consequence of adaptive tradeoffs, but rather due to the gradual accumulation of disabling mutations in unused portions of the genome.

## Introduction

Evolving populations face the fundamental dilemma that there is no single phenotype that is optimal in all environments. When an evolving population occupies the same selective environment for an extended period of time, no advantage is realized by maintaining fitness on resources it no longer encounters. Adaptation to a selective environment can result in correlated responses in

alternative environments. While these can be synergistic improvements, a response that decreases fitness in other environments is known as specialization. This prevents the rise of 'Darwinian demons': single super-genotypes that are optimized across all conditions [1]. It is critical to understand the origin of specialization because it underlies the origin and maintenance of diversity- it is why 'the jack of all trades is a master of none' [2].

Specialization can result from either selective or neutral processes. Antagonistic pleiotropy describes when natural selection favors changes that are beneficial in the current environment but reduce function in other environments. Alternatively, specialization may result from mutations that decrease fitness in alternative environments that are neutral in the selective environment. These mutations have the potential to either drift or hitchhike to fixation via "mutation accumulation". As neutral mutations accrue in proportion to the mutation rate, the clearest evidence of mutation accumulation can come from excess specialization in mutator lineages, which contain defects in mutational repair that can elevate mutation rates ~100-fold [3]. In contrast, where specialization is rapid and occurs in parallel across lineages, a pattern commonly seen for adaptation itself, this has been cited as support of selection-driven antagonistic pleiotropy [4].

The experimental evolution of 12 populations of *Escherichia coli* grown for thousands of generations on a single substrate has been used to distinguish whether selective or neutral processes drive metabolic specialization [4]. The

populations were part of the Lenski Long-Term Evolution Experiment (LTEE) [5], in which wild-type *E. coli* B have been diluted 1:100 daily and re-grown in well-mixed medium containing glucose as the sole usable carbon source. After 20,000 generations (20k), competitive fitness on glucose had increased by ~70%. However, respiration assays in static 96-well Biolog plates (Hayward, CA) [6] suggested dramatic decreases in metabolic performance on alternative substrates. These declines occurred rapidly and in parallel across populations, coincident with the largest gains in fitness. This was suggested to indicate selection-driven antagonistic pleiotropy as the main mechanism of specialization. Furthermore, because there was only a weak, non-significant excess of declines by the populations which had become mutators earlier in the experiment [3], this suggested neutral mutation accumulation played little role, if any, in glucose specialization by 20k.

In this study we have readdressed the basis of specialization in the LTEE populations, motivated by our discovery that the growth rates of isolates in well-mixed media are poorly captured by assays of cellular respiration in static, proprietary media. Given this surprising finding, we analyzed the selected (i.e., glucose) and correlated responses of isolates from both 20k and 50,000 generations (50k) from four perspectives:

(i) Generality and parallelism of specialization- Is there an overall pattern of decreased growth performance on alternative substrates? Is the pace and pattern of declines in function consistent with antagonistic pleiotropy?

(ii) Novel gains of function- Besides the prior example of evolution to catabolize citrate present in the medium [7], are there any other substrates used by evolved isolates that are not utilized by the ancestors?

(iii) Role of elevated mutation rate driving mutation accumulation- Will the 30,000 generations that have passed since 20k have lead to sufficient mutation accumulation for the mutators to exhibit significantly greater decreases in growth than non-mutators?

(iv) Predictability of catabolic declines- Can we predict which substrates would be metabolized less effectively based upon the similarity of their use to glucose? Is there a pattern to this specialization that might suggest a common biophysical basis behind them for the mutator lineages?

## Results

*Respiration assays indicated broad declines in function across substrates, including those where adaptation occurred*

As a first step to readdressing specialization in the LTEE populations, we sought to replicate the Biolog respiration results at 20k presented by Cooper and Lenski [4], as well as extend this analysis to the populations at 50k. Despite changes in the Biolog assay itself since the previous study, we recovered a similar pattern for the panel of substrates (Supplementary Figure 3.1).

The validity of Biolog assays as a proxy for strain improvement came into question after finding decreases for the very substrates on which selection occurred. (Figure 3.1A). Despite abundant evidence of improvement from competitive fitness assays [4] and growth rates [8], respiration on glucose consistently decreased over the course of the experiment. Furthermore, although one lineage in the A-3 population evolved to utilize as a carbon source the citrate included in Davis Minimal (DM) medium [7], it produced a statistically indistinguishable respiration value from the other Cit$^-$ isolates at 50k (Supplementary Figure 3.2).

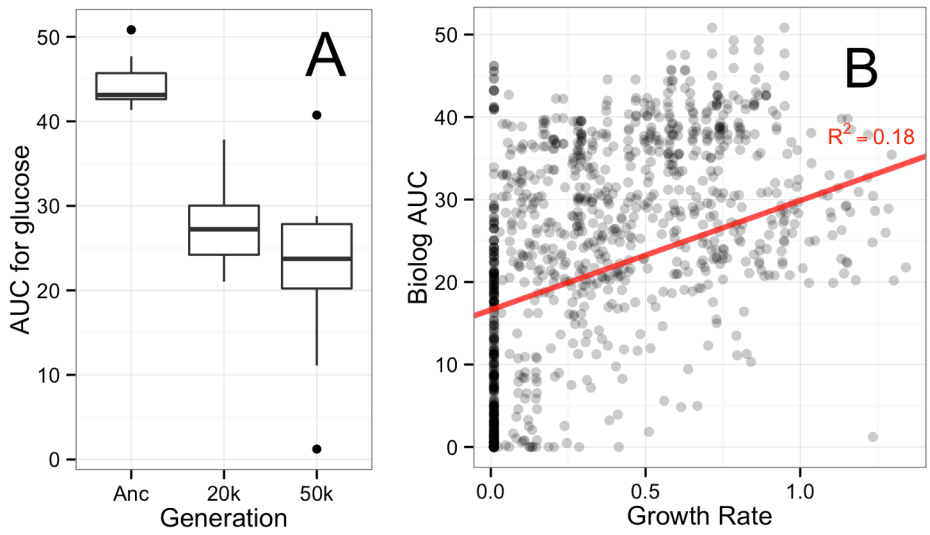



**Figure 3.1: Biolog measurements are a poor proxy for growth performance.** A) Biolog AUC as measured for the D-glucose on Biolog plates. The evolved strains have a lower AUC value than the ancestor on glucose, the carbon source available during evolution ($P<0.0001$, Welch's two sample T-test). The mean AUC for the 20k and 50k isolates on glucose are not statistically different. B) Scatter plot showing the measurement of function as Biolog AUC vs. growth rate on all substrates, for all strains at 20k and 50k generations as well as the ancestors. The regression shown is for substrates after removal of categorical disagreements (growth without respiration or respiration without growth, 167/702 in total).

*Growth rates reveal a balance of decreased and increased performance across substrates*

Given that the selective substrates with known fitness improvement had decreased cellular respiration, we turned to direct measurements of growth rates across a wide panel of substrates using a robotic growth analysis system [9,10]. By measuring growth rate we capture the demographic metric best-correlated with competitive fitness in the evolutionary environment [11]. As the LTEE was performed in shaken, fully-aerated flasks, these well-mixed 48-well plates were a closer match to the evolutionary environment than unshaken 96-well plates, as unshaken plates commonly exhibit sub-exponential growth due to oxygen limitation [10]. While Biolog uses a proprietary minimal media, for the growth rate measurements we used the same Davis Minimal (DM) media as the LTEE experiment. We omitted citrate, however, as this choice allowed us to include the $Cit^+$ A-3 population in our analysis. We chose carbon sources based upon the substrates for which significant, parallel decreases were previously observed via respiration assays [4], as well as citrate and several sugars on which growth tradeoffs were previously measured after 2,000 generations [12].

Comparing growth rates to respiration data, it becomes evident that the latter is not an accurate assay for growth (Figure 3.1B). There were many cases where respiration occurred without growth, as 156 out of the 702 strain/substrate combinations measured did not permit growth but did have measurable respiration- a known feature of Biolog assays [6]. Even after removing these

categorical disagreements, and a smaller number of instances of growth without respiration (11 strain/substrate combinations), Biolog respiration values were a poor predictor of growth rate ($R^2$=0.18, linear regression F-test(1,499)=108.8).

Growth rate data across substrates revealed a surprising degree of correlated gains in performance (Figure 3.2). Indeed, at 20k, there were actually more correlated increases in rate than decreases (165 vs. 99, respectively, P<0.0001 for binomial two-sided test with null of random gains and losses). By 50k, the picture had reversed, now with more decreases than increases (167 vs. 119, P=0.005, binomial two-sided test).

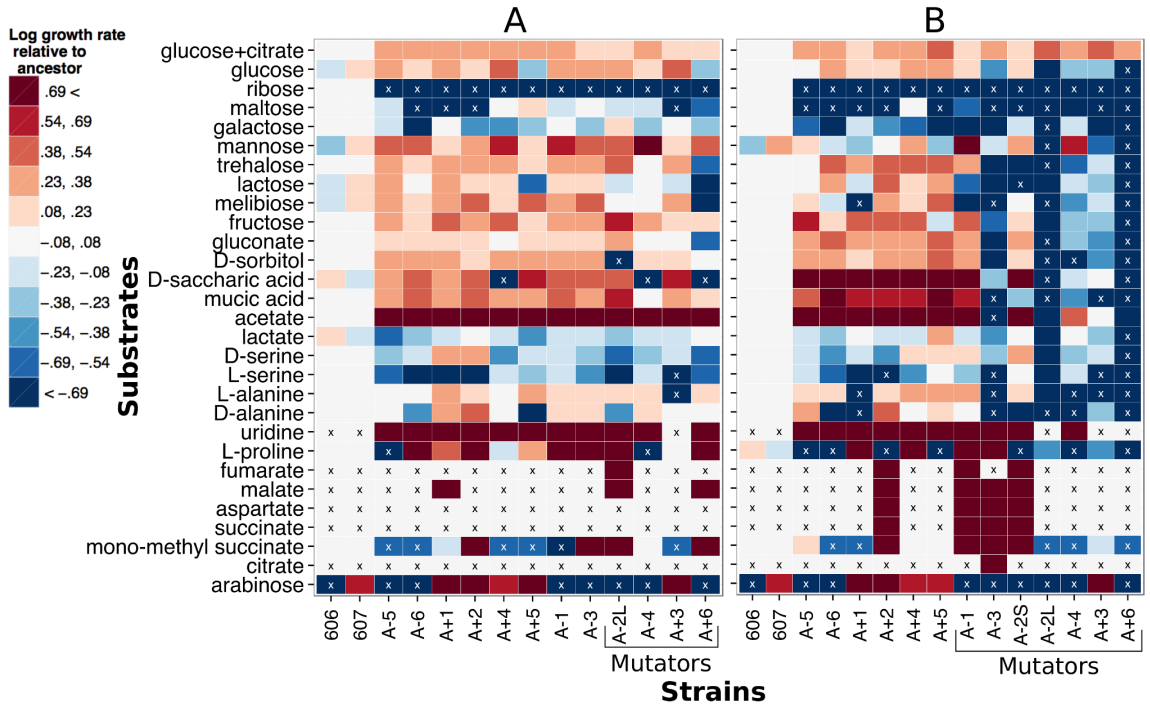



**Figure 3.2: Relative growth rates across a variety of growth substrates for evolved strains from 20k (A) or 50k generations (B).** Heatmaps indicate the log ratio of growth rates relative to the average of the two ancestors on that carbon source. White indicates a growth rate equal to that of the ancestor average, red faster, and blue slower. The growth rates are plotted on a log scale with the limits of the color range set for twice as fast and half as fast as the ancestor average. An 'x' in a box indicates that no growth was observed for that combination of strain and substrate over 48 hours. Strains that were mutators by that timepoint are indicated.

*Gains of function on alternate carbon sources*

In addition to many quantitative improvements of growth rates, there were several examples where isolates acquired the ability to grow on substrates that the ancestral strain could not utilize over the 48 hour time-course of the growth experiments. Only one such example was previously known: the aforementioned gain of citrate utilization by the A-3 population [7]. We found that this strain also gained the ability to grow within 48 hours on three $C_4$-dicarboxylate tricarboxylic acid cycle intermediates (succinate, aspartate, and malate). Three other 50k isolates from different replicate populations gained the ability to use this same set of three C4 dicarboxylate intermediates, as well as fumarate.

*Mutator strains suffered greater declines in growth on alternative substrates*

We compared mutators to non-mutators to ask whether mutation accumulation contributed to the observed decreases in growth (Figure 3.2). At 20k generations, despite increasing in growth rate more than decreasing (47 vs. 41 cases), the mutators were marginally worse, on average, than non-mutators (P=0.03, Pearson's chi-squared test comparing proportion of growth rate reductions). By 50k there was a stark pattern of mutators declining in catabolic ability compared to non-mutators (P<<0.0001, Pearson's chi-squared test). This can be seen in the large block of blue (decreases in rate) for five of the mutators. Non-mutators at 50k still increased in growth rate more often than they decreased (80 vs. 52 cases, P=0.018 binomial two-sided test). Because some strains are known to be

affected by the absence of citrate even though they cannot use it as a carbon source [13], we also tested the growth rate of the 50k strains on alternative substrates supplemented with 1.7 mM sodium citrate, as in LTEE growth media. Although the reductions in growth rate relative to the ancestor were ameliorated in some cases by the addition of citrate, the mutator strains still suffered significantly more growth rate decreases than the non-mutators (P=0.003, Pearson's chi-squared test).

*Parallelism of metabolic decreases*

Given the trend of both increased and decreased growth rate on alternative carbon sources, we assessed the degree of parallelism with which metabolic erosion occurred. We segmented the data by substrate and asked how many evolved strains decreased in growth rate or cellular respiration on each substrate. We took as a null expectation that decreases in metabolic function are equally as likely as increases, and plotted the observed pattern against this null distribution (Supplementary Figure 3.3).

In no case do our observations of metabolic decreases closely match the null distribution. As previously, cellular respiration was reduced for nearly all strains on all substrates. The observed average number of strains with reduced respiration on a substrate was 11.3 at 20k and 12.8 at 50k, 5.3 out of 12 and 6.3 out of 13 more strains than would be expected given the null distribution ($P<<0.0001$, binomial two-tailed exact test). For growth rate, at 20k on average

5.3 strains reduced in growth rate on each substrate, in fact 0.7 fewer strains than expected given the null distribution (P=0.01). However, this average somewhat masks the bimodal pattern seen in the distribution, with some substrates showing nearly no strains reducing growth rate and others nearly all. At 50k, an average of 8.3 strains lost function on each substrate, 1.8 more than expected (P<<0.0001). Clearly there is some parallelism in decreases in growth rate, but it is worth emphasizing that the substrates used in this study were those for which widespread, parallel losses in cellular respiration were previously observed.

*Metabolic similarity between substrates was a poor predictor of correlated evolved responses*

We asked whether the correlated changes in performance on alternative substrates could be predicted based on the similarity of the catabolic network for growth on that compound compared to that for glucose. There are two rationales that would support this hypothesis. First, there are more loss-of-function mutations available for a non-glucose substrate if it uses many unique enzymes, and we might expect to see metabolic specialization scale with mutational target size under mutation accumulation. These mutations may either simply be unguarded by purifying selection, or perhaps even selectively advantageous to lose. Second, the balance and direction of flux through various pathways will lead to a different optimal allocation of enzymes to balance the needs of catalysis versus expression costs. As such, antagonistic and synergistic pleiotropy suggest that highly overlapping metabolic flux patterns might be expected to suffer fewer

declines, or possibly even synergistic gains, relative to a very differently used substrate. A rough approximation of the similarity between different substrates is to simply group them as sugars or 'non-sugars' that require gluconeogenesis for anabolism. To frame this more quantitatively, however, we also used genome-scale metabolic models to make predictions about specialization.

In order to approximate internal metabolic states we used Flux Balance Analysis (FBA) to generate predicted flux patterns for each compound [14]. This approach generates a vector of internal flux values that describes the relative flow through every reaction in a cell were it to optimize biomass production per substrate molecule. Although selection in batch culture largely acts upon rate, biomass production per unit substrate has been shown to effectively capture the growth of the LTEE ancestor on glucose, and 50k evolved strains deviated only slightly from this pattern [15]. We therefore compared FBA-derived flux vectors using a number of metrics to determine their degree of dissimilarity to the flux vector for glucose (see Materials and Methods).

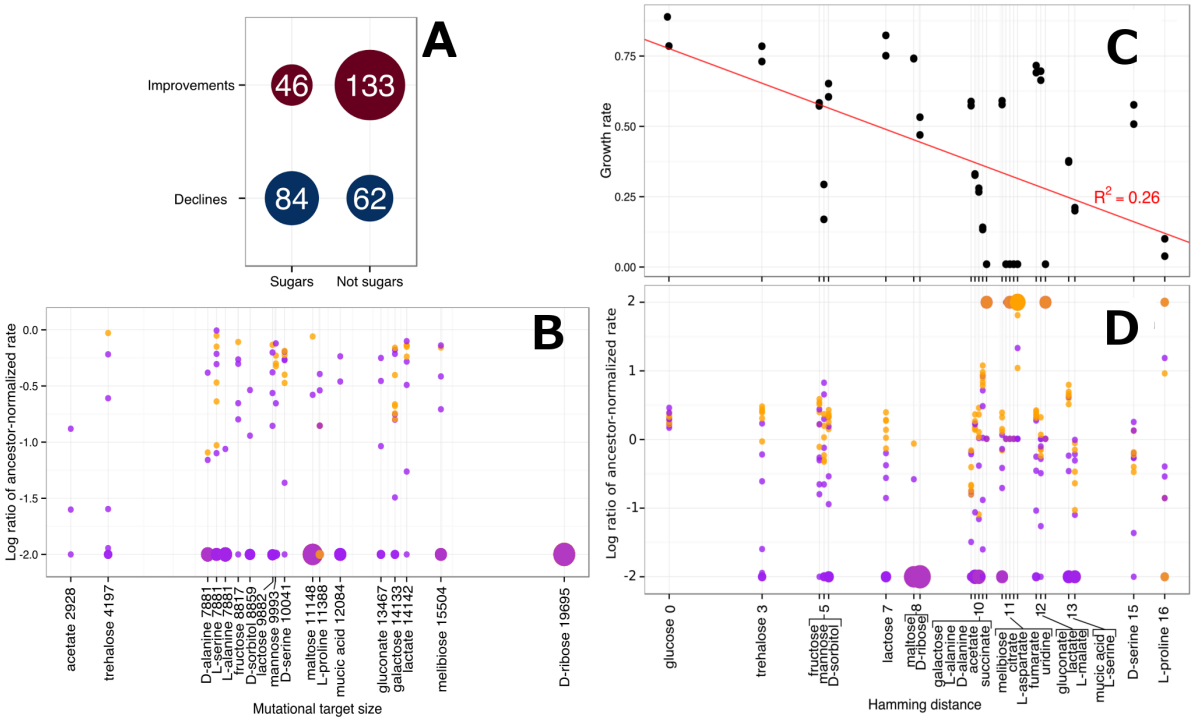



**Figure 3.3: Substrate dissimilarity does not predict metabolic erosion.** A) A simple categorization of substrates as sugars and non-sugars finds that the correlation between relatedness to glucose and evolved metabolic changes is the opposite from what is hypothesized. B) The FBA-predicted mutational target size does not correlate with decreases in growth rate. C) Hamming distance between FBA-generated flux vectors for carbon sources partially predicts ancestral growth rate. Black dots indicate the growth rate of the 2 ancestral strains. 268 reactions were predicted as necessary for optimal metabolism on glucose. D) Hamming distance between a substrate and glucose does not correlate with increases or decreases in growth rate. The y axis is the log of the ratio of growth rate relative to

**Figure 3.3, continued:** the ancestor, with all ratios greater or less than $e^2$ binned at the axis limit. For C-D, purple dots are mutator strains, orange dots are non-mutators. Larger dots at the axis extrema indicate more overlapping points, and the shading between purple and orange indicates the different proportions of mutators and non-mutators at that limit**.** For B-D, substrates with the same x axis values were plotted with a slight offset, and the true value is listed in the axis label.

We first tested whether evolved decreases in growth rate scaled with mutational target size. There is no expected behavior under this hypothesis for increases in growth rate, so we limited our analysis to combinations of strains and substrates for which growth rate had decreased. By identifying the reactions necessary for optimal growth on alternative substrates that are not necessary for growth on glucose, and determining the number of coding nucleotides necessary for those reactions, we were able to approximate the number of available mutations that would decrease growth rate on a substrate. Contrary to our hypothesis, we found no significant relationship between mutational target size and reduction in growth rate ($P=0.15$, linear regression F-test$(1, 146)=2.1$) (Figure 3.3B).

Our hypothesis also suggests that substrates used more similarly to glucose would permit more rapid growth. Starting with the simple categorization of substrates as sugars and non-sugars, we found no correlation between these groupings and changes in growth rate (Figure 3.3A). Indeed, there were many reductions in growth rate for sugars other than glucose. To frame this hypothesis in a more quantitative way, we compared the Hamming distance between the

vector of predicted fluxes for an alternative compound and that of glucose. Contrary to our hypothesis, we found no significant relationship between metabolic similarity to glucose and correlated responses (P=0.26, linear regression F-test(1, 323)=1.27), and any relationship measured was in fact in the opposite direction as predicted (Figure 3.3D). As a confirmation that Hamming distance between flux vectors for alternative substrates is biologically relevant, we found that it was a significant predictor of some of the variance in the relative growth rate of the ancestor (P=0.0001, $R^2$=0.26, linear regression F-test(1, 48)=17.2) (Figure 3.3C). Alternative metrics to Hamming distance performed similarly poorly in predicting patterns of tradeoffs (Supplementary Table 3.4). These data suggest that the similarity of overall flux patterns is a surprisingly poor predictor about which substrates would experience correlated increases or decreases in performance.

*Growth for mutator strain was more sensitive to temperature than for non-mutators*

Linking the mutator-driven metabolic specialization to their vastly elevated mutation rate itself, we hypothesized that their abundance of amino acid substitutions may generate trends indicative of the types of effects they had upon their gene products. The mutator lineages acquired mutations with a rate up to 0.06 per generation [16]. For the A-1 lineage, by 40,000 generations there were 627 SNPs, 599 of which were in coding regions, and 513 of those were non-synonymous [17]. This is a tremendous load of amino acid substitutions, which

are viewed as likely to be deleterious due to destabilizing proteins [18-20]. We therefore hypothesized that, if protein destabilization was a dominant factor affecting growth at 37 °C, we could predictably ameliorate these defects by lowering the growth temperature. Non-mutators will have a small number of such mutations, but the ~100-fold greater rate of such mutations in the mutator genomes should make growth more temperature sensitive than for non-mutators.

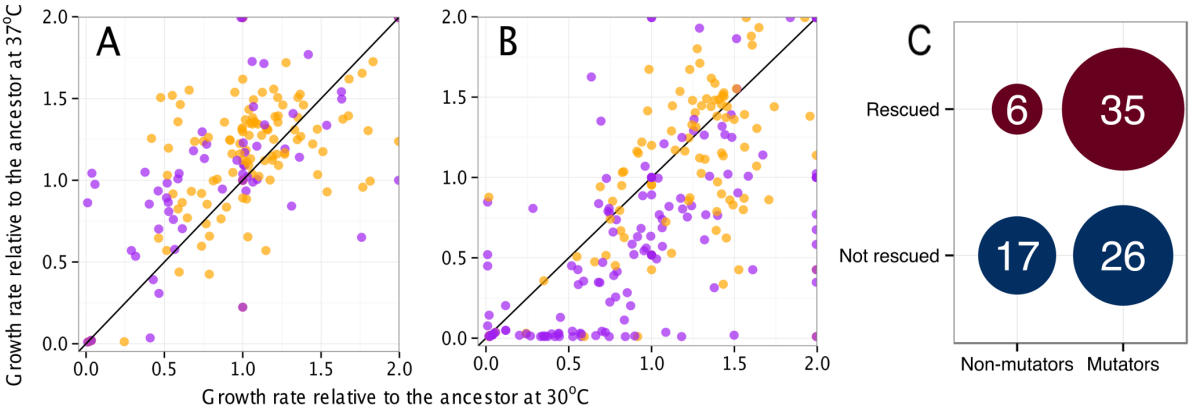



**Figure 3.4: Temperature dependence of growth rate on alternative substrates.** For all strain/substrate measurements, we determined the relative change in growth rate by changing temperature from 37 °C to 30 °C. For A-B, Purple dots are mutator strains; orange dots non-mutators. Points that fall outside of the plot range are plotted at the edge of the graph. A) Effect of temperature change on 20k isolates B) Effect on 50k isolates. C) For 50k isolates, the number of mutators and non-mutators that were rescued from no growth at 37 °C to growth at 30 °C.

Growth rate data supports the hypothesis that mutators have general temperature-sensitivity. For 50k isolates, growth rate relative to the ancestor at 30 °C was higher than at 37 °C in 98 cases, compared to only 37 cases where it was reduced relative to the ancestor ($P<<0.0001$, binomial two-tailed exact test) (Figure 3.4B). That is, despite the fact that these strains have adapted for 50,000 generations at 37 °C, the ratio of their growth rate to that of the ancestor is higher at the foreign 30 °C than their native temperature. This general improvement at the lower temperature was not present for non-mutators (56 improved relative to the ancestor by moving to 30 °C, 57 worse- $P=0.99$, binomial two-tailed exact test), and the difference between mutators and non-mutators was significant ($P=0.0002$, Pearson's chi-squared test). Furthermore, in the cases where evolved 50k isolates completely lost the ability to grow on substrates, when grown at 30 °C these losses were ameliorated more than half the time for mutators (35 of 61), significantly more than for non-mutators (6 out of 23, $P=0.01$, Pearson's chi-squared test) (Figure 3.4C).

An alternative hypothesis for the elevated temperature sensitivity of mutators is that the phenotype is directly caused by the mutation in mismatch repair rather than the accumulation of destabilizing mutations that it caused. For the 20k isolates, in most cases growth was better relative to the ancestor at the native 37 °C than at 30 °C (119 vs. 75, $P=0.002$, binomial two-tailed exact test) (Figure 3.4A). There was no significant difference at 20k between this pattern for mutators and non-mutators ($P=0.20$, Pearson's chi-squared test), and no

significant difference in the number of rescues from complete loss of growth (P=0.31, Pearson's chi-squared test), ruling out that mutator status itself generates temperature sensitivity.

**Discussion**

Replicate experimentally-evolved populations such as the Lenski LTEE are ideal for studying the processes leading to specialization. Here we directly measured growth rates in well-mixed conditions rather than cellular respiration in static media. This seemingly minor choice generated qualitatively different results, leading to the opposite conclusions from those previously made about the mechanisms and patterns of specialization during tens of thousands of generations of growth on a single substrate.

The earlier report [4] of widespread, parallel tradeoffs after extended growth on a single compound fit comfortably with the general notion that unused capacities will tend to degrade after an extended period of disuse. However, the suitability of cellular respiration assays for growth performance was seriously challenged by our data. There were many cases of 'false positives', with respiration on substrates that do not support growth, and a weak predictive ability for growth even after removing these. The parallel decreases in function observed across compounds in the respiration assay seem to be due to cultures becoming worse at reducing the dye in the Biolog assay environment. This generic effect was strong

enough to mask both adaptation on glucose and the novel gain of citrate use in the A-3 50k isolate.

*Little parallelism in metabolic erosion*

Growth rates indicated little evidence in support of widespread antagonistic pleiotropy, with more increases than decreases in growth on alternative compounds through 20k, and very few parallel declines. There were individual counter examples observed, such as the previously characterized universal loss of ribose utilization early in adaptation [21], and the tendency for reduced or loss of growth on maltose [22-24]. As such, it is clear that examples of antagonistic pleiotropy do exist in the data. However, relatively few other substrates showed this pattern at 20k or 50k, despite the fact that these substrates were those where parallel reductions in respiration were observed. Because selection drives antagonistic pleiotropy, it is commonly expected that the early period of rapid adaptation would coincide with the most tradeoffs in alternative environments, and that the frequent parallelism in the targets of amongst early beneficial mutations would drive parallel losses [4]. Given these criteria, the growth data do not support antagonistic pleiotropy as the primary driver of specialization.

There are three implicit assumptions about antagonistic pleiotropy, however, that if not met alter the expectations for specialization driven by selection. First, if different beneficial mutations occur across lineages, they will not necessarily lead to the same pleiotropic tradeoffs. As of 20k, out of the 14 genes screened in all of

the populations there were three genes with mutations in all populations and two more in a majority. The other screened genes had mutations in a minority or none of the other populations, suggesting that a variety of different beneficial mutations occurred across lineages [17]. Second, beneficial mutations in the same target may have differing pleiotropic effects in different lineages due to other mutations present. This "epistatic pleiotropy" [25] has been found to be common in multiple model systems [26-29]. Third, the early large-effect beneficial mutations may or may not be responsible for greater pleiotropic effects than later, smaller effect mutations. Yeast morphological pleiotropy scaled with fitness, for example, but the correlation explained only 17% of the variation [30]. The first and second scenarios above – distinct mutations or epistatic pleiotropy – would lead to a scenario whereby parallel metabolic declines are no longer necessarily expected from antagonistic pleiotropy. The third scenario – pleiotropy not scaling with selective effect – would mean the temporal dynamics of fitness gain in the selective conditions and the rate of performance losses in alternative environments need not be tied.

These caveats underscore our limited ability make conclusions about the role of antagonistic pleiotropy in the observed metabolic declines. The only sure determinant of whether a correlated change is the result of pleiotropy or neutral mutation is to genetically manipulate the strains to isolate the effect of individual mutations. This suggests future experiments, for example testing early and parallel mutations previously screened for epistatic effects [31] for pleiotropy.

Ultimately, as we discuss below, the key determinant of the role of mutation accumulation is whether metabolic specialization was substantially affected by mutation rate.

*Increases in growth rate on carbon sources other than glucose*

Rather than a general pattern of metabolic specialization, these data revealed an unexpected extent of correlated improvements in growth on alternative compounds. Why would *E. coli* maintain or improve performance on substrates that had not been supplied for decades? There are three general classes of explanations, two of which mirror the processes considered for pleiotropic tradeoffs.

The first explanation for the correlated improvements, and undoubtedly the least likely, would be neutral performance gains through mutations that had no selective consequence in glucose: the beneficial analog of mutation accumulation. If this were the case, mutators might have more increases in rate than non-mutators, and more improvements would have occurred by 50k than 20k, which is the opposite of what was observed.

The second explanation of the correlated improvements is that the same mutations that were beneficial during growth on glucose may have led to gains in alternative environments, i.e. “synergistic pleiotropy”. There are known examples of this occurring for early mutations in *E. coli* evolving in these conditions [12,32], and it is a common pattern seen across organisms (for example [33]).

These synergistic mutations may be generally beneficial in the laboratory environment of the LTEE and thus unrelated to carbon source metabolism. Indeed, there are examples of both adaptation to generic aspects of a selective environment, such as the trace metal formulation [34], and removal or down-regulation of costly genes or genome regions [21,35,36]. Synergistic pleiotropy could also result from mutations that directly improved glucose metabolism, such as mutations in the phosphotransferase-mediated uptake system that also increased growth on the other sugars imported by this system [12].

A third hypothesis for correlated gains of function on alternative compounds is that there were additional compounds besides glucose (and citrate) available from cell excretions or lysis. The serial transfer regime of the LTEE creates a scenario whereby populations use all of their glucose resources within the first few hours, and remain in stationary phase the remainder of the day. The ancestral *E. coli* excrete a small amount of acetate in this environment, and this increased on average two-fold by 50k generations [15]. It is thus unsurprising that the strongest, most universal gain in alternative compounds by 20k was on acetate (Figure 3.2A). In terms of cell lysis, this has allowed one population (A-2) to maintain a long-term polymorphism for over 40,000 generations. A “large” colony lineage that grows fast on glucose but lyses substantially in stationary phase cross-feeds a “small” colony lineage that is not as fast on glucose as the larges but has specialized as a “cannibal” [37-39]. This results in a stable, negative frequency-dependent fitness effect between these strategies. Although

an earlier study of the other populations at 20k failed to reject that fitness interactions were transitive through time [40], these competitions were performed at a 50:50 ratio and thus may have missed interactions that occur when one partner is rare.

*Novel gains-of-function*

The most remarkable correlated increases were the several examples of "novel" gains of function by evolved isolates on substrates where the ancestor failed to grow. The citrate example has been reported previously [7], but we did not expect to find other such substrates. These novel gains of function are distinct from what was seen for citrate, as over a longer duration (>100 hours) the ancestral *E. coli* seem to grow to measurable density on these substrates. We are currently exploring whether these long lags represent slow physiological acclimation or the emergence of evolved genomic changes. In the case of the Cit$^+$ A-3 lineage, succinate is likely excreted during citrate import [41], thus selection for its use is perhaps unsurprising. The fact that several other strains experienced similar gains across the same range of C4 dicarboxylic acids, and that this included the cross-feeding "small" phenotype clone from the A-2 population appears to suggest that these compounds may be excreted, or present during stationary phase from lysed cells.

*Predictability of correlated responses on alternative carbon sources*

With the availability of whole-genome scale metabolic models for *E. coli*, we asked whether we could predict the trend of correlated responses by comparing their pattern of use to that of glucose. We proposed a generic, seemingly obvious hypothesis that the more different the metabolism of an alternative substrate was from the metabolism of glucose, the more likely it would be that populations would have decreased (or lost) their ability to use it. As described, this logic holds regardless of whether or not selection drove metabolic specialization. In order to quantify the similarity of substrates, we applied FBA to compare the predicted optimal metabolic flux states for each compound. The data, however, did not support our hypothesis: there was almost no relationship between similarity to glucose and correlated response. Recent *in silico* attempts to predict growth capability on substrates based on metabolic similarities have had some success [42], suggesting that evolution may be acting here on functions not included in the model. For example, mutations may have occurred in functions related to differential regulation that distinguish these sugars, rather than central metabolic enzymes for which their use is nearly identical. These mutations are known to have occurred in the LTEE, for example in *spoT* and *nadR* [35,43]. These predictions are also based on the assumption that glucose is the only available carbon source. If growth on other carbon sources is under selection due to their excretion or presence after cell lysis, it may explain some of the lack of predictive power here.

*Temperature sensitivity of mutators consistent with protein destabilizing mutations*

Although the identity of the substrates that experienced tradeoffs (or improvements) were not those we expected, we reasoned that the biophysical effects of the deluge of mutations in mutators might lead to a predictable pattern of temperature sensitivity in these strains. The genomic sequences and data available to date [16,17] suggest mutators will have on the order of 500-2000 non-synonymous mutations, perhaps more. Random amino acid substitutions have been shown to be mildly deleterious in general due to destabilizing proteins [20]. To ask whether tradeoffs observed in the mutators were at least partly due to destabilizing mutations in proteins needed for alternative substrates, we tested whether mutators would be more sensitive to changes in incubation temperature than non-mutators. Consistent with this hypothesis, we found that the 50k mutators performed better relative to the ancestor at 30 °C than at the 37 °C temperature where they have evolved. One alternative hypothesis that was ruled out is that the mutator allele itself leads to temperature sensitivity, as the 20k mutators performed better at 37 °C than at 30 °C, and the changes with temperature were not distinguishable from non-mutators. Although other alternative hypotheses may explain some of the temperature sensitivity, these data are consistent with the hypothesis that neutral degradation of protein-coding sequences in these strains proceeded via partial destabilization on the way to eventual loss of function.

*Role of elevated mutation rate driving metabolic specialization*

The comparison of mutators and non-mutators at 50k strongly suggests that neutral mutation accumulation was the primary driver of metabolic specialization. The difference in metabolic erosion allows us to distinguish the overall trend from forms of antagonistic pleiotropy that could have lacked parallelism (different mutations or epistatic pleiotropy) or that may have arisen late relative to fitness gains (if pleiotropy did not scale with selective effects). Despite a significant difference from non-mutators in the proportion of growth rate reductions, after 20,000 generations and a decade of adaptation the mutators still increased growth rate in more cases than they decreased, and only by 50k did mutators as a group have more decreases than increases. By the later timepoint five of the seven mutators had decreased growth (or complete loss) for essentially every single alternative compound (except citrate and C4 dicarboxylic acids for A-3, which were under selection for this strain). Interestingly, the other two (A-1, A-2S) do not show this pattern. These counterexamples may be due to the fact that A-1 acquired its mutator status late [17], and A-2S is the cross-feeding generalist described above that adapted to grow upon lysed cell material [38].

The late appearance of metabolic erosion argues for the unparalleled utility of truly long-term experiments. A neutral process such as mutation accumulation needs time to become apparent, although hitchhiking with beneficial mutations can speed their fixation (i.e., “draft” [44,45]). With a reduced effective population

size, the window of selective effects that behave neutrally grows. As such, the effects of elevated mutation rates and mutation accumulation becomes apparent much more quickly with evolution regimes with small bottlenecks, such as single colonies [46,47]. The late appearance of specialization also contrasts sharply with abundant evidence that lineages can diversify and specialize in mere tens or hundreds of generations. In addition to population size, this difference in timescale appears to correlate with the type of selective environment, and thus the evolutionary process that was responsible. Whereas the Lenski LTEE is notable as an environment with a single nutrient resource at high concentration, the cases of rapid diversification have involved spatial heterogeneity [48], rate-limiting resources in a chemostat [49], or the presence of multiple substrates simultaneously [50]. In those scenarios, selection is actively pulling on different performance features of an organism and antagonistic pleiotropy appears to dominate. The relatively slow degradation of catabolic capacity in the LTEE suggests that *E. coli* faces comparatively little tension between improving upon glucose and maintaining performance on other substrates, even those which are predicted to be utilized in a very distinct manner. Specialization in this case appears not to have been a requisite tradeoff of adaptation, but rather a result of the degradation of unneeded proteins.

*The fate of mutator strains in the long-term*

Given the severity of metabolic erosion for mutators even in large laboratory populations, it is remarkable just how common mutator lineages are in nature.

Mutators have been isolated at frequencies over 1%, and seem to be particularly common in organisms such as pathogens [51]. The frequency of mutators in nature, despite the associated costs, may be partially explained by increased evolvability, shown in laboratory medium [52] and in mice [53].

Our results add substantially to the idea that an elevated mutation rate is an ill-fated long-term strategy even for large populations, as declines in performance in alternative environments will eliminate previously-occupied parts of the niche space. Recent findings have suggested that mutators may tend to attenuate their increased mutation rate over time [16,54,55], perhaps to avoid the harmful effects of Muller's Ratchet. Thus both tradeoffs in alternative environments and mutation load in selective environments may contribute to the paradox that over the short-term lineages often benefit from elevated mutation rates, but the long-term trend across phylogenies has been for stability in mutation rates of free-living microbes [56].

## Materials and Methods

*Strains and LTEE conditions*

*Escherichia coli* B isolates were obtained from the Lenski LTEE [5] after 20,000 and 50,000 generations. Briefly, in the evolution experiment 12 populations of *E. coli* B were founded with either the arabinose-negative strain REL606 (populations A-1 to A-6) or the otherwise isogenic arabinose-positive derivative,

REL607 (A+1 to A+6). These have been evolved since 1988 in 50 mL flasks containing 10 mL of Davis minimal media (DM) with 139 μM glucose (25 mg/L) as a carbon source. The cultures were grown at 37°C while shaking at 120 RPM, and were transferred daily via 1:100 dilutions (~6.64 net doublings per day).

The isolates analyzed in this experiment consisted of the ancestral lines REL606 and REL607, as well as the 'A' clone frozen at 50k and 20k generations for the 12 populations. The A-2A clones at 20k and 50k were from the 'large' lineage that has coexisted with a cross-feeding 'small' lineage for tens of thousands of generations [57], and thus here we refer to them as A-2L. At 50k we also examined A-2C (REL11335), a 'small' clone that we refer to here as A-2S. All evolved strains are listed in Supplementary Table 3.1.

*Growth Rate Experiments*

For growth rate measurements, we acclimated out of the freezer by inoculating 10 μL frozen cultures into 630 μL modified DM250 media in 48-well micotiter plates (Costar) and growing overnight on a plate shaking tower (Caliper). All growth rates were measured at the LTEE selective temperature (37 °C) unless otherwise described. The modified DM media is the same as previously used throughout the evolution of these strains [5], except it contained 250 mg/L glucose and no sodium citrate. Following acclimation, saturated cultures were transferred into new plates with a 1:64 dilution in DM media supplemented with 5 mM of a single carbon source. Under these conditions, growth in plates correlates well with

growth in flasks, both with and without citrate ($P<0.0001$, $R^2=0.74$, linear regression F-test(1,28)=79.3). The substrates analyzed were those where consistent reduction in cellular respiration were previously observed, as well as several sugars for which fitness changes had been previously measured after 2,000 generations [12] and citrate. Between 3 and 11 biological replicates were run for each strain/carbon source combination.

Optical densities were obtained every 30 minutes to 1 hour on a Wallac Victor 2 plate reader (Perkin-Elmer), until 48 hours had passed or cultures reached saturation, using a previously described automated measurement system [9,10]. Growth rates were determined by fitting an exponential growth model using custom analysis software, Curve Fitter (N. F. Delaney, CJM, unpublished; http://www.evolvedmicrobe. com/Software.html). Representative growth curves and fitted growth rates are shown in Supplementary Figure 3.4. The growth rate for all strains relative to the ancestor was calculated for each plate (averaging over the 2 ancestors, REL606 and REL607), and averaged across plates. Mean growth rates relative to the average of the ancestors were used throughout. When quantifying the number of increases and decreases in rate for evolved strains, we used all of the data for the substrates for which the ancestor exhibited growth- a necessary criteria for the evolved strains to demonstrate reductions.

*Cellular Respiration Assays*

Biolog assays for respiration capacity were run as in Cooper and Lenski [4]. Briefly, cultures were grown from freezer stocks for 2 cycles of 1:100 dilution and 24 hours of growth in 10 mL LB in flasks shaking at 37 °C. LB was used to avoid catabolite repression due to growth in minimal media, which could result in fewer positive results on non-glucose carbon sources. These cultures were inoculated 1:100 into fresh LB and grown for 6 hours before being spun down at 12,000 g for 10 minutes and rinsed in saline to remove residual medium. Rinsed cells were resuspended in IF-0 buffer with dye added (Biolog) to a constant density of 85% transmittance, and all wells of Biolog PM1 plates were inoculated with 100 μL of this suspension. The plates were incubated, unshaken, at 37 °C. $OD_{600}$ was measured at 0, 4, 12, 24, and 48 hours, and all well readings were adjusted by subtracting the reading of the well at 0 hours. A trapezoidal area approximation [4] combined the five measurements for each well into one value, which reflects the area under the curve (AUC) of optical density versus time. One replicate plate experiment was performed for each evolved strain at 20k and 50k generations, and 4 replicates were run for each ancestor (REL606 and REL607). Tests for tradeoffs for the evolved strains as a group on a substrate had were 1-sample T-tests against the ancestral distribution with significance cutoff of P=0.002 to adjust for multiple comparisons. Tests for individual strains were the same but with a cutoff of P=0.0005.

*Flux Balance Analysis*

Flux analysis was carried out with a genome-scale model of *E. coli* metabolism (iAF_1260 [58]). The model incorporates 2382 reactions and 1668 metabolites. The default minimal media environment and reaction bounds were used. Fluxes were predicted for each individual carbon source provided, normalized by number of carbon atoms to 10 units of glucose. Maximal biomass per substrate was used as the objective criterion as previously[15]. To determine whether a reaction was necessary for optimal growth on a substrate, each reaction flux predicted was individually constrained to zero. Only the necessary reactions, those for which constraining the flux resulted in a reduction in final biomass, were considered in the analysis of differences between flux vectors (Supplementary Table 3.2). Reaction differences between substrates, considered for the Hamming distance, are listed in Supplementary Table 3.3. Supplementary Table 3.4 summarizes alternative distance metrics that were used to assess the difference between flux vectors.

*Author Contributions*

The author(s) have made the following declarations about their contributions: Conceived and designed the experiments: NL CJM. Performed the experiments: NL. Analyzed the data: NL CJM. Contributed reagents/materials/analysis tools: NL. Wrote the paper: NL CJM.

*Acknowledgments*

NL was funded by an NSF Graduate Research Fellowship and NIH grant GM080177. We thank Rich Lenski and Neerja Hajela for the LTEE strains, and appreciate Susi Remold, Will Harcombe, and other members of the Marx laboratory for helpful discussion of the project and manuscript. Data pertaining to the experiments in this paper have been deposited at doi:10.5061/dryad.7g401 [59].

## References

1. Law R (1979) Optimal life histories under age-specific predation. Am Nat 114: 399–417.

2. MacArthur R (1972) Geographical Ecology. Princeton, NJ: Princeton University Press.

3. Sniegowski PD, Gerrish PJ, Lenski RE (1997) Evolution of high mutation rates in experimental populations of *E. coli*. Nature 387: 703–705.

4. Cooper VS, Lenski RE (2000) The population genetics of ecological specialization in evolving *Escherichia coli* populations. Nature 407: 736–739. doi:10.1038/35037572.

5. 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. doi:10.1086/285289.

6. Bochner BR, Gadzinski P, Panomitros E (2001) Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. Genome Res 11: 1246–1255. doi:10.1101/gr.186501.

7. 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 USA 105: 7899–7906. doi:10.1073/pnas.0803151105.

8. Novak M, Pfeiffer T, Lenski RE, Sauer U, Bonhoeffer S (2006) Experimental tests for an evolutionary trade-off between growth rate and yield in *E. coli*. Am Nat 168: 242–251. doi:10.1086/506527.

9. Delaney NF, Rojas Echenique JI, Marx CJ (2013) Clarity: an open-source manager for laboratory automation. J Lab Autom 18: 171–177. doi:10.1177/2211068212460237.

10. Delaney NF, Kaczmarek ME, Ward LM, Swanson PK, Lee M-C, et al. (2013) Development of an optimized medium, strain and high-throughput culturing methods for *Methylobacterium extorquens*. PloS ONE 8: e62957. doi:10.1371/journal.pone.0062957.s006.

11. Vasi F, Travisano M, Lenski RE (1994) Long-term experimental evolution in *Escherichia coli*. II. Changes in life-history traits during adaptation to a seasonal environment. Am Nat: 432–456.

12. Travisano M, Lenski RE (1996) Long-term experimental evolution in *Escherichia coli*. IV. Targets of selection and the specificity of adaptation. Genetics 143: 15–26.

13. Leiby N, Harcombe WR, Marx CJ (2012) Multiple long-term, experimentally-evolved populations of *Escherichia coli* acquire dependence upon citrate as an iron chelator for optimal growth on glucose. BMC Evol Biol 12: 151. doi:10.1186/1471-2148-12-151.

14. Orth JD, Thiele I, Palsson BO (2010) What is flux balance analysis? Nat Biotechnol 28: 245–248. doi:10.1038/nbt.1614.

15. Harcombe WR, Delaney NF, Leiby N, Klitgord N, Marx CJ (2013) The ability of flux balance analysis to predict evolution of central metabolism scales with the initial distance to the optimum. PLoS Comput Biol 9: e1003091. doi:10.1371/journal.pcbi.1003091.

16. Wielgoss S, Barrick JE, Tenaillon O, Wiser MJ, Dittmar WJ, et al. (2013) Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. Proc Natl Acad Sci USA 110: 222–227. doi:10.1073/pnas.1219574110.

17. Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature 461: 1243–1247. doi:10.1038/nature08480 -sup.

18. Bloom JD, Silberg JJ, Wilke CO, Drummond DA, Adami C, et al. (2005) Thermodynamic prediction of protein neutrality. Proc Natl Acad Sci USA 102: 606–611.

19. Tokuriki N, Stricher F, Schymkowitz J, Serrano L, Tawfik DS (2007) The stability effects of protein mutations appear to be universally distributed. J Mol Biol 369: 1318–1332. doi:10.1016/j.jmb.2007.03.069.

20. DePristo MA, Weinreich DM, Hartl DL (2005) Missense meanderings in sequence space: a biophysical view of protein evolution. Nat Rev Genet 6: 678–687. doi:10.1038/nrg1672.

21. Cooper VS, Schneider D, Blot M, Lenski RE (2001) Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. J Bacteriol 183: 2834–2841. doi:10.1128/JB.183.9.2834.

22. Meyer JR, Agrawal AA, Quick RT, Dobias DT, Schneider D, et al. (2010) Parallel changes in host resistance to viral infection during 45,000 generations of relaxed selection. Evolution 64: 3024–3034. doi:10.1111/j.1558-5646.2010.01049.x.

23. Travisano M, Vasi F, Lenski RE (1995) Long-Term Experimental Evolution in *Escherichia Coli*. III. Variation Among Replicate Populations in Correlated Responses to Novel Environments. Evolution 49: 189–200.

24. Pelosi L, Kühn L, Guetta D, Garin J, Geiselmann J, et al. (2006) Parallel changes in global protein profiles during long-term experimental evolution in Escherichia coli. Genetics 173: 1851–1869. doi:10.1534/genetics.105.049619.

25. Remold S (2012) Understanding specialism when the jack of all trades can be the master of all. Proc Biol Sci 279: 4861–4869. doi:10.1371/journal.pcbi.1000112.

26. Burch CL, Chao L (2004) Epistasis and its relationship to canalization in the RNA virus phi 6. Genetics 167: 559–567. doi:10.1534/genetics.103.021196.

27. Remold SK, Lenski RE (2004) Pervasive joint influence of epistasis and plasticity on mutational effects in *Escherichia coli*. Nat Genet 36: 423–426. doi:10.1038/ng1324.

28. Flynn KM, Cooper TF, Moore FBG, Cooper VS (2013) The environment affects epistatic interactions to alter the topology of an empirical fitness landscape. PLoS Genet 9: e1003426. doi:10.1371/journal.pgen.1003426.

29. Carroll SM, Lee M-C, Marx CJ (In press) Sign epistasis limits evolutionary tradeoffs at the confluence of single- and multi-carbon metabolism in *Methylobacterium extorquens* AM1. Evolution.

30. Cooper TF, Ostrowski EA, Travisano M (2007) A negative relationship between mutation pleiotropy and fitness effect in yeast. Evolution 61: 1495–1499. doi:10.1111/j.1558-5646.2007.00109.x.

31. Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF (2011) Negative epistasis between beneficial mutations in an evolving bacterial population. Science 332: 1193. doi:10.1126/science.1203801.

32. Ostrowski EA, Rozen DE, Lenski RE (2005) Pleiotropic effects of beneficial mutations in *Escherichia coli*. Evolution 59: 2343–2352.

33. Lee M-C, Chou H-H, Marx CJ (2009) Asymmetric, bimodal trade-offs during adaptation of *Methylobacterium* to distinct growth substrates. Evolution 63: 2816–2830. doi:10.1111/j.1558-5646.2009.00757.x.

34. 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: e1000652. doi:10.1371/journal.pgen.1000652.

35. 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 USA 100: 1072–1077.

36. Lee M-C, Marx CJ (2012) Repeated, selection-driven genome reduction of accessory genes in experimental populations. PLoS Genet 8: e1002651. doi:10.1371/journal.pgen.1002651.

37. 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 USA 109: 9487–9492. doi:10.1073/pnas.1207091109.

38. Rozen DE, Philippe N, Arjan de Visser J, Lenski RE, Schneider D (2009) Death and cannibalism in a seasonal environment facilitate bacterial coexistence. Ecol Lett 12: 34–44. doi:10.1111/j.1461-0248.2008.01257.x.

39. Rozen DE, Lenski RE (2000) Long-term experimental evolution in *Escherichia coli*. VIII. Dynamics of a balanced polymorphism. Am Nat 155: 24–35. doi:10.1086/303299.

40. de Visser JAGM, Lenski RE (2002) Long-term experimental evolution in *Escherichia coli*. XI. Rejection of non-transitive interactions as cause of declining rate of adaptation. BMC Evol Biol 2: 19.

41. 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. doi:10.1038/nature11514.

42. Barve A, Wagner A (2013) A latent capacity for evolutionary innovation through exaptation in metabolic systems. Nature 500: 203–206. doi:10.1038/nature12301.

43. 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 USA 103: 9107–9112. doi:10.1073/pnas.0602917103.

44. Barton NH (2000) Genetic hitchhiking. Philos Trans R Soc Lond, B, Biol Sci 355: 1553–1562. doi:10.1098/rstb.2000.0716.

45. Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. Genet Res 23: 23–35. doi:10.1017/S0016672300014634.

46. Funchain P, Yeung A, Stewart JL, Lin R, Slupska MM, et al. (2000) The consequences of growth of a mutator strain of Escherichia coli as measured by loss of function among multiple gene targets and loss of fitness. Genetics 154: 959–970.

47. Andersson DI, Hughes D (1996) Muller's ratchet decreases fitness of a DNA-based microbe. Proceedings of the National Academy of Sciences of the United States of America 93: 906–907.

48. Rainey PB, Travisano M (1998) Adaptive radiation in a heterogeneous environment. Nature 32: 69–72.

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

50. Friesen ML, Saxer G, Travisano M, Doebeli M (2004) Experimental evidence for sympatric ecological diversification due to frequency-dependent competition in *Escherichia coli*. Evolution 58: 245–260.

51. LeClerc JE, Li B, Payne WL, Cebula TA (1996) High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. Science 274: 1208–1211.

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

53. Giraud A, Matic I, Tenaillon O, Clara A, Radman M, et al. (2001) Costs and

benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. Science 291: 2606–2608. doi:10.1126/science.1056421.

54. McDonald MJ, Hsieh Y-Y, Yu Y-H, Chang S-L, Leu J-Y (2012) The evolution of low mutation rates in experimental mutator populations of Saccharomyces cerevisiae. Current biology : CB 22: 1235–1240. doi:10.1016/j.cub.2012.04.056.

55. Turrientes M-C, Baquero F, Levin BR, Martínez J-L, Ripoll A, et al. (2013) Normal Mutation Rate Variants Arise in a Mutator (Mut S) Escherichia coli Population. PloS ONE 8: e72963. doi:10.1371/journal.pone.0072963.s011.

56. Woese CR (1987) Bacterial evolution. Microbiol Rev 51: 221–271.

57. Rozen DE, Schneider D, Lenski RE (2005) Long-term experimental evolution in *Escherichia coli*. XIII. Phylogenetic history of a balanced polymorphism. J Mol Evol 61: 171–180. doi:10.1007/s00239-004-0322-2.

58. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, et al. (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol 3. doi:10.1038/msb4100155.

59. Leiby N, Marx CJ. (2013) Data from: Metabolic erosion primarily through mutation accumulation, and not tradeoffs, drives limited evolution of substrate specificity in *Escherichia coli*. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.7g401.

# Chapter 4

# The ability of flux balance analysis to predict evolution of central metabolism scales with the initial distance to the optimum

William R. Harcombe[1], Nigel F. Delaney[1,5], Nicholas Leiby[1,2], Niels Klitgord[3,6], and Christopher J. Marx[1,4,*]

[1] Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA, [2] Systems Biology Program, Harvard University, Cambridge, MA, USA, [3] Bioinformatics Graduate Program, Boston University, Boston, MA, USA and [4] Faculty of Arts and Sciences Center for Systems Biology, Harvard University, Cambridge, MA, USA.

[*] Corresponding author

*This chapter was originally published in PLoS Computational Biology. My role in collecting the data for this project was in the initial modeling and simulation of the LTEE population metabolism. I performed most of the statistical analysis for this work, and produced all of the figures except 4.1 and 4.2a. I was also a secondary author of the manuscript.*

The most powerful genome-scale framework to model metabolism, flux balance analysis (FBA), is an evolutionary optimality model. It hypothesizes selection upon a proposed optimality criterion in order to predict the set of internal fluxes

that would maximize fitness. Here we present a direct test of the optimality assumption underlying FBA by comparing the central metabolic fluxes predicted by multiple criteria to changes measurable by a $^{13}C$-labeling method for experimentally-evolved strains. We considered datasets for three *Escherichia coli* evolution experiments that varied in their length, consistency of environment, and initial optimality. For ten populations that were evolved for 50,000 generations in glucose minimal medium, we observed modest changes in relative fluxes that led to small, but significant decreases in optimality and increased the distance to the predicted optimal flux distribution. In contrast, seven populations evolved on the poor substrate lactate for 900 generations collectively became more optimal and had flux distributions that moved toward predictions. For three pairs of central metabolic knockouts evolved on glucose for 600-800 generations, there was a balance between cases where optimality and flux patterns moved toward or away from FBA predictions. Despite this variation in predictability of changes in central metabolism, two generalities emerged. First, improved growth largely derived from evolved increases in the rate of substrate use. Second, FBA predictions bore out well for the two experiments initiated with ancestors with relatively sub-optimal yield, whereas those begun already quite optimal tended to move somewhat away from predictions. These findings suggest that the tradeoff between rate and yield is surprisingly modest. The observed positive correlation between rate and yield when adaptation initiated further from the optimum resulted in the ability of FBA to use stoichiometric constraints

to predict the evolution of metabolism despite selection for rate.

## Introduction

Systems biology is beginning to provide insight into how interactions within complex networks give rise to the holistic behavior of biological systems, and how natural selection would shape these systems over the course of adaptation. Some mathematical models are made with the goal of translating known parameters of components of a small system into predictions of their function. This approach has been used to predict behavior ranging from the oscillation of natural or engineered genetic regulatory networks [1] to flow through small metabolic networks [2,3]. For larger, genome-scale networks there is insufficient information to generate direct predictions in the same manner. Instead, one can ask how the system *should* behave were it to have already been selected to function optimally given tradeoffs between different selective criteria. One use of mechanistically-explicit optimality models is to consider the possible optimality of current biological phenomena, such as the optimality of the genetic code [4] or of the enzymatic properties of RuBisCO [5]. On the other hand, optimality models can also be used directly to predict phenotypic changes in a system that would occur over the course of adaptation, such as the evolution of virulence [6] or enzyme expression [7].

The most broadly applied metabolic modeling framework, Flux Balance Analysis

(FBA), is a constraint-based evolutionary optimality model. It quantitatively predicts flux through a metabolic network that will maximize a given criterion thought to represent prior natural selection [8]. At the heart of FBA is a stoichiometric matrix, which is a mathematically transformed list of mass-balanced biochemical reactions that fully describes the known topology of the metabolic network of a cell (or other system). It is further assumed that the cell is in a metabolic steady-state, such that the sum of fluxes in and out of each internal metabolite are balanced. As additional constraints are considered (e.g., maximal flux values, irreversible reactions, biomass composition), this matrix can then be used to help define and constrain the space of feasible flux distributions in the cell. Within this feasible space, linear programing is subsequently used to solve for an optimality criterion -such as maximal biomass per substrate (see below)- to identify a feasible flux distribution that permits that optimum.

Evolutionary optimality models are powerful tools as they make it possible to build intuition about the forces that shape biological diversity. However, as has been pointed out most famously by Gould and Lewontin, they can also be misleading and can foster the wrong intuitions [9]. Optimality models make three assumptions: 1) selection (and not other processes) is the primary evolutionary force shaping a trait of interest, 2) we can identify the criterion upon which selection is acting, and 3) there are not underlying constraints which prevent a trait from being optimized. Optimality models are constructive for understanding

the evolution of traits only to the extent that these assumptions can be evaluated.

FBA provides an excellent framework to generate testable hypotheses as to which selective criteria are appropriate for a given set of conditions [10,11]. In environments such as batch culture, selection acts directly upon growth rate -as well as lag and survival in stationary phase- but not upon yield [12]. The most common optimality criterion for FBA is commonly referred to as maximizing growth rate [11]. Because this is performed by constraining one (or occasionally multiple) substrate uptake rate (S/time), this criterion is fully equivalent to predicting the maximum yield (i.e., BM/S) under the given, user-supplied substrate uptake rate. Since FBA cannot predict absolute rates of substrate uptake used as the key constraint, the question as to whether adaptation would optimize BM/S during batch culture critically depends upon the correlation between growth rate and yield. There are solid theoretical grounds to expect absolute limits to the maximization of both rate and yield of reactions [13], but it is often unclear how close biological systems are to these constraints.

In addition to maximization of biomass, various other cellular objectives have been suggested as alternative selective criteria. These include optimal energetic (rather than biosynthetic) efficiency whereby generation of ATP per substrate (ATP/S), or the minimization of the sum of fluxes (BM/$\Sigma v$ or ATP/$\Sigma v$). The latter are based upon the rationale that enzymes are costly, and thus a general

relationship between enzyme levels and reaction rates (although actually quite weak for any given enzyme, [14]) would lead to selection to minimize the total burden of enzymes needed. Finally it has been suggested that selection acts simultaneously upon multiple, competing criteria, leading cells to inhabit an optimal tradeoff surface known as a Pareto optimum [15,16]. This approach constructs a surface on which no single criteria can be further increased without reducing another. It is then assumed that evolution pushes biological systems to exist somewhere on this surface. Data from a variety of experiments suggested that cells operate near to the Pareto optimum defined by BM/S, ATP/S, and minimization of $\Sigma v$ [15].

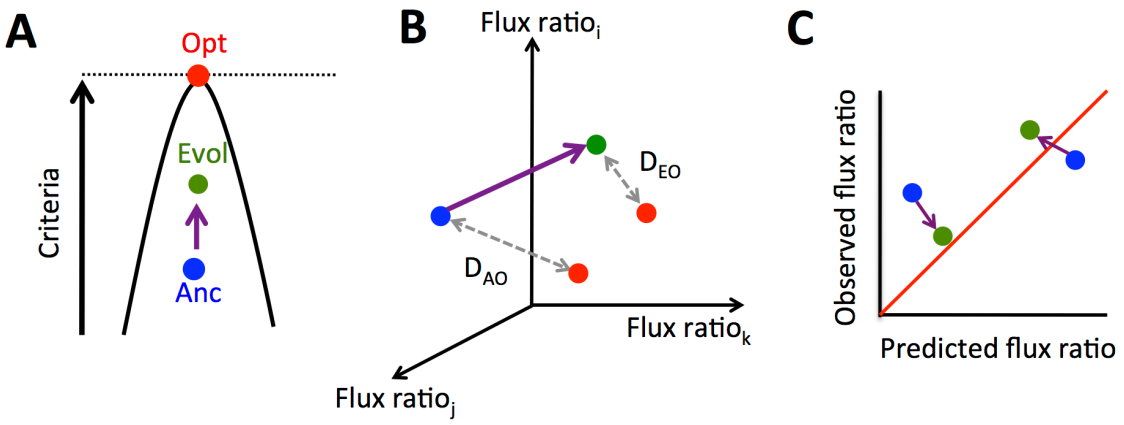



**Figure 4.1: Evolution of metabolic fluxes and measures of optimality and predictability.** We consider three ways to analyze changes in metabolism that relate an ancestor (Anc, blue) to an evolved isolate ($E_i$, green) in regard to an FBA-predicted optimum (Opt, red). A) Evolution of metabolic fluxes can be evaluated from the perspective of changes in proximity to the theoretical maximum for a given optimality criterion (Δ% Optimality). B) A vector of flux ratios defines a position in multi-dimensional flux space. One can then consider the relative Euclidian distance of a given evolved population in this space from its optimum ($D_{EO}$) compared to that of an ancestor from its optimum ($D_{AO}$; plotted as $\log(D_{EO}/D_{AO})$). C) At the most detailed level, one can compare the FBA-predicted value for a given flux ratio versus that observed via $^{13}C$ labeling.

Tests of the predictive capacity of FBA have differed in two ways depending upon: 1) whether there was *known or assumed adaptation* to the substrate in question, and 2) whether tests were a *direct or indirect comparison* of predicted internal fluxes to measured fluxes (Table 4.1). The majority of these tests have been

conducted with *Escherichia coli*, and have assumed past selection on BM/S. The direct tests of FBA compared predicted to observed flux distributions (Figure 4.1) by taking advantage of empirical data generated by $^{13}$C-labeling techniques [17]. Briefly, this method to assay relative metabolic fluxes takes advantage of the fact that the carbon atoms of the growth substrate are shuffled in different ways by alternative metabolic pathways, and that these rearrangements leave a signature in biomass. Using gas chromatography-mass spectrometry (GC-MS) to determine the $^{13}$C-labeling of protein-derived amino acids, it becomes possible to infer the flux splits in the metabolic pathways leading to their synthesis [17–23]. Notable amongst these tests was a quantitative assessment of the relative merits of a series of optimality criteria (and constraints) in their ability to predict the intracellular fluxes of *E. coli* measured in six environments [11]. Data for wild-type cultures indicated that ATP/$\Sigma v^2$, BM/S or ATP/S were more predictive depending upon the growth condition; however, in all cases there was still significant variation between predicted and measured fluxes.

| Past adaptation | Test of internal fluxes | Major approaches | Example papers |
|---|---|---|---|
| **Assumed** | **Indirect** | Growth rate and excretion. | Varma & Palsson, 1994 |
| | | Growth phenotypes or gene essentiality of knockouts. | Raghunathan *et al*, 2009 |
| | **Direct** | Comparison of wild-type or knockout flux pattern to mutants in one or more environments, usually using just BM/S or ATP/S as an optimality criterion. | Emmerling *et al*, 2002 |
| | | Explicit comparison of *E. coli* fluxes across environments to predictions from multiple optimality criteria. | Schuetz *et al*, 2007 |
| **Known** | **Indirect** | Uptake, excretion, and/or growth rates for evolved strains. | Ibarra *et al*, 2002; Teusink *et al*, 2009 |
| | | mRNA and protein levels correlated with predicted pathways in FBA. | Fong *et al*, 2005; Lewis *et al*, 2010 |
| | **Direct** | Flux changed during adaptation of *E. coli* evolved with key metabolic knockouts or on the poor substrate lactate, but no comparison made to FBA. | Fong *et al*, 2006; Hua *et al*, 2007 |
| | | Flux changes during adaptation of *E. coli* to a fluctuating environment compared to predictions of a Pareto surface. | Schuetz *et al*, 2012 |
| | | Flux measurements following 50,000 generations of *E. coli* adaptation and comparison of this and other datasets to FBA. | This study |

**Table 4.1: Major approaches to test of FBA predictions depending upon whether there was known selection under experimental conditions and whether there was direct measurement of internal fluxes.**

A key advance in the use and testing of FBA came from the realization that the best test of an optimality model is to examine whether there is movement toward predicted optimal phenotypes following adaptation under known experimental conditions (Table 4.1). In a classic paper, populations of *E. coli* were adapted to various carbon substrates for 100-700 generations [24]. The authors ran FBA for all pairwise constraints of substrate and oxygen uptake to predict the maximal BM/S within those constraints, and what metabolites might be excreted. Remarkably, adaptation on five out of six substrates conformed to the predictions, remaining on or evolving toward a 'line of optimality' representing the optimal oxygen to substrate ratio. For only one of these substrates did the population move away from the predicted optimality. A follow-up study further showed that the genes expressed in evolved lines correspond to the fluxes predicted to be active by FBA [25]. Since flux changes are only sometimes well-correlated with gene expression [26], however, it remains unclear whether FBA can predict the change in internal fluxes. Although indirect, these studies have suggested that FBA might reasonably capture the evolutionary forces acting on cellular physiology and hence would be capable of predicting the outcome of evolution [27].

To our knowledge there have been only two studies in which the internal fluxes have been measured for both ancestral and evolved strains grown in a constant environment with a single growth substrate. Both involved rapid, short-term

adaptation (<1,000 generations) of *E. coli* under conditions where the cultures were kept in continual exponential growth in batch culture by using frequent, large dilutions. Hua *et al* [28] measured fluxes following adaptation to the poorly-utilized substrate lactate, while Fong *et al* [20] measured fluxes following adaptation of a series of *E. coli* strains with knockouts (KOs) deleting individual enzymes of major branches of central carbon metabolism (e.g., glycolysis). Interestingly, both studies found rather divergent changes in flux distribution across replicates, and found that most improvement in growth rate was the result of increases in substrate uptake. These studies were not compared to FBA predictions, however, thus it remains unclear whether the assumed optimality criteria improved, or whether observed intracellular fluxes moved toward those predicted with a genome-scale FBA model.

In terms of using experimental evolution to test optimality, the cultures that have had the greatest time to adapt are those from the *E. coli* long-term experimental evolution (LTEE) populations that have been evolving in the Lenski laboratory for over 50,000 generations [29,30]. These twelve replicate populations have evolved in minimal medium with glucose since 1988, experiencing 100-fold daily dilutions that result in a short lag phase, nearly seven consecutive generations in exponential phase, and then stationary phase. The LTEE experiment has enabled an unprecedented examination of genotypic and phenotypic change over an extended period of adaptation [29,31]. Despite starting with a wild-type strain

capable of rapid growth on glucose, all populations have increased dramatically in both growth rate and competitive fitness through adaptation in batch culture [32,33]. It should be noted however, that batch culture inherently incorporates some non-steady state conditions and that improvements in lag or survival may have had pleiotropic consequences for growth. Despite this, here we ask how well FBA predictions align with the evolved changes in these populations. If FBA is unable to predict adaptation to single-nutrient, seasonal batch culture conditions we will not be able to apply it to most laboratory environments, not to mention the variable habitats experienced in nature.

The goal of the current work was to test whether the central metabolic fluxes of replicate populations of *E. coli* with known selective history in the laboratory evolved in a manner that is predictable by FBA (Figure 4.1). We compared the fluxes inferred from $^{13}C$ labeling to the ranges predicted to permit optimal performance and summarize these changes in three ways: the % *optimality* possible given the inferred fluxes, the minimal *distance* in flux space between the inferred fluxes and the optimal space of distributions, and a *flux-by-flux* comparison to see how each flux changed relative to predictions. Testing the ability of optimality criteria to predict adaptation not only provides insight into the mechanisms of evolution, but also represents a critical test of the central optimality assumption of FBA. The LTEE lines began with an ancestor operating at near-optimal BM/S, but the independent populations evolved to use central

metabolism less optimally. This was reflected in both a small, but statistically significant, decrease in the % optimal BM/S, and a corresponding increase in the distance from the observed to optimal flux state. In contrast, the seven lactate-evolved populations evolved to increase BM/S and moved closer to an optimal flux distribution. The three pairs of KOs had mixed results in terms of optimality and flux pattern. Overall these results indicate that evolved increases in growth rate largely resulted from increased substrate uptake. Furthermore, ancestral strains operating far from optimal yield evolved as suggested by FBA, whereas those close to the optimum experienced a modest decrease in optimality and evolved to be further from FBA predicted fluxes than their ancestor.

## Results

*Growth rate, cell dry weight and carbon uptake all increased after 50,000 generations of adaptation on glucose minimal medium*

Prior to measuring internal metabolic fluxes, we first examined key growth parameters for one isolate from the 50,000 generation time-point for each of 10 independent LTEE populations (Supplementary Table 4.1). Growth rate increased by 45% on average (Supplementary Table 4.1), which is concordant with the 16% increase observed in these lines after 2,000 generations [32], and the 20% increase measured after 20,000 generations [33]. All evolved lines also

increased their glucose uptake rates (individually significant for 5 of 10 lines: A+3, A-2, A-4, A-5, A-6; T-test, $p < 0.05$, two-sample, equal variance throughout unless noted otherwise, Supplementary Table 4.1), with an average increase of 18%. The cell dry weight per gram of glucose also increased by an average of 20% while max $OD_{600}$ increased by 68%. This did not come from decreasing their excretion of organic acids, however, as acetate production actually increased by an average of 50%. No other excreted ions were observed above our limit of detection of ~50 μM (Supplementary Table 4.1).

*LTEE isolates have modest, but significant changes to their relative central metabolic flux distribution*

In order to determine whether the improved performance of the LTEE isolates was reflected in changes in the relative use of central metabolic pathways, we used $^{13}C$-labeling of protein-derived amino acids [17] to infer several key flux ratios in central carbon metabolism (Figure 4.2A). Often the goal is to extrapolate from the measured flux ratios to calculate the flux for each reaction in a network [15,23]. For this study, however, we limit our discussion and analyses to the flux ratios themselves, as these represent the actual number of inferences from the $^{13}C$-labeling data and thus each cellular branch-point is given equal weight (Supplementary text 3.1). It should be noted that $^{13}C$ data for the LTEE isolates were analyzed with a program, FiatFlux [17], which is based on a simplified model of central carbon metabolism. This program was used for the previous

study comparing alternate optimality criteria mentioned above [11], as well as for obtaining the flux data about the lactate [28] and KO [20] lines we analyze below. Inferences with this commonly used program are less variable than inferences based on larger models [34].

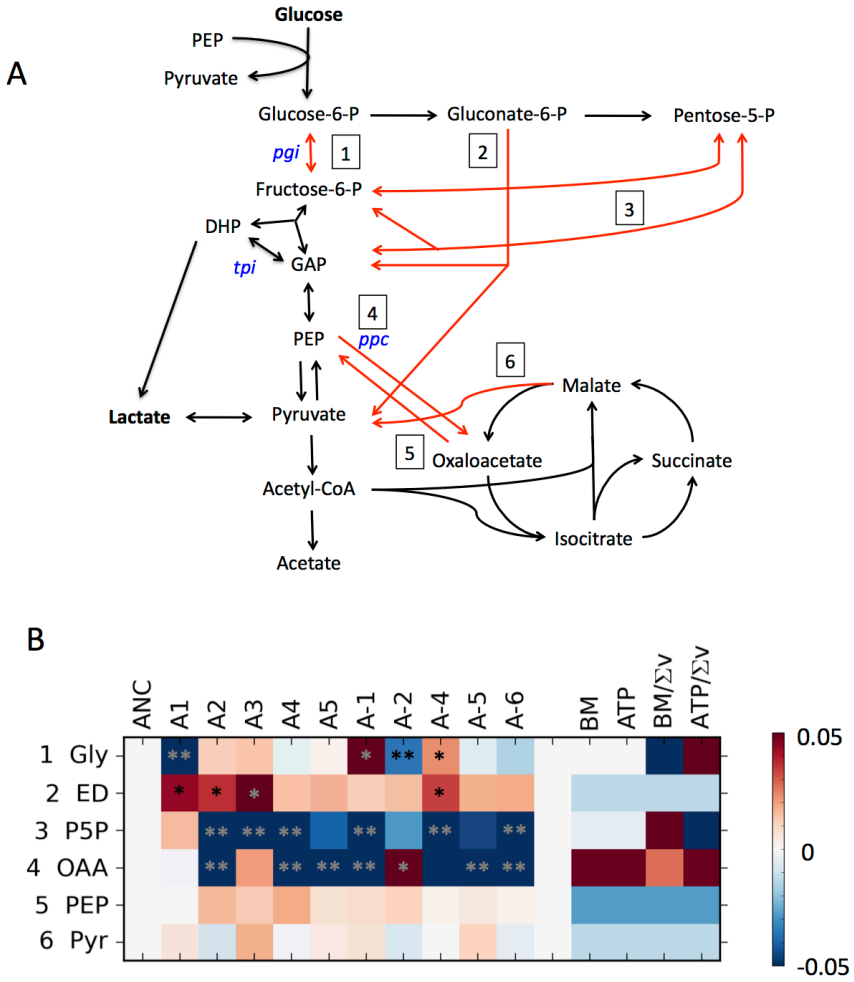



**Figure 4.2: Evolved changes in central carbon metabolism for the LTEE populations after 50,000 generations of adaptation on glucose.** A) The flux pathways measured for the LTEE lines are denoted with numbers and red arrows. The genes knocked out in the knockout data set and the entry point of

**Figure 4.2, continued:** lactate into the network are both indicated. B) A heat map of the difference between evolved and ancestral flux ratios from the LTEE populations. The right side indicates flux ratios predicted for the ancestral line according to each optimality criterion. The number of the flux ratio corresponds to the numbered pathways in A. Single asterisks denote significant changes as calculated by ANOVA, double asterisks are also significant by Tukey-HD.

We uncovered statistically significant, but modest variation in the flux ratios of evolved isolates relative to their ancestor (Figure 4.2B, Supplementary Table 4.2). In terms of the overall pattern, a MANOVA test found that flux ratios changed significantly as a function of population (Pillai's Trace = 3.80, $p<0.001$, Supplementary Figure 4.1). Additionally, ANOVA tests on the flux ratios for individual lines found at least one significantly different isolate ($p < 0.05$) for all ratios except two, and all lines had significant change in at least one flux ratio. A joint linear regression of the populations found 22 fluxes that differed from the ancestor at a $p \leq 0.05$. The False Discovery Rate (FDR) metric suggests that 18 more significant changes were found than expected by chance, whereas the more conservative Tukey HSD test finds that 10 flux changes remain significant.

A few patterns emerged in terms of the actual fluxes found to have changed in evolved isolates. First, the most parallel change was that a small, but significant portion of glucose was routed through the Entner-Doudoroff pathway (Figure 4.2,

flux 2). In all but one case this was accompanied by a similar decrease in the proportion of carbon flowing through the pentose-phosphate pathway (flux 3). On the other hand, replicate lines evolved in opposite directions for flux through glycolysis (flux 1), and for the fluxes producing oxaloacetate from phospho*enol*pyruvate (fluxes 4). Additionally, in all cases there was no significant change in the lower bound of production of pyruvate from malate via malic enzyme (flux 6) across evolved isolates.

*Long-term evolution on glucose did not increase any optimality criterion*

As a first step in testing the validity of different optimality criteria, we asked whether the flux ratios observed in evolved isolates led to increased or decreased performance with regard to each criterion (Figure 4.1A). The '*% optimality*' can be calculated by comparing the maximum value of a criterion when the model was constrained with the observed flux ratios and substrate uptake rate to the maximum value of the criterion in the absence of the flux ratio constraints. Note that because this metric simply compares values of given optimality criteria rather than a particular set of flux ratios it is not affected by the existence of alternate optima for some fluxes.

There was a slight (0.8%) but significant drop in the average percent optimal biomass production (BM/S; T-test, p=0.008), with 9 of the 10 evolved lines

decreasing relative to the ancestor (Figure 4.3A). Turning to alternative optimality criteria, we first found that ATP/S did not change significantly (Figure 4.3D), though unlike all other measures throughout, the output was not normally distributed (Shapiro-Wilk test of residuals, $p = 0.002$; for rest see Figures S2 and S3). Correspondingly, significance for changes in this criterion was tested with the non-parametric Mann-Whitney-Wilcoxon Rank Sum Test ($p = 0.79$). BM/$\Sigma v$ and ATP/$\Sigma v$ behaved qualitatively similarly to BM/S and ATP/S, respectively, but as neither change was significant these results are displayed only in supplementary material (Supplementary Figure 4.4). Finally, we calculated the nearest possible flux distribution for each evolved isolate to the Pareto optimum, and found that 9 of 10 isolates were further from an optimal tradeoff between criteria than the ancestor (Supplementary Figure 4.5).

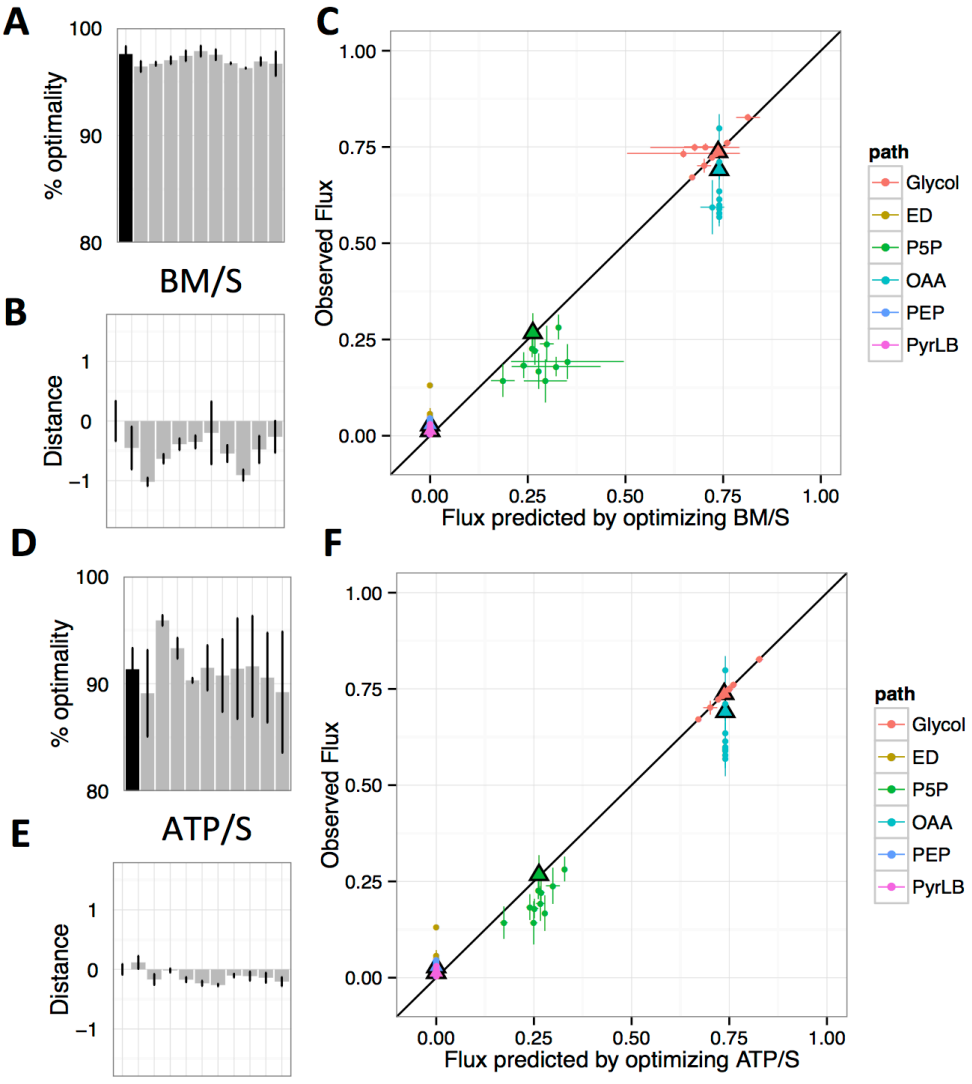



**Figure 4.3: Measures of optimality and predictability after adaptation of LTEE populations to glucose for 50,000 generations.** A,D) The % optimality of the ancestor (black) and evolved isolates (grey, same order as Fig. 2); B,E) distance to optimal flux distribution (plotted as $\log(D_{EO}/D_{AO})$); and C, F) comparison of predicted to observed flux ratios for FBA-predictions based upon BM/S (A-C) or ATP/S (D-F). Error bars represent standard errors of three biological replicates.

In order to test the sensitivity of these findings to assumptions made in using FBA, we compared the effect of changing the values used for $O_2$ limitation, maintenance energy, and the possible change in biomass composition that would result from the documented increase in average cell size [12]. None of these modulations changed the qualitative results and generally the default values outperformed the others (Supplementary Figures 3.6 and 3.7). Therefore, the conclusion that adaptation did not lead to an increase in any optimality criterion for the LTEE populations seems rather robust.

*Long-term glucose evolution resulted in movement of the flux distribution away from predicted states*

We next examined whether the flux distributions we inferred for the LTEE isolates moved toward (or away) from the flux distribution predicted to result from optimizing each criterion. We calculated the *distance* to the optimal fluxes for each evolved isolate relative to the distance between the ancestor and optimality (Figure 4.1B). Because the per-substrate criteria (e.g., BM/S, ATP/S) had many equally-optimal flux distributions, we identified the optimal solution that minimized the Euclidean distance from observed flux ratios. Choosing the FBA solution that is the closest to our empirical flux observations should, if anything, bias in favor of FBA.

Beginning with the overall pattern of fluxes, we quantified the log ratio of evolved to ancestral flux distance to their nearest optimum (Figure 4.1B). Both BM/S and ATP/S predicted optima in the opposite direction of the evolutionary flux movement, and hence evolved lines ended up significantly farther from optima than the ancestor (Figures 3.3B,E; S2; BM/S, T-test p=0.0008; ATP/S, T-test, p = 0.0004). In both cases the movement away from the optimum was primarily driven by changes in the flux of oxaloaceate from phospho*enol*pyruvate.

Turning to individual flux ratios, no criterion fared particularly well (Figure 4.2B, 3C,F). None correctly predicted the observed increased flux through the Entner-Doudoroff pathway, nor the trend of reduced oxaloacetate from phospho*enol*pyruvate in evolved lines.

*Metabolic changes in E. coli evolved on the poor substrate lactate were well-predicted by FBA using BM/S as an optimality criterion*

A second data set we considered was the seven populations of *E. coli* that evolved on the poorly-utilized substrate lactate for ~900 generations [28]. These populations improved in growth rate and cell dry weight substantially (112% and 50%, respectively) in addition to increasing lactate uptake by 40% [28].

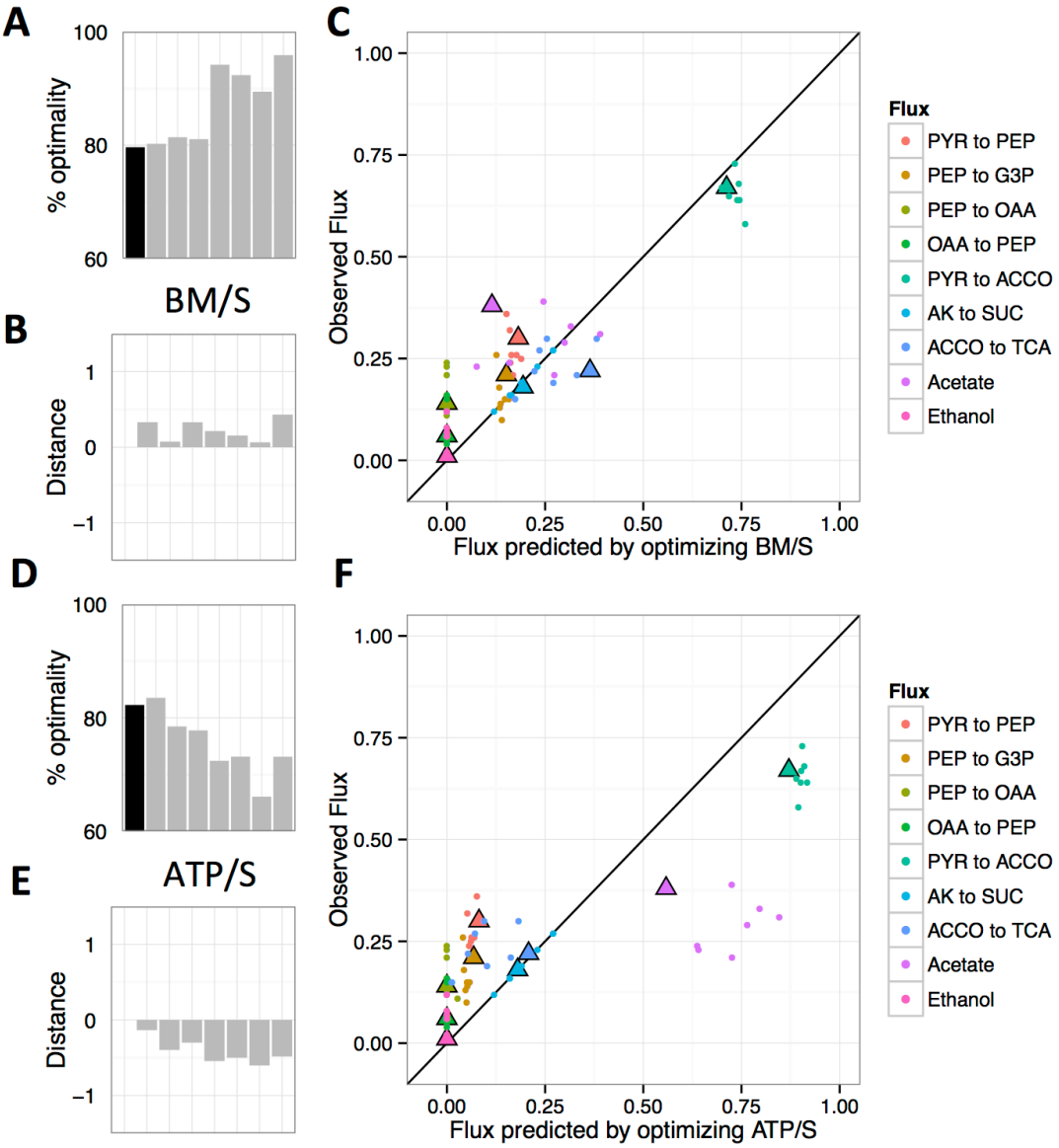



**Figure 4.4: Measures of optimality and predictability after adaptation to lactate for ~900 generations.** A,D) The % optimality of the ancestor (black) and evolved isolates (grey); B,E) distance to optimal flux distribution (plotted as $\log(D_{EO}/D_{AO})$); and C, F) comparison of predicted to observed flux ratios for FBA-predictions based upon BM/S (A-C) or ATP/S (D-F).

We found that adaptation to growth on lactate led to a significant increase of 8% in the predicted percent optimal BM/S (Figure 4.4A; T-test, p = 0.02), whereas the % optimal ATP/S decreased significantly (Figure 4.4D; T-test, p = 0.01) by 7%. The % optimality for BM/Σ*v* and ATP/Σ*v* again qualitatively followed the respective per substrate criteria (Supplementary Figure 4.4). Similarly, fluxes moved closer to the state predicted by BM/S by an average of 20% (Figure 4.4B; T-test, p = 0.005), largely as the result of changes in the predicted and observed flux to acetate (Figure 4.4C,F). In contrast, they moved away from the state predicted by ATP/S (Figure 4.4E; T-test, p = 0.0004). Additionally, 6 of the 7 lactate populations evolved to be further from the Pareto optimal surface than their ancestor (Supplementary Figure 4.5).

*E. coli central metabolic knockouts did not evolve in the direction of FBA predictions*

As a third test of whether strains evolve in a manner consistent with FBA predictions, we considered the results from evolution on glucose for KO populations with lesions in central metabolism (see Figure 4.2A). These data come from two populations each initiated with strains lacking phosphoglucose isomerase (Δ*pgi*), triose-phosphate isomerase (Δ*tpi*) or phospho*enol*pyruvate carboxylase (Δ*ppc*) and evolved for ~800, ~600, and ~750 generations respectively [20]. Considering the improvement of these populations jointly, they

increased in both growth rate and glucose uptake (172% and 157%), had large changes in central metabolic fluxes, but were largely unchanged in dry cell weight (3%). For analyzing changes in their metabolic fluxes, however, we do not present statistical tests of significance given that we only have two observations for each of these three ancestors.

Our analysis of the flux data indicated that, for BM/S, Δ*pgi,* and Δ*tpi* strains got worse while Δ*ppc* strains improved their % optimality (Figure 4.5A). This pattern largely held for ATP/S as well, though Δ*tpi* strains showed essentially no change in % optimality (Figure 4.5C). The KO data set is the only one in which minimizing Σ*v* led to qualitatively different behavior from the per substrate analyses. Minimizing flux led to increases in the % optimality for *Δpgi* and *Δtpi* when using BM/Σ*v* as a criterion (Supplementary Figure 4.4).

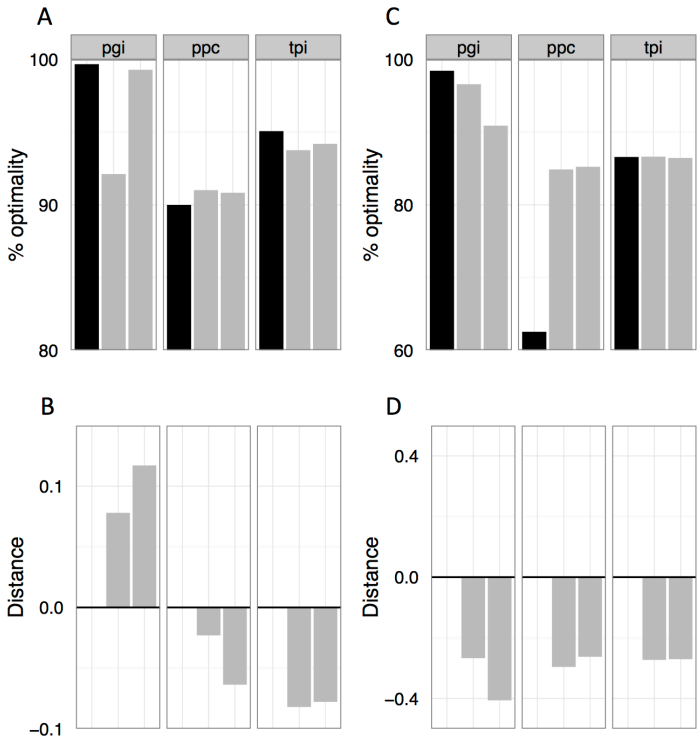



**Figure 4.5: Measures of optimality and predictability after adaptation of gene knockouts on glucose for ~600-800 generations.** A,B) The % optimality of the ancestor (black) and evolved isolates (grey); C,D) distance to optimal flux distribution for FBA-predictions based upon BM/S (A,C) or ATP/S (B,D).

Evolution pushed strains further away from optima in all cases except Δ*pgi* as predicted by BM/S (Figure 4.5B,D). Reduced distance to the optima for Δ*pgi* was driven by reduction in the flux from oxaloacetate to phosphoenolpyruvate in evolved lines. Finally, the two Δ*pgi* evolved isolates evolved to be more Pareto optimal, the Δ*tpi* isolates were essentially equivalent to their ancestor, and the

Δ*ppc* isolates became less Pareto optimal (Supplementary Figure 4.5).

## Discussion

Genome-scale metabolism is sufficiently complex that the current state of the art in predictive models uses stoichiometry and other constraints to define the space of possible flux patterns and then suggests a given state that the cell would adopt if selection had maximized a proposed optimality criterion. The application of a mechanistic evolutionary optimality model to propose a solution to an underdetermined physiological problem is elegant and has been adopted broadly. However, there is a paucity of data testing either the central assumption that intracellular fluxes are optimized by a simple criterion, or which criterion best represents the target of selection. Here we present an analysis of metabolic evolution in the Lenski LTEE populations and make the first direct comparison of observed flux evolution to genome-scale FBA predictions.

Our analysis of the evolution of metabolic fluxes during 50,000 generations of adaptation of *E. coli* on glucose revealed changes in both the absolute and relative fluxes. Concordant with faster growth rates, we observed that all lines had increases in measured glucose uptake. Beyond this, all populations altered the way in which they utilize glucose, with significant changes in flux ratios observed across the network of central carbon metabolism. The most parallel changes in flux distribution were observed in the glycolytic pathways with a universal small,

but significant increase in flux through the Entner-Doudoroff pathway, which was nearly always accompanied by a decrease through the pentose phosphate pathway. This is somewhat perplexing, as the Entner-Doudoroff pathway provides less ATP than glycolysis and no important biosynthetic intermediates. The Entner-Doudoroff pathway is shorter than glycolysis, and hence potentially less enzymatically costly. Indeed, what maintains the pathway in *E. coli* remains an open question, though it has been observed to be upregulated in *E. coli* during long-term starvation [35].

The major basis of improvement during selection upon growth rate for the LTEE populations –as was observed for the lactate and KO populations– came from increasing substrate uptake. We found that the LTEE populations continued to increase their growth rate over the 30,000 generations since it was last reported [33]. Alternative measures of yield, such as cell dry weight and $OD_{600}$, also increased despite the slight decrease in efficiency of biomass production by central metabolism. Cell dry weight depends upon both BM/S in terms of carbon, but can also change due to the relative biomass composition of elements such as nitrogen or phosphorus. $OD_{600}$ is even more indirect, depending upon all of these factors as well as changes in optical properties such as cell size, which is known to have increased in the LTEE [32]. We only measured flux ratios in central carbon metabolism, and thus would have missed significant adaptation that happened in peripheral metabolic pathways. Alternatively, either the bulk composition of

biomass itself or the maintenance energy might change. We addressed these latter two factors in additional analyses (Supplementary Figure 4.7), but neither of these factors significantly alters results.

Data on the evolution of central metabolism for the LTEE populations, combined with prior observations of flux evolution on lactate or by a series of three KO strains provided the opportunity to test several facets of whether the direction of evolutionary change was consistent with FBA predictions.

Across experimental systems we ascertained which proposed optimality criteria are most often consistent with the observed evolution in central metabolism. On average across five different ancestors, BM/S outperformed the other criteria in terms of either increasing or going unchanged (Supplementary Figure 4.8). The most dramatic example was seen for the lactate-evolved populations, for which BM/S increased while ATP/S decreased. The per flux criteria (BM/Σ$v$ and ATP/Σ$v$) behaved qualitatively the same as the per substrate criteria in all but two of the cases (*Δpgi* and *Δtpi*). BM/Σ$v$ outperformed BM/S in these two cases, but, for example, did not significantly improve in the lactate populations. The data also suggest that cultures quite often evolved to be further from their Pareto optimum representing the space of optimal tradeoffs [15], with 19 of 23 populations in total moving further from the Pareto surface than their respective ancestral genotypes. These results suggest that optimal biomass yield –which is

the most commonly utilized criterion for FBA– was the best overall stoichiometric proxy for cultures where selection was directly upon growth rate. It will be quite interesting to analyze populations grown in a manner where yield (BM/S) is directly selected.

Overall, approximately half of the flux data were consistent with FBA predictions, and half refuted the common assumption that evolution acts to optimize efficiency; what accounts for this discrepancy? The major factor that appears to account for this difference is the initial degree of optimality for the ancestor of the evolved lines (Figure 4.6). For the lactate and Δ*ppc* populations, which began at approximately 80% and 90% optimality for BM/S, all 9 total replicates increased in BM/S. On the other hand, 13 of 14 populations starting at or above 95% efficiency –LTEE and the other two KOs– decreased in BM/S. A negative correlation holds whether one performs a parametric statistical test (Pearson correlation, $p < 0.0001$) or a non-parametric Spearman correlation coefficient ($p < 0.0001$), though it should be noted that the strength of the correlation is largely driven by the lactate data set.

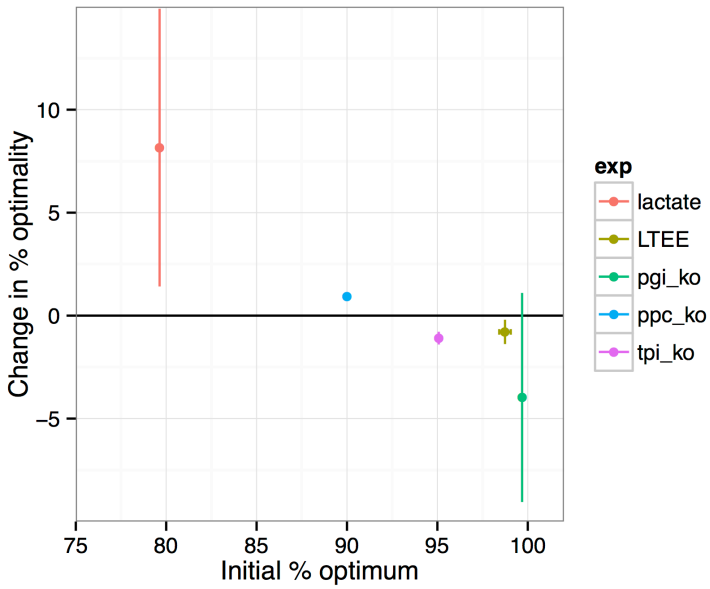



**Figure 4.6: Evolutionary change in % optimality versus initial % optimality of the ancestor across data sets for BM/S.** Error bars represent standard errors between evolved populations.

The finding that selection on optimal efficiency depends on distance to the optimum is both of practical and fundamental interest. The analysis represents the first direct demonstration that FBA can be used to predict changes in intracellular metabolism that result from adaptation on a single carbon source. This positive result comes with the caveat that strains must begin far from the optimum. Systems initially operating at high yield –like the LTEE and the Δ*pgi* strains that both began above 98% optimal– may end up evolving to be further from optimal than they began. In other words, this suggests one can either predict the initial physiological state or the direction of evolution, but not both.

What is perhaps the most remarkable about these findings is that even for cultures with a negative correlation between rate and yield, the tradeoff was quite modest. Small decreases in BM/S were more than made up for by large increases in uptake, leading to a net increase in growth rate despite mild antagonism. Given that there is no direct selection upon yield during batch culture, this perhaps suggests the existence of constraints upon the further improvement of substrate uptake. As long as uptake is held constant then changes in yield would directly translate into changes in growth rate. As such, this would maintain purifying selection upon yield, even over 50,000 generations. On the other hand, the low efficiency ancestors were able to evolve both improved substrate uptake and yield simultaneously.

Although FBA is typically applied as a practical tool to guide experiments –and it has had some remarkable successes, such as correctly predicting a rather unexpected new metabolic pathway in some cancers [36]– it also serves as a quantitative, testable, falsifiable model that connects physiology to evolution. The interplay of optimality models and laboratory adaptation will be critical as the field continues to move toward a fuller understanding of the selection and constraints that act upon biochemical networks.

## Materials and Methods

*Strains and growth conditions during selection*

*Escherichia coli* B isolates were obtained from the Lenski LTEE experiment [29] after 50,000 generations. Briefly, 12 populations of *E. coli* were founded with either the arabinose-negative strain REL606 (populations A-1 to A-6) or the arabinose-positive derivative, REL607 (A+1 to A+6). These were evolved in 10 mL of Davis-Mingioli minimal medium with 139 μM glucose (25 mg/L) as a growth substrate in 50 mL flasks since 1988. These lines have been cultured at 37 °C while shaking at 120 rpm and have been transferred daily via 1:100 dilutions (~6.64 net doublings per day).

The isolates analyzed in the current experiment consisted of the ancestral line, REL606 [29], as well as the 'A' clone from 10 of the 12 lines frozen at 50,000 generations that were used in an earlier paper (A-1 = REL11330; A-2 = REL11333; A-4 = REL11336; A-5 = REL11339; A-6 = REL11389; A+1 = 11392; A+2 = REL11342; A+3 = REL11345; A+4 = REL11348, A+5 = REL11367)[30]. The A-2 clone used is from the 'large' lineage that has coexisted with a cross-feeding 'small' lineage for tens of thousands of generations [37]. The isolate from the citrate-consuming population A-3 (REL11364) was not used because it adapted to citrate consumption in addition to glucose [38]. The A+6 isolate (REL11370) was excluded from analysis because it had inconsistent growth, and gave irregular

flux data. This population was previously excluded from a study of growth rate vs. yield at 20,000 generations for similar reasons [13].

*Measurement of key metabolic flux ratios*

Flux measurements were obtained based on the methods of Zamboni *et al* [17]. Evolved isolates were grown in 150 mL of Davis-Mingioli minimal media with 139 μM glucose without sodium citrate (excluded to ensure that it was not used as a secondary carbon source by any line). In order to obtain information from different parts of central metabolism, $^{13}$C-labeling either utilized a 20:80 ratio of [U-$^{13}$C]labeled:unlabeled glucose or 100% [1-$^{13}$C]glucose (Cambridge Isotope Laboratories, Andover, MA). The ancestral REL606 was grown in 200 mL to obtain sufficient cell material. At mid-log phase (60-80% max OD) all cells were pelleted from the media, hydrolyzed overnight in 6 M HCl, and dried. The dry cell material was then derivatized for an hour at 85 °C with 40 μL each of dimethylformamide and *N*-*tert*-butyldimethylsilyl-*N*-methyltrifluoroacetamide with 1% *tert*-butyldimethyl-chlorosilane. The derivatized cell material was injected into a Shimadzu QP2010 GCMS (Columbia, MD). The injection source was 230 °C. The oven was held at 160 °C for 1 min, ramped to 310 °C at 20 °C min$^{-1}$, and finally held at 310 °C for 0.5 min. Flow rate was 1 mL min$^{-1}$ and split was 10. The column was a 30 m Rxi-1ms (Restek, Bellefonte, PA). Three technical and three biological replicates were run for each isolate.

Data files from the GC-MS were analyzed in FiatFlux [17], as had been used for the lactate [28] and KO [20] populations we also analyzed. The data conversion files were rewritten to load the raw spectra produced by our MS. Following the established protocol, uninformative amino acid fragments were removed. Means for each biological replicate were calculated from the average of three technical replicates. Shapiro-Wilk tests were used to validate the assumption of normally-distributed errors for estimated flux ratios for each strain (Figures S2, S3). Variance in flux ratios was then analyzed with a MANOVA test using the Pillai's Trace metric with flux ratios entered as separate dependent variables (Supplementary Figure 4.1). Univariate ANOVA tests were also run to investigate which of the measured flux ratios changed significantly for individual strains.

The flux of oxaloacetate (OAA) from phospho*enol*pyruvate (PEP) was further estimated by a Monte-Carlo method to determine the contribution of the glyoxylate shunt. The method follows Waegeman *et al,* 2011 [23] and uses MATLAB code they kindly provided. In short, average mass distribution vectors and standard deviations were calculated from the measured samples. The 'normrand' function was then used to randomly draw from these mean distributions 1000 times. For each draw, a grid search was used to find the flux ratios that best fit the mass distribution vectors. Substantial variation was found for the fraction of labeled $CO_2$ and flux through the glyoxylate shunt, but in all

cases there was very strong support for the flux ratio of oxaloacetate from phospho*enol*pyruvate that had previously been calculated by FiatFlux.

*Physiological measurements of growth rate, cell dry weight, glucose uptake and acetate excretion*

Uptake and production of cell material were determined in a separate set of experiments. In these experiments glucose concentrations were increased ten-fold to 1.39 mM so that enough of the compounds would be present to measure precisely. A volume of 250 μL of overnight culture was inoculated into 50 mL of media grown in a 250 mL flask at 225 rpm. Growth rate was determined by fitting a logarithmic model to $OD_{600}$ measurements. A 10 mL sample was removed at early ($OD_{600}$ of 0.090-0.120) and late ($OD_{600}$ of 0.275-0.400) log phase. Cells were immediately removed from the media by passage through a 0.2 μM filter. Glucose concentrations were determined in the spent media using a glucose oxidase assay kit (Sigma, Saint Louis, MO). Acetate concentrations were determined by ion chromatography with a Dionex ICS-200 RFIC. The flow rate was 1.5 ml/min and the column temperature was 30 °C. Cell dry weight (CDW), was measured as the mass of the pellet from 100 mL of fully-grown culture after overnight lyophilization. Three replicates were assayed for each measurement.

*Calculation of variation in flux ratios across evolved isolates*

The degree of parallelism between replicates in the evolution of flux ratios was determined by calculating the coefficient of variation in flux ratios. For each flux ratio the standard deviation between evolved replicates was divided by the mean of that flux ratio. This value was then averaged across all flux ratios. Values close to zero indicate a high degree of similarity between evolved lines.

*Prediction of FBA optima*

Flux analysis was carried out with a genome-scale model of *E. coli* metabolism (iAF_1260 [39]). The model incorporates 2382 reactions and 1668 metabolites. Substrate uptake and excretion were constrained to that observed, otherwise the default minimal media environment was used. The lower bound on maintenance energy was left at the default value of 8.9 mmol ATP/g/hr. Oxygen uptake rates were set to those observed for the lactate strains; however these data were not available for the REL or KO strains. In these cases, oxygen uptake for the ancestor was scaled across the previously observed range of 11.5-14.75 mmol/gCDW/hr [11]. Previous work demonstrated that oxygen uptake varies as a function of evolution, but that the ratio of substrate to oxygen usage remained largely constant [24]. Oxygen constraints for evolved lines were therefore set based on evolved glucose uptake rates and the ancestral ratio of oxygen/glucose.

Changing the value of ancestral oxygen constraint, or the slope of constraint line had little qualitative effect (Supplementary Figure 4.6), so just the results based on an ancestral uptake of 14.75 mmol/g/hr and a slope maintaining the original oxygen/glucose rates are reported in the text. Gene knockouts were simulated by constraining flux through the missing gene to zero.

For all data sets we systematically tested the predictive ability of four different optimality criteria: max biomass per unit substrate (BM/S), max ATP per unit substrate (ATP/S), max biomass per unit flux (BM/$\Sigma v$) and max ATP per unit flux (ATP/$\Sigma v$). These criteria relate to the best performers in Schuetz *et al* 2007 [11] and were defined as in that study. The per-substrate criteria maximized the criterion and then subsequently chose a flux distribution that minimized the difference from the observed isolate ratios. This process always provides a flux distribution with maximal production of ATP (or biomass). The per-flux criteria optimize the ratio of ATP (or biomass) to the sum of the flux. Optimizing this ratio leads to a single optimal flux solution that often produces less than the maximal ATP (or biomass). For ATP criteria, flux to excess ATP use (via maintenance energy) was maximized while constraining the lower limit of biomass production to the ancestral growth rate.

Minimizing the distance between observed and predicted optimal flux distributions was accomplished by minimizing a distance term. Flux ratios can be

constrained by adding a row to the S matrix such that:

$$V_2 - R*V_1 = 0$$

Where $V_n$ is the flux through reaction n and R is the ratio $V_2/V_1$. To minimize distance between observed and predicted ratios the equation becomes:

$$V_2 - R*V_1 + D = 0$$

$$V_2 - R*V_1 - D = 0$$

Where D represents distance from the observed ratio and is added as two columns to the S matrix (and concomitant rows in the flux vector). Biomass or ATP can be constrained to its maximum value and then the flux distribution that is closest to observed values can be calculated by running linear optimization minimizing D as the objective function.

*Comparison of experimental flux ratios to FBA-predicted optima*

We first tested whether flux ratios evolve to increase each selective criterion. The optimal value of each criterion was compared against the maximum value of the criterion when the model was constrained to have the experimentally observed flux ratios. Percent optimality, calculated as the constrained criterion divided by optimal criterion, was determined for the ancestor and evolved lines.

For the LTEE lines the constrained flux ratios were serine through glycolysis, pyruvate through Entner-Doudoroff, oxaloacetate from phospho*enol*pyruvate,

phospho*enol*pyruvate from oxaloacetate, and the pyruvate from malate. The ratios were calculated following Fischer and Sauer 2003 [40]; the exact equations used are provided in the supplementary material (Supplementary Table 4.3). Each ratio was constrained by adding a row to the S matrix that defined the relationship between relevant fluxes (as described in the first equation of the previous section). The ratio inferred for pyruvate from malate was treated either as an absolute constraint or a lower bound but because all optimality criteria push this value towards 0 the results were equivalent.

To propagate uncertainty in glucose uptake, acetate excretion and flux ratios for the LTEE isolates, separate calculations of properties such as BM/S were made for each of 3 biological replicates, which themselves represented the average of 3 technical replicates. The mean and standard error for optimality metrics was calculated for each strain from the biological replicates.

Flux constraints for lactate and knockout data sets were implemented as upper and lower bounds, because reported flux ratios were relative to substrate uptake rather than other internal fluxes. Lactate adaptation lines were constrained to have flux ratios ± 5% of the values reported in Hua *et al* 2007 [28]. Gene knockout lines were constrained with the flux ratios and errors reported in Fong *et al*, 2006 [20].

To determine whether strains evolved towards predicted optimal intracellular physiologies we used a standardized metric to ask if evolved lines were closer to an optimal solution than the ancestor. This distance metric was calculated as:

$$\log(D_{EO}/D_{AO})$$

where $D_{EO}$ was the distance of the evolved flux ratios from the closest optimal solution, and $D_{AO}$ was the distance of the ancestor from its closest optimal solution. Distances were calculated as Euclidean distance between the flux ratios observed in each data set and those predicted. It should be noted that because optimal flux ratios change with substrate uptake the ancestral and evolved optima were different points. The metric is 0 if the evolved isolate distance has not changed relative to the ancestor, increasingly positive as the evolved strain moves nearer an optimum, and increasingly negative as it moves further away.

*Pareto optimality*

A Pareto optimal surface was calculated for each line by constraining the substrate uptake rate and then doing a nested grid search [15]. A grid search across the range of feasible biomass values was executed. At each value of biomass a grid search of ATP yields was carried out and the sum of fluxes was subsequently minimized at every interval. Conservatively, for each isolate we then determined the closest possible position to its optimal surface given the observed constraints. Distance between the isolate and the Pareto optimal surface

was calculated from the difference in standardized criteria.

*Statistical tests*

The normality assumption for physiological measurements for the LTEE populations and optimality metrics for all data sets were checked with the Shapiro-Wilk test on the residuals of the linear model fitting the metric against strains. In all but one case the null hypothesis that the distribution was normal could not be rejected at $p < 0.05$. The % optimality for the LTEE lines with ATP/S as the optimality criterion was not normally distributed. Q-Q plots are presented in the supplementary material (Figures S2 and S3).

For the LTEE lines ancestral versus evolved values were compared with two-sided, two sample T-tests assuming equal variance. For the non-normal ATP/S comparison a Mann-Whitney Wilcoxon Rank Sum Test was used instead. For the lactate populations only a single value was available for the ancestor so two-sided, one-sample T-tests were performed testing against the ancestral value as the mean.

*Author Contributions*

Conceived and designed the experiments: WRH CJM. Performed the experiments: WRH. Analyzed the data: WRH NFD NL NK CJM. Contributed reagents/materials/analysis tools: WRH NFD NK. Wrote the paper: WRH NFD NL NK CJM.

*Acknowledgements*

We are extremely grateful to Richard Lenski and Neerja Hajela for providing us with the 50,000 generation isolates of the long-term evolved *E. coli*, to Uwe Sauer, Jörg Stelling, and two anonymous reviewers for helpful criticism of our approach, Colleen Hansel for allowing us to use her ion chromatography system, Alex Bradley for assistance establishing the GC-MS method, and Hendrik Waegeman for providing code for analysis of the glyoxylate shunt. We thank Daniel Segrè, Deepa Agashe, and members of the Marx Lab for helpful comments.

## References

1. Ellis T, Wang X, Collins JJ (2009) Diversity-based, model-guided construction of synthetic gene networks with predicted functions. Nat Biotechnol 27: 465–471.
2. Carothers JM, Goler JA, Juminaga D, Keasling JD (2011) Model-driven engineering of RNA devices to quantitatively program gene expression. Science 334: 1716–1719.
3. Tan Y, Liao JC (2012) Metabolic ensemble modeling for strain engineers.

Biotechnol J 7: 343–353.

4. Jestin JL, Kempf A (2009) Optimization models and the structure of the genetic code. J Mol Evol 69: 452–457.

5. Savir Y, Noor E, Milo R, Tlusty T (2010) Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape. Proc Natl Acad Sci USA 107: 3475–3480.

6. Heineman RH, Bull JJ (2007) Testing optimality with experimental evolution: lysis time in a bacteriophage. Evolution 61: 1695–1709.

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

8. Varma A, Palsson BO (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. Appl Environ Microbiol 60: 3724–3731.

9. Gould SJ, Lewontin RC (1979) The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc R Soc Lond B Biol Sci 205: 581–598.

10. Feist AM, Palsson BO (2010) The biomass objective function. Curr Opin Microbiol 13: 344–349.

11. Schuetz R, Kuepfer L, Sauer U (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. Mol Syst Biol 3: 119.

12. Lenski RE, Mongold JA, Sniegowski PD, Travisano M, Vasi F, et al. (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.

13. Novak M, Pfeiffer T, Lenski RE, Sauer U, Bonhoeffer S (2006) Experimental tests for an evolutionary trade-off between growth rate and yield in *E. coli*. Am Nat 168: 242–251.

14. Kacser H, Burns JA (1973) The control of flux. Symp Soc Exp Biol 27: 65–104.

15. Schuetz R, Zamboni N, Zampieri M, Heinemann M, Sauer U (2012) Multidimensional optimality of microbial metabolism. Science 336: 601–

604.

16. Nagrath D, Avila-Elchiver M, Berthiaume F, Tilles AW, Messac A, et al. (2007) Integrated energy and flux balance based multiobjective framework for large-scale metabolic networks. Ann Biomed Eng 35: 863–885.

17. Zamboni N, Fendt S-M, Rühl M, Sauer U (2009) (13)C-based metabolic flux analysis. Nat Protoc 4: 878–892.

18. Fischer E, Zamboni N, Sauer U (2004) High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived 13C constraints. Anal Biochem 325: 308–316.

19. Van Dien SJ, Strovas T, Lidstrom ME (2003) Quantification of central metabolic fluxes in the facultative methylotroph *Methylobacterium extorquens* AM1 using 13C-label tracing and mass spectrometry. Biotechnol Bioeng 84: 45–55.

20. Fong SS, Nanchen A, Palsson BO, Sauer U (2006) Latent pathway activation and increased pathway capacity enable *Escherichia coli* adaptation to loss of key metabolic enzymes. J Biol Chem 281: 8024–8033.

21. Nanchen A, Fuhrer T, Sauer U (2007) Determination of metabolic flux ratios from 13C-experiments and gas chromatography-mass spectrometry data: protocol and principles. Methods Mol Biol 358: 177–197.

22. Chen X, Alonso AP, Allen DK, Reed JL, Shachar-Hill Y (2011) Synergy between (13)C-metabolic flux analysis and flux balance analysis for understanding metabolic adaptation to anaerobiosis in *E. coli*. Metab Eng 13: 38–48.

23. Waegeman H, Beauprez J, Moens H, Maertens J, De Mey M, et al. (2011) Effect of *iclR* and *arcA* knockouts on biomass formation and metabolic fluxes in *Escherichia coli* K12 and its implications on understanding the metabolism of *Escherichia coli* BL21 (DE3). BMC Microbiol 11: 70.

24. Ibarra RU, Edwards JS, Palsson BO (2002) *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature 420: 186–189.

25. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, et al. (2010) Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. Mol Syst Biol 6: 390.

26. Rossell S, van der Weijden CC, Lindenbergh A, van Tuijl A, Francke C, et al. (2006) Unraveling the complexity of flux regulation: a new method demonstrated for nutrient starvation in *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA 103: 2166–2171.

27. Papp B, Notebaart RA, Pál C (2011) Systems-biology approaches for predicting genomic evolution. Nat Rev Genetics 12: 591–602.

28. Hua Q, Joyce AR, Palsson BØ, Fong SS (2007) Metabolic characterization of *Escherichia coli* strains adapted to growth on lactate. Appl Environ Microbiol 73: 4639–4647.

29. Lenski R, Rose M, Simpson S (1991) Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. Am Nat 138: 1315–1341.

30. Leiby N, Harcombe WR, Marx CJ (2012) Multiple long-term, experimentally-evolved populations of *Escherichia coli* acquire dependence upon citrate as an iron chelator for optimal growth on glucose. BMC Evol Biol 12: 151.

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

32. Vasi F, Travisano M, Lenski R (1994) Long-term experimental evolution in *Escherichia coli*. II. Changes in life-history traits during adaptation to a seasonal environment. Am Nat 144: 432–456.

33. Cooper VS, Bennett AF, Lenski RE (2001) Evolution of thermal dependence of growth rate of *Escherichia coli* populations during 20,000 generations in a constant environment. Evolution 55: 889–896.

34. Suthers PF, Burgard AP, Dasika MS, Nowroozi F, Van Dien S, et al. (2007) Metabolic flux elucidation for large-scale models using 13C labeled isotopes. Metab Eng 9: 387–405.

35. Pin C, Rolfe MD, Muñoz-Cuevas M, Hinton JCD, Peck MW, et al. (2009) Network analysis of the transcriptional pattern of young and old cells of *Escherichia coli* during lag phase. BMC Syst Biol 3: 108.

36. Frezza C, Zheng L, Folger O, Rajagopalan KN, MacKenzie ED, et al. (2011) Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. Nature 477: 225–228.

37. Rozen DE, Schneider D, Lenski RE (2005) Long-term experimental evolution in *Escherichia coli*. XIII. Phylogenetic history of a balanced polymorphism. J Mol Evol 61: 171–180.

38. 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 USA 105: 7899–7906.

39. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, et al. (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol 3: 121.

40. Fischer E, Sauer U (2003) Metabolic flux profiling of *Escherichia coli* mutants in central carbon metabolism using GC-MS. J FEBS 270: 880–891.

41. Raghunathan A, Reed J, Sookil S, Palsson B, Daefler S (2009) Constraint-based analysis of metabolic capacity of *Salmonella typhimurium* during host-pathogen interaction. BMC Syst Biol 3:38.

## Supplementary Material

**Supplementary Table 3.1:** List of Lenski LTEE isolates used.

| 20,000 generation strains | | 50,000 generation strains | |
|---|---|---|---|
| A-1 | REL8593 | A-1 | REL11330 |
| A-2L | REL8594 | A-2L (Large- A clone) | REL11333 |
| | | A-2S (Small- C clone) | REL11335 |
| A-3 | REL8595 | A-3 | REL11364 |
| A-4 | REL8596 | A-4 | REL11336 |
| A-5 | REL8597 | A-5 | REL11339 |
| A-6 | REL8598 | A-6 | REL11389 |
| A+1 | REL9282 | A+1 | REL11392 |
| A+2 | REL8601 | A+2 | REL11342 |
| A+3 | REL8602 | A+3 | REL 11345 |
| A+4 | REL8603 | A+4 | REL11348 |
| A+5 | REL8604 | A+5 | REL11367 |
| A+6 | REL8605 | A+6 | REL11370 |

**Supplementary Table 3.2:** Reactions predicted by FBA as necessary for optimal growth on substrates. These are the reactions used to compare substrate differences with FBA. The substrates for which the reactions are necessary are listed, as well as the percent of optimal growth that is possible without the reactions and related genes.

| Substrate | Reaction name | % optimality without reaction | Necessary genes |
|---|---|---|---|
| fumarate | Fumarate transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| fumarate | malic enzyme NADP | 98.9 | maeB |
| fumarate | phosphoenolpyruvate carboxykinase | 98.0 | pckA |
| melibiose | galactokinase | 0.0 | galK wcaK |
| melibiose | a galactosidase melibiose | 0.0 | melA |
| melibiose | melibiose transport in via symport periplasm | 0.0 | melB |
| melibiose | melibiose transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| melibiose | phosphoglucomutase | 94.8 | pgm yqaB |
| melibiose | pyruvate kinase | 96.8 | pykF pykA |
| melibiose | UDPglucose hexose 1 phosphate uridylyltransferase | 0.0 | galT |
| acetate | Isocitrate lyase | 0.0 | aceA |
| acetate | phosphoenolpyruvate carboxykinase | 96.9 | pckA |
| D-alanine | D Alanine transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| D-alanine | Isocitrate lyase | 97.4 | aceA |
| D-alanine | phosphoenolpyruvate synthase | 96.4 | ppsA |
| L-proline | iron Fe2 transport out via proton antiport periplasm | 0.0 | yiiP |
| L-proline | Fe III reduction | 79.3 | fre |
| L-proline | 1 pyrroline 5 carboxylate dehydrogenase | 0.0 | putA |
| L-proline | phosphoenolpyruvate carboxykinase | 95.7 | pckA |
| L-proline | Proline dehydrogenase | 0.0 | putA |
| L-proline | L proline transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| L-aspartate | L aspartate transport in via proton symport periplasm | 97.6 | gltP |
| L-aspartate | L aspartate transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| L-aspartate | phosphoenolpyruvate carboxykinase | 98.0 | pckA |
| mucic acid | galactarate dehydratase | 0.0 | garD |

**Supplementary Table 3.2, Continued**

| | | | |
|---|---|---|---|
| mucic acid | D galactarate transport via proton symport reversible periplasm | 0.0 | gudP garP |
| mucic acid | D galactarte transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| mucic acid | 5 dehydro 4 deoxyglucarate aldolase | 0.0 | garL |
| mucic acid | pyruvate kinase | 96.8 | pykF pykA |
| L-malate | Malate transport via proton symport 2 H periplasm | 97.5 | dctA |
| L-malate | Malate transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| L-malate | malic enzyme NADP | 98.9 | maeB |
| L-malate | phosphoenolpyruvate carboxykinase | 98.0 | pckA |
| D-ribose | pyruvate kinase | 93.8 | pykF pykA |
| D-ribose | ribokinase | 0.0 | rbsK |
| D-ribose | D ribose transport via ABC system periplasm | 0.0 | rbsD rbsA rbsC rbsB alsC alsA alsB ytfQ ytfT yjfF ytfR |
| D-ribose | ribose transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| glucose | 3 hydroxyacyl acyl carrier protein dehydratase n C100 | 0.0 | fabZ fabA |
| glucose | 3 hydroxyacyl acyl carrier protein dehydratase n C121 | 0.0 | fabZ fabA |
| glucose | 3 hydroxyacyl acyl carrier protein dehydratase n C141 | 0.0 | fabZ fabA |
| glucose | 3 hydroxyacyl acyl carrier protein dehydratase n C161 | 0.0 | fabZ fabA |
| glucose | 3 hydroxyacyl acyl carrier protein dehydratase n C40 | 0.0 | fabZ fabA |
| glucose | 3 hydroxyacyl acyl carrier protein dehydratase n C60 | 0.0 | fabZ fabA |
| glucose | 3 hydroxyacyl acyl carrier protein dehydratase n C80 | 0.0 | fabZ fabA |
| glucose | 3 oxoacyl acyl carrier protein reductase n C100 | 0.0 | fabG |
| glucose | 3 oxoacyl acyl carrier protein reductase n C121 | 0.0 | fabG |
| glucose | 3 oxoacyl acyl carrier protein reductase n C140 | 0.0 | fabG |
| glucose | 3 oxoacyl acyl carrier protein reductase n C141 | 0.0 | fabG |
| glucose | 3 oxoacyl acyl carrier protein reductase n C161 | 0.0 | fabG |
| glucose | 3 oxoacyl acyl carrier protein reductase n C40 | 0.0 | fabG |
| glucose | 3 oxoacyl acyl carrier protein reductase n C60 | 0.0 | fabG |
| glucose | 3 oxoacyl acyl carrier protein reductase n C80 | 0.0 | fabG |
| glucose | 3 oxoacyl acyl carrier protein synthase n C100 | 0.0 | fabF fabB |
| glucose | 3 oxoacyl acyl carrier protein synthase n C121 | 0.0 | fabB |

**Supplementary Table 3.2, Continued**

| | | | |
|---|---|---|---|
| glucose | 3 oxoacyl acyl carrier protein synthase n C140 | 0.0 | fabF fabB |
| glucose | 3 oxoacyl acyl carrier protein synthase n C141 | 0.0 | fabB |
| glucose | 3 oxoacyl acyl carrier protein synthase n C161 | 0.0 | fabB |
| glucose | 3 oxoacyl acyl carrier protein synthase n C60 | 0.0 | fabF fabB |
| glucose | 3 oxoacyl acyl carrier protein synthase n C80 | 0.0 | fabF fabB |
| glucose | arabinose 5 phosphate isomerase | 0.0 | gutQ yrbH |
| glucose | acetyl CoA carboxylase | 0.0 | accA accD accB accC |
| glucose | acetylglutamate kinase | 0.0 | argB |
| glucose | N acetylglutamate synthase | 0.0 | argA |
| glucose | 2 aceto 2 hydroxybutanoate synthase | 0.0 | ilvI ilvH ilvN ilvB |
| glucose | acetolactate synthase | 0.0 | ilvI ilvH ilvN ilvB |
| glucose | acetylornithine deacetylase | 0.0 | argE |
| glucose | aconitase half reaction A Citrate hydro lyase | 0.0 | acnB acnA |
| glucose | aconitase half reaction B Isocitrate hydro lyase | 0.0 | acnB acnA |
| glucose | 4 aminobenzoate synthase | 0.0 | pabC |
| glucose | 4 amino 4 deoxychorismate synthase | 0.0 | pabB pabA |
| glucose | adenylyl sulfate kinase | 0.0 | cysC |
| glucose | adenylsuccinate lyase | 0.0 | purB |
| glucose | adenylosuccinate lyase | 0.0 | purB |
| glucose | adenylosuccinate synthase | 0.0 | purA |
| glucose | 1 hexadecanoyl sn glycerol 3 phosphate O acyltransferase n C160 | 0.0 | plsC |
| glucose | 1 hexadec 7 enoyl sn glycerol 3 phosphate O acyltransferase n C161 | 0.0 | plsC |
| glucose | S adenosylhomocysteine nucleosidase | 0.0 | pfs |
| glucose | phosphoribosylaminoimidazolecarboxamide formyltransferase | 0.0 | purH |
| glucose | phosphoribosylaminoimidazole carboxylase | 0.0 | purK |
| glucose | D alanine D alanine ligase reversible | 0.0 | ddlB ddlA |
| glucose | alanine racemase | 0.0 | dadX alr |
| glucose | 4 amino 2 methyl 5 phosphomethylpyrimidine synthetase | 0.0 | thiC |
| glucose | anthranilate phosphoribosyltransferase | 0.0 | trpD |
| glucose | anthranilate synthase | 0.0 | trpD trpE |
| glucose | 5 amino 6 5 phosphoribosylamino uracil reductase | 0.0 | ribD |
| glucose | argininosuccinate lyase | 0.0 | argH |
| glucose | argininosuccinate synthase | 0.0 | argG |
| glucose | aspartate 1 decarboxylase | 0.0 | panD |
| glucose | aspartate carbamoyltransferase | 0.0 | pyrI pyrB |

**Supplementary Table 3.2, Continued**

| | | | |
|---|---|---|---|
| glucose | aspartate kinase | 0.0 | thrA metL lysC |
| glucose | ATP phosphoribosyltransferase | 0.0 | hisG |
| glucose | ATP synthase four protons for one ATP periplasm | 35.9 | atpC atpD atpG atpA atpH atpF atpE atpB atpI |
| glucose | calcium Ca2 transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| glucose | Carbamate kinase | 98.9 | yahI arcC yqeA |
| glucose | 4 cytidine 5 diphospho 2 C methyl D erythritol kinase | 0.0 | ispE |
| glucose | chorismate mutase | 0.0 | pheA tyrA |
| glucose | chorismate synthase | 0.0 | aroC |
| glucose | Chorismate pyruvate lyase | 0.0 | ubiC |
| glucose | chloride transport out via proton antiport 21 periplasm | 0.0 | yadQ ynfJ |
| glucose | chloride Cl 1 transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| glucose | cobalt Co2 transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| glucose | cobalt transport in via permease no H | 0.0 | ygiE corA |
| glucose | citrate synthase | 0.0 | gltA |
| glucose | CTP synthase glutamine | 0.0 | pyrG |
| glucose | copper Cu2 transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| glucose | copper transport in via permease no H | 0.0 | ygiE |
| glucose | cysteine synthase | 0.0 | cysK cysM |
| glucose | cystathionine b lyase | 0.0 | malY metC |
| glucose | cytochrome oxidase bo3 ubiquinol 8 4 protons periplasm | 80.9 | cyoD cyoC cyoB cyoA |
| glucose | diaminopimelate decarboxylase | 0.0 | lysA |
| glucose | diaminopimelate epimerase | 0.0 | dapF |
| glucose | CDP diacylglycerol synthetase n C160 | 0.0 | cdsA |
| glucose | CDP diacylglycerol synthetase n C161 | 0.0 | cdsA |
| glucose | 3 4 Dihydroxy 2 butanone 4 phosphate synthase | 0.0 | ribB |
| glucose | 3 deoxy D arabino heptulosonate 7 phosphate synthetase | 0.0 | aroG aroH aroF |
| glucose | dihydroxy acid dehydratase 2 3 dihydroxy 3 methylbutanoate | 0.0 | ilvD |
| glucose | Dihydroxy acid dehydratase 2 3 dihydroxy 3 methylpentanoate | 0.0 | ilvD |
| glucose | dihydrodipicolinate reductase NADPH | 0.0 | dapB |
| glucose | dihydrodipicolinate synthase | 0.0 | dapA |

**Supplementary Table 3.2, Continued**

| | | | |
|---|---|---|---|
| glucose | dihydrofolate reductase | 0.0 | folA ydgB |
| glucose | dihydrofolate synthase | 0.0 | folC |
| glucose | dihydroneopterin aldolase | 0.0 | ygiG |
| glucose | diaminohydroxyphosphoribosylaminopryrimidine deaminase 25drapp | 0.0 | ribD |
| glucose | dihydropteroate synthase | 0.0 | folP |
| glucose | 4 5 dihydroxy 2 3 pentanedione cyclization spontaneous | 0.0 | thrL |
| glucose | 3 dehydroquinate synthase | 0.0 | aroB |
| glucose | 3 dehydroquinate dehydratase irreversible | 0.0 | aroD |
| glucose | dimethylallyltranstransferase | 0.0 | ispA |
| glucose | Dihydroneopterin triphosphate pyrophosphatase | 0.0 | mutT ntpA |
| glucose | dephospho CoA kinase | 0.0 | yacE |
| glucose | 2 dehydropantoate 2 reductase | 0.0 | apbA ilvC |
| glucose | dTMP kinase | 0.0 | tmk |
| glucose | 1 deoxy D xylulose reductoisomerase | 0.0 | ispC |
| glucose | 1 deoxy D xylulose 5 phosphate synthase | 0.0 | dxs |
| glucose | Erythrose 4 phosphate dehydrogenase | 0.0 | epd |
| glucose | enolase | 80.4 | eno |
| glucose | fructose bisphosphate aldolase | 98.4 | ydjI fbaB fbaA |
| glucose | Ferrochelatase | 0.0 | hemH |
| glucose | FMN adenylyltransferase | 0.0 | ribF |
| glucose | fumarase | 92.3 | fumC fumA fumB |
| glucose | glucosamine 1 phosphate N acetyltransferase | 0.0 | glmU |
| glucose | glutamate 1 semialdehyde aminotransferase | 0.0 | hemL |
| glucose | glycerol 3 phosphate acyltransferase C160 | 0.0 | plsB |
| glucose | glycerol 3 phosphate acyltransferase C161 | 0.0 | plsB |
| glucose | L glutamate 5 semialdehyde dehydratase spontaneous | 0.0 | thrL |
| glucose | glucose 6 phosphate dehydrogenase | 98.0 | zwf |
| glucose | glyceraldehyde 3 phosphate dehydrogenase | 68.8 | gapA |
| glucose | Glycolaldehyde dehydrogenase | 0.0 | aldA |
| glucose | glutamine fructose 6 phosphate transaminase | 0.0 | glmS |
| glucose | glycine hydroxymethyltransferase reversible | 96.9 | glyA |
| glucose | guanylate kinase GMPATP | 0.0 | gmk |
| glucose | D glucose transport via PEPPyr PTS periplasm | 97.4 | ptsG malX manX manY manZ ptsH ptsI crr |
| glucose | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible | 0.0 | lamB |

**Supplementary Table 3.2, Continued**

| | | | |
|---|---|---|---|
| glucose | glutamine synthetase | 0.0 | ycjK glnA |
| glucose | glutamine phosphoribosyldiphosphate amidotransferase | 0.0 | purF |
| glucose | glutamyl tRNA reductase | 0.0 | hemA |
| glucose | Glutamyl tRNA synthetase | 0.0 | gltX |
| glucose | GMP synthase | 0.0 | guaA |
| glucose | phosphogluconate dehydrogenase | 98.0 | gnd |
| glucose | geranyltranstransferase | 0.0 | ispA |
| glucose | GTP cyclohydrolase I | 0.0 | folE |
| glucose | GTP cyclohydrolase II 25drapp | 0.0 | ribA |
| glucose | Hydroxybenzoate octaprenyltransferase | 0.0 | ubiA |
| glucose | HCO3 equilibration reaction | 0.0 | yadF cynT |
| glucose | histidinol dehydrogenase | 0.0 | hisD |
| glucose | histidinol phosphatase | 0.0 | hisB |
| glucose | hydroxymethylbilane synthase | 0.0 | hemC |
| glucose | 6 hydroxymethyl dihydropterin pyrophosphokinase | 0.0 | folK |
| glucose | homoserine kinase | 0.0 | thrB |
| glucose | homoserine O succinyltransferase | 0.0 | metA |
| glucose | histidinol phosphate transaminase | 0.0 | hisC |
| glucose | isocitrate dehydrogenase NADP | 0.0 | icdA |
| glucose | Imidazole glycerol 3 phosphate synthase | 0.0 | hisH hisF |
| glucose | imidazoleglycerol phosphate dehydratase | 0.0 | hisB |
| glucose | indole 3 glycerol phosphate synthase | 0.0 | trpC |
| glucose | 3 isopropylmalate dehydrogenase | 0.0 | leuB |
| glucose | 2 isopropylmalate synthase | 0.0 | leuA |
| glucose | KDO 2 lipid IV A transport via ABC system periplasm | 0.0 | msbA |
| glucose | ketol acid reductoisomerase 2 Acetolactate | 0.0 | ilvC |
| glucose | 3 deoxy manno octulosonate cytidylyltransferase | 0.0 | kdsB |
| glucose | 3 deoxy manno octulosonate 8 phosphatase | 0.0 | yrbI |
| glucose | 3 deoxy D manno octulosonic acid 8 phosphate synthase | 0.0 | kdsA |
| glucose | potassium transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| glucose | leucine transaminase irreversible | 0.0 | ilvE tyrB |
| glucose | Lipid A disaccaride synthase | 0.0 | lpxB |
| glucose | Malonyl CoA ACP transacylase | 0.0 | fabD acpP |
| glucose | murein crosslinking transpeptidase 1A A2pm D ala periplasm | 0.0 | ftsI mrcB mrdA mrcA |
| glucose | malate dehydrogenase | 96.8 | mdh |

**Supplementary Table 3.2, Continued**

| | | | |
|---|---|---|---|
| glucose | 2C methyl D erythritol 2 4 cyclodiphosphate dehydratase | 0.0 | ispG |
| glucose | 2 C methyl D erythritol 2 4 cyclodiphosphate synthase | 0.0 | ispF |
| glucose | 2 C methyl D erythritol 4 phosphate cytidylyltransferase | 0.0 | ispD |
| glucose | methionine adenosyltransferase | 0.0 | metK |
| glucose | methionine synthase | 0.0 | metE metH |
| glucose | magnesium Mg2 transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| glucose | Manganese Mn2 transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| glucose | 3 deoxy D manno octulosonic acid transferase | 0.0 | kdtA |
| glucose | 3 deoxy D manno octulosonic acid transferase | 0.0 | kdtA |
| glucose | molybdate transport via ABC system periplasm | 0.0 | modA modB modC cysA cysW cysU cysP sbp |
| glucose | molybdate transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| glucose | 3 methyl 2 oxobutanoate hydroxymethyltransferase | 0.0 | panB |
| glucose | murein polymerizing transglycosylase | 0.0 | mrcB pbpC mrcA |
| glucose | methenyltetrahydrofolate cyclohydrolase | 97.4 | folD |
| glucose | methylenetetrahydrofolate dehydrogenase NADP | 97.4 | folD |
| glucose | 5 10 methylenetetrahydrofolate reductase NADH | 0.0 | metF |
| glucose | NAD kinase | 0.0 | yfjB |
| glucose | NAD synthase nh3 | 0.0 | nadE |
| glucose | nucleoside diphosphate kinase ATPUDP | 0.0 | adk ndk |
| glucose | nucleoside diphosphate kinase ATPdTDP | 0.0 | adk ndk |
| glucose | ammonia transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| glucose | ammonia reversible transport periplasm | 0.0 | amtB thrL |
| glucose | nicotinate nucleotide adenylyltransferase | 0.0 | nadD |
| glucose | nicotinate nucleotide diphosphorylase carboxylating | 0.0 | nadC |
| glucose | oxygen transport via diffusion extracellular to periplasm | 21.2 | phoE ompF ompN ompC |
| glucose | o2 transport via diffusion periplasm | 21.2 | thrL |
| glucose | ornithine carbamoyltransferase | 0.0 | argF argI |
| glucose | Octaprenyl pyrophosphate synthase | 0.0 | ispB |
| glucose | O Phospho 4 hydroxy L threonine2 oxoglutarate aminotransferase | 0.0 | serC |

**Supplementary Table 3.2, Continued**

| | | | |
|---|---|---|---|
| glucose | 2 Oxo 4 methyl 3 carboxypentanoate decarboxylation | 0.0 | leuB |
| glucose | orotidine 5 phosphate decarboxylase | 0.0 | pyrF |
| glucose | Octaprenyl hydroxybenzoate decarboxylase | 0.0 | ubiX ubiD |
| glucose | pyrroline 5 carboxylate reductase | 0.0 | proC |
| glucose | pantothenate synthase | 0.0 | panC |
| glucose | phospho N acetylmuramoyl pentapeptide transferase meso 2 6 diaminopimelate | 0.0 | mraY |
| glucose | pyruvate dehydrogenase | 96.2 | aceE aceF lpdA |
| glucose | Pyridoxine 5 phosphate synthase | 0.0 | pdxA pdxJ |
| glucose | phosphatidylethanolamine transport via ABC system n C160 periplasm | 0.0 | msbA |
| glucose | phosphatidylethanolamine transport via ABC system n C161 periplasm | 0.0 | msbA |
| glucose | Erythronate 4 phosphate 4per dehydrogenase | 0.0 | pdxB |
| glucose | phosphofructokinase | 98.4 | pfkB pfkA |
| glucose | phosphoglycerate dehydrogenase | 95.3 | serA |
| glucose | glucose 6 phosphate isomerase | 98.6 | pgi |
| glucose | 6 phosphogluconolactonase | 98.0 | ybhE |
| glucose | phosphate transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| glucose | phosphomethylpyrimidine kinase | 0.0 | thiD |
| glucose | pantothenate kinase | 0.0 | coaA |
| glucose | porphobilinogen synthase | 0.0 | hemB |
| glucose | phosphopantothenoylcysteine decarboxylase | 0.0 | dfp |
| glucose | phosphopantothenate cysteine ligase | 0.0 | dfp |
| glucose | prephenate dehydrogenase | 0.0 | tyrA |
| glucose | prephenate dehydratase | 0.0 | pheA |
| glucose | phosphoribosylglycinamide synthase | 0.0 | purD |
| glucose | phosphoribosylaminoimidazole synthase | 0.0 | purM |
| glucose | phosphoribosylanthranilate isomerase irreversible | 0.0 | trpC |
| glucose | phosphoribosyl AMP cyclohydrolase | 0.0 | hisI |
| glucose | phosphoribosylaminoimidazolesuccinocarboxamide synthase | 0.0 | purC |
| glucose | phosphoribosyl ATP pyrophosphatase | 0.0 | hisI |
| glucose | phosphoribosylformylglycinamidine synthase | 0.0 | purL |
| glucose | 1 5 phosphoribosyl 5 5 phosphoribosylamino methylideneamino imidazole 4 carboxamide isomerase | 0.0 | hisA |
| glucose | 3 phosphoshikimate 1 carboxyvinyltransferase | 0.0 | aroA |
| glucose | Phosphatidylserine decarboxylase n C160 | 0.0 | psd |

**Supplementary Table 3.2, Continued**

| | | | |
|---|---|---|---|
| glucose | Phosphatidylserine decarboxylase n C161 | 0.0 | psd |
| glucose | phosphoserine transaminase | 95.3 | serC |
| glucose | phosphoserine phosphatase L serine | 95.3 | serB |
| glucose | Phosphatidylserine syntase n C160 | 0.0 | pssA |
| glucose | Phosphatidylserine syntase n C161 | 0.0 | pssA |
| glucose | pantetheine phosphate adenylyltransferase | 0.0 | coaD |
| glucose | quinolinate synthase | 0.0 | nadA |
| glucose | riboflavin kinase | 0.0 | ribF |
| glucose | riboflavin synthase | 0.0 | ribC |
| glucose | riboflavin synthase | 0.0 | ribE |
| glucose | S ribosylhomocysteine cleavage enzyme | 0.0 | ygaG |
| glucose | ribulose 5 phosphate 3 epimerase | 98.4 | rpe sgcE |
| glucose | Sulfate adenyltransferase | 0.0 | cysN cysD |
| glucose | succinyl diaminopimelate desuccinylase | 0.0 | dapE |
| glucose | serine O acetyltransferase | 0.0 | cysE |
| glucose | sirohydrochlorin dehydrogenase NAD | 0.0 | cysG |
| glucose | sirohydrochlorin ferrochetalase | 0.0 | cysG |
| glucose | shikimate dehydrogenase | 0.0 | ydiB aroE |
| glucose | shikimate kinase | 0.0 | aroL aroK |
| glucose | O succinylhomoserine lyase L cysteine | 0.0 | metB |
| glucose | sulfate transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| glucose | succinate dehydrogenase irreversible | 95.5 | sdhC sdhD sdhA sdhB |
| glucose | sulfite reductase NADPH2 | 0.0 | cysI cysJ |
| glucose | sulfate transport via ABC system periplasm | 0.0 | modA modB modC cysA cysW cysU cysP sbp |
| glucose | trans 2 decenoyl ACP isomerase | 0.0 | fabA |
| glucose | transaldolase | 98.2 | talB talA |
| glucose | Tetraacyldisaccharide 4 kinase | 0.0 | lpxK |
| glucose | tetrahydrodipicolinate succinylase | 0.0 | dapD |
| glucose | L threonine deaminase | 0.0 | sdaA sdaB tdcB ilvA |
| glucose | threonine synthase | 0.0 | thrC |
| glucose | thiazole phosphate synthesis | 0.0 | thiI iscS thiH thiG thiF thiS |
| glucose | transketolase | 98.6 | tktB tktA |
| glucose | transketolase | 98.6 | tktB tktA |
| glucose | thymidylate synthase | 0.0 | thyA |
| glucose | thiamine phosphate kinase | 0.0 | thiL |

**Supplementary Table 3.2, Continued**

| | | | |
|---|---|---|---|
| glucose | thiamine phosphate diphosphorylase | 0.0 | thiE |
| glucose | triose phosphate isomerase | 91.5 | tpiA |
| glucose | UDP 3 O 3 hydroxymyristoyl glucosamine acyltransferase | 0.0 | lpxD |
| glucose | UDP N acetylmuramoyl L alanyl D glutamyl meso 2 6 diaminopimelate synthetase | 0.0 | murE |
| glucose | UDP N acetylglucosamine acyltransferase | 0.0 | lpxA acpP |
| glucose | UDP N acetylglucosamine 1 carboxyvinyltransferase | 0.0 | murA |
| glucose | UDP N acetylglucosamine diphosphorylase | 0.0 | glmU |
| glucose | UDP N acetylglucosamine N acetylmuramyl pentapeptide pyrophosphoryl undecaprenol N acetylglucosamine transferase | 0.0 | murG |
| glucose | UDP N acetylmuramoyl L alanyl D glutamate synthetase | 0.0 | murD |
| glucose | UDP N acetylmuramoyl L alanine synthetase | 0.0 | murC |
| glucose | UDP N acetylenolpyruvoylglucosamine reductase | 0.0 | murB |
| glucose | undecaprenyl diphosphatase | 0.0 | ybjG pgpB bacA |
| glucose | Undecaprenyl diphosphate synthase | 0.0 | uppS |
| glucose | UDP N acetylmuramoyl L alanyl D glutamyl meso 2 6 diaminopimeloyl D alanyl D alanine synthetase | 0.0 | murF |
| glucose | UDP 3 O acetylglucosamine deacetylase | 0.0 | lpxC |
| glucose | UMP kinase | 0.0 | pyrH cmk |
| glucose | uroporphyrinogen methyltransferase | 0.0 | cysG hemX |
| glucose | uroporphyrinogen III synthase | 0.0 | hemD |
| glucose | uroporphyrinogen decarboxylase uroporphyrinogen III | 0.0 | hemE |
| glucose | UDP sugar hydrolase | 0.0 | ybbF |
| glucose | zinc Zn2 transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| uridine | Uracil exchange | 60.7 | |
| uridine | pyruvate kinase | 98.2 | pykF pykA |
| uridine | pyrimidine nucleoside phosphorylase uracil | 95.1 | udp |
| uridine | uridine transport via diffusion extracellular to periplasm | 0.0 | tsx |
| D-sorbitol | sorbitol 6 phosphate dehydrogenase | 0.0 | srlD |
| D-sorbitol | D sorbitol transport via PEPPyr PTS periplasm | 0.0 | ptsH ptsI srlA srlE srlB |
| D-sorbitol | D sorbitol transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| fructose | Fructose transport via PEPPyr PTS f6p generating periplasm | 97.6 | manX manY manZ ptsH ptsI |

**Supplementary Table 3.2, Continued**

| | | | |
|---|---|---|---|
| fructose | D fructose transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| L-serine | Isocitrate lyase | 98.7 | aceA |
| L-serine | phosphoenolpyruvate synthase | 98.6 | ppsA |
| L-serine | L serine transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| trehalose | trehalose transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| D-serine | D serine dehydrogenase | 96.4 | ydfG |
| D-serine | D serine transport in via proton symport periplasm | 0.0 | cycA |
| D-serine | D serine transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| D-serine | Isocitrate lyase | 98.7 | aceA |
| D-serine | phosphoenolpyruvate synthase | 98.6 | ppsA |
| galactose | galactokinase | 0.0 | galK wcaK |
| galactose | D galactose transport in via proton symport periplasm | 92.0 | galP |
| galactose | D galactose transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| galactose | phosphoglucomutase | 89.4 | pgm yqaB |
| galactose | pyruvate kinase | 96.7 | pykF pykA |
| galactose | UDPglucose hexose 1 phosphate uridylyltransferase | 0.0 | galT |
| lactose | galactokinase | 51.2 | galK wcaK |
| lactose | Lactose transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| lactose | phosphoglucomutase | 94.8 | pgm yqaB |
| lactose | UDPglucose hexose 1 phosphate uridylyltransferase | 51.2 | galT |
| gluconate | D gluconate transport via proton symport reversible periplasm | 0.0 | gntT idnT gntP gntU |
| gluconate | D gluconate transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| gluconate | gluconokinase | 0.0 | gntK idnK |
| gluconate | pyruvate kinase | 95.3 | pykF pykA |
| mannose | mannose 6 phosphate isomerase | 0.0 | manA |
| mannose | D mannose transport via PEPPyr PTS periplasm | 0.0 | manX manY manZ ptsH ptsI |
| mannose | D mannose transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| maltose | maltose transport via ABC system periplasm | 0.0 | malG malF malE malK |

**Supplementary Table 3.2, Continued**

| | | | |
|---|---|---|---|
| maltose | maltoseMaltotriose transport via diffusion extracellular to periplasm irreversible | 0.0 | lamB |
| maltose | phosphoglucomutase | 94.8 | pgm yqaB |
| maltose | pyruvate kinase | 96.9 | pykF pykA |
| lactate | Isocitrate lyase | 98.8 | aceA |
| lactate | L lactate reversible transport via proton symport periplasm | 0.0 | yghK lldP |
| lactate | L lactate transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| lactate | phosphoenolpyruvate synthase | 96.3 | ppsA |
| lactate | NAD P transhydrogenase periplasm | 98.3 | pntB pntA |
| succinate | malic enzyme NADP | 95.2 | maeB |
| succinate | phosphoenolpyruvate carboxykinase | 95.5 | pckA |
| succinate | succinate transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| citrate | Citrate transport via succinate antiport periplasm | 0.0 | citT |
| citrate | citrate transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| citrate | phosphoenolpyruvate carboxykinase | 98.0 | pckA |
| L-alanine | L alanine transport via diffusion extracellular to periplasm | 0.0 | phoE ompF ompN ompC |
| L-alanine | Isocitrate lyase | 97.4 | aceA |
| L-alanine | phosphoenolpyruvate synthase | 96.4 | ppsA |

**Supplementary Table 3.3:** Reactions that are not predicted to be shared between the metabolism of glucose and alternative substrates, determining the Hamming distance between substrates for FBA comparisons.

| Substrate | Reaction Name (reaction not in common with glucose utilization) |
|---|---|
| succinate | glucose 6 phosphate dehydrogenase |
| succinate | D glucose transport via PEPPyr PTS periplasm |
| succinate | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| succinate | phosphogluconate dehydrogenase |
| succinate | malic enzyme NADP |
| succinate | phosphofructokinase |
| succinate | glucose 6 phosphate isomerase |
| succinate | 6 phosphogluconolactonase |
| succinate | phosphoenolpyruvate carboxykinase |
| succinate | succinate transport via diffusion extracellular to periplasm |
| D-serine | D serine dehydrogenase |
| D-serine | D serine transport in via proton symport periplasm |
| D-serine | D serine transport via diffusion extracellular to periplasm |
| D-serine | glucose 6 phosphate dehydrogenase |
| D-serine | D glucose transport via PEPPyr PTS periplasm |
| D-serine | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| D-serine | phosphogluconate dehydrogenase |
| D-serine | Isocitrate lyase |
| D-serine | phosphofructokinase |
| D-serine | phosphoglycerate dehydrogenase |
| D-serine | glucose 6 phosphate isomerase |
| D-serine | 6 phosphogluconolactonase |
| D-serine | phosphoenolpyruvate synthase |
| D-serine | phosphoserine transaminase |
| D-serine | phosphoserine phosphatase L serine |
| L-serine | glucose 6 phosphate dehydrogenase |
| L-serine | D glucose transport via PEPPyr PTS periplasm |
| L-serine | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| L-serine | phosphogluconate dehydrogenase |
| L-serine | Isocitrate lyase |
| L-serine | phosphofructokinase |
| L-serine | phosphoglycerate dehydrogenase |
| L-serine | glucose 6 phosphate isomerase |
| L-serine | 6 phosphogluconolactonase |
| L-serine | phosphoenolpyruvate synthase |

**Supplementary Table 3.3, continued**

| | |
|---|---|
| L-serine | phosphoserine transaminase |
| L-serine | phosphoserine phosphatase L serine |
| L-serine | L serine transport via diffusion extracellular to periplasm |
| acetate | glucose 6 phosphate dehydrogenase |
| acetate | D glucose transport via PEPPyr PTS periplasm |
| acetate | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| acetate | phosphogluconate dehydrogenase |
| acetate | Isocitrate lyase |
| acetate | pyruvate dehydrogenase |
| acetate | phosphofructokinase |
| acetate | glucose 6 phosphate isomerase |
| acetate | 6 phosphogluconolactonase |
| acetate | phosphoenolpyruvate carboxykinase |
| uridine | aspartate carbamoyltransferase |
| uridine | Acetate exchange |
| uridine | Uracil exchange |
| uridine | fructose bisphosphate aldolase |
| uridine | D glucose transport via PEPPyr PTS periplasm |
| uridine | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| uridine | orotidine 5 phosphate decarboxylase |
| uridine | phosphofructokinase |
| uridine | pyruvate kinase |
| uridine | pyrimidine nucleoside phosphorylase uracil |
| uridine | uridine transport via diffusion extracellular to periplasm |
| L-malate | Acetate exchange |
| L-malate | glucose 6 phosphate dehydrogenase |
| L-malate | D glucose transport via PEPPyr PTS periplasm |
| L-malate | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| L-malate | phosphogluconate dehydrogenase |
| L-malate | Malate transport via proton symport 2 H  periplasm |
| L-malate | Malate transport via diffusion extracellular to periplasm |
| L-malate | malic enzyme NADP |
| L-malate | phosphofructokinase |
| L-malate | glucose 6 phosphate isomerase |
| L-malate | 6 phosphogluconolactonase |
| L-malate | phosphoenolpyruvate carboxykinase |
| L-proline | Acetate exchange |
| L-proline | iron Fe2 transport out via proton antiport periplasm |
| L-proline | Fe III reduction |

**Supplementary Table 3.3, continued**

| | |
|---|---|
| L-proline | L glutamate 5 semialdehyde dehydratase spontaneous |
| L-proline | glucose 6 phosphate dehydrogenase |
| L-proline | D glucose transport via PEPPyr PTS periplasm |
| L-proline | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| L-proline | phosphogluconate dehydrogenase |
| L-proline | 1 pyrroline 5 carboxylate dehydrogenase |
| L-proline | pyrroline 5 carboxylate reductase |
| L-proline | phosphofructokinase |
| L-proline | glucose 6 phosphate isomerase |
| L-proline | 6 phosphogluconolactonase |
| L-proline | phosphoenolpyruvate carboxykinase |
| L-proline | Proline dehydrogenase |
| L-proline | L proline transport via diffusion extracellular to periplasm |
| fumarate | Acetate exchange |
| fumarate | Fumarate transport via diffusion extracellular to periplasm |
| fumarate | glucose 6 phosphate dehydrogenase |
| fumarate | D glucose transport via PEPPyr PTS periplasm |
| fumarate | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| fumarate | phosphogluconate dehydrogenase |
| fumarate | malic enzyme NADP |
| fumarate | phosphofructokinase |
| fumarate | glucose 6 phosphate isomerase |
| fumarate | 6 phosphogluconolactonase |
| fumarate | phosphoenolpyruvate carboxykinase |
| D-ribose | fructose bisphosphate aldolase |
| D-ribose | D glucose transport via PEPPyr PTS periplasm |
| D-ribose | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| D-ribose | phosphofructokinase |
| D-ribose | pyruvate kinase |
| D-ribose | ribokinase |
| D-ribose | D ribose transport via ABC system periplasm |
| D-ribose | ribose transport via diffusion extracellular to periplasm |
| mucic acid | Acetate exchange |
| mucic acid | glucose 6 phosphate dehydrogenase |
| mucic acid | galactarate dehydratase |
| mucic acid | D galactarte transport via proton symport reversible periplasm |
| mucic acid | D galactarte transport via diffusion extracellular to periplasm |
| mucic acid | 5 dehydro 4 deoxyglucarate aldolase |
| mucic acid | D glucose transport via PEPPyr PTS periplasm |

**Supplementary Table 3.3, continued**

| | |
|---|---|
| mucic acid | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| mucic acid | phosphogluconate dehydrogenase |
| mucic acid | phosphofructokinase |
| mucic acid | glucose 6 phosphate isomerase |
| mucic acid | 6 phosphogluconolactonase |
| mucic acid | pyruvate kinase |
| D-alanine | D Alanine transport via diffusion extracellular to periplasm |
| D-alanine | glucose 6 phosphate dehydrogenase |
| D-alanine | D glucose transport via PEPPyr PTS periplasm |
| D-alanine | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| D-alanine | phosphogluconate dehydrogenase |
| D-alanine | Isocitrate lyase |
| D-alanine | phosphofructokinase |
| D-alanine | glucose 6 phosphate isomerase |
| D-alanine | 6 phosphogluconolactonase |
| D-alanine | phosphoenolpyruvate synthase |
| L-alanine | L alanine transport via diffusion extracellular to periplasm |
| L-alanine | glucose 6 phosphate dehydrogenase |
| L-alanine | D glucose transport via PEPPyr PTS periplasm |
| L-alanine | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| L-alanine | phosphogluconate dehydrogenase |
| L-alanine | Isocitrate lyase |
| L-alanine | phosphofructokinase |
| L-alanine | glucose 6 phosphate isomerase |
| L-alanine | 6 phosphogluconolactonase |
| L-alanine | phosphoenolpyruvate synthase |
| L-aspartate | L aspartate transport in via proton symport periplasm |
| L-aspartate | L aspartate transport via diffusion extracellular to periplasm |
| L-aspartate | Acetate exchange |
| L-aspartate | glucose 6 phosphate dehydrogenase |
| L-aspartate | D glucose transport via PEPPyr PTS periplasm |
| L-aspartate | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| L-aspartate | phosphogluconate dehydrogenase |
| L-aspartate | phosphofructokinase |
| L-aspartate | glucose 6 phosphate isomerase |
| L-aspartate | 6 phosphogluconolactonase |
| L-aspartate | phosphoenolpyruvate carboxykinase |
| D-sorbitol | D glucose transport via PEPPyr PTS periplasm |
| D-sorbitol | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |

**Supplementary Table 3.3, continued**

| | |
|---|---|
| D-sorbitol | sorbitol 6 phosphate dehydrogenase |
| D-sorbitol | D sorbitol transport via PEPPyr PTS periplasm |
| D-sorbitol | D sorbitol transport via diffusion extracellular to periplasm |
| lactate | glucose 6 phosphate dehydrogenase |
| lactate | D glucose transport via PEPPyr PTS periplasm |
| lactate | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| lactate | phosphogluconate dehydrogenase |
| lactate | Isocitrate lyase |
| lactate | L lactate reversible transport via proton symport periplasm |
| lactate | L lactate transport via diffusion extracellular to periplasm |
| lactate | phosphofructokinase |
| lactate | glucose 6 phosphate isomerase |
| lactate | 6 phosphogluconolactonase |
| lactate | phosphoenolpyruvate synthase |
| lactate | NAD P transhydrogenase periplasm |
| citrate | Citrate transport via succinate antiport periplasm |
| citrate | citrate transport via diffusion extracellular to periplasm |
| citrate | Acetate exchange |
| citrate | glucose 6 phosphate dehydrogenase |
| citrate | D glucose transport via PEPPyr PTS periplasm |
| citrate | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| citrate | phosphogluconate dehydrogenase |
| citrate | phosphofructokinase |
| citrate | glucose 6 phosphate isomerase |
| citrate | 6 phosphogluconolactonase |
| citrate | phosphoenolpyruvate carboxykinase |
| galactose | fructose bisphosphate aldolase |
| galactose | galactokinase |
| galactose | D galactose transport in via proton symport periplasm |
| galactose | D galactose transport via diffusion extracellular to periplasm |
| galactose | D glucose transport via PEPPyr PTS periplasm |
| galactose | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| galactose | phosphofructokinase |
| galactose | phosphoglucomutase |
| galactose | pyruvate kinase |
| galactose | UDPglucose hexose 1 phosphate uridylyltransferase |
| gluconate | Acetate exchange |
| gluconate | fructose bisphosphate aldolase |
| gluconate | glucose 6 phosphate dehydrogenase |

**Supplementary Table 3.3, continued**

| | |
|---|---|
| gluconate | D gluconate transport via proton symport reversible periplasm |
| gluconate | D gluconate transport via diffusion extracellular to periplasm |
| gluconate | D glucose transport via PEPPyr PTS periplasm |
| gluconate | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| gluconate | gluconokinase |
| gluconate | phosphofructokinase |
| gluconate | glucose 6 phosphate isomerase |
| gluconate | 6 phosphogluconolactonase |
| gluconate | pyruvate kinase |
| lactose | fructose bisphosphate aldolase |
| lactose | galactokinase |
| lactose | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| lactose | Lactose transport via diffusion extracellular to periplasm |
| lactose | phosphofructokinase |
| lactose | phosphoglucomutase |
| lactose | UDPglucose hexose 1 phosphate uridylyltransferase |
| mannose | D glucose transport via PEPPyr PTS periplasm |
| mannose | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| mannose | mannose 6 phosphate isomerase |
| mannose | D mannose transport via PEPPyr PTS periplasm |
| mannose | D mannose transport via diffusion extracellular to periplasm |
| maltose | fructose bisphosphate aldolase |
| maltose | D glucose transport via PEPPyr PTS periplasm |
| maltose | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| maltose | maltose transport via ABC system periplasm |
| maltose | maltoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| maltose | phosphofructokinase |
| maltose | phosphoglucomutase |
| maltose | pyruvate kinase |
| fructose | Fructose transport via PEPPyr PTS f6p generating  periplasm |
| fructose | D fructose transport via diffusion extracellular to periplasm |
| fructose | D glucose transport via PEPPyr PTS periplasm |
| fructose | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| fructose | phosphofructokinase |
| trehalose | D glucose transport via PEPPyr PTS periplasm |
| trehalose | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| trehalose | trehalose transport via diffusion extracellular to periplasm |
| melibiose | fructose bisphosphate aldolase |
| melibiose | galactokinase |

**Supplementary Table 3.3, continued**

| | |
|---|---|
| melibiose | a galactosidase melibiose |
| melibiose | D glucose transport via PEPPyr PTS periplasm |
| melibiose | D glucoseMaltotriose transport via diffusion extracellular to periplasm irreversible |
| melibiose | melibiose transport in via symport periplasm |
| melibiose | melibiose transport via diffusion extracellular to periplasm |
| melibiose | phosphofructokinase |
| melibiose | phosphoglucomutase |
| melibiose | pyruvate kinase |
| melibiose | UDPglucose hexose 1 phosphate uridylyltransferase |

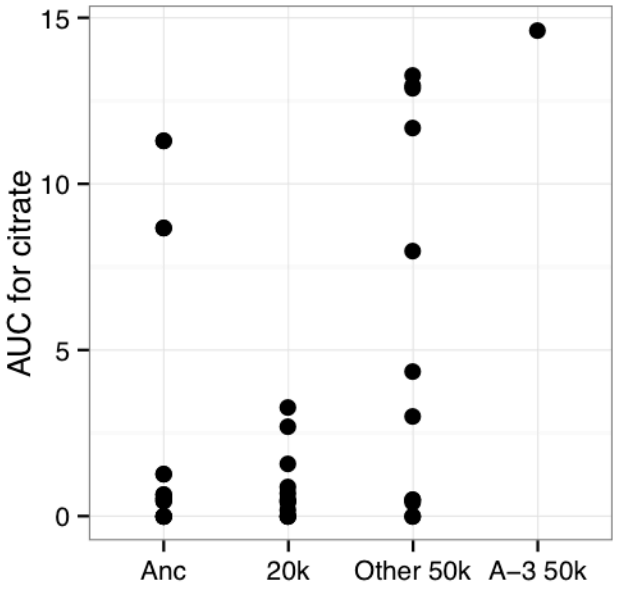
**Supplementary Table 3.4:** Summary of alternative metrics used to assess the difference between substrate utilization using flux balance analysis predictions

| Flux Vector Comparison Metric | Sign of relationship matches prediction? | Significance ? | $R^2$ | Description and Notes |
|---|---|---|---|---|
| Euclidean distance | No | P=0.49 | 0 | Includes amount of flux per reaction, not just reaction use. Coefficient of relationship is positive (i.e., in the opposite direction predicted). |
| 1-Correlation | Yes | P=0.91 | 0 | Includes amount of flux per reaction, not just reaction use, and uses one minus the Pearson's R coefficient as the metric (glucose is a score of 0- perfectly correlated with itself) |
| Hamming distance | No | P=0.26 | 0.004 | Counts the number of reactions used for optimal growth on glucose and not for an alternative substrate as well as those used for an alternative substrate and not for glucose. |
| Mutational target size | Yes | P=0.15 | 0.01 | Counts the number of coding nucleotides for the genes involved in reactions predicted as necessary for optimal growth on a carbon source. |
| Count of unused reactions | Yes | P=0.04 | 0.02 | Unlike Hamming distance, which counts reactions used for glucose and not for alternative substrates, this only counts reactions used for alternative substrates and not glucose |

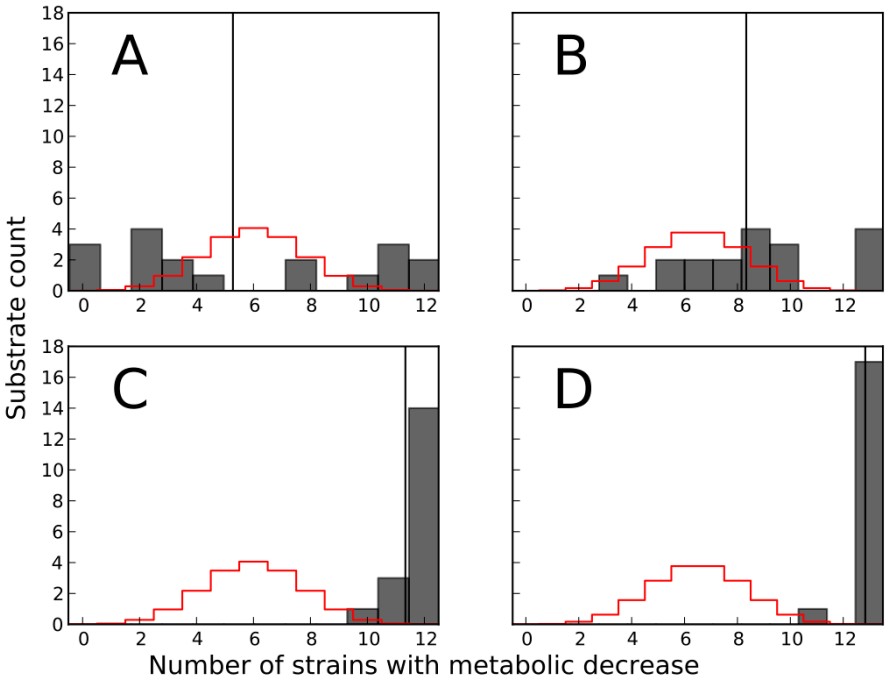
**Supplementary Figure 3.1:** Summary of parallel changes observed using Biolog assays. Blue shading indicates catabolic function that consistently decayed across the parallel populations (statistically significant loss of function for the evolved strains as a group compared to ancestor). Red shading represents statistically significant gains of function. The number in each cell is the number of populations that significantly lost catabolic function on that carbon source relative to the ancestor ($P<0.0005$). There were twelve evolved isolates tested for all timepoints except 50k, for which 13 strains were tested (including A-2S). Substrates in green and italicized allowed no growth of the ancestor within 48 hours in growth rate assays in DM media.

| | Cooper and Lenski 2000 Time (generations) | | | Leiby and Marx 2013 Time (generations) | |
|---|---|---|---|---|---|
| **Carbon Source** | **2k** | **10k** | **20k** | **20k** | **50k** |
| D-Saccharic Acid | 9 | 11 | 11 | 4 | 12 |
| *Succinic Acid* | 9 | 12 | 12 | 8 | 11 |
| *L-Aspartic Acid* | 9 | 12 | 12 | 9 | 9 |
| D-Alanine | 1 | 3 | 6 | 3 | 11 |
| D-Serine | 12 | 11 | 10 | 3 | 12 |
| D-Sorbitol | 12 | 11 | 11 | 11 | 13 |
| L-Lactic acid | 11 | 12 | 10 | 6 | 13 |
| D-Glucose-6- Phosphate | 11 | 12 | 8 | 2 | 11 |
| *D,L-Malic Acid* | 5 | 12 | 12 | 9 | 9 |
| D-Ribose | 12 | 12 | 12 | 12 | 13 |
| D-Fructose | 11 | 10 | 9 | 11 | 12 |
| *L-Asparagine* | 8 | 12 | 12 | 5 | 11 |
| *Uridine* | 12 | 12 | 10 | 9 | 13 |
| L-Glutamine | 12 | 12 | 12 | 10 | 13 |
| D-Glucose-1- Phosphate | 12 | 11 | 10 | 10 | 13 |
| *Fumaric Acid* | 9 | 12 | 12 | 10 | 9 |
| Bromo Succinic Acid | 7 | 11 | 12 | 0 | 0 |
| Mucic Acid | 12 | 8 | 9 | 7 | 13 |
| *L-Malic Acid* | 7 | 12 | 12 | 8 | 9 |
| Mono Methyl Succinate | 2 | 12 | 12 | 0 | 6 |
| p-Hydroxy Phenyl Acetic Acid | 5 | 12 | 11 | 2 | 7 |
| Glucuronamide | 0 | 4 | 8 | 0 | 3 |

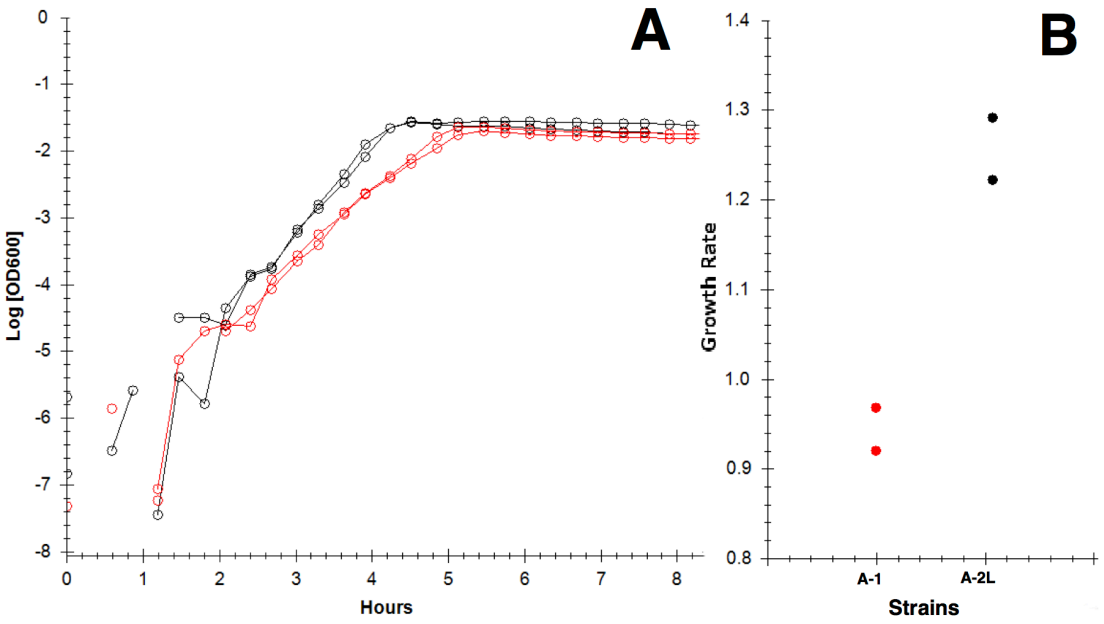
**Supplementary Figure 3.2:** Respiration assay for ancestral and evolved strains on citrate. Points show biological replicate measurements for different groups of strains. The signal for the Cit$^+$ A-3 50k isolate is statistically indistinguishable from other 50k isolates.



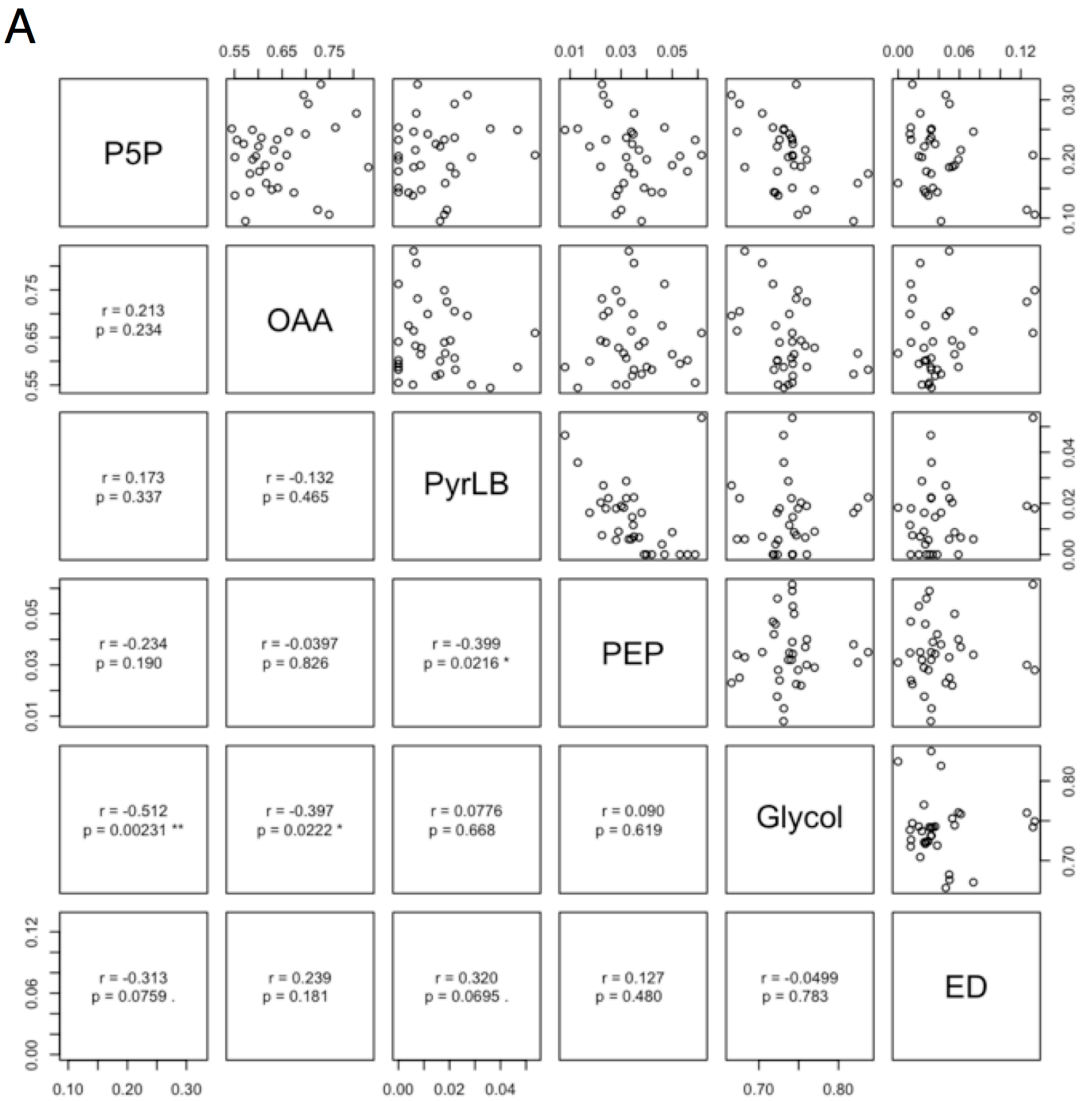
**Supplementary Figure 3.3:** Histogram showing less parallelism of metabolic declines in growth rate than respiration. The x axis indicates the number of strains that exhibited a metabolic decline on a substrate. The grey bars are observed metabolic decreases, the black line is the mean observed number of decreases, and the red outline is the null distribution for a single observation given random increases and decreases. Growth rate changes for 20k (A) and 50k (B) isolates did not show the same degree of parallelism as cellular respiration declines at 20k (C) and 50k (D). The substrates considered were all those for which growth rate and respiration data were both available and for which the ancestor exhibited growth or respiration- necessary for the evolved strains to demonstrate reductions. These substrates were: acetate, D-alanine, D-saccharic acid, D-serine, D-sorbitol, galactose, L-alanine, L-proline, L-serine, lactate, lactose, maltose, mannose, melibiose, mono-methyl succinate, mucic acid, ribose, and trehalose.



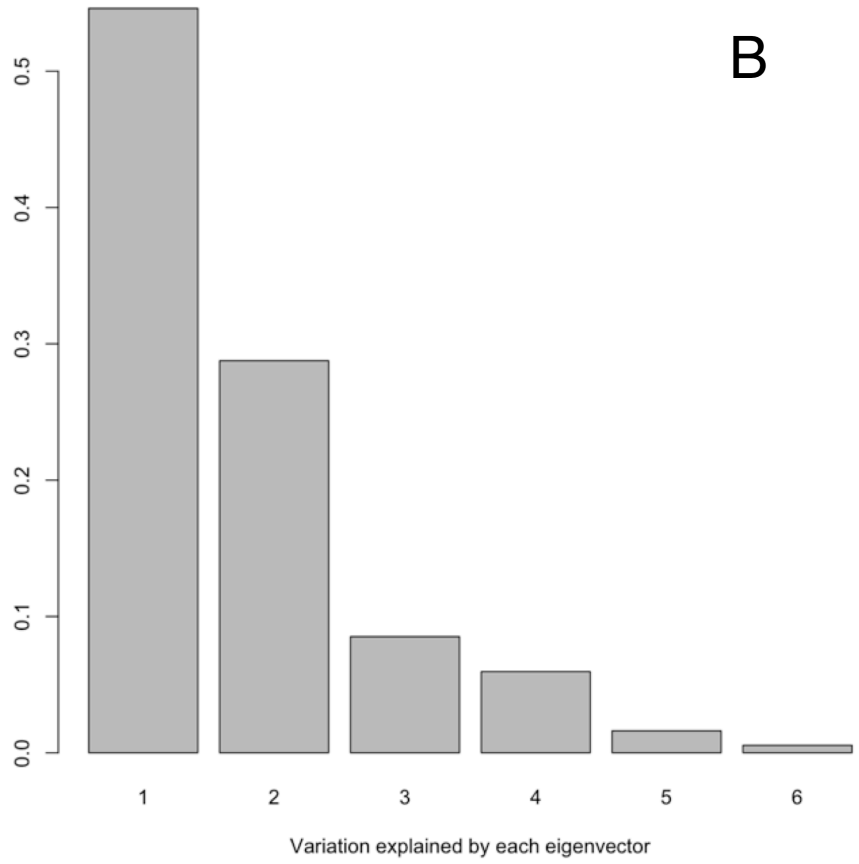
**Supplementary Figure 3.4:** Representative growth curves and fitted growth rates. A) Measured growth curves for 50k isolates of A-2L (red), A-1 (black). B) Fitted growth rates from the measured growth curves.



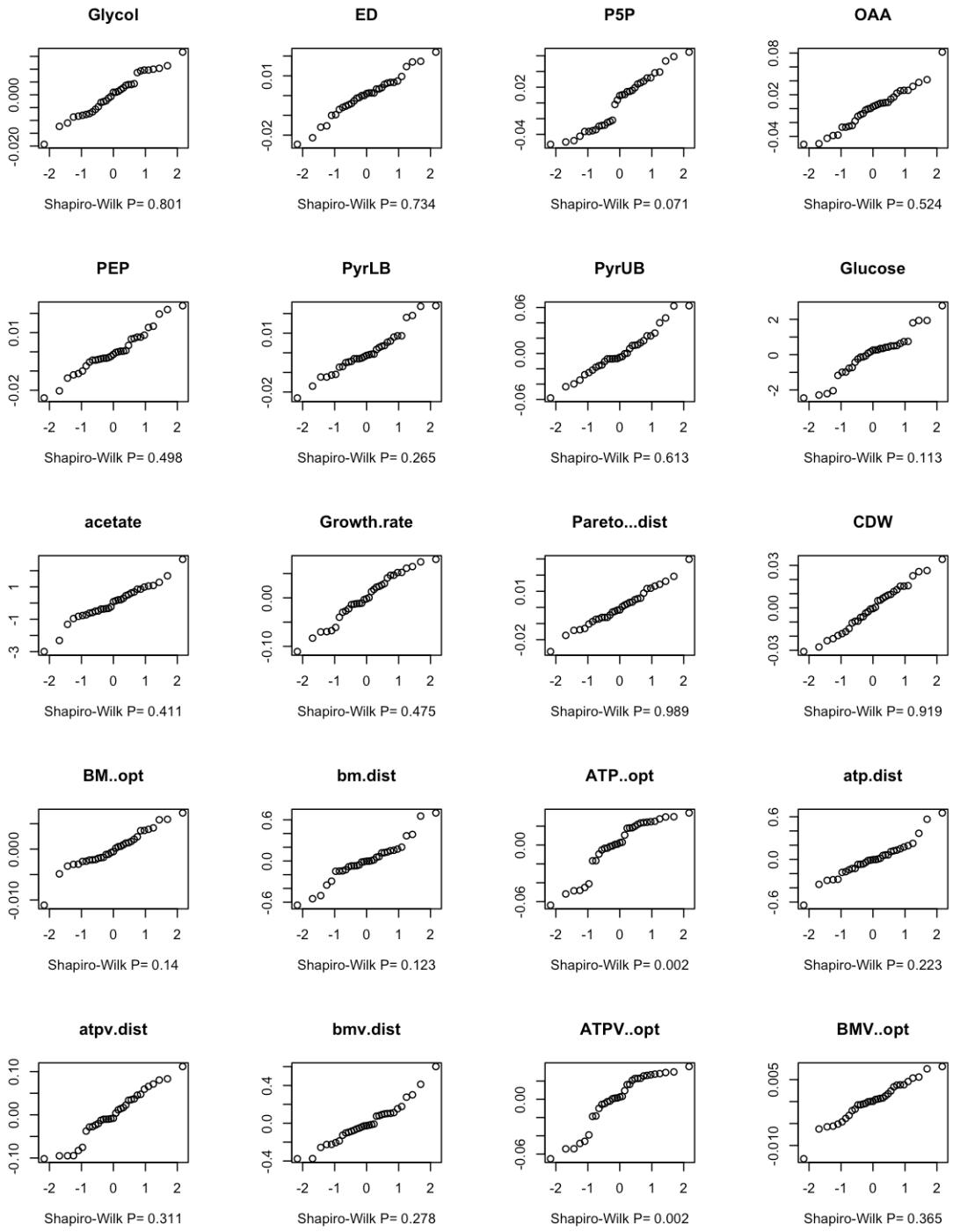
**Supplementary Figure 4.1: Covariance of fluxes inferred for the LTEE.** To determine whether there was a significant change in flux ratios between populations of the LTEE we ran a MANOVA as described in the text; however, to provide further insight into the basis of the significant differences that we observed we present a chart of the correlations between all fluxes. A) The value of the correlation and the significance are presented on the bottom half of the chart. B) The proportion of variation explained by each eigenvector.



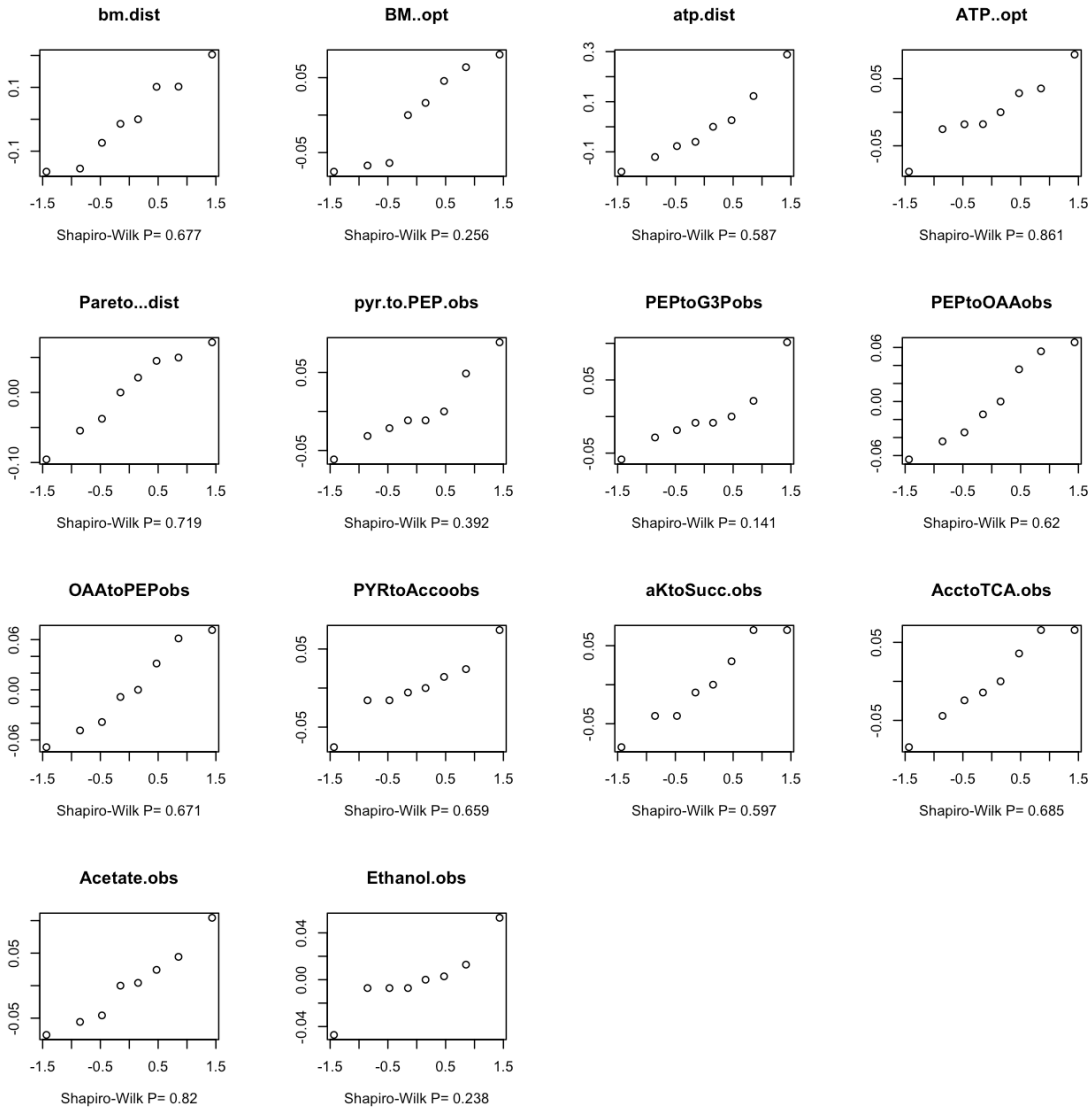
**Supplementary Figure 4.1, continued**



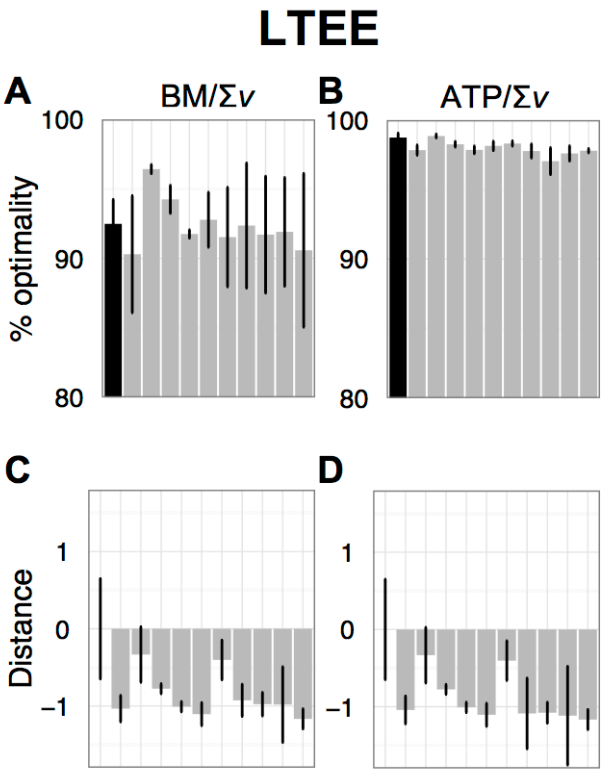
**Supplementary Figure 4.2: Normality tests for data associated with the LTEE.** Q-Q plots and Shapiro-Wilk values are displayed for growth parameters, and flux ratios. Additionally, data is displayed about the normality of % optimality and distance for different criteria.



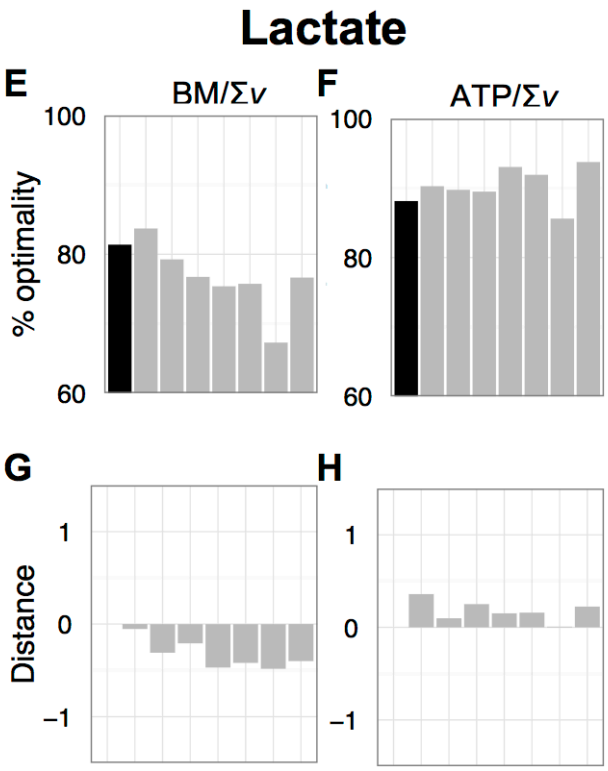
**Supplementary Figure 4.3: Normality tests for data associated with the lactate strains.** Q-Q plots and Shapiro-Wilk values are displayed for growth parameters, and flux ratios. Additionally, data is displayed about the normality of % optimality and distance for different criteria.



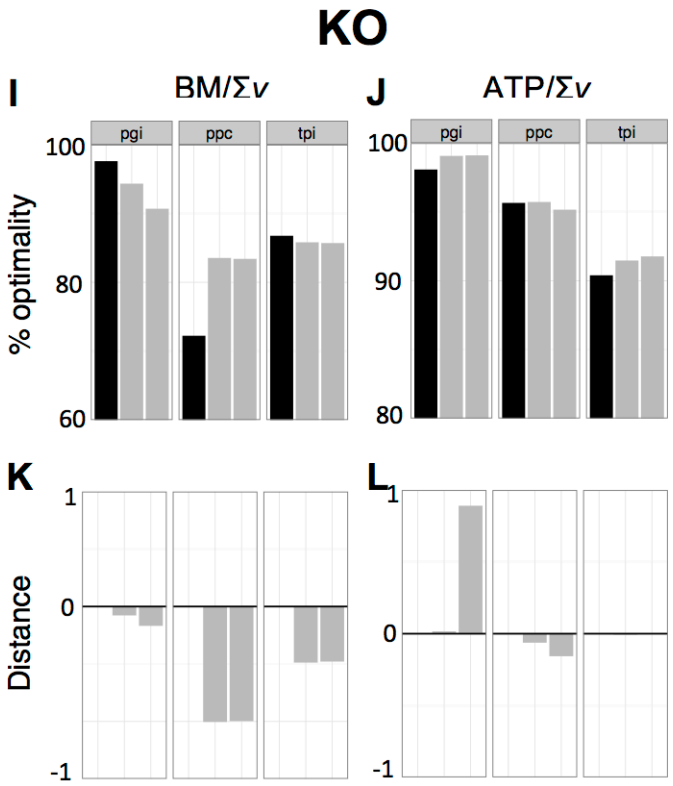
**Supplementary Figure 4.4: Measures of optimality based upon BM/Σ *v* or ATP/Σ *v* for all data sets.** (A,B,E,F,I,J) The % optimality of the ancestor (black) and evolved isolates (grey); (C,D,G,H,K,L) distance to optimal flux distribution for FBA-predictions (plotted as $\log(D_{EO}/D_{AO})$). These were performed based upon BM/Σ*v* (A,C,E,G,I,K) or ATP/Σ*v* (B,D,F,H,J,L). The data sets are LTEE (A-D), lactate (E-H), and KO (I-L). Error bars for LTEE represent standard errors of three biological replicates.



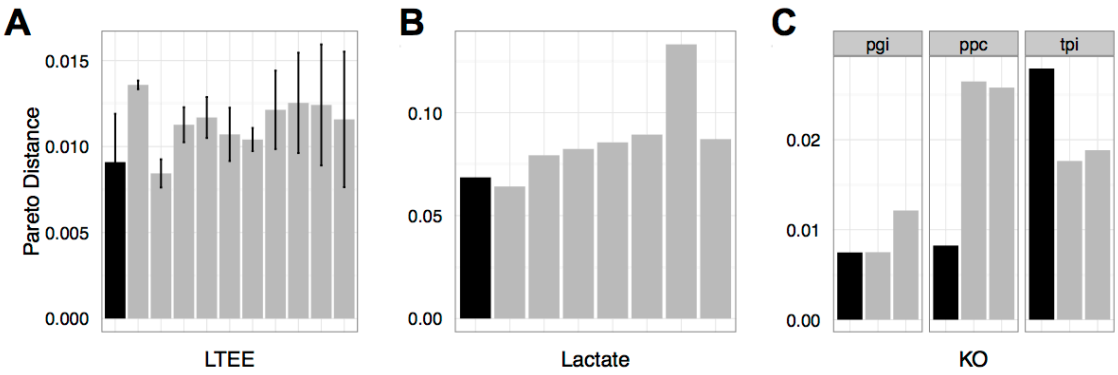
**Supplementary Figure 4.4, continued**



**Supplementary Figure 4.4, continued**

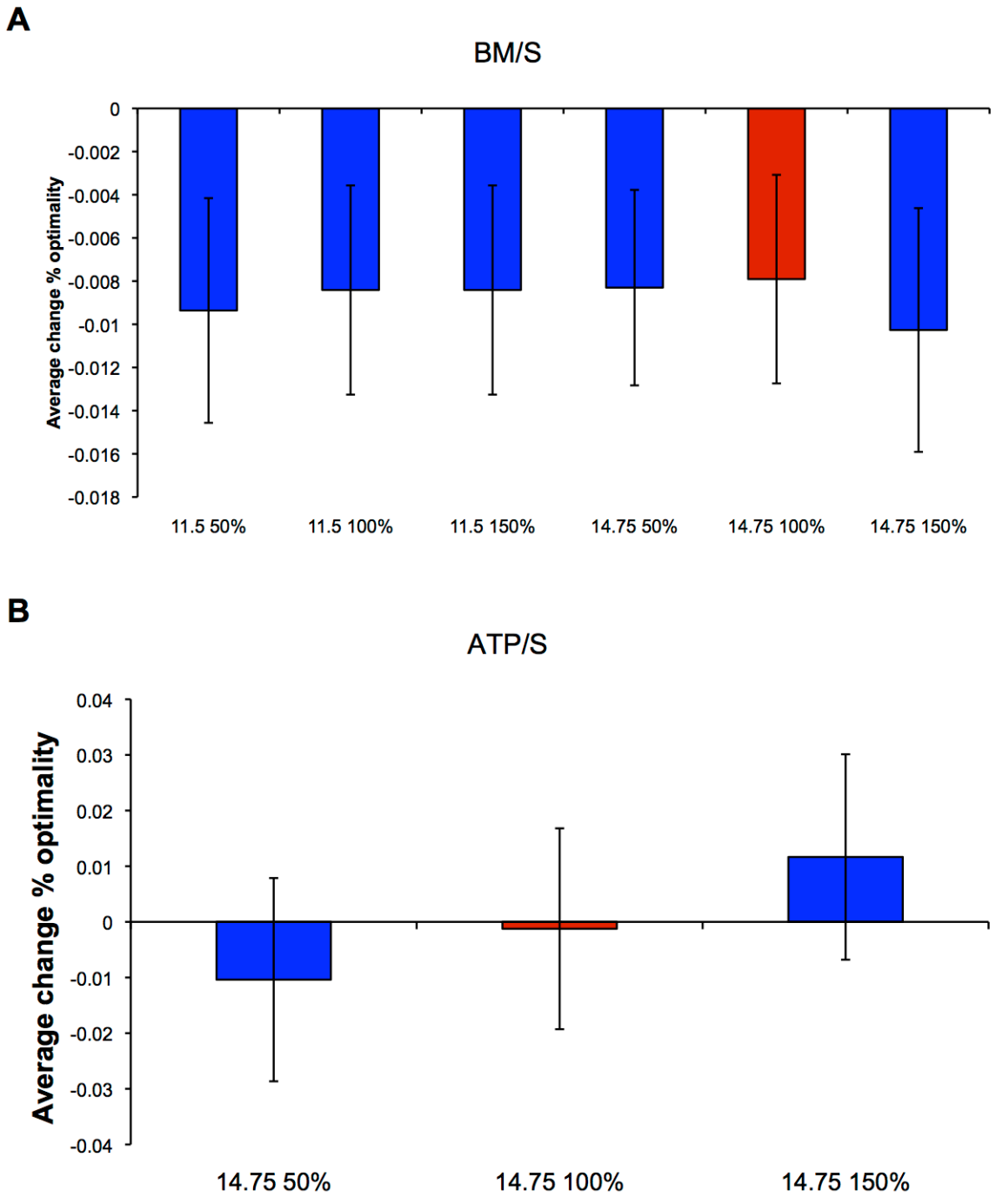


**Supplementary Figure 4.5: Measures of optimality based on maximizing the tradeoff between BM, ATP and Σ *v* for all data sets.** The Pareto distance of the ancestor (black) and evolved isolates (grey) for LTEE (A), lactate (B), and KO (C). Error bars represent standard errors of three biological replicates.



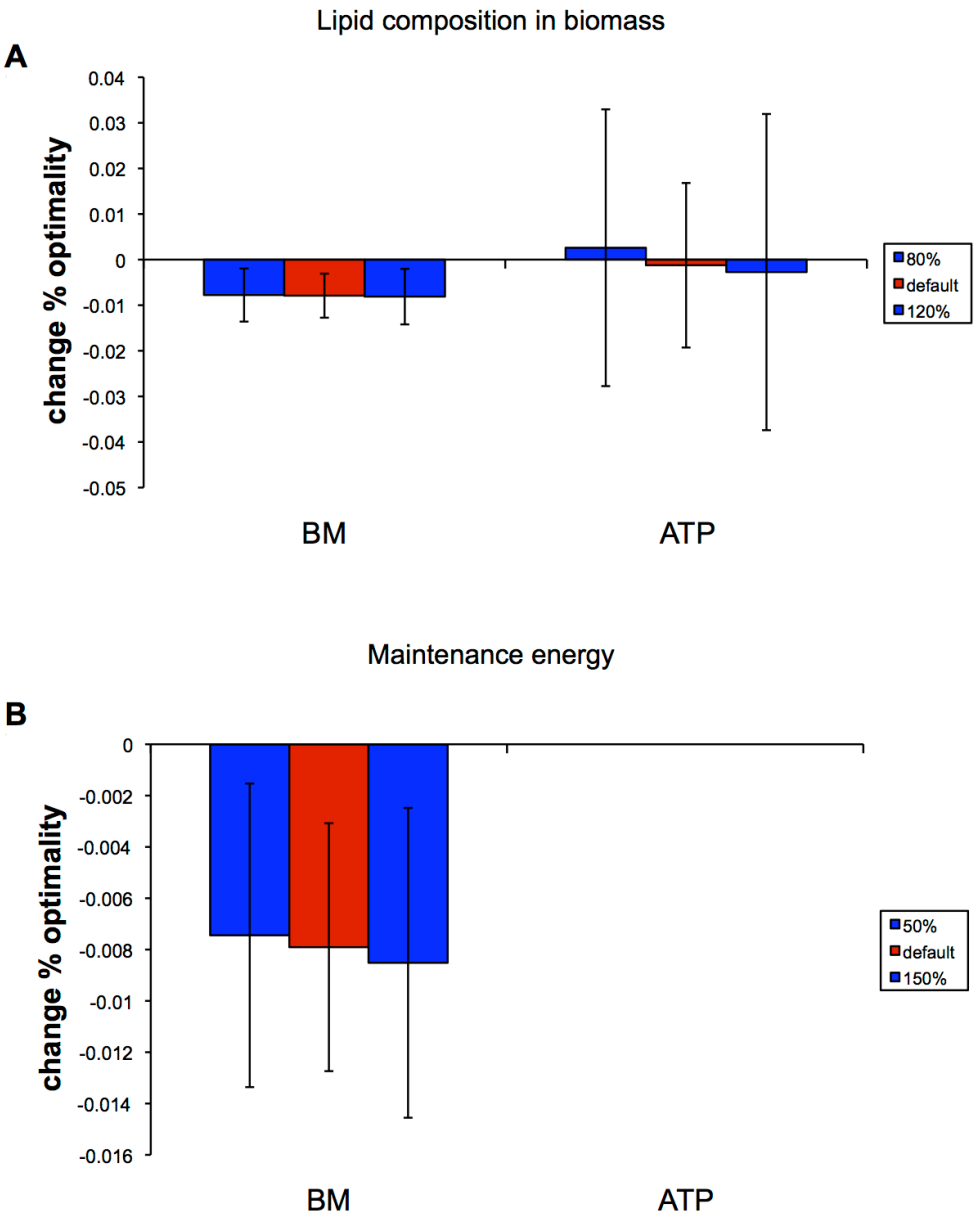
**Supplementary Figure 4.6: Implementation of oxygen constraints.** Following the example of Schuetz *et al* 2007 [11] we varied the ancestral oxygen uptake rate across the range reported in the literature (11.5-14.75 mmol/g hr). Ibarra *et al* 2002 [24] report that the ratio of oxygen to glucose uptake remains largely constant as cells evolve. We tested the impact of varying ancestral oxygen/glucose ratio as well as the slope of evolutionary change from 0.5 to 1.5. There was no significant difference in the change in % optimality for either BM/S (A) or ATP/S (B) across this wide range of parameter values. Results are not presented for an ancestral oxygen uptake rate of 11.5 for ATP/S because this constraint caused infeasible solutions for several evolved populations. Results obtained with the default values used throughout the manuscript, an ancestral uptake of 14.75 mmol/g hr and a slope of 1, are highlighted in red.

**Supplementary Figure 4.6, continued**

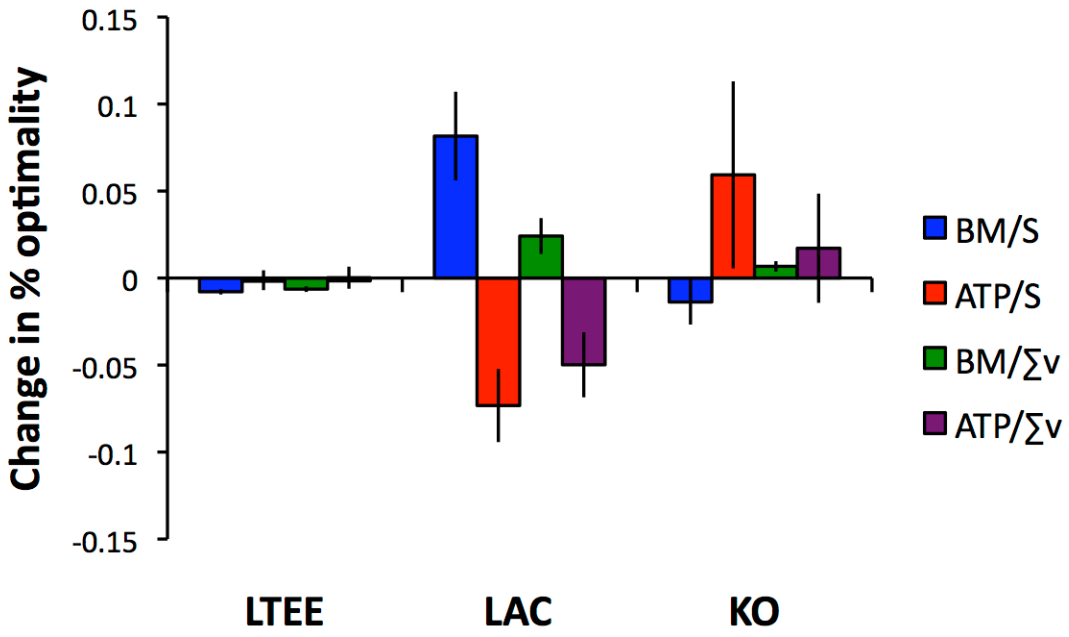


**Supplementary Figure 4.7: The effect that potential evolution of constraints would have on average change in % optimality between ancestor and evolved lines.** A) Lipid content was altered in evolved lines from 80-120% of the default values. B) Maintenance energy in evolved lines was altered from 50-150% of the default value of 8.39 mmol/g hr. Analyses for ATP/S are not shown, as setting a lower bound on maintenance energy has no effect if ATP production is being maximized. Results for simulations run with default (red) and altered (blue) constraints are shown for the LTEE set when optimized for either BM/S or ATP/S. Error bars represent standard errors between replicate lines.

**Supplementary Figure 4.7, continued**



**Supplementary Figure 4.8: Average difference in % optimality between ancestor and evolved lines for each data set for each criterion.** The criteria tested were BM/S (blue), ATP/S (red), BM/Σ*v* (green) and ATP/Σ*v* (purple). Error bars represent standard deviations of replicate lines.



**Supplementary Table 4.1: Growth parameters for ancestral and evolved LTEE isolates.**

| | | Glucose uptake g/g CDW $h^{-1}$ | Acetate excretion g/g CDW $h^{-1}$ | Growth rate $h^{-1}$ | CDW g/g glucose |
|---|---|---|---|---|---|
| Anc | REL606 | 12.83 ± 0.23 | 3.27 ± 0.75 | 0.80 ± 0.10 | 0.36 ± 0.02 |
| A+1 | REL11392 | 13.59 ± 0.87 | 6.69 ± 0.30 | 1.09 ± 0.04 | 0.44 ± 0.02 |
| A+2 | REL11342 | 15.74 ± 2.24 | 6.69 ± 0.96 | 1.18 ± 0.03 | 0.47 ± 0.03 |
| A+3 | REL11345 | 16.43 ± 0.90 | 4.32 ± 2.86 | 1.22 ± 0.02 | 0.46 ± 0.03 |
| A+4 | REL11348 | 15.78 ± 2.49 | 6.71 ± 0.94 | 1.18 ± 0.05 | 0.45 ± 0.03 |
| A+5 | REL11367 | 14.80 ± 2.04 | 5.64 ± 1.16 | 1.16 ± 0.07 | 0.38 ± 0.02 |
| A-1 | REL11330 | 14.49 ± 2.13 | 3.67 ± 0.65 | 1.10 ± 0.01 | 0.43 ± 0.01 |
| A-2 | REL11333 | 16.21 ± 0.22 | 2.85 ± 2.07 | 1.29 ± 0.07 | 0.39 ± 0.02 |
| A-4 | REL11336 | 15.35 ± 1.03 | 2.92 ± 0.73 | 0.99 ± 0.06 | 0.41 ± 0.01 |
| A-5 | REL11339 | 14.22 ± 0.46 | 6.35 ± 0.87 | 1.15 ± 0.07 | 0.43 ± 0.01 |
| A-6 | REL11389 | 14.25 ± 0.67 | 3.16 ± 1.15 | 1.19 ± 0.06 | 0.44 ± 0.01 |

**Supplementary Table 4.2: Experimentally determined flux ratios for ancestral and evolved LTEE isolates.** PEP through PPP is an upper bound (ub); PYR from MAL is a lower bound (lb).

| | Serine through glycolysis | PYR through ED pathway | PEP through PPP (ub) | OAA from PEP | PEP from OAA | PYR from MAL (lb) |
|---|---|---|---|---|---|---|
| Anc | 0.74 ± 0.01 | 0.01 ± 0.00 | 0.27 ± 0.05 | 0.69 ± 0.05 | 0.03 ± 0.01 | 0.01 ± 0.01 |
| A+1 | 0.67 ± 0.01 | 0.06 ± 0.01 | 0.28 ± 0.03 | 0.69 ± 0.02 | 0.03 ± 0.01 | 0.02 ± 0.01 |
| A+2 | 0.75 ± 0.01 | 0.05 ± 0.01 | 0.18 ± 0.03 | 0.61 ± 0.03 | 0.04 ± 0.01 | 0.00 ± 0.01 |
| A+3 | 0.75 ± 0.01 | 0.13 ± 0.00 | 0.14 ± 0.06 | 0.71 ± 0.05 | 0.04 ± 0.02 | 0.03 ± 0.02 |
| A+4 | 0.73 ± 0.01 | 0.03 ± 0.00 | 0.19 ± 0.05 | 0.59 ± 0.07 | 0.05 ± 0.01 | 0.01 ± 0.02 |
| A+5 | 0.74 ± 0.01 | 0.03 ± 0.01 | 0.23 ± 0.02 | 0.57 ± 0.03 | 0.03 ± 0.02 | 0.02 ± 0.02 |
| A-1 | 0.83 ± 0.01 | 0.02 ± 0.02 | 0.14 ± 0.04 | 0.59 ± 0.02 | 0.03 ± 0.00 | 0.02 ± 0.00 |
| A-2 | 0.70 ± 0.02 | 0.03 ± 0.02 | 0.24 ± 0.05 | 0.80 ± 0.04 | 0.04 ± 0.01 | 0.00 ± 0.00 |
| A-4 | 0.76 ± 0.01 | 0.05 ± 0.02 | 0.18 ± 0.03 | 0.63 ± 0.01 | 0.03 ± 0.01 | 0.01 ± 0.01 |
| A-5 | 0.73 ± 0.01 | 0.03 ± 0.00 | 0.22 ± 0.04 | 0.60 ± 0.01 | 0.03 ± 0.02 | 0.02 ± 0.02 |
| A-6 | 0.72 ± 0.00 | 0.03 ± 0.01 | 0.17 ± 0.05 | 0.58 ± 0.03 | 0.03 ± 0.01 | 0.01 ± 0.01 |

**Supplementary text 4.1: Equations used to calculate flux ratios for the LTEE.**

The notation v(x) represents the flux through reaction x of the iaf1260 genome-scale model of metabolism.

serine through glycolysis

=2*(v(1006)+v(1043)-v(2246)-
v(2203))/(v(688)+2*(v(1006)+v(1043))+v(2245)+v(2246));

pyruvate through Entner-Doudoroff

=v(688)/(v(688)+v(575)+v(1233)+v(1626)+v(1627));

oxaloacetate from phospho*enol*pyruvate
=v(2020)/(v(2020)+v(1622)+v(1623)+v(1624));

phospho*enol*pyruvate from oxaloacetate =v(2022)/(v(2022)+v(695));

pyruvate from malate

PYR=(v(1626)+v(1627))/((v(1626)+v(1627)+v(575)+v(1233)+v(1626)+v(1627)));