UNIVERSITY OF CALIFORNIA,
IRVINE

Genotypic and Phenotypic Dynamics of Adaptation in Experimentally Evolved *Escherichia coli*

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Shaun Michael Hug

Dissertation Committee:
Professor Brandon Gaut, Chair
Professor Anthony Long
Associate Professor Adam Martiny

2016

DEDICATION

To

the trillions of

Escherichia coli

who lived and died

to make this work possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my committee chair, principal investigator, advisor, and mentor, Dr. Brandon Gaut. Never have I encountered a faculty member so unconditionally supportive of my progress and goals. Through my successes and setbacks, he was always excited for what I had done and what I was capable of doing. Thanks to his constant guidance and encouragement, I do not just feel like a scientist; I *know* that I am a scientist. I could not be luckier to have found a home in his lab.

I would like to thank my committee members, Dr. Adam Martiny and Dr. Anthony Long. They always provided me with useful feedback, thoughtful comments, and support for my professional goals.

I would like to thank the students, postdocs, and visiting scientists of the Gaut lab who not only provided me with helpful insights and encouragement along the way, but who are also genuinely lovely friends and colleagues: Jonas Aguirre, Jesse Alas, Andrea Gonzalez-Gonzalez, Qingpo Liu, Flavia Mascagni, Concepcion Munoz-Diaz, Jinhua Ran, Alejandra Rodriguez-Verdugo, Kyria Roessler, Danelle Seymour, Shohei Takuno, Bridgett VonHoldt, and Yongfeng Zhu.

I would like to thank the Gaut lab's two absolutely incredible lab technicians, Rebecca Gaut and Pam McDonald. There is no way I would have completed my degree without their time, energy, expertise, and training, and I consider myself very lucky to have been able to absorb even a tiny fraction of their experimental wisdom and finesse. Words can't express just how much I appreciate everything they did for me and what wonderful people they are.

I would like to thank the Department of Ecology and Evolution and the Center for Complex Biological Systems for all of the support they provided over the years, as well as all of their free meals that gave me the energy and good vibes to keep moving forward. The faculty, administrators, and students in both groups have always been positive, supportive, and helpful, and I am forever appreciative of all that they do.

I would like to thank all of my friends, new and old, for being a part of my grad school journey. There's no way I could list all of you, but know how much I love and appreciate you. From dinner to movies to games to just sitting around and chatting, you have kept me positive, relaxed, and balanced. Thank you for all the adventures we've had, and for all the adventures we'll undoubtedly be having soon. I've got some more free time now!

I would like to thank my best friend in the whole wide world, Marissa Macchietto. We have been classmates and friends since the first grade, and I do not think I would have gone on this grad school adventure if she had not also been adventuring right there by my side. I cannot wait to see where we go next because I know we will be in it together until the end. Italy and France: together forever! We did it, my lovely! Now let's blow this joint!

I would like to thank my family for their nonstop love, support, encouragement, and feigned interest during my time as a grad student. They were always there to put up with the highs and lows of my grad school experience, and I absolutely would not have made it this far

without them supporting me the whole way. Again, words fail to describe just how much I love them and appreciate having them with me on my journey through life. Now that I'm done, we can finally catch up on all that TV I missed! I love you!

This work was supported by the National Institute of Biomedical Imaging and Bioengineering, National Research Service Award EB009418 from the University of California, Irvine, Center for Complex Biological Systems.

Chapter 1 was published with the permission of BioMed Central. The text of this chapter is a reprint of the material as it appears in *BMC Evolutionary Biology*.

CURRICULUM VITAE

Shaun M. Hug
Department of Ecology and Evolutionary Biology
University of California, Irvine
Email: shug@uci.edu
Phone: (949)-824-2963

Education:

2012 – 2016	**Ph.D. – Biological Sciences** University of California, Irvine Advisor: Brandon S. Gaut
2012 – 2015	**M.S. – Biological Sciences** University of California, Irvine
2011 – 2012	**Mathematical, Computational, and Systems Biology Graduate Gateway Program** University of California, Irvine
2009 – 2011	**B.S. – Biological Sciences (*summa cum laude*)** University of California, Irvine
2007 – 2009	**A.A. – Biology (high honors)** Fullerton College

Professional and Teaching Positions:

2013 – 2016	**Teaching Assistant** BIO SCI E131L: Image Analysis Laboratory BIO SCI 94: From Organisms to Ecosystems (Course Coordinator) BIO SCI E109: Human Physiology BIO SCI E153: Genome Evolution BIO SCI 94: From Organisms to Ecosystems BIO SCI 1A: Introduction to Life Sciences BIO SCI E131L: Image Analysis Laboratory BIO SCI 94: From Organisms to Ecosystems BIO SCI 1A: Introduction to Life Sciences
2011 – 2016	**Graduate Student Researcher** Ecology and Evolutionary Biology Graduate Program Mathematical, Computational, and Systems Biology Graduate Gateway Program

Fellowships, Grants, and Awards:

2014 – 2015	**Edward Steinhaus Teaching Award** ($750)
2012 – 2013	**UC Irvine Center for Complex Biological Systems Opportunity Award** ($11,000)
2011 – 2013	**National Institute of Biomedical Imaging and Bioengineering Predoctoral Training Grant** ($25,000 / year)

2009 – 2010 **Elizabeth V. Wright Life Science Scholarship** ($150)

Publications:

Hug, S.M., & B.S. Gaut. 2015. The phenotypic signature of adaptation to thermal stress in *Escherichia coli*. BMC Evolutionary Biology 15:177.

Presentations:

Hug, S.M. 2016. The Lazarus Effect: When Bacteria Return from the Brink of Extinction. [Presentation]. Winter Ecology and Evolutionary Biology Graduate Student Symposium, University of California, Irvine. Irvine, CA.

Hug, S.M. 2014. Phenotypic Diversity of Experimentally Evolved *Escherichia coli*. [Invited Presentation]. Ecology and Evolutionary Biology New Student Recruitment Event, University of California, Irvine. Irvine, CA.

Hug, S.M. 2014. Phenotypic Diversity of Experimentally Evolved *Escherichia coli*. [Invited Presentation]. Mathematical, Computational, and Systems Biology New Student Recruitment Event, University of California, Irvine. Irvine, CA.

Hug, S.M. 2014. Phenotypic Diversity of Experimentally Evolved *Escherichia coli*. [Presentation]. Winter Ecology and Evolutionary Biology Graduate Student Symposium, University of California, Irvine. Irvine, CA.

Hug, S.M., & B.S. Gaut. 2014. Phenotypic Dynamics of Adaptation in Genetically Divergent Populations of Heat-tolerant *Escherichia coli*. [Poster]. 4th Annual Southern California Systems Biology Conference, Irvine, CA.

Hug, S.M., & B.S. Gaut. 2013. Phenotypic Dynamics of Adaptation in Genetically Divergent Populations of Heat-tolerant *Escherichia coli*. [Poster]. Southern California Evolutionary Genetics and Genomics Meeting, Irvine, CA.

Hug, S.M., & B.S. Gaut. 2013. Phenotypic Dynamics of Adaptation in Genetically Divergent Populations of Heat-tolerant *Escherichia coli*. [Poster]. Society for Molecular Biology and Evolution Meeting, Chicago, IL.

Hug, S.M., A.G. Kent, M.G. Macchietto, B.S. Gaut, A.C. Martiny, & S.A. Mortazavi. 2013. Fitness Tradeoffs in Experimentally Evolved *E. coli*. [Presentation]. Center for Complex Biological Systems Annual Retreat, University of California, Irvine. Los Angeles, CA.

Hug, S.M. 2013. Understanding Evolutionary Outcomes and Processes in *E. coli*. [Presentation]. Mathematical, Computational, and Systems Biology New Student Recruitment Event, University of California, Irvine. Irvine, CA.

Hug, S.M., T.D. Long, R.L. Gaut, & B.S. Gaut. 2013. Assessing Phenotypic Convergence in Genetically Divergent Subsets of Heat-tolerant *Escherichia coli*. [Poster]. 3rd Annual Southern California Systems Biology Conference, Irvine, CA.

Hug, S.M. 2012. Graduate School for Biology Majors. [Invited Presentation]. Science and Health Careers Symposium, Fullerton College. Fullerton, CA.

Hug, S.M., T.D. Long, R.L. Gaut, & B.S. Gaut. 2012. Assessing Phenotypic Convergence in Genetically Divergent Subsets of Heat-tolerant *Escherichia coli*. [Poster]. National Institute of Biomedical Imaging and Bioengineering Training Grantees Meeting, Bethesda, MD.

Hug, S.M. 2012. Understanding Evolutionary Outcomes and Processes in *E. coli*.

[Invited Presentation]. Mathematical, Computational, and Systems Biology New Student Recruitment Event, University of California, Irvine. Irvine, CA.

Professional Service:

2016	**Biology Tutor**, MCSB Summer Boot Camp, UC Irvine.
2015 – 2016	**Graduate Student Representative**, Department of Ecology and Evolutionary Biology, UC Irvine.
2014	**Biology Tutor**, MCSB Summer Boot Camp, UC Irvine.
2014	**Math and Computation Tutor**, MCSB Summer Boot Camp, UC Irvine.
2013	**Visiting Scientist**, Ask-A-Scientist Night, Irvine Unified School District.
2013	**Math and Computation Tutor**, MCBU Summer Research Experience, UC Irvine.
2013	**Biology Tutor**, MCSB Summer Boot Camp, UC Irvine.
2012	**Teaching Assistant**, California State Summer School for Mathematics and Science (COSMOS), UC Irvine.
2012	**Visiting Scientist**, Ask-A-Scientist Night, Irvine Unified School District.
2012	**Math and Computation Tutor**, MCSB Summer Boot Camp, UC Irvine.
2012	**Biology Tutor**, MCSB Summer Boot Camp, UC Irvine.

Society Membership:

2013 – 2014	Society for Molecular Biology and Evolution

Graduate Coursework:

2013	Writing Proposals
2013	Special Topics in Evolution
2012	Ecology and Evolutionary Biology Seminar
2012	Responsible Conduct in Research
2012	Systems Developmental Biology
2012	Computational Systems Biology
2012	Systems Biology Journal Club
2012	Mathematical and Computational Biology II (Partial Differential Equations)
2012	Systems Cell Biology
2011	Mathematical and Computational Biology I (Ordinary Differential Equations)
2011	Biophysics of Molecules and Molecular Machines
2011	Critical Thinking in Systems Biology

ABSTRACT OF THE DISSERTATION

Genotypic and Phenotypic Dynamics of Adaptation in Experimentally Evolved *Escherichia coli*

By

Shaun Michael Hug

Doctor of Philosophy in Biological Sciences

University of California, Irvine, 2016

Professor Brandon Gaut, Chair

Despite its centrality to Darwin's theory of evolution by natural selection, the process of adaptation is still not fully understood. In particular, the dynamics of the genotypes and phenotypes associated with an adaptive response remain to be fully elucidated. In my dissertation, I utilized laboratory evolution experiments to study how the genotypes and phenotypes of *Escherichia* coli change over time as they adapted to high temperature.

Chapter 1 explored how metabolic phenotypes of 115 evolved *E. coli* clones changed as a result of 2,000 generations of adaptation to 42.2°C. Using phenotypic microarrays (Biolog plates), I quantified 94 phenotypes of these evolved clones, as well as their ancestor under stressed (42.2°C) and unstressed (37.0°C) conditions. Comparing the evolved phenotypes to the ancestral phenotypes revealed that adaptation was predominantly restorative, shifting evolved phenotypes from the stress state toward the unstressed state. I also uncovered associations among common genotypic changes found in the evolved clones and their phenotypes.

Chapter 2 investigated the different mutational dynamics in populations traversing two different adaptive pathways typified by mutations in the *rpoB* and *rho* genes, respectively. These genes were predicted to be differentially pleiotropic, and were therefore expected to create

differences in compensatory evolution when mutated. I used temporal sequencing data of four *rpoB* and four *rho* populations to reconstruct their mutational trajectories over the course of adaptation to 42.2°C. These trajectories revealed that *rpoB* and *rho* mutations occurred early on during adaptation, canalizing the adaptive process. Furthermore, *rpoB* populations accumulated more mutations and experienced more clonal interference over the course of adaptation than *rho* populations.

Chapter 3 was a study of the Lazarus effect, a phenomenon of population recovery under lethal selection conditions. I evolved ~300 *E. coli* populations to the lethal temperature of 43.0°C and measured their cell density over five days. I sequenced those populations that recovered and found mutations in two operons—*hslUV* and *rpoBC*—to be the major drivers of Lazarus events. These mutations differed in their frequency in the experiment, degree of parallelism within and between weeks, and fitness tradeoffs at 37.0°C, suggesting different origins and adaptive dynamics between them.

INTRODUCTION

Adaptive evolution is the process by which a population improves its phenotype(s) to better fit the pressures of the environment, and it encompasses two complementary processes: the removal of maladapted individuals, and the spread of better-adapted individuals. This first process, known as purifying selection, has been studied for a long time, and as such, much is known about it, especially in the context of disease (Muller, 1950). Purifying selection also makes sense intuitively; there are almost limitless ways in which can organism can be defective in its environment, and those organisms are less likely than their compatriots to survive and reproduce. Out of all possible new mutations that might affect an organism, most are deleterious (Eyre-Walker and Keightley, 2007), and algorithms have been developed to predict whether or not mutations will be maladaptive (Ng and Henikoff, 2001; Adzhubei *et al.*, 2010).

The second process—positive selection—is not as straightforward. New beneficial mutations are rare (Eyre-Walker and Keightley, 2007), making them intrinsically more difficult to study. Moreover, the characteristics that make a mutation beneficial are often dependent on its genetic context and the environment. What exactly gives an individual an advantage over other members of the population? Are there limitless ways to adapt, just as there are so many ways to fail, or is there a finite set of changes that enhance survival and reproduction? If evolution could be replayed, would adaptation repeat itself, or would populations find new and different ways to adapt each time?

Studying adaptive evolution is not always easy. Natural environments can impose different and varying selective pressures, making it difficult to connect the genotypic, phenotypic, and fitness effects of adaptation to a specific cause. The long generation times of some organisms can make it nearly impossible to observe evolutionary changes within one

human lifetime. Furthermore, studying extant populations only provides insight into the outcome of adaptation; the dynamic process of adaptation cannot be fully accounted for using a single snapshot of populations as they currently are. Laboratory evolution experiments, combined with the rise of cheap, high-throughput genome sequencing, now have the power to address some of these issues.

Bacteria such as *Escherichia coli* provide an excellent model system for testing evolutionary hypotheses for several reasons. First, they grow rapidly, allowing evolutionary changes to be observed at human-friendly timescales. Second, they maintain relatively large population sizes, reducing the importance of genetic drift and enhancing the ability to observe beneficial mutations. Third, they are amenable to being frozen and revived, allowing intermediate stages of the evolutionary process to be observed, studied, and even restarted. Lastly, *E. coli* has a relatively small genome size, and its genome is thoroughly annotated, facilitating the processes of sequencing and putting mutations into a biologically meaningful context. Subjecting *E. coli* populations to singular, controlled environmental pressures in the laboratory offers us the best opportunity to link genotypes, phenotypes, and fitness, and to observe how the adaptive process unfolds over time.

Arguably the most famous bacterial evolution experiment began in 1988 at the University of California, Irvine, under the direction of Richard Lenski. In this experiment, 12 replicate populations of *E. coli*, which were derived from a single ancestor, were serially propagated under low-glucose conditions (Lenski *et al.*, 1991). This experiment is still running nearly 30 years and over 60,000 bacterial generations later, and it continues to provide insights into the workings of adaptive evolution. Albert Bennett, also at the University of California, Irvine, carried out a similar evolution experiment with Lenski, evolving six replicate *E. coli* populations at high temperature for one month (Bennett *et al.*, 1990).

To better understand the genetics underlying an adaptive response, our lab devised and carried out an evolution experiment with Bennett and Anthony Long that was akin to previous studies, but with a 10-fold greater level of replication (Tenaillon *et al.*, 2012). This amount of replication provided the statistical power necessary to draw useful conclusions about the parallelism, diversity, and interactions of mutations generated during adaptive evolution. Following previous established methods (Bennett and Lenski, 1993), this experiment began with single colonies of *E. coli* B possessing a neutral Ara- marker (REL1206) that were inoculated into each of 115 independent culture tubes containing 10 mL of Luria-Bertani medium (LB) and grown overnight at 37.0°C. These cultures were each transferred in a 1:100 dilution into 9.9 mL of Davis minimal medium supplemented with 25 mg/L glucose (DM25) and grown overnight at 37.0°C. For the remainder of the experiment, each culture was transferred daily in a 1:100 dilution into fresh DM25 and maintained in a shaking water bath at 42.2°C. This process continued for 2,000 generations, or approximately one year. Additionally, population samples of each of the 115 lines were taken at regular intervals (generations 100, 200, and every 200 generations thereafter) and frozen.

After 2,000 generations of adaptation, single clones were isolated from each of the 115 lines for genome sequencing and fitness measurements relative to the ancestor at 42.2°C. On average, evolved clones experienced increases in fitness of ~40% and possessed an average of ~11 mutations each, most of which were determined to be beneficial (Tenaillon *et al.*, 2012). Additionally, two groups of highly parallel genetic changes were found to be in negative statistical epistasis with one another. These mutations constituted two distinct genetic pathways to adaptation. The first was typified by mutations in the *rpoB* gene, which encodes a subunit of the RNA polymerase complex, and the second was typified by mutations in the *rho* gene, which encodes a major transcriptional terminator (Tenaillon *et al.*, 2012). Although these two

pathways led to improved thermal stress tolerance after 2,000 generations, questions about their differences, side effects, and mechanisms of action remained largely unanswered.

Recently, it has been suggested and shown that a major outcome of adaptation in evolution experiments is restoration of an organism's gene expression from a stressed state to an unstressed state (Fong *et al.*, 2005; Carroll and Marx, 2013; Sandberg *et al.*, 2014). The first chapter of my dissertation investigates whether or not restorative adaptation extends beyond gene expression to the level of metabolic phenotypes. Following evolution to thermal stress, are *E. coli* clones more similar metabolically to their stressed or unstressed ancestor, or do they behave in completely novel ways? Moreover, can instances of restoration or novelty be associated to specific mutations that occurred during the course of adaptation? To address these questions, I employed Biolog phenotypic microarrays, 96-well plates that colorimetrically measure microbial metabolism on a variety of carbon sources and in the presence of a variety of inhibitory compounds. I then connected these phenotypes to genotypes obtained from whole-genome sequencing of *E. coli* clones derived from 115 independent populations evolved at 42.2°C for 2,000 generations.

My second chapter explores the dynamics of adaptation over 2,000 generations, building specifically upon our lab's finding of two pathways to thermal stress adaptation—*rpoB* and *rho* (Tenaillon *et al.*, 2012). Work by other members of our lab has suggested that the *rpoB* and *rho* pathways differ in their fitness tradeoffs at low temperature (Rodríguez-Verdugo *et al.*, 2014) and their effects on gene expression (Rodríguez-Verdugo *et al.*, 2016; González-González *et al*., in prep.). Furthermore, work I performed for my first chapter revealed that clones of these two different pathways were distinguishable by their phenotypes. To understand whether these two adaptive pathways show distinct evolutionary histories and dynamics, I collected and sequenced genomic DNA from eight evolved populations (four *rpoB*, four *rho*) across 11 evolutionary time

points. In particular, I was interested in how differing degrees of pleiotropy between these two adaptive pathways might lead to differential compensation, and the effect of that phenomenon on the number of mutations each type of population accumulates and how much they interact with one another via clonal interference.

In nature, adaptation does not necessarily occur under the relatively mild stressors imposed by scientists in the laboratory. Environmental changes can sometimes be too great for populations to maintain their numbers, leading to crashes and potential extinctions. Nonetheless, some populations do survive these initially lethal pressures, adapting rapidly and recovering before going extinct. This phenomenon has been observed incidentally in laboratory evolution experiments and is known as the "Lazarus effect" (Mongold *et al.*, 1999). My third chapter shines a light on the Lazarus effect and the very initial stages of the adaptive process. How often does the Lazarus effect occur? What are the identity and diversity of mutations that enable population recovery? To answer these questions, I evolve approximately 400 populations of *E. coli* to lethally high temperatures (43.0°C and 44.0°C) over the course of five days. By measuring their cell densities on each day, I can determine whether or not any populations have recovered, and if so, I can save those populations for later whole-genome sequencing.

Overall, the work I carry out in my thesis elucidates important features of the adaptive process and the utility of laboratory evolution experiments. It reveals that the first beneficial mutations to take hold in a population are key to how the remainder of adaptation unfolds, and that multiple pathways to adaptation can lead to distinct differences in evolutionary dynamics and phenotypic outcomes.

REFERENCES

Adzhubei, I. A. et al., 2010 A method and server for predicting damaging missense mutations. Nat Methods **7**: 248-249.

Bennett, A. F., K. M. Dao, and R. E. Lenski, 1990 Rapid evolution in response to high temperature selection. Nature **346**: 79-81.

Bennett, A. F., and R. E. Lenski, 1993 Evolutionary adaptation to temperature II. Thermal niches of experimental lines of *Escherichia coli*. Evolution 1-12.

Carroll, S. M., and C. J. Marx, 2013 Evolution after introduction of a novel metabolic pathway consistently leads to restoration of wild-type physiology. PLoS genetics **9**: e1003427.

Eyre-Walker, A., and P. D. Keightley, 2007 The distribution of fitness effects of new mutations. Nat Rev Genet **8**: 610-618.

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

Lenski, R. E., M. R. Rose, S. C. Simpson, and S. C. Tadler, 1991 Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. American naturalist 1315-1341.

Mongold, J. A., A. F. Bennett, and R. E. Lenski, 1999 Evolutionary adaptation to temperature. VII. Extension of the upper thermal limit of *Escherichia coli*. Evolution 386-394.

Muller, H. J., 1950 Our load of mutations. Am J Hum Genet **2**: 111-176.

Ng, P. C., and S. Henikoff, 2001 Predicting deleterious amino acid substitutions. Genome Res **11**: 863-874.

Rodríguez-Verdugo, A., O. Tenaillon, and B. S. Gaut, 2016 First-Step Mutations during Adaptation Restore the Expression of Hundreds of Genes. Mol Biol Evol **33**: 25-39.

Rodríguez-Verdugo, A., D. Carrillo-Cisneros, A. González-González, B. S. Gaut, and A. F. Bennett, 2014 Different tradeoffs result from alternate genetic adaptations to a common environment. Proceedings of the National Academy of Sciences **111**: 12121-12126.

Sandberg, T. E. et al., 2014 Evolution of Escherichia coli to 42 degrees C and Subsequent Genetic Engineering Reveals Adaptive Mechanisms and Novel Mutations. Mol Biol Evol

Tenaillon, O. et al., 2012 The molecular diversity of adaptive convergence. Science **335**: 457 461.

CHAPTER 1 - The phenotypic signature of adaptation to thermal stress in *Escherichia coli*

ABSTRACT

In the short-term, organisms acclimate to stress through phenotypic plasticity, but in the longer term they adapt to stress genetically. The mutations that accrue during adaptation may contribute to completely novel phenotypes, or they may instead act to restore the phenotype from a stressed to a pre-stress condition. To better understand the influence of evolution on the diversity and direction of phenotypic change, we used Biolog microarrays to assay 94 phenotypes of 115 *Escherichia coli* clones that had adapted to high temperature (42.2°C). We also assayed these same phenotypes in the clones' ancestor under non-stress (37.0°C) and stress (42.2°C) conditions. We explored associations between Biolog phenotypes and genotypes, and we also investigated phenotypic differences between clones that have one of two adaptive genetic trajectories: one that is typified by mutations in the RNA polymerase β-subunit (*rpoB*), and another that is defined by mutations in the *rho* termination factor. Most (58%) phenotypic variation was restorative, shifting the phenotype from the acclimated state back toward the unstressed state. Novel phenotypes were more rare, comprising between 7% and 20% of informative phenotypic variation. Genetic variation associated statistically with phenotypic variation, demonstrating a genetic basis for shifted phenotypes. Finally, clones with *rpoB* mutations differed in phenotype from those with *rho* mutations, largely due to differences in chemical sensitivity. Our results contribute to previous observations showing that a major component of adaptation in microbial evolution experiments is toward restoration to the unstressed state. In addition, we found that a large deletion strongly affected phenotypic variation. Finally, we demonstrated that the two genetic trajectories leading to thermal adaptation encompass different phenotypes.

INTRODUCTION

Our understanding of the dynamics of adaptation in populations is incomplete (Orr, 2005), particularly with respect to the repeatability and the direction of adaptation. For repeatability, the major questions are, first, whether replicated evolutionary events converge on a single adaptive phenotype and, second, whether convergent phenotypes are caused by the same set of underlying genetic changes. For the direction of adaptation, the major question is whether adaptation commonly leads to novel phenotypes or instead acts to restore phenotypes to pre-stress states. To understand this last point, it is important to recognize that adaptation often begins with a physiological stress in a new environment. In the short term, there may be acclimation to stress through a physiological response, but genetic and phenotypic adaptation occurs in the longer term. The question is whether adaptation typically restores phenotypes to a pre-stress state or more often leads to phenotypic novelty (Figure 1.1).

Questions about novelty, restoration, and convergence have been addressed in the context of experimental evolution (Fong *et al.*, 2005; Carroll and Marx, 2013; Sandberg *et al.*, 2014). These studies have found that evolution typically proceeds toward the restoration of the pre-stress condition. For example, Carroll and Marx (Carroll and Marx, 2013) evolved eight replicate bacterial lineages under stress conditions and then measured gene expression. They found that 93% of all adaptive changes in gene expression restored expression from the acclimated (stressed) state back to the wild-type (pre-stress) condition. Of these restorative changes, 70% occurred in parallel across all eight populations. These studies make the important point that characterizing the intermediate acclimation process is essential to understanding the repeatability and direction of adaptation. However, these studies have also been limited to a low number (< 10) of experimental replicates.

We recently performed a highly replicated experiment in which *Escherichia coli* evolved to high temperature (Tenaillon *et al.*, 2012). To begin, we inoculated a clone of *E. coli* strain REL1206 (Lenski *et al.*, 1991) into 115 replicate populations, and then allowed the populations to evolve independently for 2,000 generations at 42.2°C. At the end of the experiment, we sequenced the genomes of single clones from each of the replicates. Our sequencing efforts revealed a total of 1,331 mutations across a set of 115 evolved clones. Roughly half of these mutations were shared among two or more clones, with many falling into one of two different adaptive genetic trajectories. The first of these includes mutations in *rpoB*, which codes for the β-subunit of RNA polymerase, along with associated mutations in other RNA polymerase subunits and the *rod* genes that define cell shape. The second trajectory includes mutations in *rho*, which codes for a major transcription termination factor, along with mutations in *iclR*, a transcriptional regulator of the glyoxylate shunt of the Krebs cycle, and *cls*, a cardiolipin synthase gene important for regulating membrane fluidity and permeability. Mutations in the *rpoB* and *rho* adaptive trajectories are not mutually exclusive, but they are strongly negatively associated, presumably due to negative epistatic interactions (Tenaillon *et al.*, 2012).

Overall, we have observed adaptive genetic convergence—i.e., mutations in two or more independent clones—in ~80 genes (Tenaillon *et al.*, 2012), of which the two adaptive trajectories represent only a subset. We have been left, then, with a large amount of unexplained genetic diversity that is presumed to be adaptive at 42.2°C, hundreds more mutations that are unique to single evolved clones, and the observation that all 115 of our populations independently evolved the ability to persist in the same high-temperature environment. While our study has provided a description of the breadth of genetic change underpinning an adaptive response (Tenaillon *et al.*, 2012), the extent of phenotypic convergence remains unclear, as does the direction of phenotypic evolution.

In this study, we assess phenotypic diversity among our 115 evolved *E. coli* clones using high-throughput Biolog arrays. Biologs are 96-well plates that test metabolic phenotypes (Bochner *et al.*, 2001), including 71 carbon utilization assays and 23 chemical sensitivity assays. Biologs have been used to discover new links in microbial biochemical pathways (Loh *et al.*, 2006), to associate genotypes with phenotypes (Zhou *et al.*, 2003; Pommerenke *et al.*, 2010), to uncover a decoupling between genotypic and phenotypic diversity across *E. coli* strains (Sabarly *et al.*, 2011), and to validate patterns of long-term phenotypic evolution in diverse groups of bacteria (Plata *et al.*, 2015). They have also been employed in an evolution experiment to characterize ecological dynamics and niche displacement in coevolving subpopulations of *E. coli* (Le Gac *et al.*, 2012).

Using Biolog assays, we have measured 94 phenotypes for each of our 115 evolved clones at 42.2°C and for their REL1206 ancestor at two treatment temperatures (37.0°C and 42.2°C). With this dataset of phenotypes, our study has four interacting objectives. The first is to assess phenotypic variation among clones based on a Biolog 'fingerprint'. The second is to measure the direction of phenotypic adaptation in our 115 evolved clones relative to the stressed (42.2°C) and non-stressed (37.0°C) ancestor. Based on previous studies (Fong *et al.*, 2005; Carroll and Marx, 2013), we hypothesize that many of the phenotypic changes restore phenotypes from the stressed toward the pre-stress state. The third objective is to assess whether phenotypic shifts have a genetic component—i.e., to ascertain that genetic adaptation has contributed to the observed phenotypic shifts rather than phenotypic plasticity. Finally, we contrast the two adaptive trajectories typified by *rho* and *rpoB* mutations. Do these two genetic trajectories vary in their resultant phenotypes?

MATERIALS AND METHODS

Evolution to Thermal Stress

Our thermal stress experiment was reported elsewhere (Tenaillon *et al.*, 2012), but we cover the experimental design here for the sake of clarity. Following previously established methods (Bennett and Lenski, 1993), we inoculated our ancestor Ara- *E. coli* clone (REL1206) into 115 culture tubes that contained 10 mL of Davis minimal medium supplemented with 25 mg/L glucose (DM25). The ancestral REL1206 clone had been propagated previously at 37.0°C for 2,000 generations in DM25 and thus was likely to be well adapted to the media. The 115 cultures were maintained in a shaking water bath at 42.2°C for 2,000 generations and were transferred daily into fresh media via 100-fold dilution.

At the end of the experiment, we isolated one clone from each of the 115 populations. Each genome was sequenced and the clones were also assessed for their fitness relative to the ancestor at 42.2°C (Tenaillon *et al.*, 2012); on average, the evolved clones were ~40% more fit than the ancestor. For further details about genotypes and relative fitness values, please refer to (Tenaillon *et al.*, 2012).

Biolog Assays

We streaked each of the 115 clones from the *E. coli* thermal stress experiment from frozen glycerol stocks onto tetrazolium and arabinose (TA) agar plates and grew them for one day at 37.0°C. Although the clones had evolved at 42.2°C in DM25, it is common practice in thermal stress studies to allow clones to recover from freezing under less stressful conditions (Bennett and Lenski, 1993; Lenski and Travisano, 1994; Rodriguez-Verdugo *et al.*, 2014).

For each clone, we chose colonies to assay on a GEN III Biolog MicroPlate. To perform the assay, we followed the manufacturer's protocol, which included: *i*) inoculating bacterial

colonies into Inoculating Fluid A (Biolog) to a turbidity between 97% and 99% transmittance, as determined by optical density (OD) at 600 nm on a Synergy H1 Hybrid Multi-Mode Microplate Reader (Biotek); *ii*) adding 100 μL of inoculum to each of the 96 wells of a Biolog plate; *iii*) incubating each plate for 22.25 hours at 42.2°C, and *iv*) developing the assay by measuring optical density (OD) at 590 nm. The OD at 590 nm measured the amount of reduced tetrazolium redox dye in each well, providing a quantitative measurement of respiratory activity in each well of the plate. We performed three Biolog assays for each of the 115 evolved clone, for a total of $115 \times 3 = 345$ plates.

We applied the same Biolog procedures to the REL1206 ancestor, but it was incubated at one of two different treatment temperatures: 37.0°C and 42.2°C. Moreover, for each temperature we performed two sets of three replicates, with the replicate sets performed on different days in order to incorporate potential 'day effects' into the experimental design. Thus, we performed $6 \times 2 = 12$ assays with the REL1206 ancestor.

Because REL1206 evolved at 37.0°C in DM25 from a lab strain of *E. coli* B, we assumed that 37.0°C represents a non-stress condition, while 42.2°C was a stressful environment.

Statistical and Directional Analysis of Phenotypes

Each Biolog plate contained 94 assay wells (or 'tests') and two control wells. For each plate, the OD for each test was normalized to the OD of the appropriate control well. For example, the 71 tests that measure carbon utilization were normalized to a negative control lacking any added metabolic substrates, and the 23 tests that measure chemical sensitivity were normalized to a positive control lacking an inhibitor but permitting a baseline level of assay development. We term the normalized OD values as 'phenotype values,' or PVs.

We first examined Pearson correlation coefficients between all pairs of 94 Biolog tests, based on the average PV ($\overline{\text{PV}}$) for each clone. Because $\overline{\text{PV}}$s were highly correlated among tests, we reduced the dimensionality of log-transformed PV data using principle components analysis (PCA). PCA analyses were based on the R (Team, 2014) module prcomp, with the flags retx=TRUE, center=TRUE, and scale=TRUE. Thereafter, we considered only principal components with significant eigenvectors, as determined by the 'random average under permutation' metric of Peres-Neto et al. (2005) (Peres-Neto *et al.*, 2005), which was based on 1,000 permutated datasets. The significance of loadings was examined with the bootstrap eigenvector metric of Peres-Neto et al. (2003) (Peres-Neto *et al.*, 2003), based on 1,000 resamplings. Both metrics have been shown to be well behaved on a range of simulated datasets (Peres-Neto *et al.*, 2003; Peres-Neto *et al.*, 2005).

For each retained component of the PCA, we compared the average of scores $\bar{S}_{\text{x}}$ of each evolved clone *x* to the average scores of REL1206 at 37.0°C ($\bar{S}_{37°\text{C}}$) and at 42.2°C ($\bar{S}_{42°\text{C}}$). We used *t*-tests for these pairwise comparisons, under the null hypothesis that the REL1206 scores did not differ from those of an evolved clone. The *p*-value for individual *t*-tests were determined by an empirical null distribution, based on 10^5 permutations. For each set of comparisons, the resultant *p*-values were adjusted using a false discover rate (FDR) of $q < 0.01$, based on the p.adjust module of *R*.

Within each principal component, we categorized the direction of phenotypic evolution for each evolved clone by comparing the magnitude and significance of pairwise comparisons among $\bar{S}_{37°\text{C}}$, $\bar{S}_{42°\text{C}}$ and $\bar{S}_{\text{x}}$. Following previous literature (Carroll and Marx, 2013), we defined a total of six directional categories, which represent the phenotypic consequences of evolution for clone *x* (Table 1.1 and Figure 1.1).

Hierarchical clustering of the clones was based on $\bar{S}_x$. Clustering utilized Euclidean distances and an unweighted pair group method with arithmetic mean (UPGMA) (Sneath and Sokal, 1973) and was implemented in MATLAB. Multivariate analysis of variance (MANOVA) was used to test for differences in phenotypes between pre-defined groups of clones. The analyses used PV data from each test and each replicate as dependent variables and the groups as the independent variables, resulting in the model (PVs ~ groups). MANOVA was implemented in the R function manova, based on the Pillai test of significance.

Associations between Phenotype and Genotype

To test for associations between phenotypes and genetic mutations, we grouped mutations found in our evolved clones into 'mutational objects.' These groupings arose by classifying mutations into three broad classes: genic, intergenic, and multigenic. Genic mutations comprised all point mutations, small indels, and IS insertions that affected a single gene, and we grouped these into one mutational object whose identifier was the name of the affected gene. For example, an evolved clone possessing a point mutation in the *cls* gene and another evolved clone possessing an IS insertion in the *cls* gene each received a single identifier, 'cls,' to describe their mutations. Intergenic mutations comprised point mutations, small indels, and IS insertions that fell in noncoding regions between two genes, and we split these into two objects, one associated with each neighboring gene. Lastly, multigenic mutations comprised deletions and insertions spanning two or more genes; we classified these as their own objects whose identifiers were not associated with any specific gene. All genotypic data were from Tenaillon et al. (Tenaillon *et al.*, 2012).

We grouped mutations into objects because most mutations discovered within the thermal stress experiment were found in only a single clone, and hence provided no basis for associating

genotype with phenotype. These groupings likely increased statistical power but may have had an unintended trade-off in statistical power if there was allelic heterogeneity.

For each of the mutational objects present in two or more evolved clones, the PCA scores from each evolved clone were placed into one of two groups: those possessing the mutational object (the cases), or those lacking the mutational object (the controls). A *t*-test assuming equal variance was used to determine whether the cases and controls differed significantly in each of the nine principal components. Results were corrected to $q < 0.01$.

RESULTS

Phenotypic Space

To better understand phenotypic evolution during a previously published evolution experiment (Tenaillon *et al.*, 2012), we performed a total of 357 Biolog assays on 115 evolved clones and two ancestral treatments. Each assay included 94 discrete tests. After normalizing OD readings, we first calculated $\overline{PV}$ values for each test and each clone and then measured Pearson pairwise correlations between these tests. Of 8,836 (= 94 × 94) pairwise comparisons between tests, 20.8% (1,836) were significantly correlated after sequential Bonferroni correction at $\alpha = 0.01$ (Figure S1.1). Given substantial correlation between tests, we reduced the complexity of PV data by PCA transformation into orthogonal components. The first component of the PCA represented 31% of the variance (Figure S1.2), and the eigenvector of the first nine components was significant (Peres-Neto *et al.*, 2005). Each of the first nine components had eigenvectors > 2.0 and together explained 68.7% of variation. We retained the first nine components for further analysis.

Figure 1.2 plots the first and second principal components and helps convey two pieces of information about PCA scores. First, the ancestral data were typically well differentiated by

treatment (37.0°C or 42.2°C). For example, the first component visually separated the sets of six ancestral replicates by treatment (Figure 1.2). While the separation was less obvious for the second component, *t*-test comparisons between $\bar{S}_{37°C}$ and $\bar{S}_{42°C}$ indicated that the two ancestral treatments were significantly differentiated in seven of nine principal components (pc1, pc2, pc5, pc6, pc7, pc8 and pc9; *t*-test, unequal variances; sequential Bonferroni correction for $\alpha = 0.01$). This differentiation represents the phenotypic effects of acclimation (Figure 1.1).

A plot of the first two components also provides an opportunity to illustrate features of the direction and magnitude of evolutionary change (Figure 1.2). In the first principal component, most score values clustered near the stressed (42.2°C) ancestor, suggesting that most of the evolved clones were phenotypically unrestored in pc1. However, the scores of several clones fell intermediate between the two ancestral treatments or near the unstressed (37.0°C) ancestor, indicating partial or full phenotypic restoration, respectively. We consider the direction of phenotypic change more formally below.

PCA also estimated loadings on each axis (Table S1.1); these loadings provide information about individual tests that may have contributed variation to an axis. We tested for 'significant' loadings using a published bootstrapping heuristic, but none were significant at $p < 0.05$. The lack of significance reflects the fact that the loadings were fairly even among tests. For example, in pc1 the highest loading—in terms of the percent of the total loading values—was 2.0% for metabolic activity on 'glucuronamide,' but altogether 31 of 94 tests contributed between 1.5% and 2% to loadings in pc1. However, 27 of these 31 (87%) were tests that measure OD on sugar substrates, suggesting that pc1 primarily reflects variation related to carbohydrate metabolism. The top loadings in pc2 were related primarily to tests either on Krebs cycle compounds or on amino acids, while the top loadings in pc3 included assays for chemical sensitivity (Table S1.1).

The Direction of Adaptation

The two ancestral treatments were significantly differentiated for seven of nine principal components. These observations lay the foundation for assessing the direction of evolution—i.e., did evolution tend to restore phenotypic traits to the non-stressed (37.0°C) state, or did it lead to novel phenotypes?

To assess directionality more formally, we first tested for differences in $\bar{S}_{\mathrm{x}}$ between an evolved clone and each of the two ancestral treatments. For example, we compared $\bar{S}_{\mathrm{x}}$ of each of the 115 clones to $\bar{S}_{42^{\circ}\mathrm{C}}$ in each of the nine axes, for a total of 115 × 9 = 1,035 contrasts, and found that 27.2% (or 282 out of 1,035) of tests were significant at $q < 0.01$. There were nonetheless more differences between the evolved clones and the 37.0°C control, because 53.3% (552 of 1,035) of contrasts between $\bar{S}_{\mathrm{x}}$ and $\bar{S}_{37^{\circ}\mathrm{C}}$ were significant ($q < 0.01$). These results were similar to Figure 1.2 in giving an overall impression that the evolved clones tended to be more similar in phenotype to the stressed ancestor than to the non-stressed ancestor.

We classified the results of *t*-tests into six categories based on the direction and significance of comparisons among $\bar{S}_{\mathrm{x}}$, $\bar{S}_{37^{\circ}\mathrm{C}}$ and $\bar{S}_{42^{\circ}\mathrm{C}}$ (Table 1.1). Of 1,035 comparisons, the highest number of tests (330 of 1,035) fell into the 'uninformative' category, due to a lack of significance among comparisons. Among informative categories, the most comparisons were in the 'unrestored' (230) category, followed by 'partially restored' (151), 'restored' (124), 'reinforced' (77), 'novel' (61) and 'over-restored' (24) (Table 1.1).

These categorical numbers reflect directionality, but they do not account for the fact that the nine principle components explained different proportions of variance (Figure S1.2). To estimate the total proportion of variance explained by each directional category, we weighted

results by the proportion of variance explained in each axis (Table S1.2). Summing across all informative comparisons, weighting revealed that the biggest contributor to phenotypic variance was 'partial restoration' of the unstressed phenotype, which explained 34.6% of observed phenotypic variation (Figure 1.3). The category of 'partial restoration' was followed by the unrestored (23.2%) and restored phenotypes (22.5%). In contrast, novel, over-restored, and reinforced phenotypes combined to explain 19.8% of variation.

Phenotype-Genotype Associations

The predominant phenotypic response during our experiment was toward the partial or full restoration of the pre-stress condition. To verify that these shifts in phenotype had a genetic component—and therefore resulted from adaptive change rather than phenotypic plasticity—we used a case-control approach to associate scores with 165 mutational objects. In total, we found 117 significant ($q < 0.01$) associations with 70 mutational objects distributed across eight of the nine principal components (Table 1.2). We explored the validity of our phenotype-genotype associations by performing *t*-tests on random permutations of our case/control categories for each Biolog assay. When the absolute values of the *t*-statistics obtained from our observed genotypic groups were sorted and plotted against those obtained from randomized groups, there was a strong signal of more extreme *t*-statistics in the observed data (Figure S1.2), suggesting a biological signal in our results.

Among the many significant genotype-phenotype associations (Table 1.2), a few were especially notable. For example, a large deletion variant (ECB_00503_large) that was common to 35 of the 115 evolved lines also had the lowest *p*-values in associations to the first two principle components, suggesting it had a major effect on adaptive phenotypes. Moreover, *rpoB*

and *rho*, two genes that represent the two major adaptive trajectories (Tenaillon *et al.*, 2012), each exhibited significant associations to four and five principal components, respectively.

Contrasting the *rho* and *rpoB* Trajectories

Like previous experiments (Fong *et al.*, 2005; Sandberg *et al.*, 2014), we have shown that most phenotypic variation in our experiment was due to partial or full restoration of the unstressed phenotype. We have gone further to show that some of this variation associates with an underlying genetic component. However, we have not yet addressed the question as to whether different adaptive trajectories—particularly those that include *rho* and *rpoB*—differ in phenotype. We used two approaches to compare these two trajectories.

The first was to test for differences between clones with *rpoB* mutations and clones with *rho* mutations, using MANOVA applied directly to PV data. The results indicated that the two groups differ in phenotype ($p < 2.2 \times 10^{-16}$). MANOVA also assessed the significance of individual tests (or factors) between groups; 23 of the 94 tests differed significantly between the *rpoB* and *rho* groups at $\alpha = 0.01$ (sequential Bonferroni correction) (Table 1.3). Among these 23 factors, the five with the lowest *p*-values were tests of chemical sensitivity.

The second approach was hierarchical clustering of the 115 clones by phenotype, followed by visual examination of the distribution of *rho* and *rpoB* mutants on the dendrogram. We reasoned clones should group phenotypically according to genetic trajectory if the *rpoB* and *rho* trajectories lead to different Biolog phenotypes. The results were intriguing, if not completely clear (Figure 1.4). The dendrogram showed that clones with either mutational object fell into clusters; that is, clones with *rpoB* mutations clustered into discrete groups, and likewise for clones bearing *rho* mutations. In addition, the clusters of *rho*- and *rpoB* clones tended to be mutually exclusive, as expected given that few clones carried mutations in both genes (Tenaillon

et al., 2012). However, for each mutational object, there were multiple clusters, without a clear delineation between the two genetic groups.

DISCUSSION

Adaptation moves an organism toward a phenotypic optimum, but the question remains as to whether there is a single or several genetic trajectories to one or several optima. Previously, we evolved 115 separate *E. coli* lines under thermal stress (42.2°C) for approximately one year (2,000 generations), with the intent to measure the diversity of an adaptive response. This experiment revealed that each of the experimental lines improved in fitness by an average of 40% across clones isolated from each line, fueled by the accumulation of ~11 mutations per clone on average (Tenaillon *et al.*, 2012). The most frequently mutated genes were related to DNA transcription, particularly the *rpoB* and *rho* genes. Mutations within these two genes tended to be negatively associated.

It is an open question whether these two adaptive trajectories—or indeed, the > 1,200 mutations observed during the experiment—lead to convergent phenotypes beyond an increased ability to grow at 42.2°C. Accordingly, this study has been designed to measure phenotypic diversity among these 115 *E. coli* clones using Biolog plates. These plates assess phenotypic characteristics by assaying metabolic activities and chemical sensitivities, but they have at least three important limitations. The first is that many of the phenotypes measured by Biolog plates may not have a direct relationship to fitness during the evolutionary experiment; they may represent pleiotropic effects. It is nonetheless an important task to characterize the phenotypic diversity generated during an adaptive response, as diversity may impact evolvability (Pigliucci, 2008; Barrick *et al.*, 2010; Woods *et al.*, 2011). The second is that many of the assays are not independent (Figure S1.1). The lack of independence necessitated orthogonal transformation of

the data, but these transformations resulted in the loss of information and lessened the ability to associate a discrete phenotype (i.e., a specific Biolog test) to a causative genotype. Lastly, although Biolog technology measures utilization of carbon sources and resistance to inhibitors, bacterial growth and metabolism are complex and environment-dependent; as a result, changes in OD (or a lack thereof) are not always reliable indicators of bacterial growth and metabolism in each assay (Leiby and Marx, 2014). Nonetheless, changes in OD are consistent within our system, both across replicates and on different days, making the Biolog data a useful indicator of a phenotypic 'fingerprint.'

Restoration, Not Novelty, Predominates in Our Experiment

Previous experiments have found that a major component of adaptation to a stressful state is the restoration of phenotypes to a non-stressed state (Fong *et al.*, 2005; Carroll and Marx, 2013; Sandberg *et al.*, 2014). Similar to these experiments, we find that the predominant phenotypic shift in our experiment was toward a restored state like that of the 37.0°C ancestor. Together, full and partial restoration of phenotypes represents 58% of the phenotypic variation among our evolved clones (Figure 1.3).

In contrast, evolutionary novelty is less common, but the proportion of novel phenotypic variation varies by definition. Writ narrowly, novelty may be defined as an evolved state that differs from ancestral treatments that do not differ from each other (Table 1.1). Under this definition, novelty accounts for 7% of phenotypic variation (Figure 1.3). However, novelty can also be described more broadly as a phenotype beyond the limits of the two ancestral treatments, so that over-restoration and reinforcement also encompass novelty (Figure 1.1). With this broader definition, novelty encompasses ~20% of variation but is still dwarfed by both partial and full restoration.

This general result—i.e., that novelty is a less frequent component of adaptation than restoration—is also consistent with previous studies. For example, Sandberg et al. (Sandberg *et al.*, 2014) have found that just 13% (101 / 804) of parallel gene expression shifts in their thermal stress experiment are reinforcements, a classification that can be considered a type of novelty in their system. Likewise, Carroll and Marx (Carroll and Marx, 2013) have documented that cases of parallel novelty are rare in their gene expression data, occurring in just five out of thousands of genes.

Generalizing across studies, the predominant effect within microbial evolution experiments appears to be restoration, at least in the short term. As such, this directional response likely indicates pressure to compensate for the metabolic and energy requirements of the stress response. In the case of *E. coli* thermal stress, the immediate response to thermal stress—i.e., the heat shock response (HSR)—has been well characterized. The HSR up-regulates expression of the transcription factor σ^{32}, thereby driving increased expression of heat shock and other chaperone proteins (Yura *et al.*, 1993) that then help to guide proper folding of crucial cellular proteins at high temperature (Richter *et al.*, 2010). However, while there are myriad studies of HSR in the short term (i.e., on the scale of minutes), the sets of genes that contribute to *E. coli* thermal acclimation over the space of hours and days are not well known. Acclimation may prove to be a distinct physiological state, with specific energetic costs that merit further study.

Genetic Associations and the Effects of a Large Deletion

To assess whether phenotypic variation is driven by genetic variation that accrued during our adaptation experiment, we have associated genotypes with phenotypes. Our association analyses do not find a consequent phenotype for all of the mutational objects. For example, the gene *ybaL*, which was mutated in 65 of 115 lines, does not associate with any phenotypic axes.

Nonetheless, we do find 117 genetic associations across eight PCA axes (Table 4), providing compelling evidence that at least some of the measured phenotypic variation has underlying genetic causations.

Among the many associations, the ECB_00503_large deletion is particularly surprising, because it is the major associate with the first two principal components of variation. In total, it associates with six of the nine principal components under study (Table 1.2), and it exhibits a strong signal of phenotypic differentiation when evolved clones are clustered hierarchically (Figure 1.4). The ECB_00503_large deletion is also unique because it is the most common single mutation from the thermal evolution experiment; 35 of 115 evolved clones share this mutation. We previously speculated that the deletion has a high mutation rate due to homologous recombination between flanking IS insertions (Tenaillon *et al.*, 2012). No matter the mutation rate, it is likely to have been under strong selection to reach high frequency in 35 independent populations.

The ECB_00503_large deletion is 71 kb in length and removes 64 genes (Figure 1.5). These genes include the *cus* operon, which has been shown to be down-regulated in response to osmotic and heat stress (Gunasekera *et al.*, 2008), as well as the *fep* and *ent* operons, which regulate iron acquisition and are regulated by the iron-dependent master transcriptional regulator Fur (Escolar *et al.*, 1999). Interestingly, Fur also regulates enzymes of glycolysis and the Krebs cycle, as well as enzymes that combat oxidative stress (McHugh *et al.*, 2003). It seems possible that the deletion of iron acquisition genes and their Fur binding sites could, in theory, lead to pleiotropic effects by affecting the activation state of Fur or its titration on remaining binding sites. Single genes in the region could also play a role in the stress response, such as the transcription factors encoded by *appY* and *envY*, and the heat shock protease encoded by *ompT* (Figure 1.5).

The ECB_00503_large deletion is one in a series of four overlapping deletions that permit preliminary dissection of the phenotypic effects of the *appY*/*envY*/*ompT* cluster, the *cus* operon, and the *fep*/*ent* operons (Figure 1.5). For example, the hokE_large mutation removes the *fep*/*ent* operons, and this mutation associates with both pc2 and pc5 (Table 1.2). Hence, deletion of the *fep/ent* operon appears to be sufficient to generate some of the phenotypic variation caused by the larger deletion. Similarly, the ECB_00503_small mutation associates with pc4 (Table 1.2), suggesting that the region near *appY*/*envY*/*ompT* also contributes to phenotypic variation in our system.

Two Adaptive Trajectories: *rho* vs. *rpoB*

One of our motivating questions is whether the two adaptive trajectories defined by *rho* and *rpoB* lead to identical phenotypes and fitness optima. To that end, Rodriguez-Verdugo et al. (Rodriguez-Verdugo *et al.*, 2014) have documented that the two trajectories (as well as single mutations in the *rho* and *rpoB* genes) lead to different fitness trade-offs at low temperatures. Thus, the two trajectories do differ in phenotype in a low temperature environment. However, Rodriguez-Verdugo et al. (2014) were also unable to detect a difference in relative fitness between the two sets of clones at 42.2°C, suggesting that the two trajectories may ascend 'fitness peaks' of similar height under thermal stress.

To better understand differences between the two trajectories, we applied MANOVA to our phenotypic data. The analyses revealed significant overall differences between the *rho* and *rpoB* trajectories and also identified factors that contribute to the difference. Based on these factors, the two trajectories appear to differ most substantially in chemical sensitivity but also in other aspects (Table 3).

Unfortunately, we cannot at this point infer the molecular causes of these phenotypic

differences. We can, however, posit reasonable hypotheses. For example, the *rho* and *rpoB* trajectories differ in their associations with the *cls* gene; 23 of 30 clones with mutations in *rho* also contain a *cls* mutation, most of which interrupt *cls* function. In contrast, mutations within *cls* and *rpoB* are associated less often than expected by chance (Tenaillon *et al.*, 2012); only 19 of 60 *rpoB* clones contain a *cls* mutation. These associations may be important because the *cls* gene produces a membrane lipid (Nishijima *et al.*, 1988), and changes in membrane lipid composition are known to alter sensitivity to antibiotics and other chemicals (Handwerger and Tomasz, 1985; Arias *et al.*, 2011). Hence, the two trajectories may differ in chemical sensitivity assays in part because of their different level of association with *cls* mutations. We note, however, that we have no insights as to why mutations within the *rho* and *cls* genes are statistically positively associated while mutations in *rpoB* and *cls* are not (Tenaillon *et al.*, 2012).

Another reasonable explanation for differences between the *rho* and *rpoB* trajectories is pleiotropy, because *rho* and *rpoB* mutations are expected to have different pleiotropic effects (Rodriguez-Verdugo *et al.*, 2014). *rpoB* mutants have the capacity to affect the expression of every gene, but *rho* influences termination in only a subset of genes (Peters *et al.*, 2009; Hollands *et al.*, 2014). Even if the two trajectories do differ in pleiotropic effects, the phenotypic differences we have documented here may not affect fitness under the conditions of the initial thermal stress experiment. However, they are likely to have consequential fitness effects in other environments, such as has been shown at low temperature (Rodriguez-Verdugo *et al.*, 2014).

CONCLUSIONS

Overall, our data reveal that phenotypes converged predominantly toward states like those of the unstressed ancestor during our evolution experiment. This observation supports previous studies, which also document that adaptation in laboratory experiments consists largely

of restorations toward the wild-type, pre-stress phenotype. Either plasticity or adaptation could drive phenotypic shifts, but phenotype-genotype associations confirm that phenotypic change has a genetic component. In contrast to restoration, phenotypic novelty was less common, but did explain as much as ~20% of phenotypic variation. It remains an open question whether such novelty is merely a pleiotropic side effect of restorative evolution, or whether it provides some adaptive function of its own. Finally, our contrast of the *rho* and *rpoB* adaptive trajectories shows that the two represent different phenotypic spaces, but the interpretation of their effects is complicated by the compounded effects of several overlapping deletions as well as the genetic mutations associated with each trajectory.

FIGURES

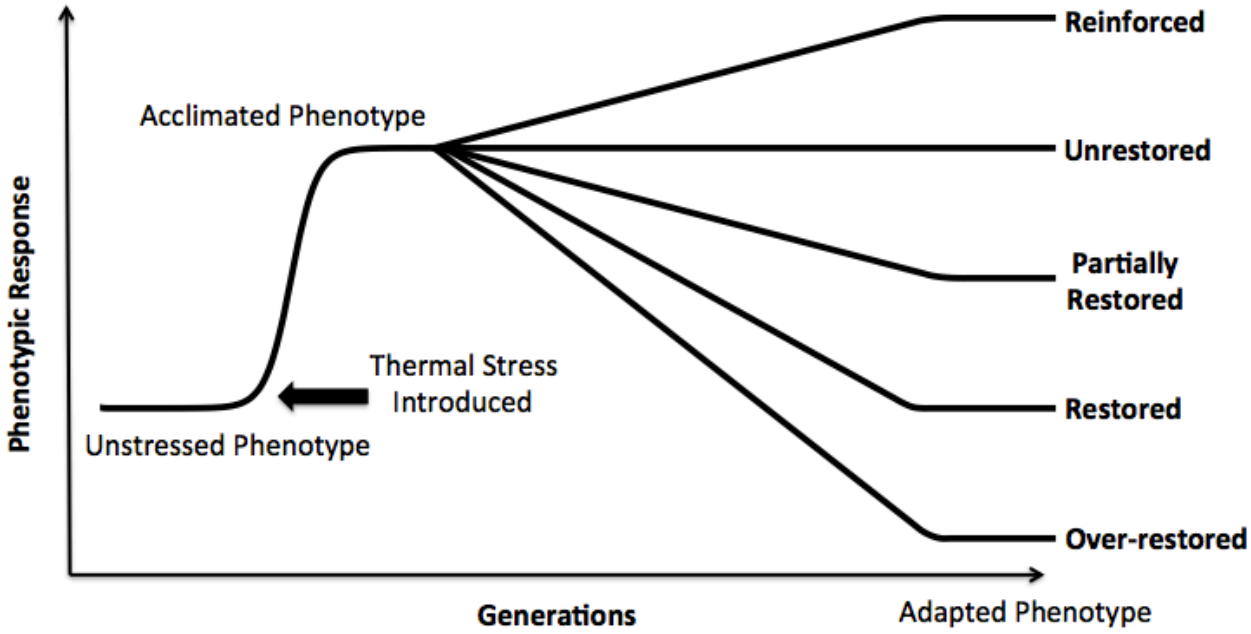

Figure 1.1. Schematic of acclimation and the potential directional outcomes of adaptation. In addition to restored and unrestored states, which reflect the phenotype of the unstressed and stressed ancestor, respectively, evolved clones may exhibit partially restored, over-restored or reinforced phenotypes. Not shown are cases of novelty, in which evolved clones differ from ancestral treatments that do not differ.

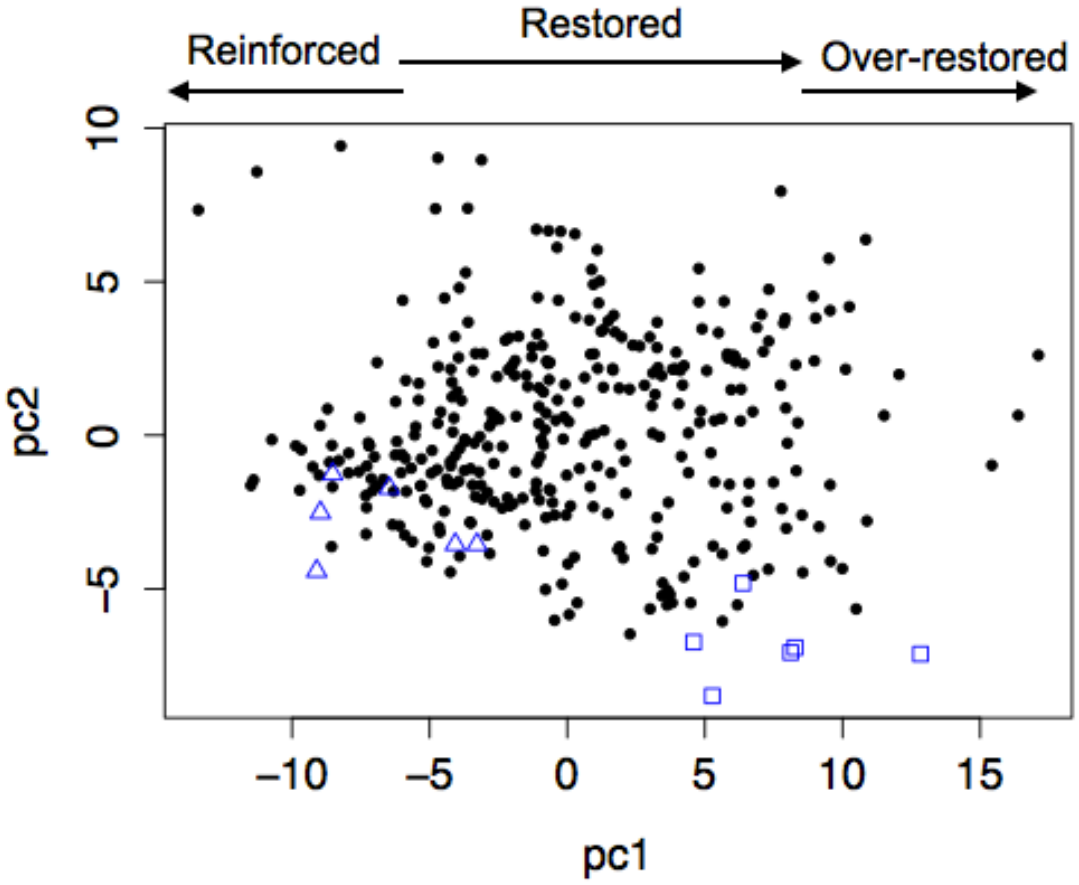

Figure 1.2. Plot of the first two principal components. The dots represent scores from the 115 evolved clones, each of which was replicated three times. The triangles represent the six replicates of the REL1206 ancestral strain at 42.2°C; squares denote the ancestor at 37.0°C. The arrows at the top of the plot illustrate directions of change relative to the two ancestral treatments (see Figure 1.1 and Table 1.1).

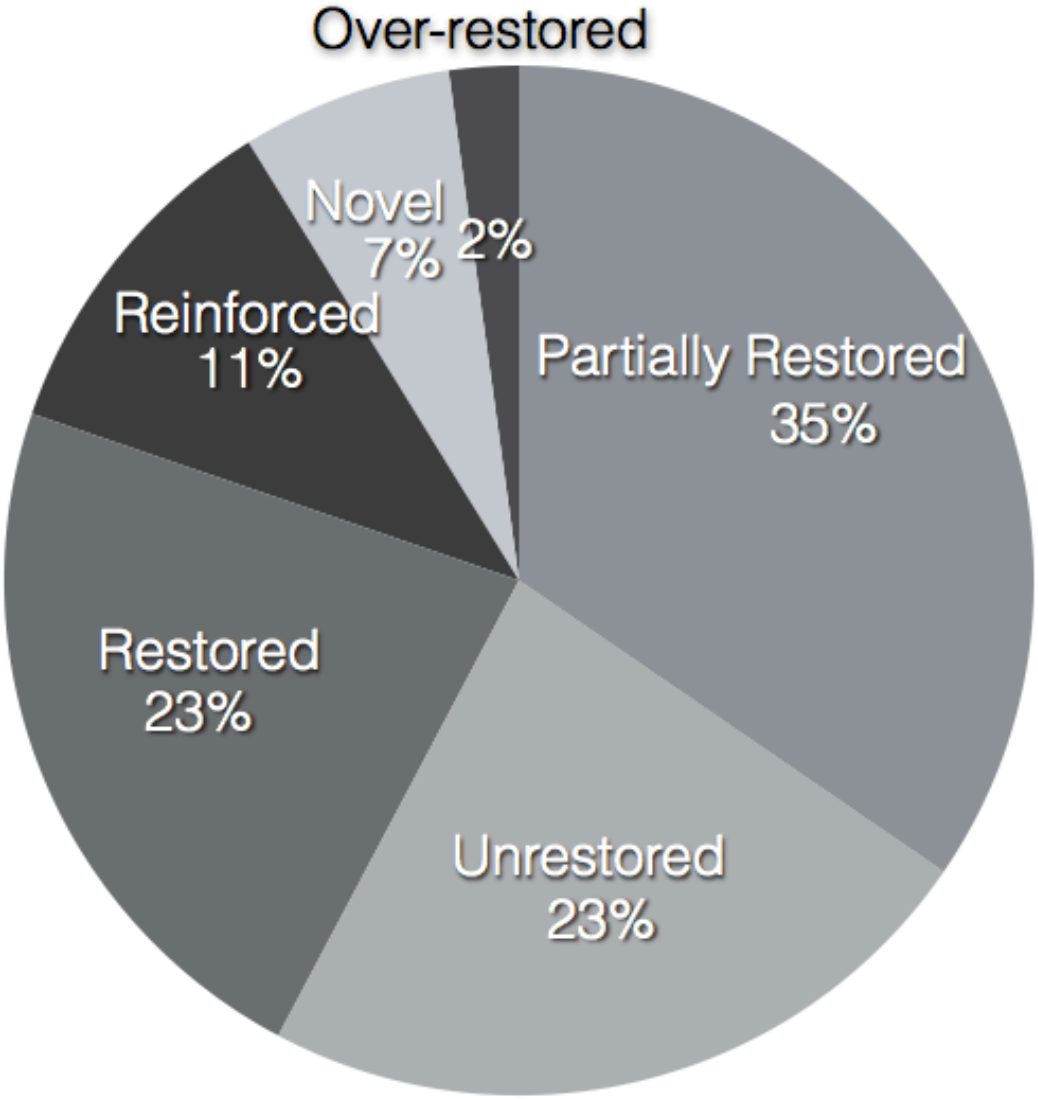

Figure 1.3. Pie chart reporting estimates of the proportion of phenotypic variation attributable to directions of adaptation.

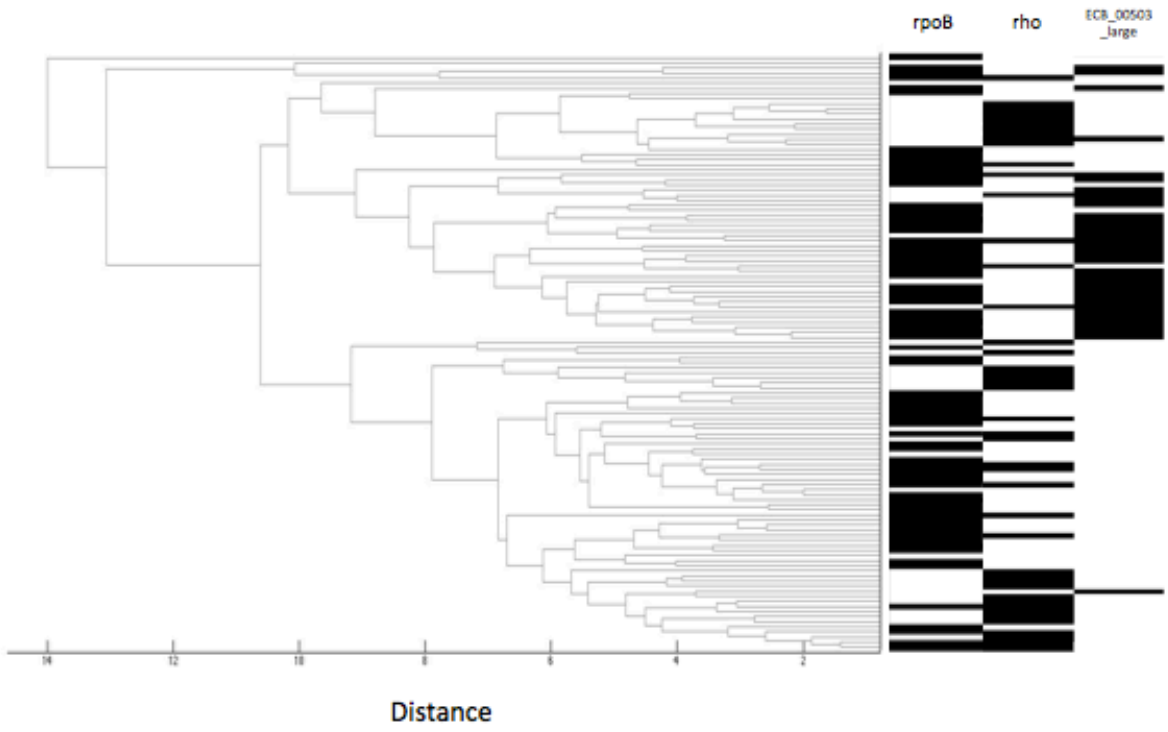

Figure 1.4. Hierarchical clustering of evolved lines by phenotypes. Dendrograms are labeled with the presence (black) or absence (white) of each of mutation in the *rho* gene, the *rpoB* gene and the large deletion (ECB_00503_large).

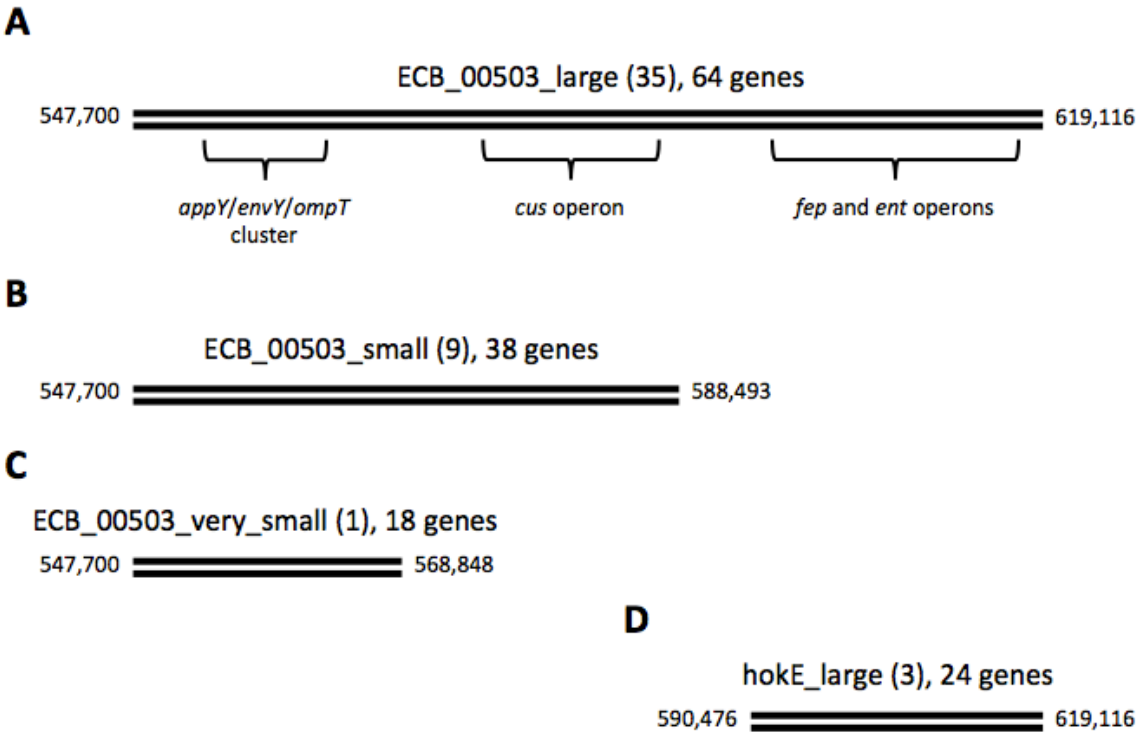

Figure 1.5. **Schematic of the four overlapping deletion events found during the thermal evolution experiments.** A) The largest deletion (ECB_00503_large) removes 64 genes and was found in 35 of 115 evolved clones. B) The smaller deletion (ECB_00503_small) was found in 9 of 115 clones and removed 38 genes. C) An even smaller variant was found in 1 of 115 clones and removed 18 genes. D) The hokE_large mutation removed 24 genes and was found in 3 of 115 clones. All variants are described in Tenaillon et al. (Tenaillon *et al.*, 2012). In all diagrams, numbers at the 5' and 3' end of the schematic represent the base position on the reference genome (Jeong *et al.*, 2009).

TABLES

Table 1.1: Categorizations of phenotypic magnitude and direction.

Category	Condition[1]	Number[2]
Partially Restored	$\bar{S}_{37°C} < \bar{S}_x < \bar{S}_{42°C}$ $\bar{S}_{37°C} > \bar{S}_x > \bar{S}_{42°C}$	151
Reinforced	$\bar{S}_x < \bar{S}_{42°C} < \bar{S}_{37°C}$ $\bar{S}_x > \bar{S}_{42°C} > \bar{S}_{37°C}$	77
Over-restored	$\bar{S}_x > \bar{S}_{37°C} > \bar{S}_{42°C}$ $\bar{S}_x < \bar{S}_{37°C} < \bar{S}_{42°C}$	24
Unrestored	$\bar{S}_{42°C} \cong \bar{S}_x < \bar{S}_{37°C}$ $\bar{S}_{42°C} \cong \bar{S}_x > \bar{S}_{37°C}$	230
Restored	$\bar{S}_{37°C} \cong \bar{S}_x < \bar{S}_{42°C}$ $\bar{S}_{37°C} \cong \bar{S}_x > \bar{S}_{42°C}$	124
Novel	$\bar{S}_{42°C} \cong \bar{S}_{37°C} > \bar{S}_x$ $\bar{S}_x < \bar{S}_{42°C} \cong \bar{S}_{37°C}$	61
Uninformative	$\bar{S}_{42°C} \cong \bar{S}_x \cong \bar{S}_{37°C}$	330
Inconsistent	All Remaining Relationships[3]	38

[1] Throughout the table, the symbol '≅' reflects a comparison between two $\bar{S}$ values that are similar enough that they do not differ statistically by *t*-test; however, '>' and '<' refer to significantly different values.

[2] Of 1,035 total comparisons (115 clones × 9 principal components)

[3] Mostly resulting from non-transitive pairwise significance

Table 1.2: Significant ($q < 0.01$) associations between genetic and phenotypic variation.

Principal Component	Associated Mutational Objects (Number of Affected Clones)[1]
1	ECB_00503_large (35), ESCRE1901 (13), trkD (3), dctA (5), yhjK (5), folM (2), yccE (2), metB (2)
2	ECB_00503_large (35), hokE_large (3), rpoB (76), rssB (10), yraL (2), rho (45), pta (5), fbp (7), cls (56), oppD_large (3), rpsR (5), ilvL (30), iclR (37), sdaC (2), kpsD (13), nth (3)
3	*(none)*
4	mrdA (23), ECB_00503_large (35), rpoB (76), fbaB (26), gatY (26), yhdJ (2), rho (45), lon (4), gmr (2), rnb (2), ygjF (6), cysH (2), fis (5), rpoH (2), ompR (6), pta (5), kpsE (11), metA (4), ECB_00503_small (9)
5	dnaG (15), ompF (10), rpoD (36), trkD (3), atoC (2), asnS (8), hokE_large (3), dctA (5), yhjK (5), ESCRE1902 (3), yebO (3), nusA (12), IS150_insJ (4), ECB_00503_large (35), yghD (2), yccE (2), mrdA (23), cls (56)
6	cls (56), rpoB (76), rho (45), glpF (13), mreB (12), rpsA (3), ECB_00503_large (35), IS3_insF (3), iclR (37), kpsM (7), pta (5), ygjF (6), kpsE (11), ycbC (2), nusA (12)
7	fbp (7), ompF (10), asnS (8), glpF (13), cysB (6), dctA (5), yhjK (5), dnaG (15), cpsG_large (21), dusB (4), ECB_00503_large (35), dnaA (4), IS1_insB (2)
8	fbp (7), rpsR (5), mreB (12), rho (45), folM (2), rssB (10)
9	dctA (5), yhjK (5), rho (45), uspA (8), rssB (10), cls (56), gltP (7), nrfG (7), iclR (37), glpT (8), kpsD (13), yifB (25), rhoL (2), ilvL (30), ECB_01993 (2), ECB_01994 (2), IS186_insL (2), ycbC (2), rpoB (76), ompF (10), IS150_insK (2), glyS (2)

[1] Mutational objects are organized by *p*-value, in ascending order.

Table 1.3: Individual Biolog tests that contribute significantly to differences between clones that contain *rpoB* vs. *rho* mutations.

Biolog Test	***p*-value**	**Category**
4% NaCl	1.04E-25	Chemical Sensitivity
Sodium Butyrate	2.85E-22	Chemical Sensitivity
Lincomycin	1.61E-18	Chemical Sensitivity
Tetrazolium Blue	2.34E-14	Chemical Sensitivity
Nalidixic Acid	4.49E-10	Chemical Sensitivity
p-Hydroxy-Phenylacetic Acid	6.83E-10	Carboxylic Acids, Esters, Fatty Acids
Gelatin	1.53E-09	Amino Acids
D-Sorbitol	3.72E-08	Carbohydrates, Carbohydrate Derivatives
D-Salicin	9.24E-07	Carbohydrates, Carbohydrate Derivatives
Beta-Hydroxy-D,L-Butyric Acid	1.08E-06	Carboxylic Acids, Esters, Fatty Acids
Aztreonam	1.46E-06	Chemical Sensitivity
L-Arginine	1.57E-06	Amino Acids
D-Malic Acid	1.62E-06	Carboxylic Acids, Esters, Fatty Acids
Quinic Acid	1.90E-06	Carbohydrates, Carbohydrate Derivatives
L-Histidine	2.05E-06	Amino Acids
Inosine	3.17E-06	Carbohydrates, Carbohydrate Derivatives
L-Pyroglutamic Acid	5.92E-06	Amino Acids
N-Acetyl-Neuraminic Acid	6.21E-06	Carbohydrates, Carbohydrate Derivatives
pH 5	7.78E-06	Chemical Sensitivity
Tween 40	1.01E-05	Carboxylic Acids, Esters, Fatty Acids
D-Raffinose	2.70E-05	Carbohydrates, Carbohydrate Derivatives
Bromo-Succinic Acid	3.85E-05	Carboxylic Acids, Esters, Fatty Acids

SUPPORTING INFORMATION

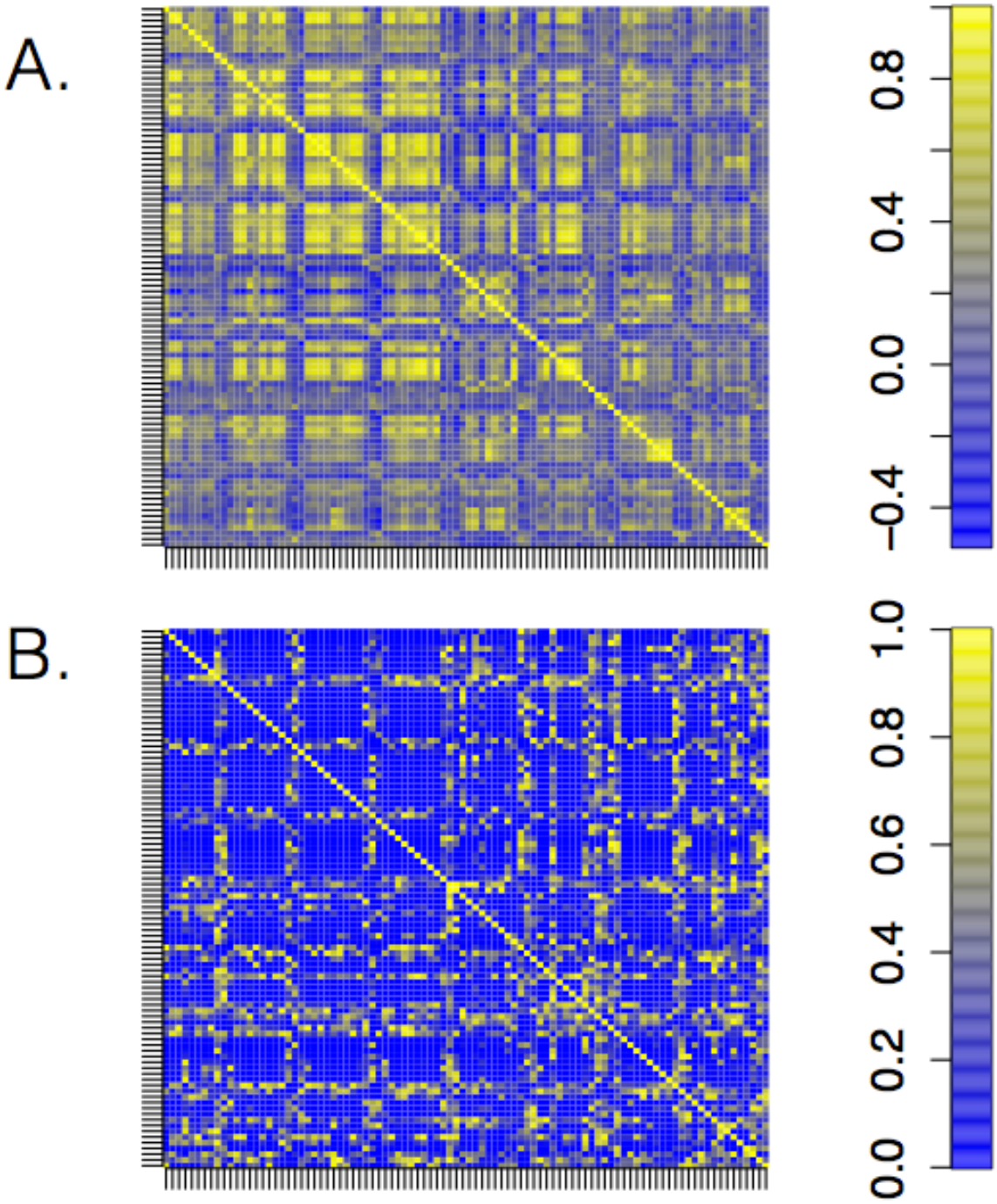

Figure S1.1. Correlations between tests on the Biolog plates. A) The Pearson correlation coefficient between individual tests across all measured clones. B) The corresponding *p*-value for correlation coefficients. For both graphs, the axes represent the 94 tests on the Biolog plates.

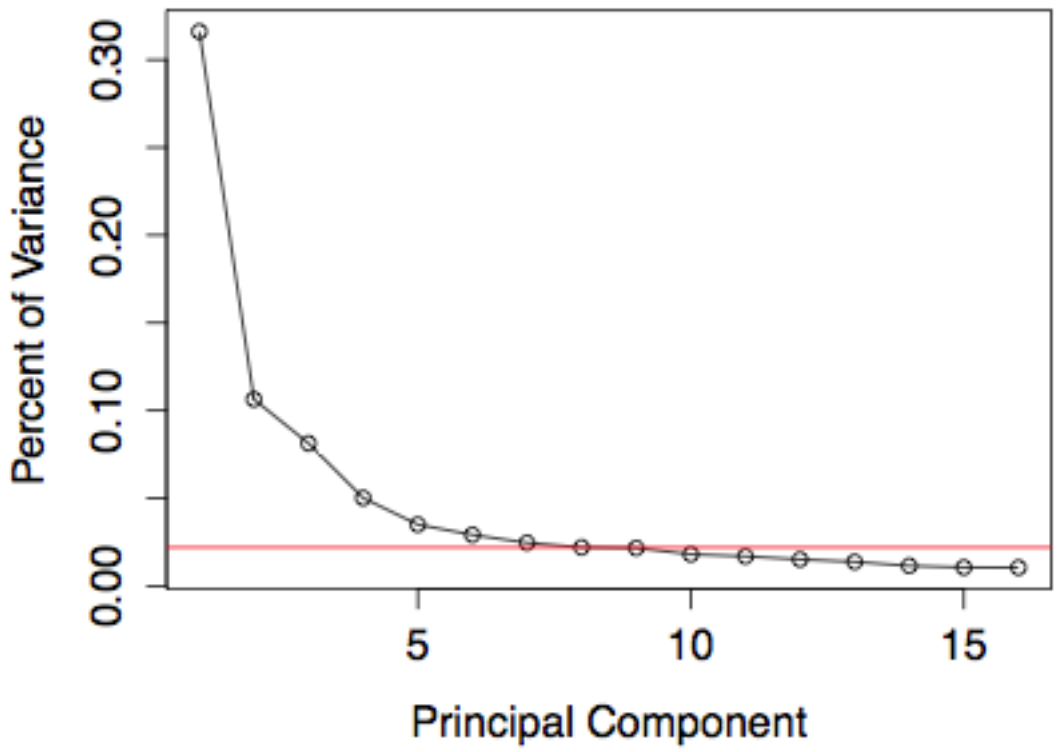

Figure S1.2. Scree plot of the percent of variation explained by each principal component.

The line corresponds to significance level, as determined by a bootstrapping heuristic.

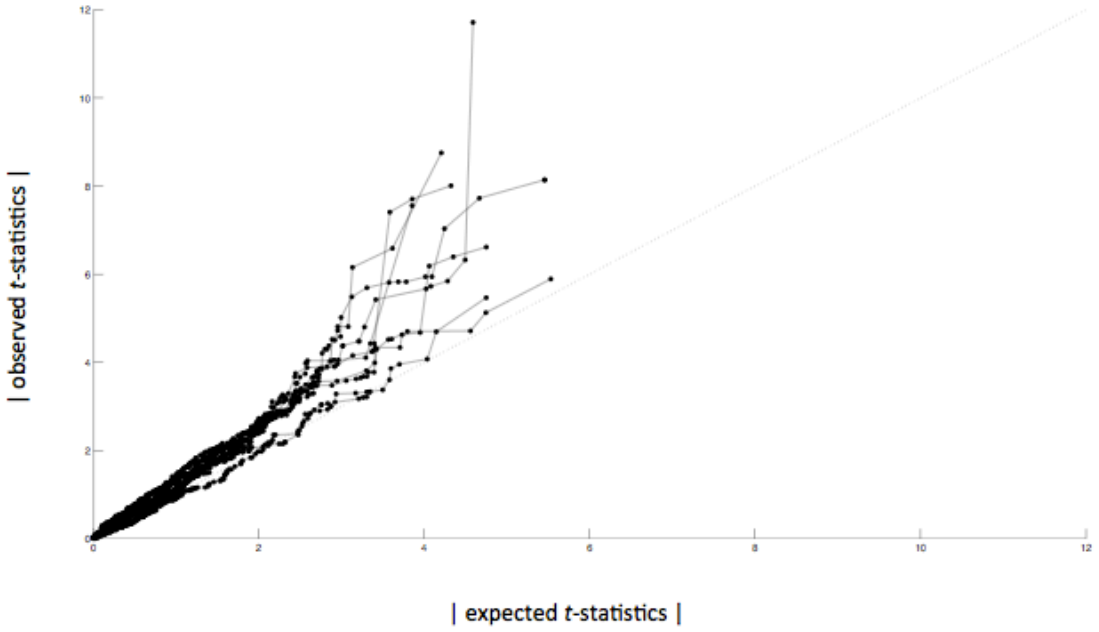

Figure S1.3. Q-Q plots for the results of association analyses. Each line represents one of the nine principal components. The diagonal represents the line for which there is no enrichment for tests with low *p*-values and therefore no evident biological signal.

Table S1.1. Loadings for each Biolog test in the first nine principal components (PCs).

Assays related to carbohydrates/carbohydrate derivatives, amino acids, carboxylic acids/esters/fatty acids, antibiotics, and other chemical sensitivities are colored yellow, red, dark blue, light blue, and purple, respectively.

Assay	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Dextrin	-0.11048 3403	-0.07993 2609	-0.06439 7521	-0.02129 7971	0.00210 3801	-0.00226 0122	0.20349 6153	-0.19020 3693	0.18330 7624
D-Maltose	-0.15071 1901	0.12299 1675	0.03726 9933	-0.02197 806	-0.05111 3167	0.00796 2457	0.02653 7019	0.05208 8649	-0.01200 4206
D-Trehalose	-0.14718 3126	0.10753 4064	0.06808 5927	-0.03104 235	-0.01536 4787	0.03172 0942	0.05650 47	0.08120 6026	-0.00654 9077
D-Cellobiose	-0.06648 6232	0.05005 1706	-0.08266 2719	0.04654 4123	-0.32833 5221	-0.05694 6204	0.03074 7373	-0.04707 2819	-0.06114 8995
Gentiobiose	-0.09288 6126	0.04378 5139	-0.01971 328	-0.03677 3307	-0.05451 254	-0.08205 8558	0.07074 2526	0.24753 9369	-0.11775 7342
Sucrose	-0.10437 5094	0.02909 2219	-0.06182 4642	0.07261 9671	-0.12389 6229	-0.06682 3298	0.16937 8676	0.04053 1292	0.01361 9188
D-Turanose	-0.08746 904	0.04481 6898	-0.03671 5941	0.05629 4191	-0.32006 7512	0.09186 6738	0.01570 7865	0.06940 5839	-0.03279 849
Stachyose	-0.11098 5114	0.05410 5569	-0.05058 4262	0.12296 1979	-0.15691 6418	-0.08203 5661	0.00489 5939	0.08408 6062	-0.02781 2489
pH 6	0.00765 9175	0.03888 2054	-0.24058 7386	-0.07507 9893	0.05484 5264	0.14339 9945	-0.17704 3587	-0.02514 7561	0.10934 6312
pH 5	0.00466 3512	0.09939 6298	-0.13199 3012	-0.08592 0827	-0.05700 1837	-0.04885 9386	-0.12443 5407	-0.17945 83	-0.07135 7177
D-Raffinose	-0.05905 0527	-0.02521 8256	-0.04881 5634	-0.15051 4153	-0.22639 0357	-0.15008 6677	-0.04764 3946	-0.11097 9995	-0.00502 5153
Alpha-D-Lactose	-0.15501 1323	0.09201 6001	0.03064 274	-0.04304 5848	-0.04478 1366	0.02291 5722	0.06485 0718	-0.04460 6319	-0.04826 2726
D-Melibiose	-0.16238 6374	0.10266 8404	0.03094 6696	-0.01677 2087	0.00029 4402	0.01316 2476	0.10174 1052	0.02083 7588	0.01216 0659
Beta-Methyl-	-0.09923	-0.15805	-0.00796	-0.19954	-0.12282	0.00296 7244	0.03670 9713	0.08593 9922	-0.04787

D-Glucoside	4145	3636	9701	5039	0152				5208
D-Salicin	-0.097989123	-0.061033444	-0.13724706	0.10530617	-0.043970507	-0.053007921	0.134149937	-0.109362317	-0.082216633
N-Acetyl-D-Glucosamine	-0.159053955	0.109801576	0.036631214	-0.063064518	0.017958338	-0.025073451	0.044197602	0.02034417	-0.041597641
N-Acetyl-Beta-D-Mannosamine	-0.116751429	-0.039255556	-0.023740678	-0.079946261	-0.198654003	0.083241036	0.038542577	-0.076382616	0.116564294
N-Acetyl-D-Galactosamine	-0.163070728	0.080387996	0.025259133	-0.062601364	0.037057037	-0.007333848	0.040026458	-0.002354993	-0.026980354
N-Acetyl Neuraminic Acid	-0.136081845	0.147196516	0.044522958	-0.073333233	-0.070778217	-0.040524823	-0.036445285	0.113916254	-0.019665573
1% NaCl	-0.016119706	0.101898183	-0.209033009	-0.117599238	-0.050481657	0.106507387	-0.119499318	0.046540377	0.058816772
4% NaCl	0.005572637	0.149777129	-0.066435564	-0.068955398	-0.056666507	-0.278301902	-0.082014324	0.090567908	0.23855689
8% NaCl	0.031069191	0.032579853	-0.148261618	0.000869661	0.035888728	0.021538268	0.077162835	-0.125177091	0.194457833
Alpha-D-Glucose	-0.151722118	0.1198142	0.054612159	0.002783161	-0.030343863	0.029746311	0.050556013	0.018818006	-0.067490872
D-Mannose	-0.156423597	0.116822182	0.021324684	-0.009409756	0.021339331	-0.016599695	0.104736724	-0.015502084	0.084271972
D-Fructose	-0.160791579	0.072797327	0.053416396	-0.03505431	0.015930028	0.027692383	0.064621523	0.036316146	-0.061244296
D-Galactose	-0.152917466	0.114836932	0.007589047	-0.021348865	0.023006652	-0.026461521	0.126011769	-0.036565167	0.022157821
3-Methyl Glucose	-0.112694606	-0.120028109	-0.045070185	-0.115051385	-0.058696925	-0.085832969	0.005613238	0.074814513	-0.195458934
D-Fucose	-0.131463001	-0.062986375	0.002878383	-0.141398605	0.092674591	0.005339505	0.049217835	0.037469257	-0.080355369
L-Fucose	-0.137390607	0.011950346	0.062481594	-0.023362496	-0.043410257	0.021746078	-0.178344735	-0.094299535	0.089565452

L-Rhamnose	-0.16289159	0.089594799	0.027837041	0.008231337	0.001769247	0.00500601	-0.001216912	-0.054111459	0.071477782
Inosine	-0.152104275	0.101564967	0.021597893	-0.016202283	0.094608753	-0.06795828	0.010363192	-0.047567667	0.048434695
1% Sodium Lactate	-0.036487279	0.147706656	-0.030138546	-0.073906751	-0.014327729	0.168710579	-0.217156435	0.239688505	0.018971193
Fusidic Acid	0.000991327	0.112530948	-0.219339015	-0.098324201	-0.065513707	0.060261919	0.178056453	0.101284881	0.06529367
D-Serine (Sensitivity)	0.042465258	0.11918858	-0.235921044	0.042093526	0.010403628	-0.051811061	0.014353785	0.029645496	0.090256016
D-Sorbitol	-0.123997692	0.122910642	0.048716396	-0.004394991	0.041359367	-0.039434974	0.017819748	0.077047131	0.061577525
D-Mannitol	-0.159756995	0.088852857	0.045796085	-0.022750531	0.022218509	0.04894584	0.085685993	0.02831566	-0.018788243
D-Arabitol	-0.127755707	-0.058407015	-0.097377157	0.045450302	-0.071593247	-0.055893701	0.006459632	0.108643681	-0.085955524
myo-Inositol	-0.153714484	-0.036909334	-0.047884947	-0.012833616	0.080373571	0.007370235	0.045303483	0.051303988	-0.100535615
Glycerol	-0.138171848	-0.004168348	-0.006121412	0.01978894	0.035318569	-0.018431022	0.032852311	-0.157036237	0.143549175
D-Glucose-6-Phosphate	-0.159821841	0.064275416	0.031082377	0.090047306	0.089365092	0.041459826	-0.016913337	-0.055055321	0.00151253
D-Fructose-6-Phosphate	-0.162118743	0.078644157	0.021835238	0.078786737	0.098762441	0.020989702	-0.021253611	-0.058347574	-0.018625663
D-Aspartic Acid	-0.115394887	-0.076773385	-0.013813691	-0.026815355	0.129564396	-0.08493986	0.076525859	0.066777254	-0.077921832
D-Serine (Metabolism)	-0.158298966	0.023581176	0.014959605	0.079964651	0.117080411	0.045005704	-0.083683498	-0.124638947	-0.010328087
Troleandomycin	0.019385892	0.065444398	-0.255899290	-0.048536628	0.099766626	0.151686988	-0.132513793	0.103705238	0.04150884
Rifamycin SV	-0.010389667	0.073686596	-0.133221344	-0.060574650	0.048598507	0.254896274	-0.198315779	0.197863641	0.067544678

Minocycline	0.043570132	-0.003239375	-0.237830462	0.027587814	0.082921196	0.171584148	0.056977965	-0.012402329	0.004390089
Gelatin	-0.061400718	-0.118262661	-0.087033548	0.237138705	0.039430209	0.036174821	-0.055000208	0.059794367	-0.084842338
Glycyl-L-Proline	-0.095121391	-0.151715332	-0.006693965	-0.110234376	-0.051109641	-0.000106312	-0.203270326	-0.069439341	-0.039959133
L-Alanine	-0.072323327	-0.229464648	-0.024707	-0.151789292	-0.019653974	-0.016794719	-0.096581187	0.018776724	-0.028009075
L-Arginine	-0.004601946	-0.163883056	-0.141468578	0.242679949	-0.117065366	-0.092124957	-0.082300409	0.02770616	0.008377801
L-Aspartic Acid	-0.081258598	-0.212387388	-0.028443909	-0.154803821	-0.005474261	0.091746091	0.003548809	0.007763161	0.133930654
L-Glutamic Acid	-0.100704127	-0.160170122	-0.008567273	-0.10863631	-0.073512523	0.069059409	-0.199114818	-0.083466025	0.100023485
L-Histidine	-0.101612112	-0.170986326	-0.053192171	0.011175326	0.135337419	-0.006930644	-0.02826521	0.107157486	-0.165628401
L-Pyroglutamic Acid	-0.030652784	-0.146835395	-0.137574951	0.196061634	-0.195268713	-0.128947764	-0.050384034	-0.003264839	0.007328994
L-Serine	-0.160356204	0.036628796	0.011009829	0.046055878	0.057611735	0.007870135	-0.151244779	-0.110801529	0.051421055
Lincomycin	0.014515896	0.10568252	-0.158005761	-0.1411768	0.136853408	-0.224300354	0.006694112	-0.015571755	0.025383422
Guanidine HCl	0.006485782	0.06747326	-0.276790186	-0.091214437	0.040060868	0.070137361	-0.018529314	0.128353304	0.154140979
Niaproof 4	0.026085875	-0.054321348	-0.188125016	0.027618125	0.086880605	0.084463191	-0.029310389	-0.109490413	-0.073645468
Pectin	-0.114632361	0.056961021	0.058924162	-0.025144213	-0.152869382	-0.011822711	-0.10753884	0.23196326	-0.058929486
D-Galacturonic Acid	-0.12026995	0.02381434	0.028373683	0.051683417	0.110979427	0.091411154	-0.071240229	-0.041053377	-0.067997495
L-Galactonic Acid Lactone	-0.038959674	-0.096509577	-0.059158023	0.093496551	0.249093952	-0.171539604	-0.136281347	0.082089674	-0.122430178
D-	-	0.01580	-	0.08746	0.19216	0.03989	0.01569	-	-

Gluconic Acid	0.1504769	6856	0.01224246	5528	5768	3146	2126	0.070256518	0.035760981
D-Glucuronic Acid	-0.15745313	0.063893629	0.025161892	0.092437206	0.127061259	0.024517971	-0.093547155	-0.09343818	0.033885842
Glucuronamide	-0.16331874	0.057284762	0.005771775	0.091479815	0.076860415	-0.00328344	-0.037686264	-0.082778465	0.043511181
Mucic Acid	-0.125770516	-0.033824528	0.010218218	0.000785491	-0.023660289	0.019438227	-0.197862449	-0.192417528	0.149676608
Quinic Acid	-0.045374814	-0.099845187	-0.136591944	0.24741578	-0.123319757	-0.099511282	-0.054218521	0.060574466	-0.058648886
D-Saccharic Acid	-0.045233472	-0.165220762	-0.023220976	-0.108116836	-0.049476738	-0.091765974	-0.113378581	-0.147059962	0.079731424
Vancomycin	-0.005388842	0.049487293	-0.148107292	-0.065648829	0.016089889	0.029162066	-0.028239061	-0.092367522	-0.310582513
Tetrazolium Violet	-0.021171058	0.063373896	-0.209467385	-0.090576414	0.034595543	0.176998256	0.046893306	-0.078333888	-0.146587283
Tetrazolium Blue	0.010210435	0.098978792	-0.145765378	-0.148520391	0.085178365	-0.287649448	-0.030373637	-0.036415498	-0.023069975
p-Hydroxy-Phenylacetic Acid	-0.028605801	-0.069618796	-0.015568506	0.201287303	-0.05120792	-0.00139609	-0.280660505	-0.08696074	0.029413182
Methyl Pyruvate	-0.101178866	-0.01272285	0.03614807	0.006690493	-0.077829208	0.048512835	-0.130421053	-0.075183889	-0.045649546
D-Lactic Acid Methyl Ester	-0.092734953	-0.007051542	-0.021752931	-0.180604863	-0.257021598	-0.060502682	-0.130936243	-0.111467679	0.008191788
L-Lactic Acid	-0.153921764	0.034858428	0.032076534	0.042104528	0.06885451	0.094485545	-0.192544993	-0.095993247	-0.002925699
Citric Acid	-0.117632486	-0.034168576	-0.065614346	0.175132145	-0.044130245	-0.046154555	0.020992181	0.125850065	-0.043097517
Alpha-Keto-Glutaric Acid	-0.095697568	-0.182666259	-0.035627117	-0.04274268	-0.007605396	0.050595262	0.064892173	0.136261364	0.098208574
D-Malic Acid	-0.074733031	-0.178034401	-0.065192355	0.015867975	0.088169784	0.088445034	0.122294006	0.08536924	0.209166304

L-Malic Acid	-0.078537132	-0.184211834	-0.033154016	-0.034059761	-0.026233869	0.079412993	0.128376917	0.02549308	0.300800079
Bromo-Succinic Acid	-0.110582483	-0.167925461	-0.033835135	-0.042722674	0.073959993	0.127549617	0.175236733	0.071621397	0.161801423
Nalidixic Acid	0.00672083	0.038541528	-0.222672322	0.003875932	-0.106692337	0.155705757	0.193156981	-0.173949428	-0.147836349
Lithium Chloride	-0.009765624	0.092015381	-0.117526337	-0.092387043	0.074975576	-0.169954816	-0.01692385	-0.085220043	-0.07955573
Potassium Tellurite	0.018881882	0.013654884	-0.229644151	0.093007795	0.105944598	-0.002544017	-0.010465776	0.035813926	-0.098002784
Tween 40	-0.093718219	-0.059674287	-0.078631575	0.214330645	-0.066371889	-0.024353977	0.003404443	0.105634834	0.049809214
Gamma-Amino-Butyric Acid	-0.104802743	0.019930752	-0.057649829	0.18956831	-0.021253849	-0.0329682	-0.096090065	0.107952808	-0.002277036
Alpha-Hydroxy-Butyric Acid	-0.083246373	-0.131688063	-0.070271786	0.015867539	-0.003689239	-0.098938493	0.167051418	-0.081661232	0.146718533
Beta-Hydroxy-D,L-Butyric Acid	-0.038342197	-0.027430448	-0.051954608	0.098471214	0.112030066	-0.27153053	0.000644185	0.212812167	0.20985256
Alpha-Keto-Butyric Acid	-0.068261598	0.025708515	-0.067757935	0.225277998	0.084671682	0.038583951	0.073003627	-0.081347416	0.017788787
Acetoacetic Acid	-0.084025384	-0.1897294	-0.030911204	-0.155466071	0.091821145	-0.035672261	0.06785792	0.037159227	-0.127667797
Propionic Acid	-0.055431446	-0.160698297	-0.052591188	-0.104111808	0.176245961	-0.183772329	0.000299206	0.033201008	-0.023019756
Acetic Acid	-0.060535182	-0.201173844	0.005649263	-0.171940579	0.057935186	0.003846871	-0.046426597	0.091826908	-0.157148681
Formic Acid	-0.124723875	-0.112134132	-0.028413545	-0.104745795	0.017118454	0.006542333	0.00606774	-0.017720188	-0.157566614
Aztreonam	0.014422589	0.038257278	-0.199534802	0.01272401	-0.048332946	-0.105350961	0.134803494	-0.278959594	-0.147822542

Sodium Butyrate	0.00409 2425	0.10220 6829	- 0.08589 2021	- 0.11485 2623	0.08310 8036	- 0.35002 2761	- 0.13357 7685	0.00191 0982	0.12005 8474
Sodium Bromate	- 0.01404 1743	0.08535 397	- 0.13403 455	- 0.02569 6714	- 0.08064 0025	0.03022 9106	- 0.00452 4054	0.16852 8408	0.01092 0265

Table S1.2. **The number and type of directional comparisons found in each principal component, along with the proportion of phenotypic variance (PV%) captured by each axis.**

Category[1]	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
	The full dataset of 115 clones								
Partial	49	8	0	0	18	18	31	0	31
Unrestored	6	52	0	0	76	53	32	1	20
Restored	28	4	0	0	0	3	6	51	31
Reinforced	0	44	0	0	20	5	7	0	3
Inconsistent	1	0	0	0	0	4	2	0	4
Uninformative	29	7	103	64	0	32	33	52	14
Novel	0	0	12	51	0	0	0	0	0
Over	2	0	0	0	1	0	4	11	12
PV%	0.3160	0.1064	0.0813	0.0502	0.03503	0.02917	0.02464	0.02201	0.02164
	The dataset of 67 clones without a deletion overlapping ECB_00503_large								
Partial	36	4	0	0	3	13	0	51	NS[1]
Unrestored	5	23	0	0	26	2	36	7	NS
Restored	11	1	0	0	2	26	0	0	NS
Reinforced	0	12	0	0	31	2	7	0	NS
Inconsistent	0	0	0	0	0	0	0	0	NS
Uninformative	14	27	53	10	5	12	20	1	NS
Novel	0	0	14	57	0	0	0	0	NS
Over-restored	1	0	0	0	0	12	4	8	NS
PV%	0.3486	0.08642	0.08228	0.04804	0.03593	0.03426	0.02907	0.02503	NS

[1] The directional categories are described Table 1.1 and Figure 1.1.

[2] NS = not significant. With the reduced data set, only the first 8 principal components had a significant eigenvector, and so the ninth component was not considered.

REFERENCES

Arias, C. A. et al., 2011 Genetic basis for in vivo daptomycin resistance in enterococci. New England Journal of Medicine **365**: 892-900.

Barrick, J. E., M. R. Kauth, C. C. Strelioff, and R. E. Lenski, 2010 Escherichia coli rpoB mutants have increased evolvability in proportion to their fitness defects. Molecular biology and evolution **27**: 1338-1347.

Bennett, A. F., and R. E. Lenski, 1993 Evolutionary adaptation to temperature II. Thermal niches of experimental lines of Escherichia coli. Evolution 1-12.

Bochner, B. R., P. Gadzinski, and E. Panomitros, 2001 Phenotype microarrays for high throughput phenotypic testing and assay of gene function. Genome research **11**: 1246-1255.

Carroll, S. M., and C. J. Marx, 2013 Evolution after introduction of a novel metabolic pathway consistently leads to restoration of wild-type physiology. PLoS genetics **9**: e1003427.

Escolar, L., J. Pérez-Martín, and V. De Lorenzo, 1999 Opening the iron box: transcriptional metalloregulation by the Fur protein. Journal of bacteriology **181**: 6223-6229.

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

Gunasekera, T. S., L. N. Csonka, and O. Paliy, 2008 Genome-wide transcriptional responses of Escherichia coli K-12 to continuous osmotic and heat stresses. Journal of bacteriology **190**: 3712-3720.

Handwerger, S., and A. Tomasz, 1985 Antibiotic tolerance among clinical isolates of bacteria. Reviews of infectious diseases 368-386.

Hollands, K., A. Sevostiyanova, and E. A. Groisman, 2014 Unusually long-lived pause required for regulation of a Rho-dependent transcription terminator. Proceedings of the National Academy of Sciences **111**: E1999-E2007.

Jeong, H. et al., 2009 Genome sequences of Escherichia coli B strains REL606 and BL21 (DE3). Journal of molecular biology **394**: 644-652.

Le Gac, M., J. Plucain, T. Hindré, R. E. Lenski, and D. Schneider, 2012 Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with Escherichia coli. Proceedings of the National Academy of Sciences **109**: 9487-9492.

Leiby, N., and C. J. Marx, 2014 Metabolic erosion primarily through mutation

accumulation, and not tradeoffs, drives limited evolution of substrate specificity in Escherichia coli. PLoS biology **12**: e1001789.

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

Lenski, R. E., and M. Travisano, 1994 Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. Proceedings of the National Academy of Sciences **91**: 6808-6814.

Loh, K. D. et al., 2006 A previously undescribed pathway for pyrimidine catabolism. Proceedings of the National Academy of Sciences of the United States of America **103**: 5114-5119.

Mchugh, J. P. et al., 2003 Global iron-dependent gene regulation in Escherichia coli: a new mechanism for iron homeostasis. Journal of Biological Chemistry

Nishijima, S., Y. Asami, N. Uetake, and S. Yamagoe, 1988 Disruption of the Escherichia coli cls gene responsible for cardiolipin synthesis. Journal of Bacteriology

Orr, H. A., 2005 The genetic theory of adaptation: a brief history. Nature Review Genetics **6**: 119-127.

Peres-Neto, P. R., D. A. Jackson, and K. M. Somers, 2003 Giving meaningful interpretation to ordination axes: assessing loading significance in principal component analysis. Ecology

Peres-Neto, P. R., D. A. Jackson, and K. M. Somers, 2005 How many principal components? Stopping rules for determining the number of non-trivial axes revisited. Computational Statistics & Data …

Peters, J. M. et al., 2009 Rho directs widespread termination of intragenic and stable RNA transcription. Proceedings of the National Academy of Sciences **106**: 15406-15411.

Pigliucci, M., 2008 Is evolvability evolvable? Nature Reviews Genetics **9**: 75-82.

Plata, G., C. S. Henry, and D. Vitkup, 2015 Long-term phenotypic evolution of bacteria. Nature **517**: 369-372.

Pommerenke, C. et al., 2010 Global genotype-phenotype correlations in Pseudomonas aeruginosa. PLoS pathogens **6**: e1001074.

Richter, K., M. Haslbeck, and J. Buchner, 2010 The heat shock response: life on the verge of death. Molecular cell **40**: 253-266.

Rodriguez-Verdugo, A., D. Carrillo-Cisneros, A. Gonzalez-Gonzalez, B. S. Gaut, and A.

F. Bennett, 2014 Different tradeoffs result from alternate genetic adaptations to a common environment. Proc Natl Acad Sci U S A

Sabarly, V. et al., 2011 The decoupling between genetic structure and metabolic phenotypes in Escherichia coli leads to continuous phenotypic diversity. Journal of evolutionary biology **24**: 1559-1571.

Sandberg, T. E. et al., 2014 Evolution of Escherichia coli to 42 degrees C and Subsequent Genetic Engineering Reveals Adaptive Mechanisms and Novel Mutations. Mol Biol Evol

Sneath, P. H. A., and R. R. Sokal, 1973 *Numerical taxonomy. The principles and practice of numerical classification.*

Team, R. C., 2014 R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2012. Open access available at: http://cran.r-project.org

Tenaillon, O. et al., 2012 The molecular diversity of adaptive convergence. Science **335**: 457-461.

Woods, R. J. et al., 2011 Second-order selection for evolvability in a large Escherichia coli population. Science **331**: 1433-1436.

Yura, T., H. Nagai, and H. Mori, 1993 Regulation of the heat-shock response in bacteria. Annual Reviews in Microbiology **47**: 321-350.

Zhou, L., X.-H. Lei, B. R. Bochner, and B. L. Wanner, 2003 Phenotype microarray analysis of Escherichia coli K-12 mutants with deletions of all two-component systems. Journal of bacteriology **185**: 4956-4972.

CHAPTER 2 - Antagonistic pleiotropy and the compensatory landscapes of distinct adaptive trajectories

ABSTRACT

Adaptation rarely produces perfect solutions. Instead, it proceeds through a series of mutational steps that often produce pleiotropic side effects. Hence, a crucial component of adaptation is compensating for the antagonistic effects of preceding mutations. To investigate how different degrees of antagonistic pleiotropy shape the dynamics of adapting populations, we performed temporal sequencing of eight populations of *Escherichia coli* at 11 time points throughout 2,000 generations of adaptation to high temperature (42.2°C). The eight populations represent two different pathways to high-temperature adaptation typified by mutations in *rpoB*, which encodes a subunit of the RNA polymerase complex, and *rho*, which encodes a major transcriptional terminator. Previous work has shown that *rpoB* and *rho* differ in their pleiotropic effects, with single mutations in *rpoB* causing more gene expression shifts than single mutations in *rho*. We therefore predicted that *rpoB* populations should accumulate more mutations than *rho* populations due to a greater capacity for compensatory evolution. We also predict that the greater number of mutations increase the amount of clonal interference in *rpoB* populations relative to *rho* populations. By reconstructing and quantifying the mutational trajectories of our populations, we found support for both predictions: *rpoB* populations indeed accumulate more mutations and experience greater clonal interference than their *rho* counterparts.

INTRODUCTION

Laboratory evolution experiments provide a powerful way to test both new and longstanding questions about how organisms evolve. In such studies, the use of high-throughput DNA sequencing and model microbial organisms, such as the bacterium *Escherichia coli* or the yeast *Saccharomyces cerevisiae*, facilitates a deeper and biologically meaningful understanding of the genomic changes underpinning adaptation to new environments. Perhaps the most exciting prospect of experimental microbial evolution is the ability to freeze populations at intermediate stages of adaptation, creating a sort of "fossil record" (Lenski and Travisano, 1994). These evolutionary time points can later be revived for a variety of purposes—from sequencing to fitness assays to replaying evolution—and studies of saved evolutionary time points have illuminated the evolutionary dynamics of microbial systems. For instance, they have captured the emergence and maintenance of ecological interactions (Le Gac *et al.*, 2012; Herron and Doebeli, 2013), clonal interference among beneficial mutations (Kao and Sherlock, 2008; Lang *et al.*, 2013; Frenkel *et al.*, 2014), diminishing returns epistasis (Chou *et al.*, 2011; Khan *et al.*, 2011; Kryazhimskiy *et al.*, 2014), the dynamic maintenance of different phenotypes (Traverse *et al.*, 2013), and long-term changes in population fitness over time (Wiser *et al.*, 2013). They have even hinted at the predictability and reproducibility of adaptation (Blount *et al.*, 2008; Lang *et al.*, 2013; Levy *et al.*, 2015).

Here we utilize evolutionary time points to study the potential effects of compensatory evolution on the dynamics of adaptation. Our study relies on a previous experiment by Tenaillon et al. (2012) in which 115 populations of *E. coli* were evolved at high temperature for 2,000 generations (Tenaillon *et al.*, 2012). At the end of the experiment, a single clone was sequenced from each population to describe the genetic outcome of evolution in independent populations. Intermediate time points were saved from each of the 115 populations at generations 100, 200,

and every 200 generations thereafter.

A particularly important facet of the experiment of Tenaillon et al. (2012) was the identification of two adaptive pathways to high-temperature adaptation. These pathways were typified by mutations in *rpoB* and *rho*, which encode an RNA polymerase subunit and a transcriptional terminator, respectively (Tenaillon *et al.*, 2012). Statistically, mutations in *rpoB* and *rho* were less likely to occur together in a clone than expected. This pattern suggests negative epistasis between the two genes, possibly due to their related functional impacts on transcription (Tenaillon *et al.*, 2012). Further work has shown that clones possessing certain *rpoB* and *rho* mutations differ in both fitness tradeoffs at low temperatures (Rodríguez-Verdugo *et al.*, 2014) and phenotypic signatures (Hug and Gaut, 2015).

Although they may both directly impact the process of transcription, mutations in *rpoB* and *rho* are likely to differ in their pleiotropic side effects, too. In theory, *rpoB* mutations have the potential to affect the expression of every gene in the genome directly, because it encodes part of the core RNA polymerase complex. By contrast, the Rho protein influences the transcriptional termination of between 200 and 2,000 genes (Peters *et al.*, 2009; Hollands *et al.*, 2014). Hence, mutations in *rho* are potentially pleiotropic, but perhaps not to the same extent as mutations in *rpoB*.

Previous work has addressed the extent of pleiotropic effects by engineering single mutations into the *rpoB* and *rho* genes and then measuring their effects on fitness and gene expression. The work has examined four single mutants in *rpoB* (I572F, I572L, I572N, I966S) and four mutations in *rho* (I15F, I15N, A43T, T231A), all of which arose during the thermal stress experiment. Analyses have shown that the fitness benefits of single *rpoB* and *rho* mutations vary. All four *rpoB* mutations are beneficial at 42.2°C, with an average fitness advantage of ~20% relative to the REL1206 ancestor (Rodríguez-Verdugo *et al.*, 2014).

However, *rho* mutations have a less straightforward effect on fitness, with two of the four (A43T and T231A) conferring ~25% fitness benefits at 42.2°C, and two others (I15F and I15N) being effectively neutral (Rodríguez-Verdugo *et al.*, 2014; González-González *et al*., in prep.). This last fact is surprising, because I15N was observed in 15 out of 115 populations of the thermal stress experiment and is thus unambiguously adaptive by the criterion of parallelism.

Just as single *rpoB* and *rho* mutants differ in fitness, they also confer different effects on gene expression (GE). GE in *rpoB* and *rho* mutants has been compared to GE in REL1206 at 42.2°C, demonstrating that *rpoB* and *rho* mutations affect expression in a common set of 957 genes—23% of the genome. While they affect a similar core group of genes, they also have differential effects. For example, *rho* mutations affect expression in a set of 183 genes that are unaffected by *rpoB* mutants. Similarly, 769 genes are differentially expressed only in the *rpoB* mutants. These differences in the number of differentially expressed genes (183 vs. 769) suggest that *rpoB* mutations cause a greater number of pleiotropic side effects (Rodríguez-Verdugo *et al.*, 2016; González-González *et al*., in prep.).

Overall, fitness and GE data suggest that the benefits of single *rpoB* and *rho* mutations generally outweigh potentially deleterious pleiotropic effects on GE. Nonetheless, the greater potential for pleiotropic side effects in populations that traverse an *rpoB*-mediated adaptive pathway may expose a larger space of possible compensatory mutations—i.e. genetic changes that are adaptive because they correct deleterious side effects of earlier mutations. Indeed, previous work has found evidence for compensation through additional changes to the *rpoB* gene itself (Tenaillon *et al.*, 2012; Rodríguez-Verdugo *et al.*, 2014), and to other genes, such as the *flg* operon that regulates flagellar motility (Rodríguez-Verdugo *et al.*, 2016). The *flg* operon exhibits the kind of GE dynamics that one might expect from compensatory evolution. The *flg* operon is downregulated in the REL1206 ancestor under heat stress. Single mutations in *rpoB*

restore its expression to pre-stress levels, but this expression is likely an energetically costly change given that flagellar motility may be unnecessary under the conditions of the experiment. Following 2,000 generations of high-temperature evolution, however, the *flg* operon is once again downregulated in some evolved clones due to presumably compensatory mutations. This pattern of GE change over time suggests both that re-expression of the *flg* operon is a costly side effect of a beneficial mutation in *rpoB* and that subsequent evolution compensates for antagonistic pleiotropy by again downregulating *flg* expression via other mutational events (Rodríguez-Verdugo *et al.*, 2016).

Based on these observations, we hypothesize that *rpoB* mutations, while strongly beneficial overall, cause more antagonistic side effects. As a result, we predict that populations that contain *rpoB* mutations have access to a larger pool of compensatory mutations. If this is true, it leads to two major predictions. First, if mutations in *rpoB* and *rho* are among the first steps a population takes during adaptation, *rpoB* populations should accumulate more mutations than *rho* populations over the course of adaptation, because they have access to more beneficial compensatory genetic changes (Figure 2.1). Second, *rpoB* populations should experience more clonal interference than *rho* populations due to competition among a greater number of beneficial compensatory mutations (Figure 2.1). Here we test these predictions by sequencing temporal samples from the evolution experiment of Tenaillon et al. (2012) (Tenaillon *et al.*, 2012), with the larger goal of elucidating features of the relationship between pleiotropy and compensatory evolution.

MATERIALS AND METHODS

Evolution Experiment

Full details of the thermal stress experiment are reported elsewhere (Tenaillon *et al.*,

2012), but for clarity, we describe it here. Following previously established methods (Bennett and Lenski, 1993), a loop was used to inoculate 10 mL of Luria-Bertani medium (LB) with a freezer stock of *E. coli* B possessing a neutral Ara- marker (REL1206), and the culture was grown overnight at 37.0°C. A 0.1 mL aliquot of this culture was plated onto an LB plate and grown at 37.0°C, and individual colonies from this plate were used to inoculate each of 115 independent culture tubes containing 10 mL LB. These cultures were grown overnight at 37.0°C, transferred in a 1:100 dilution to 9.9 ml of Davis minimal medium supplemented with 25 mg/L glucose (DM25), and grown overnight again at 37.0°C. Thereafter, each culture was transferred daily into fresh media via a 1:100 dilution and maintained in a shaking water bath at 42.2°C for 2,000 generations after the application of thermal stress. Throughout the experiment, population samples of each of the 115 lines were frozen at regular intervals (generations 100, 200, and every 200 generations thereafter, up to 2,000). At the end of the 2,000-generation experiment, one clone was isolated from each of the 115 populations for genome sequencing, and these clones were assessed for their fitness relative to the ancestor at 42.2°C (Tenaillon *et al.*, 2012). Note that the ancestral REL1206 clone had been propagated previously at 37.0°C for 2,000 generations in DM25 and was thus well adapted to the growth medium (DM25), making temperature (42.2°C) the major stress of the experiment. For additional information, including relative fitness values and a table of genotypes, please refer to Tenaillon et al. (2012) (Tenaillon *et al.*, 2012).

Populations, DNA Extraction and Sequencing

We identified eight populations for sequencing based on the genotype of their evolved clones. We chose four populations with mutations in *rpoB* and four with mutations in *rho*. In addition, the eight populations were selected to include a range of specific mutations (i.e. *rpoB*

I966S, *rho* 15N). The line numbers referenced in this work correspond to Tenaillon et al. (2012) (Tenaillon *et al.*, 2012). We chose to study Lines 3, 9, 16, 26, 33, 56, 60, and 82.

For each of our eight populations, we inoculated each frozen time point population into four separate culture tubes containing 10 mL DM25 and incubated them in an Innova 3100 water bath shaker (New Brunswick Scientific) overnight at 42.2°C and 120 RPM. Cells from all four tubes were pooled, and genomic DNA was extracted from these samples using Wizard Genomic DNA Purification Kits (Promega). Genomic DNA libraries were prepared using the TruSeq DNA PCR-Free HT Library Preparation Kit (Illumina). The 88 (8 lines × 11 time points) samples were multiplexed and sequenced on one lane of an Illumina HiSeq 3000 at the Bioinformatics Core Facility at the UC Davis Genome Center. All sequence data have been submitted to the NCBI Sequence Read Archive (Accession number: PRJNA339569).

We called mutations and their frequencies using breseq (Deatherage and Barrick, 2014) in polymorphism mode. The *E. coli* B REL606 genome was used as a reference. Six regions (*topA*, *spoT*, *glmU*/*atpC*, *pykF*, *yeiB*, and the *rbs* operon) harbor mutations that differ between REL606 and the REL1206 strain used to initiate our evolution experiment (Barrick *et al.*, 2009; Tenaillon *et al.*, 2012), so these were excluded from our analyses. Plotting mutation frequencies over the 11 time points revealed a few obvious anomalies: certain mutations appeared to be fixed or extinct for several time points, only to go extinct or become fixed later. We manually corrected the frequencies of these anomalous points using pileup information for these mutations in samtools 1.3 with default nucleotide quality score cutoffs (Li *et al.*, 2009). Frequencies were estimated by dividing number of mutant nucleotide calls at that position by the total number of nucleotide calls at that position. Additionally, multiple mutations called within the same region in very close proximity (<10 nucleotides apart) and following identical trajectories were manually collapsed into single mutations.

Larger structural mutations (i.e. duplications and deletions) are reported by breseq using novel junction information. However, the precise nature of the junction sequences (e.g. whether they occur cleanly at both ends or in regions of repetitive sequence), together with the difficulty of determining these events in mixed populations from sequencing coverage, makes it difficult to estimate meaningful mutational frequencies over time. Therefore, these events were excluded from our construction of mutational trajectories. We did, however, include these events in our assessment of mutational numbers. We did so by calling large duplications and deletions in each population by comparing unique reads (mapping quality >5 in samtools 1.3) across 10 kb windows of the genome at generation 2,000, when they would most likely have reached appreciable frequencies discernable from sequencing coverage alone. We defined regions with more than 1.5× average genome coverage across two or more windows as duplications, and less than 0.5× average genome coverage across two or more windows as deletions.

Mutations, Mutational Cohorts, and Mutational Parameters

Many mutations were called at very low—and thus unreliable—frequencies, and it was necessary to focus only on those mutations with clear signals of population persistence. Mutations were considered only if they were called at a nonzero frequency for more than a single time point, and if at least two time points exceeded frequencies of 25%. This cutoff was chosen to account for the pooling of four independent cultures for DNA extraction. We reasoned that a new mutation could, in theory, arise and reach appreciable frequency in a single culture during growth, but would be less likely to do so independently in multiple cultures, providing an empirical frequency cutoff of 1/4 = 25%. Note that his 25% cutoff is not the only check we have against chance artifacts, because these are also mitigated by the fact that we rely on inferences from consecutive time points.

Sets of mutations with similar temporal trajectories were grouped into cohorts using established methods (Lang *et al.*, 2013). Briefly, each mutation was represented by a vector of its frequencies at each of the 11 time points sampled. These vectors were clustered hierarchically using the unweighted pair group method with arithmetic mean (UPGMA) (Sneath and Sokal, 1973), and the resulting hierarchy for each population was grouped at a cutoff of 0.5.

For each mutation, we determined four parameters used in previous studies (Lang *et al.*, 2011; Rodríguez-Verdugo *et al.*, 2013): s_{up}, τ_{up}, f_{max}, and τ_{max} (Figure S2.1). Among these four parameters, s_{up} reflects the initial rate of increase in frequency of a mutation; τ_{up} is the generation at which a mutation achieves a frequency of 1% in the population, providing an estimate of when a mutation first appeared; f_{max} is the maximum frequency achieved by a given mutation; and τ_{max} is the generation at which f_{max} was initially achieved. We determined s_{up} by calculating the slope of the line passing through the first point at which a mutation was observed at nonzero frequency and the previous point. τ_{up} is the generation at which a mutation was at a frequency of 0.01 in the population, calculated using its particular value of s_{up}. f_{max} is the maximum frequency achieved by a given mutation. τ_{max} is the generation at which a mutation initially hits its maximum frequency.

RESULTS

rpoB and *rho* Mutations are Beneficial and Arise Early

Our differential compensation hypothesis presupposes that *rpoB* and *rho* mutations arise early in the course of adaptation. To test this presumption, we turned to population sequencing, which revealed a striking degree of parallelism in the timing and rapid rise of *rpoB* and *rho* mutations (Figure 2.2, Figure S2.2). In all four *rpoB* populations, *rpoB* mutations were the first to appear and to rise to appreciable frequency, either singly or as a cohort with other mutations.

Fixation of *rpoB* mutations occurred within < 600 generations for all four populations.

The same pattern of early, rapid sweeping held true for *rho* mutations in two of the four *rho* populations. For these two (Lines 26 and 60), their *rho* mutations fixed in < 600 generations. The other two populations were exceptions, however. One exception was Line 33, in which the defining *rho* mutation appeared approximately halfway through the population's history (τ_{up} = 804 generations), but the mutation still swept rapidly to high frequency once present. It is worth noting that Line 33 was included in our experiment specifically for its lack of an accompanying *cls* mutation, which may play a crucial epistatic role in the *rho* adaptive pathway (Tenaillon *et al.*, 2012) (See Discussion).

The fourth *rho* population (Line 82) harbored two subpopulations—one *rho* and one *rpoB*, with the *rpoB* subpopulation being numerically dominant. Because it did not definitively contain either a fixed *rpoB* or *rho* mutation, we excluded this population from most analyses, but it still yielded some interesting qualitative insights. In this mixed population, an *rpoB* mutation rose rapidly to high frequency initially, only to crash quickly thereafter as a *rho* mutation appeared and swept through the population to near fixation (Figure 2.2). Thereafter the two subpopulations coexisted until the end of the evolution experiment, although at generation 2,000 the *rho* subpopulation appears to be nearing extinction.

To assess the adaptive benefit of mutations in *rpoB* and *rho* compared to other genes, we determined four parameters from the mutational frequency trajectories of our populations: s_{up}, τ_{up}, f_{max}, and τ_{max} (Table S2.1, Figure S2.1). All four parameters were significantly different between the first *rpoB* and *rho* mutations to fix in each population and all other mutations that occurred. Using one-sided Mann-Whitney U tests, s_{up} was significantly greater (p = 0.0030), τ_{up} significantly lower (p = 0.0035), f_{max} significantly higher (p = 0.0017), and τ_{max} significantly earlier (p = 0.0186) for the first fixed *rpoB* and *rho* mutations compared to all others. Put

another way, *rpoB* and *rho* mutations tended to appear early (τ_{up}), rise in frequency rapidly (s_{up}), and achieve their highest frequencies (f_{max}) quickly (τ_{max}), supporting their status as adaptive mutations (Figure 2.2, Figure S2.2). That this is true even with a population (Line 60) containing a *rho* I15N mutation, which was previously shown to be neutral by itself at 42.2°C (Rodríguez-Verdugo *et al.*, 2014; González-González *et al*., in prep.), implying that this mutation might not be adaptive alone but as a partner of a mutation in *cls* (See Discussion).

rpoB Populations Acquire More Mutations than *rho* Populations

Our differential compensation hypothesis makes two predictions. First, because *rpoB* mutations are more pleiotropic than *rho* mutations (at least in their effects on GE), it predicts that *rpoB* populations have access to a greater number of compensatory mutations. Second, as an extension of this larger compensatory mutation space, clonal interference may be more severe in *rpoB* populations than *rho* populations, because more beneficial mutations are available to compete.

We tested the prediction that *rpoB* populations have more mutations in two ways. First, we examined the genomic data from clones at generation 2,000. Comparing all clones that possess *rpoB* mutations only ($n = 60$) to those possessing *rho* mutations only ($n = 29$), we found that *rpoB* clones contained an average of 13.9 mutations, whereas *rho* clones contained an average of 11.5 mutations, a significant difference (one-sided Mann-Whitney U test: $p = 0.0016$) (Figure 2.4). Second, we examined the temporal sequencing data. Populations traversing an *rpoB* adaptive pathway had nearly significantly more mutational trajectories (one-sided Mann-Whitney U test: $p = 0.0572$), as well as nearly significantly more mutational cohorts (one-sided Mann-Whitney U test: $p = 0.0857$), than *rho* populations. On average, *rpoB* populations contained 12.75 mutational trajectories and 6.5 mutational cohorts, whereas *rho* populations

contained only 8.33 mutational trajectories and 5.3 mutational cohorts. Lastly, while we cannot make firm conclusions from the dynamics of a single population, Line 82 also supports the hypothesis that *rpoB* trajectories have greater numbers of mutations on average; by generation 2,000, the *rpoB* subpopulation possessed eight mutations, whereas the *rho* subpopulation possessed only five (Figure 2.2).

Correlations Among Mutational Parameters in *rpoB* vs. *rho* Populations

The four mutational parameters (s_{up}, τ_{up}, f_{max}, and τ_{max}) provide a quantitative way of describing the adaptive dynamics of our populations and a means to compare those dynamics between *rpoB* and *rho* populations. Overall, we found that mutational parameters were highly correlated in both groups (Figure 2.3, Table 2.1, Table S2.1). For example, we found that s_{up} and τ_{up} were both significantly correlated with τ_{max}, and that s_{up} was significantly correlated with f_{max} (Table 2.1). Put another way, in both *rpoB* and *rho* populations, mutations that were more beneficial (s_{up}) were more likely to hit their highest frequencies (f_{max}) earlier (τ_{max}). In addition, earlier mutations (τ_{up}) were more likely to hit their maximum frequencies earlier (τ_{max}). Note, however, f_{max} was not significantly correlated with τ_{max} in either group.

Although the patterns of correlation were similar between *rpoB* and *rho* populations, there were notable differences. s_{up} and τ_{up}, as well as τ_{up} and f_{max}, were significantly correlated with one another only in *rpoB* populations, but not in *rho* populations (Table 2.1). A test of homogeneity of correlation coefficients between s_{up} and τ_{up} in *rpoB* vs. *rho* populations was borderline significant after Fisher's z-transformation (Fisher, 1950) (one-sided Z-test; p = 0.0791).

Because we analyzed more *rpoB* populations than *rho* populations (four vs. three), and because *rpoB* populations generally contained more mutations than *rho* populations (Figure 2.4),

we wanted to be sure that differences in the significance of correlations were due to the type of population (*rpoB* vs. *rho*), rather than sampling effects. To address the problem of uneven sampling, we treated the data in two ways. First, we systematically removed each of the four *rpoB* populations from correlation calculations, giving a sample size of three—equivalent to that of the *rho* populations. In all four comparisons, s_{up} remained significantly correlated with τ_{up}, which is a relationship not observed in the *rho* populations. Second, to correct for the fact that *rpoB* populations possess more mutations, we combined all of the mutations from all four *rpoB* populations and randomly subsampled to 23 mutations, which was equivalent to the total number of observed mutations in all of our *rho* populations. We performed this subsampling randomly 100 times and determined pairwise correlations and their *p*-values among all of our mutational parameters. The *p*-values and correlations of the *rho* populations were at the extreme end of the values obtained for the subsampled *rpoB* populations for two of the six pairwise parameter comparisons (Figure 2.5), again supporting the idea that population type (*rpoB* vs. *rho*), not sampling, was responsible for the observed differences in correlations.

rpoB Populations Experience More Clonal Interference than *rho* Populations

If the compensatory mutation space of *rpoB* populations is larger than that of *rho* populations, we expect them to experience more clonal interference as more compensatory mutations arise and compete with one another. Clonal interference is known to slow the rate of adaptation and reverse mutational trajectories, cutting them short of sweeping fully to high frequency (Lang *et al.*, 2011). We found evidence for this limitation of adaptation in f_{max}, the maximum frequency achieved by any given mutation. When we compared f_{max} between all mutations in *rpoB* and *rho* populations (excluding the first *rpoB* and *rho* mutations to fix), we found that mutations in *rpoB* populations reached significantly lower maximum frequencies than

those in *rho* populations (one-sided Mann-Whitney U test: $p = 0.0036$). On average, mutations in *rpoB* populations attained a maximum frequency (mean ± standard error) of 66 ± 3.9%, whereas mutations in *rho* populations reached 80 ± 5.5%.

Additionally, when beneficial mutations are placed in competition with one another during clonal interference, their effective benefits are reduced, which should be reflected in their s_{up} values. Excluding the first *rpoB* and *rho* mutations to fix, we found a difference in s_{up} between *rpoB* and *rho* populations, with mutations in *rpoB* populations having lower average values of s_{up}. This difference was not statistically significant (one-sided Mann-Whitney U test: $p = 0.1741$), but it was in the direction we predicted. Mutations in *rpoB* populations had s_{up} values (mean ± standard error) of 0.00166 ± 0.00029, whereas mutations in rho populations had sup values of 0.00192 ± 0.00040.

DISCUSSION

Microbial evolution experiments have revealed that a major outcome of adaptive evolution is the restoration of gene expression and phenotypes from a stress state to an ancestral, pre-stress state (Fong *et al.*, 2005; Carroll and Marx, 2013; Sandberg *et al.*, 2014; Hug and Gaut, 2015; Rodríguez-Verdugo *et al.*, 2016; González-González *et al*., in prep.). This process, however, almost certainly entails pleiotropic costs. Adapting populations therefore face two obstacles: the primary external stress of a changing environment, and the secondary internal stress imposed by less-than-perfect genetic solutions. How populations evolve to compensate is an open question (Fisher, 1930; Poon and Chao, 2005; Costanzo *et al.*, 2010; Szamecz *et al.*, 2014; Rodríguez-Verdugo *et al.*, 2016) and one that is important for understanding adaptive evolution as a whole.

The evolution experiment of Tenaillon et al. (2012) (Tenaillon *et al.*, 2012) provides a

unique opportunity to test questions of antagonistic pleiotropy and compensation because adaptation occurred primarily by two independent genetic pathways: *rpoB* and *rho*. The conditions of the evolution experiment were identical until either an *rpoB* or *rho* mutation canalized the adaptive trajectories of the populations in which they arose. At that point, adaptation took place upon differential compensatory landscapes. Based on the effects of *rpoB* and *rho* mutations on GE, we have predicted that there are more options for compensatory evolution available to *rpoB* populations than to *rho* populations, and that two expectations follow from this. First, *rpoB* populations should accumulate more mutations during adaptation than *rho* populations. Second, the greater number of mutations in *rpoB* populations should result in more clonal interference during adaptation than in *rho* populations.

rpoB Populations Accumulate More Mutations During Adaptation than *rho* Populations

Temporal sequencing reveals that populations traversing either the *rpoB* or *rho* adaptive pathway share one important dynamic in common: a mutation in either gene often occurs early in the course of adaptation and rises rapidly to high frequency. This trend can be seen in the mutational trajectories of six of the seven populations we examined, and even in the mixed population (Line 82) that contains separate *rpoB* and *rho* subpopulations (Figure 2.2, Figure S2.2). The parameters s_{up} and τ_{up} for these mutations also confirm that *rpoB* and *rho* tend to arise earlier and increase in frequency faster than other mutations. Depending on which of these two genes mutates and fixes first, the remainder of the adaptive process may be canalized, owing to the different pleiotropic effects of *rpoB* and *rho* mutations (Rodríguez-Verdugo *et al.*, 2016; González-González *et al*., in prep.). The subsequent compensatory landscapes of the *rpoB* and *rho* populations differ, with the greater pleiotropy of *rpoB* mutations creating a wider variety of potentially compensatory mutations for the population to explore. This statement is supported by

the greater number of mutations that accumulated in *rpoB* clones than in *rho* clones in the original evolution experiment (Tenaillon *et al.*, 2012), and also by the greater number of both individual mutational trajectories and mutational cohorts in *rpoB* vs. *rho* populations (Figure 2.4). As a result, *rpoB* populations accumulate more mutations on average than *rho* populations.

If *rpoB* and *rho* differ in the amount by which they move a population towards a new fitness optimum, then Fisher's geometric model (Fisher, 1930) predicts that the number of additional beneficial mutations available to each adaptive pathway should differ (e.g. *rho* mutations move populations closer to a fitness optimum than *rpoB* populations, so there are fewer possible beneficial mutations for them to accumulate). While this model provides a possible explanation for the difference in the number of mutations accumulated in each time of population, fitness and gene expression results from previous studies argue against it, for two reasons.

First, the fitness benefits conferred by single engineered mutations in *rpoB* and *rho* are variable. Some *rpoB* single mutations (I572F, I572L, I572N) confer fitness advantages of ~17% over REL1206, whereas another (I966S) offers an advantage of ~37%, each with a 95% confidence interval of ~10% (Rodríguez-Verdugo *et al.*, 2014). Similarly, some mutations in *rho* (A43T, T231A) confer advantages of ~24-27%, whereas others (I15F, I15N) are neutral, with a 95% confidence interval of ~6-11% (González-González *et al*., in prep.). Given this variation, our knowledge to date suggests that *rpoB* mutations confer higher maximum benefits. Neither of these situations fit with the greater mutation accumulation we observed in *rpoB* populations, although we must admit a major caveat that we do not know the fitness effects of joint *cls-rho* mutations on a single haplotype. The combined effects of *rho* mutations with *cls* knockouts requires further study.

Second, single mutations in *rpoB* restore gene expression to an unstressed ancestral state

to a greater degree than *rho* mutations (González-González *et al*., in prep.). Fisher's model would again predict that *rho* populations should accumulate more mutations because they possess more genes whose expression can be restored, yet we observe the opposite trend. Therefore, Fisher's model alone cannot account for the differences we see. We conclude that differential antagonistic pleiotropy between these two genetic pathways is a more likely explanation.

Evidence for More Clonal Interference in *rpoB* Compared to *rho* Populations

We have predicted that the higher number of compensatory mutations in an *rpoB* background leads to more clonal competition. This prediction is consistent with our observations that *rpoB* populations have lower average f_{max} values, as well as suggestively lower average value of s_{up}. Furthermore, *rpoB* and *rho* populations appear to differ in the strength of correlations between s_{up} and τ_{up}, even after correcting for sampling differences (Figure 2.3, Figure 2.5). This is consistent with the idea that s_{up} values are lower in *rpoB* populations later in the adaptive process due to the presence of competing beneficial mutations. Although some of the contrasts between *rpoB* and *rho* populations are only borderline significant (e.g. the test of the different correlation coefficients between and s_{up} and τ_{up}), the multiplicity of tests are convincing when combined using Fisher's method (Fisher, 1950). For example, when we combine three putatively independent tests (p-values from the contrasts in f_{max}, the average number of mutations within clones, and the correlation coefficient bases on s_{up} and τ_{up}), the resulting p-value suggests that *rpoB* and *rho* populations are truly distinct in their behavior ($p = 5.57 \times 10^{-5}$). Despite their similar overall fitness values at the end of the thermal stress experiment (Tenaillon *et al.*, 2012), populations traversing either of these two adaptive pathways have their own unique evolutionary histories that shape their genomes. Mutations in either *rpoB*

or *rho* serve as bellwethers for each population's adaptive trajectory.

It remains a mystery as to why *rpoB* and *rho*, two genes with hundreds to thousands of genetic interactions, are such strong and consistent targets of adaptive evolution (Long *et al.*, 2015). In *S. cerevisiae*, it has been observed that genes with more genetic interactions tend to affect more cellular functions (and are hence more pleiotropic), and they also tend to have greater fitness defects when mutated (Costanzo *et al.*, 2010). As others have noted, however, important genes may be mutated when populations are far from their fitness optima, despite the negative side effects (Fisher, 1930; Poon and Chao, 2005; Szamecz *et al.*, 2014). Even though they are highly connected, we know that *rpoB* and *rho* mutations are beneficial overall from competition experiments at 42.2°C (Rodríguez-Verdugo *et al.*, 2014; González-González *et al*., in prep.), and our present study shows that they arise early on in evolution. Interestingly, Costanzo et al. (2010) found that genes with many interactions may be more evolvable if the proteins they produce are more disordered (Costanzo *et al.*, 2010), and it has recently been suggested that the *rho*-encoded termination factor may have disordered, prion-like domains that facilitate adaptation to new environments (Pallarès *et al.*, 2015). Therefore, molecular structure, pleiotropy, and distance from optimal fitness in our system may intersect in the *rpoB* and *rho* genes to produce beneficial mutations in these genetic hubs.

Epistasis with *cls* Complicates Our Understanding of the *rho* Pathway

Although we have spoken of a *rho* adaptive pathway here and previously (Tenaillon *et al.*, 2012), new data continue to reveal that *rho* mutations actually have a complex epistatic relationship with mutations in *cls*, which encodes a cardiolipin synthase. Among the 29 *rho*-only clones in our evolution experiment, all except one (Line 33, one of the *rho* populations in this study) possess mutations in *cls*, and among the 45 total clones possessing *rho* mutations, only six

lack *cls* mutations.

This association is too widespread to be a chance occurrence, and two of our three *rho* populations (Line 26, Line 60), along with our mixed population (Line 82), provide evidence that *rho* and *cls* sweep through populations together (Figure 2.2, Figure S2.2). In the original sequenced clones, an IS insertion is called in *cls* in Line 60. Although it is difficult to estimate the frequency of this mutation over time, the sequence data do indicate the presence a novel junction in the *cls* region by as early as generation 100, when *rho* is nearly fixed.

Other work using single mutants of *rho* has shown that some *rho* mutations (i.e. I15F, I15N) confer no fitness benefit, hinting that *rho* might interact epistatically with *cls* to produce a fitness advantage (González-González *et al*., in prep.). Other *rho* mutants (i.e. A43T, T231A), however, do confer a demonstrable fitness benefit on their own, yet they still occur with *cls* in evolved clones (González-González *et al*., in prep.). Thus, the *rho*-*cls* relationship may be more complex than a simple synergistic benefit, and its precise nature still eludes us.

Mechanistically, it is unclear how a transcriptional terminator and a cardiolipin synthase might interact. The *cls* gene is not *rho*-terminated, and gene expression analyses in four different engineered *rho* mutants (I15F, I15N, A43T, T231A) indicate that *cls* is not differentially expressed in three of the four (González-González *et al*., in prep.), discounting a regulatory connection. One possibility is that *cls* mutations alter membrane permeability in a way that positively supports the gene expression changes created by *rho* mutations, or vice versa. Further work using double mutants will be necessary to determine the fitness relationship between *rho* and *cls*.

Clonal Interference vs. Diminishing Returns Epistasis

Unfortunately, clonal interference and diminishing returns epistasis are confounding

phenomena in our system, making it difficult to discern whether one, the other, or both contribute to our observed differences between *rpoB* and *rho* populations. Both phenomena have been observed in microbial evolution experiments, and they each have the potential to decelerate the rate of adaptation (Kao and Sherlock, 2008; Chou *et al.*, 2011; Khan *et al.*, 2011; Lang *et al.*, 2013; Wiser *et al.*, 2013; Frenkel *et al.*, 2014; Kryazhimskiy *et al.*, 2014).

Previous studies have tested diminishing returns epistasis in bacteria and yeast that have evolved under a variety of conditions (Chou *et al.*, 2011; Khan *et al.*, 2011; Wiser *et al.*, 2013; Kryazhimskiy *et al.*, 2014). The common conclusions of these experiments are that diminishing returns epistasis is a predictable feature of evolution that depends only upon the overall fitness of an organism, and that mutations arising later in a population's history confer progressively smaller adaptive benefits (Chou *et al.*, 2011; Khan *et al.*, 2011; Wiser *et al.*, 2013; Kryazhimskiy *et al.*, 2014). There is no reason for us to expect that either *rpoB* or *rho* populations to be exceptions to the ubiquity of this phenomenon, especially because some studies of diminishing returns epistasis utilized strains of *E. coli* similar to ours (Khan *et al.*, 2011; Wiser *et al.*, 2013). Furthermore, we observe a deceleration of adaptation in both *rpoB* and *rho* populations in the form of negative correlations between s_{up} and τ_{up}: mutations arising later in a population's history initially increase in frequency more slowly (Figure 2.3). Although this correlation is not significant in the *rho* populations, it is nonetheless in the direction predicted by diminishing returns epistasis. Given that diminishing returns epistasis may be general, we prefer to interpret differences in the rate of adaptation in *rpoB* vs. *rho* populations as stemming from differences in clonal interference.

We have shown that *rpoB* and *rho* populations differ in their adaptive dynamics, and whether this is due to differences in clonal interference, diminishing returns epistasis, or both, the pattern is striking and raises important questions about how pleiotropy and compensation

impact the course of evolution. If we assume that our results are unrelated to clonal interference effects, and that they are due entirely to diminishing returns epistasis, it would suggest that different compensatory landscapes impact the strength and timing of later fitness increases, with some compensatory landscapes generating beneficial mutations that do not diminish in strength in increasingly fit genetic backgrounds. Something akin to this pattern may be achievable if the severity of deleterious side effects differs between *rpoB* and *rho* populations (e.g. *rho* populations must compensate for particularly severe side effects compared to *rpoB* populations, increasing the fitness benefit of compensatory mutations). Indeed, our differential compensation hypothesis makes no predictions about differences in the severity of negative pleiotropic side effects. However, this should still not completely erase the signal of diminishing returns epistasis, because it depends only on global fitness, not the identity of each mutation (Kryazhimskiy *et al.*, 2014).

Given the ubiquity of diminishing returns epistasis in other microbial evolution experiments (Chou *et al.*, 2011; Khan *et al.*, 2011; Wiser *et al.*, 2013; Kryazhimskiy *et al.*, 2014) and our finding that *rpoB* and *rho* populations both show negative correlations between s_{up} and τ_{up} (despite this relationship not being significant in *rho* populations), we find differential diminishing returns epistasis to be a less plausible explanation than differential clonal interference.

FIGURES

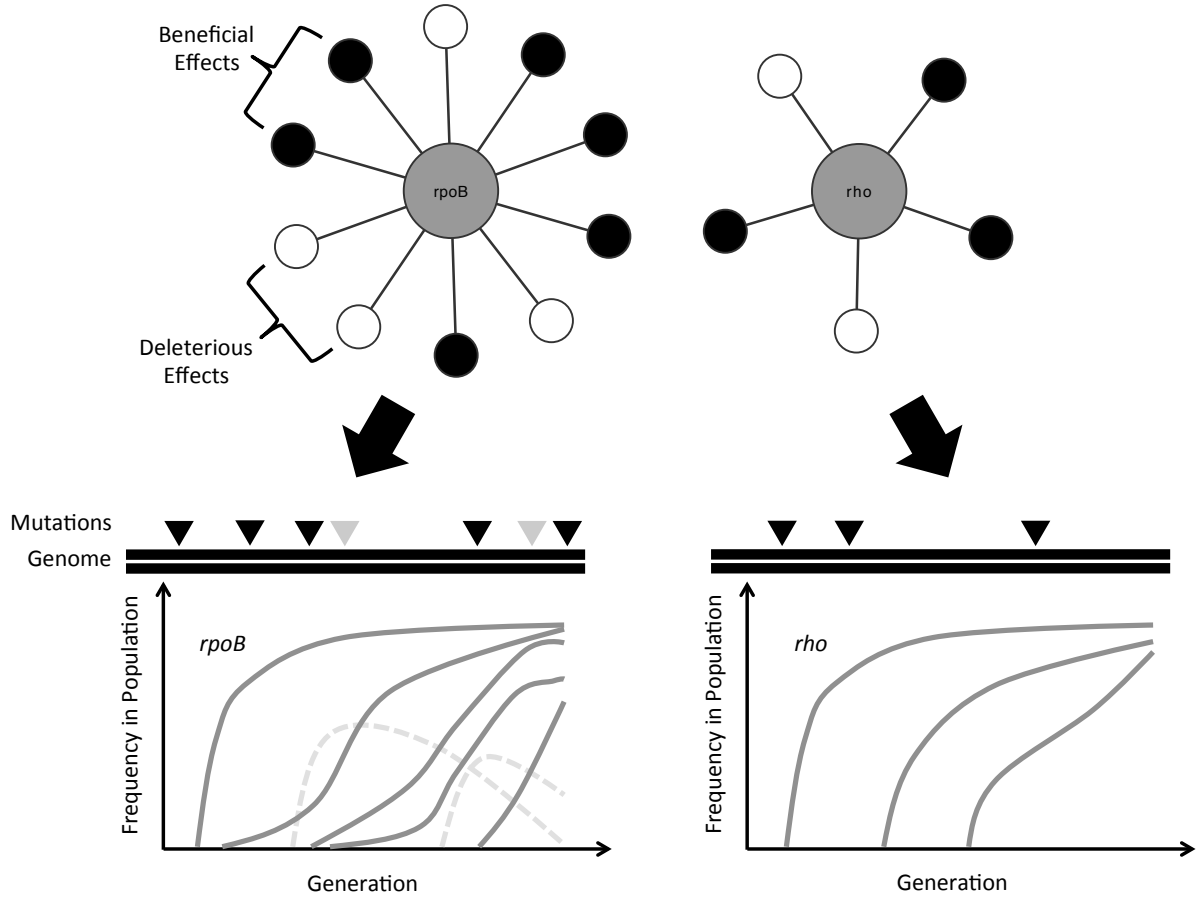

Figure 2.1: Pleiotropy of *rpoB* and *rho* and the differential compensation hypothesis. Mutations in *rpoB* and *rho* impact different numbers of targets (e.g. genes, pathways, phenotypes), and are thus differentially pleiotropic. Each target (outer circles) may be affected in a way that improves (black) or lowers (white) overall fitness. Both *rpoB* and *rho* confer similar net fitness benefits (more black circles than white circles) at 42.2°C, but differ in how many deleterious side effects they may cause (four white circles vs. two white circles). This creates two different spaces of compensatory mutations for populations traversing either the *rpoB* or *rho* adaptive pathway. According to our differential compensation hypothesis, populations that mutate *rpoB* are predicted to accumulate more mutations overall (triangles) and

to experience more clonal interference among these mutations (grey triangles, dashed trajectories). Conversely, populations that mutate *rho* are predicted to accrue fewer compensatory mutations and to experience less clonal interference among them.

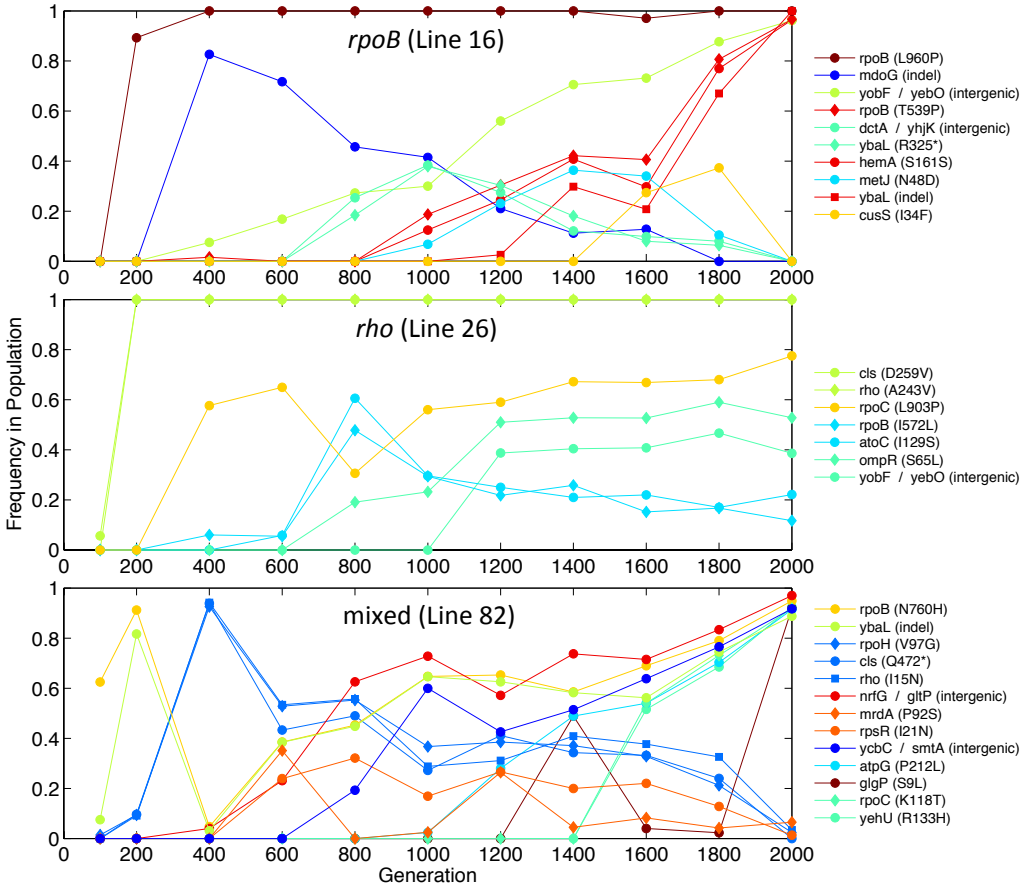

Figure 2.2: Mutational trajectories for three of eight populations, representing an *rpoB*, *rho*, and mixed population. Trajectories are colored by mutational cohort. From top to bottom, these populations correspond to Line 16, Line 26, and Line 82, respectively. Additional populations can be found in Figure S2.2.

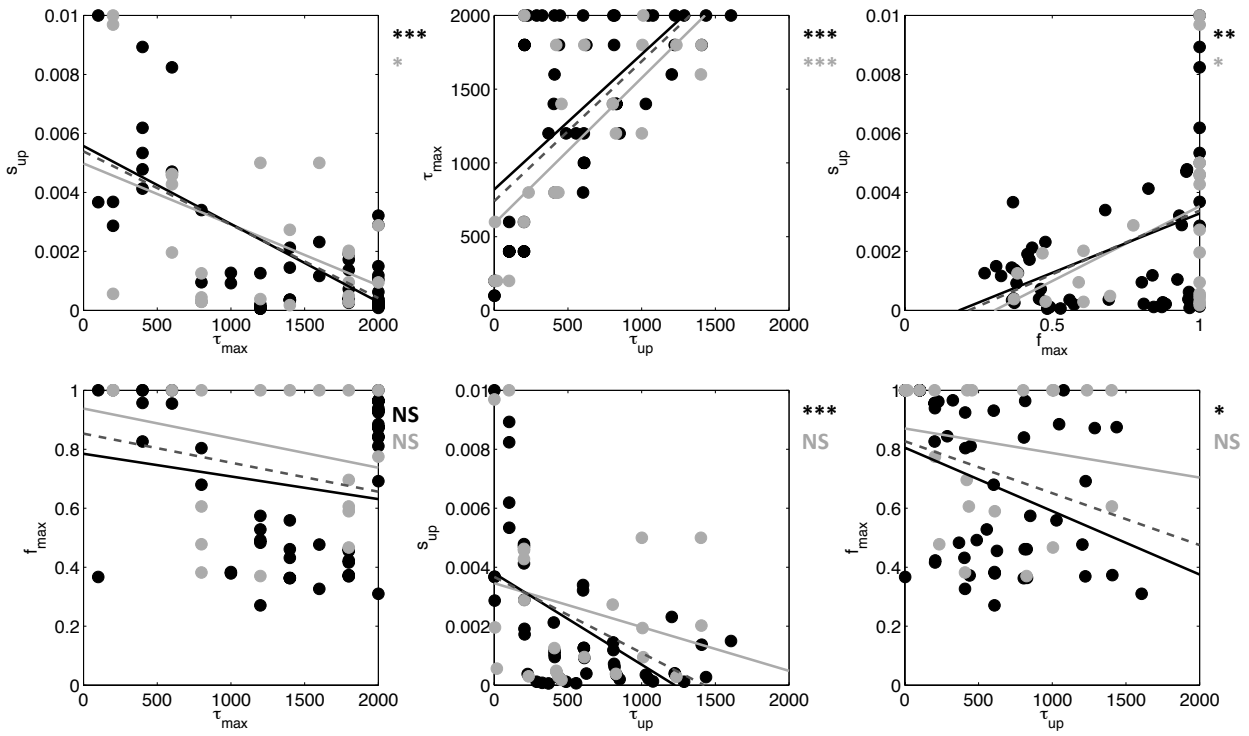

Figure 2.3: Correlations among mutational parameters s_{up}, τ_{up}, f_{max}, and τ_{max}. Values for *rpoB* populations are shown in black, and values for *rho* populations are shown in light grey. Regression lines are shown for these two types of populations are shown in their respective colors. The dark grey dashed line represents the best fit for all seven populations. Significance of the regression for *rpoB* and *rho* populations is shown to the top-right of each graph in their respective colors. NS, *, **, and *** mean not significant, $p < 0.05$, $p < 0.005$, and $p < 0.0005$, respectively.

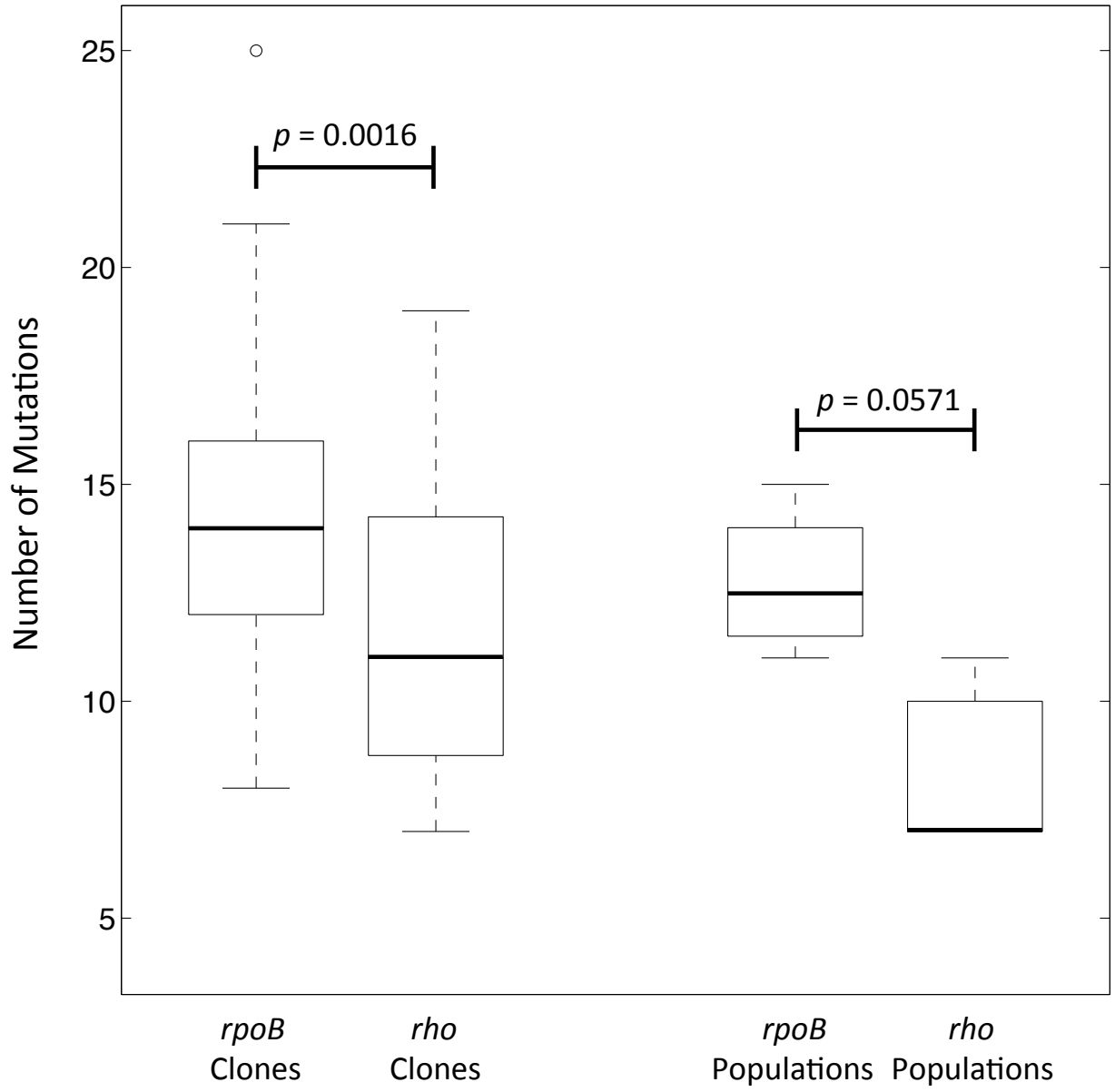

Figure 2.4: Box plot of number of mutations found in *rpoB* and *rho* clones and populations. Significance of one-sided Mann-Whitney U tests between groups is indicated.

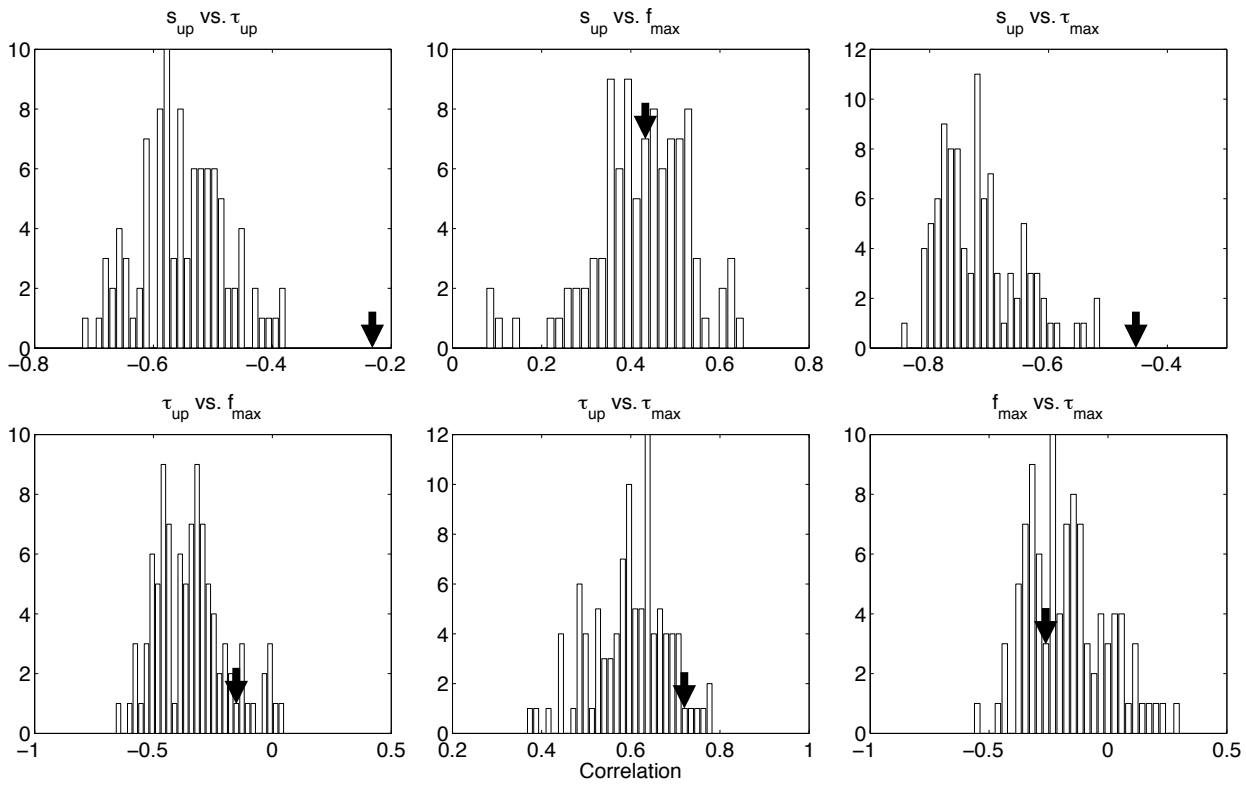

Figure 2.5: Histograms of mutational parameter correlations from 100 random subsamplings (n = 23) of *rpoB* mutations. The bin containing the value obtained for *rho* mutations is indicated with an arrow.

TABLES

Table 2.1: Pearson correlations and their significance among four mutational parameters. Significant ($p < 0.05$) correlations are shaded in light grey. Correlations that remain significant after Bonferroni correction are shaded in dark grey. "Both Populations" includes *rpoB* and *rho* populations together, excluding our mixed population (Line 82).

	***rpoB* Populations**		**rho Populations**		**Both Populations**	
	r	***p***	***r***	***p***	***r***	***p***
s_{up} **vs.** τ_{max}	-0.712	9.97E-09	-0.449	0.0317	-0.627	3.91E-09
τ_{up} **vs.** τ_{max}	0.607	3.79E-06	0.722	9.90E-05	0.639	1.57E-09
s_{up} **vs.** f_{max}	0.439	0.00160	0.428	0.0418	0.445	8.97E-05
f_{max} **vs.** τ_{max}	-0.190	0.192	-0.258	0.234	-0.241	0.0417
s_{up} **vs.** τ_{up}	-0.551	4.10E-05	-0.236	0.278	-0.438	0.000121
τ_{up} **vs.** f_{max}	-0.351	0.0134	-0.157	0.475	-0.290	0.0134

SUPPORTING INFORMATION

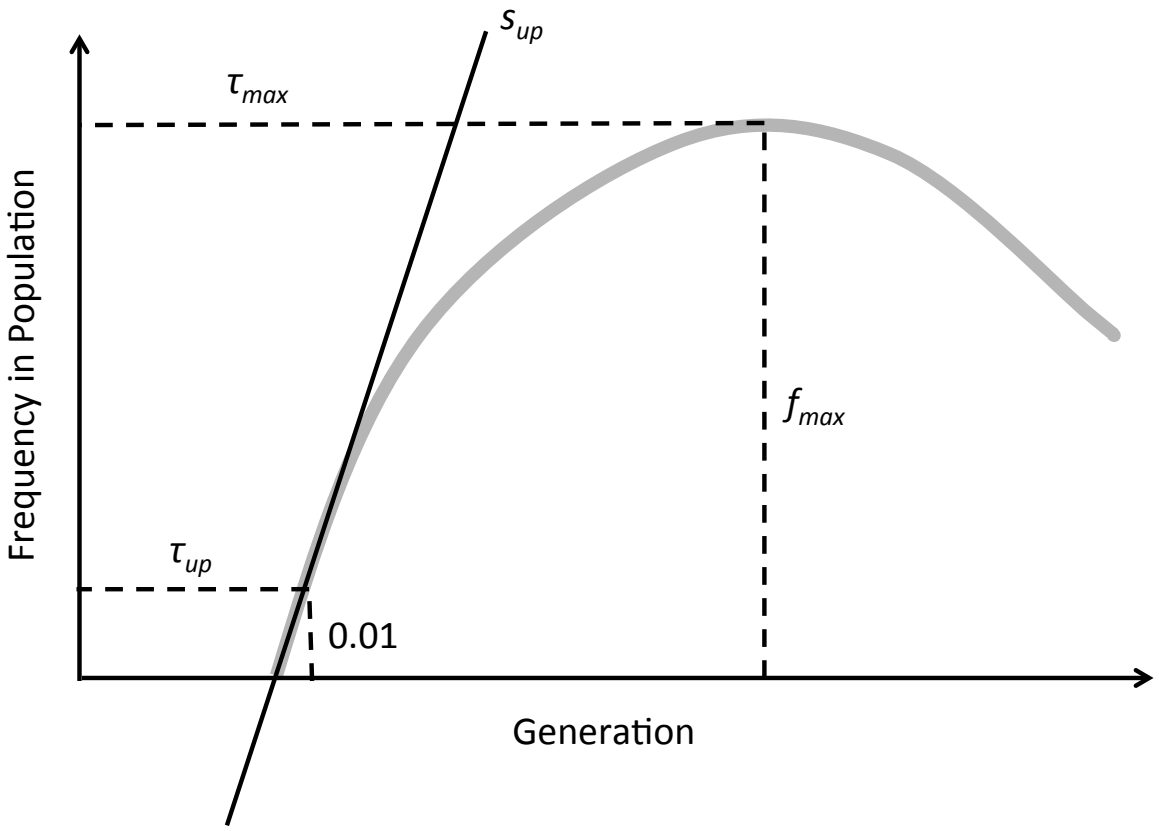

Figure S2.1: Mutational parameters measured in populations. s_{up} constitutes the initial rate of increase in frequency of a mutation. τ_{up} is the generation at which a mutation was at a frequency of 1% in the population, calculated using its particular value of s_{up}. f_{max} is the maximum frequency achieved by a given mutation. τ_{max} is the generation at which a mutation initially reaches its maximum frequency.

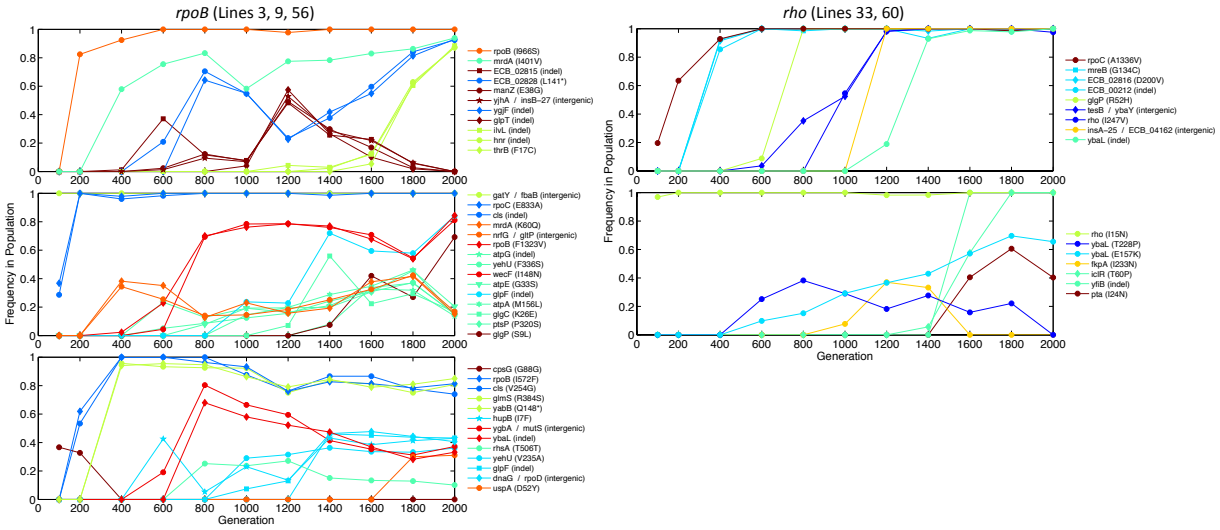

Figure S2.2: Mutational trajectories of all remaining *rpoB* and *rho* populations.

Trajectories are colored by mutational cohort. From top to bottom, the three *rpoB* populations correspond to Line 3, Line 9, and Line 56, respectively. From top to bottom, the two *rho* populations correspond to Line 33 and Line 60, respectively.

Table S2.1: Mutational parameters for each mutation in all eight populations. Mutations are grouped and colored by cohort.

Line3 (*rpoB*)	s_{up}	τ_{up}	f_{max}	τ_{max}
mrdA (I401V)	0.002895	203	0.939	2000
rpoB (I966S)	0.00824	101	1	600
ECB_02815 (indel)	0.00006	368	0.483	1200
glpT (indel)	0.0002	851	0.574	1200
yjhA / insB-27 (intergenic)	0.000065	555	0.528	1200
manZ (E38G)	0.000115	488	0.492	1200
ygjF (indel)	0.003215	603	0.931	2000
ECB_02828 (L141*)	0.001045	410	0.925	2000
thrB (F17C)	0.00027	1437	0.875	2000
ilvL (indel)	0.000215	1047	0.885	2000
hnr (indel)	0.000115	1288	0.872	2000
Line9 (*rpoB*)				
rpoB (F1323V)	0.000115	288	0.844	2000
wecF (I148N)	0.00022	446	0.811	2000
glpF (indel)	0.001185	809	0.84	2000
glgP (S9L)	0.00037	1227	0.692	2000
mrdA (K60Q)	0.00191	205	0.416	1800
nrfG / gltP (intergenic)	0.00172	206	0.423	1800
glgC (K26E)	0.000355	1028	0.559	1400
ptsP (P320S)	0.0004	1225	0.369	1800
atpG (indel)	0.00117	409	0.327	1600
atpA (M156L)	0.000725	814	0.461	1800
atpE (G33S)	0.000395	626	0.455	1800
yehU (F336S)	0.000255	440	0.372	1800
gatY / fbaB (intergenic)	0.01	1	1	100
rpoC (E833A)	0.00368	3	1	200
cls (indel)	0.00287	4	1	200
Line16 (*rpoB*)				
rpoB (L960P)	0.00893	101	1	400
yobF / yebO (intergenic)	0.000375	227	0.962	2000
ybaL (indel)	0.00013	1078	1	2000
rpoB (T539P)	8.00E-05	326	0.966	2000
hemA (S161S)	0.000625	816	0.964	2000
mdoG (indel)	0.00413	202	0.826	400
cusS (I34F)	0.00137	1407	0.373	1800

metJ (N48D)	3.40E-04	830	0.364	1400
ybaL (R325*)	0.00092	611	0.379	1000
dctA / yhjK (intergenic)	0.00127	608	0.384	1000
Line26 (*rho*)				
rpoC (L903P)	0.002885	204	0.775	2000
rpoB (I572L)	0.0003	234	0.478	800
atoC (I129S)	0.00029	435	0.606	800
ompR (S65L)	0.000955	611	0.59	1800
yobF / yebO (intergenic)	0.001935	1005	0.467	1800
rho (A243V)	0.01	101	1	200
cls (D259V)	0.00056	18	1	200
Line33 (*rho*)				
ybaL (indel)	0.000945	1011	1	2000
insA-25 / ECB_04162 (intergenic)	0.005	1002	1	1200
tesB / ybaY (intergenic)	0.00018	456	1	1400
rho (I247V)	0.00273	804	1	1400
glgP (R52H)	0.00044	423	1	800
rpoC (A1336V)	0.00196	5	1	600
ECB_00212 (indel)	0.004275	202	1	600
mreB (G134C)	0.00462	202	1	600
ECB_02816 (D200V)	0.004575	202	1	600
Line56 (*rpoB*)				
ybaL (indel)	0.0034	603	0.68	800
ygbA / mutS (intergenic)	0.000955	411	0.804	800
hupB (I7F)	0.002125	405	0.432	1400
glpF (indel)	0.00037	827	0.461	1400
dnaG / rpoD (intergenic)	0.002315	1204	0.477	1600
yehU (V235A)	0.00145	807	0.363	1400
rhsA (T506T)	0.00126	608	0.271	1200
uspA (D52Y)	0.0015	1607	0.31	2000
cpsG (G88G)	0.00367	3	0.367	100
rpoB (I572F)	0.00619	102	1	400
cls (V254G)	0.00534	102	1	400
yabB (Q148*)	0.004705	202	0.955	600
glmS (R384S)	0.004785	202	0.957	400
Line60 (*rho*)				

rho (I15N)	0.00969	1	1	200
ybaL (T228P)	0.001255	408	0.382	800
fkpA (I233N)	0.000385	826	0.37	1200
ybaL (E157K)	0.00049	421	0.696	1800
pta (I24N)	0.00202	1405	0.606	1800
iclR (T60P)	0.00028	1236	1	1800
yfiB (indel)	0.005	1402	1	1600
Line82 (mixed)				
mrdA (P92S)	0.001755	406	0.351	600
rpsR (I21N)	0.001195	408	0.321	800
rho (I15N)	0.00091	111	0.942	400
rpoH (V97G)	0.00014	72	0.926	400
cls (Q472*)	0.00097	110	0.939	400
ycbC / smtA (intergenic)	0.000965	610	0.917	2000
nrfG / gltP (intergenic)	0.0002	251	0.97	2000
ybaL (indel)	0.00075	13	0.888	2000
rpoB (N760H)	0.00625	2	0.949	2000
glgP (S9L)	0.002445	1204	0.919	2000
atpG (P212L)	0.00012	884	0.914	2000
rpoC (K118T)	0.0027	1404	0.918	2000
yehU (R133H)	0.00258	1404	0.925	2000

REFERENCES

Barrick, J. E. et al., 2009 Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature **461**: 1243-1247.

Bennett, A. F., and R. E. Lenski, 1993 Evolutionary adaptation to temperature II. Thermal niches of experimental lines of *Escherichia coli*. Evolution 1-12.

Blount, Z. D., C. Z. Borland, and R. E. Lenski, 2008 Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. Proc Natl Acad Sci U S A **105**: 7899-7906.

Carroll, S. M., and C. J. Marx, 2013 Evolution after introduction of a novel metabolic pathway consistently leads to restoration of wild-type physiology. PLoS genetics **9**: e1003427.

Chou, H. H., H. C. Chiu, N. F. Delaney, D. Segrè, and C. J. Marx, 2011 Diminishing returns epistasis among beneficial mutations decelerates adaptation. Science **332**: 1190-1192.

Costanzo, M. et al., 2010 The genetic landscape of a cell. Science **327**: 425-431.

Deatherage, D. E., and J. E. Barrick, 2014 Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq, pp. 165-188 in *Engineering and Analyzing Multicellular Systems*, edited by Springer

Fisher, R. A., 1930 *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press

Fisher, R. A., 1950 Statistical methods for research workers. Biological monographs and manuals. No. V. Statistical methods for research workers. Biological mono-graphs and manuals. No. V.

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

Frenkel, E. M., B. H. Good, and M. M. Desai, 2014 The fates of mutant lineages and the distribution of fitness effects of beneficial mutations in laboratory budding yeast populations. Genetics **196**: 1217-1226.

González-González, A., S. M. Hug, A. Rodríguez-Verdugo, J. Patel, and B. S. Gaut, *in preparation* The effect of mutations in the transcription factor Rho in *Escherichia coli* under thermal stress

Herron, M. D., and M. Doebeli, 2013 Parallel evolutionary dynamics of adaptive diversification in *Escherichia coli*. PLoS Biol **11**: e1001490.

Hollands, K., A. Sevostiyanova, and E. A. Groisman, 2014 Unusually long-lived pause required

for regulation of a Rho-dependent transcription terminator. Proc Natl Acad Sci U S A **111**: E1999-2007.

Hug, S. M., and B. S. Gaut, 2015 The phenotypic signature of adaptation to thermal stress in *Escherichia coli*. BMC Evol Biol **15**: 177.

Kao, K. C., and G. Sherlock, 2008 Molecular characterization of clonal interference during adaptive evolution in asexual populations of *Saccharomyces cerevisiae*. Nature genetics **40**: 1499-1504.

Khan, A. I., D. M. Dinh, D. Schneider, R. E. Lenski, and T. F. Cooper, 2011 Negative epistasis between beneficial mutations in an evolving bacterial population. Science **332**: 1193-1196.

Kryazhimskiy, S., D. P. Rice, E. R. Jerison, and M. M. Desai, 2014 Global epistasis makes adaptation predictable despite sequence-level stochasticity. Science **344**: 1519-1522.

Lang, G. I., D. Botstein, and M. M. Desai, 2011 Genetic variation and the fate of beneficial mutations in asexual populations. Genetics **188**: 647-661.

Lang, G. I. et al., 2013 Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. Nature **500**: 571-574.

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

Lenski, R. E., and M. Travisano, 1994 Dynamics of adaptation and diversification: a 10,000 generation experiment with bacterial populations. Proc Natl Acad Sci U S A **91**: 6808-6814.

Levy, S. F. et al., 2015 Quantitative evolutionary dynamics using high-resolution lineage tracking. Nature **519**: 181-186.

Li, H. et al., 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**: 2078-2079.

Long, A., G. Liti, A. Luptak, and O. Tenaillon, 2015 Elucidating the molecular architecture of adaptation via evolve and resequence experiments. Nat Rev Genet **16**: 567-582.

Pallarès, I., V. Iglesias, and S. Ventura, 2015 The Rho Termination Factor of *Clostridium botulinum* Contains a Prion-Like Domain with a Highly Amyloidogenic Core. Front Microbiol **6**: 1516.

Peters, J. M. et al., 2009 Rho directs widespread termination of intragenic and stable RNA transcription. Proc Natl Acad Sci U S A **106**: 15406-15411.

Poon, A., and L. Chao, 2005 The rate of compensatory mutation in the DNA bacteriophage ϕX174. Genetics **170**: 989-999.

Rodríguez-Verdugo, A., B. S. Gaut, and O. Tenaillon, 2013 Evolution of *Escherichia coli* rifampicin resistance in an antibiotic-free environment during thermal stress. BMC Evol Biol **13**: 50.

Rodríguez-Verdugo, A., O. Tenaillon, and B. S. Gaut, 2016 First-Step Mutations during Adaptation Restore the Expression of Hundreds of Genes. Mol Biol Evol **33**: 25-39.

Rodríguez-Verdugo, A., D. Carrillo-Cisneros, A. González-González, B. S. Gaut, and A. F. Bennett, 2014 Different tradeoffs result from alternate genetic adaptations to a common environment. Proceedings of the National Academy of Sciences **111**: 12121-12126.

Sandberg, T. E. et al., 2014 Evolution of *Escherichia coli* to 42 degrees C and Subsequent Genetic Engineering Reveals Adaptive Mechanisms and Novel Mutations. Mol Biol Evol

Sneath, P. H. A., and R. R. Sokal, 1973 *Numerical taxonomy. The principles and practice of numerical classification.*

Szamecz, B. et al., 2014 The genomic landscape of compensatory evolution. PLoS Biol **12**: e1001935.

Tenaillon, O. et al., 2012 The molecular diversity of adaptive convergence. Science **335**: 457 461.

Traverse, C. C., L. M. Mayo-Smith, S. R. Poltak, and V. S. Cooper, 2013 Tangled bank of experimentally evolved *Burkholderia* biofilms reflects selection during chronic infections. Proc Natl Acad Sci U S A **110**: E250-9.

Wiser, M. J., N. Ribeck, and R. E. Lenski, 2013 Long-term dynamics of adaptation in asexual populations. Science **342**: 1364-1367.

CHAPTER 3 - Lazarus effects: the frequency and genetic causes of *Escherichia coli* population recovery under lethal heat stress

ABSTRACT

Sometimes populations crash and yet recover before being lost completely. Such recoveries have been observed incidentally in evolution experiments using *Escherichia coli*, and this phenomenon has been termed the "Lazarus effect." To investigate how often recovery occurs and the genetic changes that drive it, we evolved ~300 populations of *E. coli* at lethally high temperatures (43.0°C) for five days and sequenced the genomes of recovered populations. Our results revealed that the Lazarus effect is uncommon, but frequent enough, at ~9% of populations, to be a potent source of evolutionary innovation. Population sequencing uncovered a set of mutations adaptive to lethal 43.0°CC that were mostly distinct from those that were beneficial at a high but nonlethal temperature (42.2°C). Mutations within two operons—the heat shock *hslUV* operon and the RNA polymerase *rpoBC* operon—drove adaptation to lethal temperature. Mutations in *hslUV* exhibited little antagonistic pleiotropy at 37.0°CC and may have arisen neutrally prior to subjection to lethal temperature. In contrast, *rpoBC* mutations provided greater fitness benefits than *hslUV* mutants, but were less prevalent and caused stronger fitness tradeoffs at lower temperatures. Recovered populations fixed mutations in only one operon or the other, but not both, indicating that epistatic interactions between beneficial mutations were important even at the earliest stages of adaptation.

INTRODUCTION

The fossil record contains numerous examples of organisms that appear to have gone extinct for long stretches of geologic time, only to reappear in the fossil record later. One common explanation for these so-called "Lazarus" taxa is that the fossil record is incomplete, and fossil data are missing for lineages that always existed (Jablonski, 1986). It has also been suggested that the incompleteness of the fossil record for Lazarus taxa is not due to the loss of fossils or their inability to form, but rather extremely low population sizes. That is, Lazarus taxa do not suddenly reemerge in the fossil record due to missing data, but because their populations were actually on the verge of extinction, only to rebound later (Wignall and Benton, 1999).

A similar phenomenon has been observed in laboratory evolution experiments using bacteria, and it has been termed the Lazarus effect (Bennett and Lenski, 1993; Mongold *et al.*, 1999; Rodríguez-Verdugo *et al.*, 2014). When grown under lethal temperatures, bacterial populations decline in size over time, often to levels that are nearly or completely unquantifiable. However, some among the remaining survivors may acquire beneficial mutations that rescue the population and enable survival. These individuals reproduce, spreading the beneficial mutations through the population and restoring the population to sizes of similar magnitude to those observed before near-extinction.

Following the chance evolution of such a Lazarus population in an evolution experiment (Bennett and Lenski, 1993), Mongold et al. (1999) characterized patterns of evolutionary recovery at a lethal temperature of 44°C using ancestral strains of *Escherichia coli* B that had adapted to low-glucose medium at 32°C, 37°C, and 41-42°C (Mongold *et al.*, 1999). They found that Lazarus events at 44°C occurred only in populations derived from ancestors already adapted to high temperature (41-42°C), suggesting that pre-adaptation aids adaptive recovery under lethal temperature conditions. Moreover, they found that Lazarus mutants exhibited a fitness cost at

elevated, but nonlethal temperatures, supporting the idea that some Lazarus mutations confer a fitness benefit in one environment, but a fitness cost in another: a phenomenon known as antagonistic pleiotropy (Williams, 1957; Cooper and Lenski, 2000; MacLean *et al.*, 2004).

Here we revisit the Lazarus effect, in part to determine whether Lazarus mutations are similar to, or distinct from, adaptive mutations that accumulate under thermal stress. For the comparison to thermal stress, we rely on an experiment by Tenaillon et al. (2012) that uncovered the genetic basis of adaptation to a high but sustainable temperature (42.2°C). The experiment of Tenaillon et al. (2012) began with a strain of *E. coli* B that was adapted to low-glucose medium and 37.0°C (*E. coli* B strain REL1206) (Tenaillon *et al.*, 2012). The ancestral REL1206 strain was inoculated into 115 independent populations that were then evolved for 2,000 generations at 42.2°C. Genome sequencing of a single clone from each population revealed >1,000 putatively adaptive mutations that fell into two major genetic pathways to high-temperature adaptation. The two pathways were exemplified by mutations in either *rpoB*, which encodes a subunit of RNA polymerase, or *rho*, which encodes a major transcriptional termination factor.

Subsequent work has focused on some of the *rpoB* mutations that occurred during adaptation to thermal stress. Rodríguez-Verdugo et al. (2014) engineered single *rpoB* mutations into the REL1206 background and found that they imparted an average fitness gain of ~22% at 42.2°C. However, these same mutations tended to exhibit a fitness tradeoff at lower temperatures (<20.0°C), consistent with the effects of antagonistic pleiotropy (Rodríguez-Verdugo *et al.*, 2014). It was during the course of this experiment, while investigating the fitness of clones at high temperature, that Rodríguez-Verdugo et al. (2014) observed Lazarus events at otherwise lethal temperatures in ancestral lines, despite the fact that the ancestor had not been pre-adapted to extreme temperature conditions.

In this study, we perform *E. coli* growth experiments to better understand the dynamics,

genetics, and fitness consequences of Lazarus events. Beginning with an ancestor derived from a single-colony of *E. coli* B strain REL1206, we carry out replicated evolution experiments at two temperatures (43.0°C and 44.0°C) that typically result in population extinction under our growth conditions. After observing and noting the frequency of Lazarus events, we save populations that have rebounded from the brink of extinction, sequence their genomes from population samples, and determine the fitness of the Lazarus populations relative to the ancestor. Armed with genetic and fitness data, we seek to address three questions. The first is whether population recovery involves a distinct set of mutations relative to experiments at high but sustainable temperatures, as suggested by the fitness dynamics of Mongold et al. (1999). The second concerns the likely genetic drivers of adaptation, as opposed to hitchhikers. The discrimination of drivers and hitchhikers is a major challenge facing experimental evolution studies (Rosenzweig and Sherlock, 2014), but the short-term nature of our experiment permits novel insights into this distinction. The third question reflects the fitness effects of Lazarus populations and whether they exhibit tradeoffs at elevated, but nonlethal temperature (42.2°C) or at the ancestor's optimal temperature of 37.0°C. Finally, combining genetic and fitness data allow us to comment on three important phenomena: antagonistic pleiotropy of adaptive mutations, the role of standing variation in population rescue, and the potential mechanism of population recovery.

MATERIALS AND METHODS

Lazarus Ancestral Stock

A frozen glycerol stock was prepared from a single colony of *Escherichia coli* B strain REL1206 possessing a neutral Ara- marker. This strain had been propagated previously at 37.0 °C for 2,000 generations in Davis minimal medium supplemented with glucose at 25 mg/L

(DM25), and was thus well adapted to the growth medium (Lenski *et al.*, 1991). To isolate the single colony, REL1206 was streaked from frozen onto a tetrazolium-arabinose (TA) plate and incubated overnight at 37.0°C. The single colony was inoculated into Luria-Bertani medium (LB) and grown overnight. To prepare a frozen reference stock, 900 µL of culture was mixed with 900 µL of 80% glycerol and frozen at -80°C. We term this REL1206 frozen stock the "Lazarus ancestor" (Figure 3.1). A backup Lazarus ancestor stock was prepared from the same LB culture.

Lazarus Growth Experiments

As is common practice (Bennett and Lenski, 1993; Lenski and Travisano, 1994; Rodríguez-Verdugo *et al.*, 2014; Hug and Gaut, 2015), we first acclimated the Lazarus ancestor (REL1206) to mild laboratory conditions to allow it to recover from being frozen. The Lazarus ancestral stock was inoculated into 100 mL LB and grown for eight hours in an Infors HT Minitron incubator at 37.0°C and 120 RPM (Figure 3.1). 10 µL of this culture was then inoculated into 100 mL DM25 and grown for 24 hours in an Infors HT Minitron incubator at 37.0°C and 120 RPM. We inoculated 100 µL of the 37.0°C, DM25 culture into each of 44 culture tubes containing 9.9 mL DM25. An additional four culture tubes containing 9.9 mL DM25 were used as contamination controls and cell density blanks. The total set of 48 tubes were placed into an Innova 3100 water bath shaker (New Brunswick Scientific) and grown for 24 hours at 120 RPM and at the experimental temperature of 43.0°C or 44.0°C. Tube cultures were serially propagated over the course of five days by inoculating 100 µL of culture into 9.9 mL DM25 after each day of growth (Figure 3.1).

On each of the five days, 50 µL of each culture (including blanks) was also inoculated into cuvettes containing 9.9 µL Isoton II Diluent (Beckman Coulter). These samples were

analyzed using a Multisizer 3 Coulter counter (Beckman Coulter) to determine cell densities (particles/mL). The average particle density of the four blanks was subtracted from each sample's cell density. We defined a Lazarus event as a population whose cell density increased by at least an order of magnitude over the previous day's measurements.

Following cell density measurement, Lazarus populations were saved as frozen glycerol stocks on their initial day of recovery, as well as any subsequent days. To prepare frozen stocks of Lazarus recovery events, 900 μL of culture was mixed with 900 μL of 80% glycerol and frozen at -80°C.

DNA Extraction and Sequencing

Most samples for DNA extraction and sequencing were derived from day five of the experiment; two samples (#1 and #19) were derived from day four. Each of the frozen Lazarus samples was inoculated into ten separate culture tubes containing 9.9 mL DM25 and incubated in an Innova 3100 water bath shaker (New Brunswick Scientific) overnight at 37.0°C and 120 RPM. Cells from all ten tubes were pooled, and genomic DNA was extracted from these samples using Wizard Genomic DNA Purification Kits (Promega). This method of pooling independently derived cultures provided a way to later filter out any mutations that might have risen to high frequency during the process of recovery from being frozen. Our reasoning was that a specific mutation might arise in a single tube, but would be less likely to arise independently in multiple tubes, providing an empirical frequency cutoff of 10%, based on the number of independent tubes pooled. DNA from the Lazarus ancestor was extracted in the same manner, but pooled from four tubes rather than 10.

Genomic DNA libraries were prepared using the TruSeq DNA PCR-Free Library Preparation Kit (Illumina). The 26 Lazarus samples were multiplexed and sequenced in two

lanes of an Illumina HiSeq 2500 in rapid mode at UC Irvine's Genomics High-Throughput Facility. Two ancestral samples—one working stock (Figure 3.1) and one backup stock—were also sequenced on an Illumina HiSeq 3000 at the Bioinformatics Core Facility at the UC Davis Genome Center.

Mutations and mutation frequencies were called using breseq (Deatherage and Barrick, 2014) in polymorphism mode, using the *E. coli* B REL606 genome as a reference. Six regions (*topA*, *spoT*, *glmU*/*atpC*, *pykF*, *yeiB*, and the *rbs* operon) possess mutations that differ between REL606 and REL1206 (Barrick *et al.*, 2009; Tenaillon *et al.*, 2012), so there regions were excluded from our analyses. In theory, breseq provides information about duplications and deletions by reporting novel junctions. No evidence of novel junctions or sequencing coverage was found for large deletions in our data set, but breseq did provide some novel junction evidence for the presence of large duplications. To assess duplications more formally, we compared unique reads (mapping quality >5 in samtools 1.3) across 10 kb regions of the genome, defining duplications as regions with more than twice the average genome coverage.

Fitness Assays

Relative fitness values were assessed by performing standard competition assays (Lenski *et al.*, 1991; Tenaillon *et al.*, 2012). Samples of frozen Lazarus populations and an ancestral stock containing a neutral Ara+ mutation (REL1207) were grown overnight in 9.9 mL DM25 using an Innova 3100 water bath shaker (New Brunswick Scientific) at 37.0°C and 120 RPM. 100 μL of Lazarus culture was added to each of six replicate tubes containing 9.9 mL DM25. Likewise, 100 μL of ancestor culture was added to each of three tubes containing 9.9 mL DM25. These were incubated overnight at 42.2°C and 120 RPM. All three ancestral tubes were pooled and vortexed. On day zero (t_0) 75 μL of ancestor culture and 25 μL of a Lazarus culture were

added to a tube of 9.9 mL DM25 and vortexed. 100 µL of this competition culture was diluted 1/100, and 100 µL of the dilution was plated on a TA plate for overnight incubation at 37.0°C. The remaining competition culture was grown overnight in the water bath at 42.2°C and 120 RPM. After one day of competition (t_1), 100 µL of competition culture was diluted 1/10,000 and plated as above. Ara+ and Ara- colonies were counted for both t_0 and t_1. Fitness values were calculated using the methods of Lenski et al. (1991), but results were qualitatively identical using the methods of Chevin (2011).

Data Availability: All sequence data have been submitted to the NCBI Sequence Read Archive (Accession number: PRJNA326455). Cell density measurements can be found in Table S3.1. Details about mutations present at >10% frequencies within a Lazarus population are reported in Table S3.2. Colony counts from fitness assays are reported in Table S3.3.

RESULTS

Frequency of Lazarus Events

Over the course of five days, we measured the cell densities of 308 populations as they evolved (or more commonly, went extinct) at 43.0°C (Table S3.1). 12 of these populations (one in one week, 11 in another week) were excluded from our final analysis due to mechanical issues during particle counting. Thus, 296 populations were considered for determining the frequency of Lazarus events. In total, we observed 26 Lazarus events (Figure 3.2), placing their frequency in our system at 8.8% (26/296), comparable to the frequency of 10% (3/30) observed by Mongold et al. (1999) in their pre-adapted populations grown at 44.0°C (Mongold *et al.*, 1999). It is worth noting, however, that our populations were not pre-adapted to high temperature conditions. When we grew 88 populations for five days at 44.0°C, none survived.

Most Lazarus events occurred by day four, with events on day five also being common (Figure 3.2). Only three Lazarus events occurred on day three, suggesting a limit to how quickly beneficial mutations can become established within populations. Lazarus events occurred more often in some weeks than others. The highest incidence of Lazarus events in one week was 25% (11/44), and the lowest was 0% (0/33). Initial cell density bore no relationship to the number of Lazarus events observed in a given week ($r = -0.24$, $p = 0.61$), and initial cell density was not significantly correlated with the timing of the first population recovery in a given week ($r = 0.50$, $p = 0.32$).

Mutations Associated with Lazarus Events

To characterize the genomic changes underpinning Lazarus events, we sequenced each of the 26 Lazarus populations and identified the frequencies of mutations using breseq (Deatherage and Barrick, 2014). Among the 26 Lazarus populations, we identified 419 simple mutations (i.e., single nucleotide changes and small indels) within 122 unique genic and intergenic regions. Of these, 100 were called at frequencies >10% in their populations, and these comprised 32 unique genic and intergenic regions (Table 3.1, Table S3.2, Figure 3.3).

For a mutation to rescue a population from the brink of extinction, it must sweep through the population to an appreciable frequency. Focusing on mutations at frequencies > 85% in a population as "fixed", we observed 46 fixed mutations within 20 unique genic and intergenic regions among 23 of the 26 Lazarus populations. On average, each population possessed 1.8 fixed mutations. 42% (11/26) of populations had just one fixed mutation, making these genetic changes the likely drivers of population recovery. Additionally, one of our Lazarus populations (#2; Table 3.1) evolved a mutator phenotype due to a small deletion in the *mutT* gene. Not surprisingly, this population contained six fixed mutations, the most of any population in the

experiment.

Of the three remaining populations lacking fixed mutations, one (#15) possessed mutations in both *hslU* and *hslV* (see below) at intermediate frequencies, and two (#22 and #23) possessed large duplications at detectable frequencies. In total, four populations contained large duplications based on their sequencing coverage profiles, with three of these (#22, #23, and #25) occurring in parallel within a single week. These three parallel duplications included the genes *groEL* and *groES*, which encode a major chaperonin complex that is both necessary for growth under normal conditions and induced during growth at elevated temperatures (Fayet *et al.*, 1989; Hayer-Hartl *et al.*, 2016). The fourth population (#21) contained a duplication that includes some or all of the *hslUV* operon (see below).

To be assured that the fixed mutations arose as a consequence of thermal treatment, we also sequenced two control cultures. The first culture was inoculated from the frozen ancestor stock used throughout the experiment, and the second was inoculated from a reserve ancestor stock. We sequenced both cultures to >2,000x and identified no fixed (>85% frequency) mutations relative to the REL1206 genome. We did identify 1,105 potential variants across both cultures at an average frequency was ~1.3%. There were, however, three mutations that exceeded 10% frequency. Of these, one was shared with Lazarus populations: a four-nucleotide indel within the *ECB_01992* gene found at a frequency of 19% and 21% in the two ancestral samples. Interestingly, this mutation was found at frequencies between 32% and 34% across eight of the 26 Lazarus populations, but it never reached fixation. The lack of overlapping mutations between control and Lazarus populations (with the notable exception of the *ECB_01992* indel), combined with the low frequency of mutations in control populations, argues that the fixed mutations in Table 3.1 are indeed a consequence of population recovery under a temperature that is typically lethal.

Genetic Targets of Parallel Mutations

Experimental evolution derives its power from replication. If genetic changes occur multiple times across replicates, it provides strong evidence that these changes are adaptive (Woods *et al.*, 2006). In our experiment, we found consistent fixation within a pair of genes, *hslU* and *hslV*, which together comprise an operon that encodes a heat shock protease system (Missiakas *et al.*, 1996; Bochtler *et al.*, 2000). Fixed mutations in *hslUV* affected 62% (16/26) of Lazarus populations: nine populations possessed fixed mutations in *hslU*, and another seven possessed fixed mutations in either *hslV* or its upstream region. We also noticed a striking pattern of parallelism in the *hslUV* operon, because parallel mutations occurred at the level of individual nucleotides within single weeks, but not between weeks (Figure 3.4A), and this parallelism was most prevalent in weeks that showed the most Lazarus events. These parallel events included small frameshift indels. In total, 69% (9/13) of *hslUV* mutations caused frameshifts, suggesting that interruption of *hslUV* function may be adaptive.

Like *hslUV*, the *rpoBC* operon also accumulated multiple fixed mutations (Figure 3.4B), which were found across six populations (#2, #3, #16, #20, #24, and #26). Four populations had fixed mutations in *rpoC*, another two had fixed mutations in *rpoB*, and all were nonsynonymous mutations. Interestingly, the populations bearing *rpoBC* mutations were distinct from those with *hslUV* mutations; in other words, no populations contained fixed mutations in both operons. Populations with fixed *rpoC* mutations also had nucleotide-level parallelism, but unlike *hslUV* mutations, this parallelism occurred across weeks, rather than within a single week (Figure 3.4B).

A third region, *clpA*/*serW*, accumulated the same fixed point mutation in five different populations (#2, #6, #7, #16, and #20) across four independent weeks. This region lies between

clpA—one component of a protease system of similar function to that encoded by *hslUV* (Kwon *et al.*, 2004)—and *serW*, which encodes a serine-bearing tRNA.

Mutations within *hslUV*, *rpoBC*, and *clpA*/*serW* account for 59% (27/46) of fixed mutations. In many cases, *hslUV*, *rpoBC*, and *clpA*/*serW* mutations were fixed with other mutations. For example, population #4 had a total of five fixed mutations, including a mutation in *hslV*. Importantly, a single mutation in *hslUV* or *rpoBC* represents the *only* fixed mutation in several populations (#1, #3, #9, #11, #12, #13, #14, #17, #19, #25, and #26), suggesting that some mutations within these operons are sufficient to rescue populations and drive a Lazarus effect.

Finally, because Lazarus events occurred at different times during the course of a week, we were interested to see if there was any relationship between the timing of Lazarus events and the identities of fixed mutations, but we found no obvious relationship (Mann-Whitney U test comparing *hslUV* and *rpoBC* populations: $p = 0.34$).

Density and Fitness

The hallmark of a Lazarus event is the marked increase in cell density that occurs following a population crash, but not all populations rebound to the same final cell density (Table S3.1). We tested for a relationship between cell densities after population recovery and the fixed mutations present within those populations. At both days four and five, populations with fixed *rpoBC* mutations were among those with the highest cell densities. All six populations containing *rpoBC* mutations were among the seven populations with the highest cell densities at day five, and three more had the highest cell densities on day four. *rpoBC* populations had significantly higher cell densities than all other recovered populations on both days (Mann-Whitney U test: day four $p = 0.0044$, day five $p = 3.71 \times 10^{-4}$). In addition,

populations with large duplications yielded cell densities approximately half as great as other populations that first experienced recoveries on the same day (Figure 3.2).

Given apparent relationships between mutations and cell densities, we quantified the relative fitnesses (w) of our Lazarus populations. Using standard competition assays (Lenski *et al.*, 1991; Tenaillon *et al.*, 2012; Rodríguez-Verdugo *et al.*, 2014), we competed each Lazarus population against REL1207, a strain identical to our Lazarus ancestor (Table S3.3). These relative fitness assays were performed at 42.2°C rather than 43.0°C, because REL1207 populations are sustainable at the former temperature but crash at the latter, making w measurement impossible.

We estimated w each of the 26 populations. On average, the Lazarus populations showed a w increase of 24% relative to the ancestor, with 22 of 26 populations having $w > 1.0$ ($p < 0.05$, Table 3.1). Although the average w increase of the *rpoBC* populations (28%) exceeded that of all the other populations (22%) and the *hslUV* populations alone (23%), it was not significantly higher than either (one-tailed t-test, unequal variance: $p = 0.15$ and $p = 0.18$, respectively).

Strangely, although the relative fitness values of *rpoBC* populations were not statistically higher than those of *hslUV* populations, *rpoBC* populations produced significantly more colonies during the competition experiments than *hslUV* populations at both time points (t_0 and t_1) in the competition (one-tailed t-test, unequal variance: t_0 $p = 2.42 \times 10^{-5}$, t_1 $p = 2.81 \times 10^{-4}$). The ancestor colony counts were unaffected between these two groups (one-tailed t-test, unequal variance: t_0 $p = 0.42$, t_1 $p = 0.49$). While this did not translate into a difference in w, it did agree with the cell density measurements at day five.

We also assessed whether Lazarus populations showed any fitness tradeoffs at their ancestral growth temperature of 37.0°C. On average, the Lazarus populations exhibited a fitness decrease of 3% relative to the ancestor, with 14 of 26 populations having relative fitness values

significantly < 1.0 (one-tailed t-test: $p < 0.05$, Table 3.1). Populations with fixed *rpoBC* mutations had an average *w* decrease of 8% relative to the ancestor (one-tailed t-test: $p = 0.0082$), which was significantly lower than all other populations (one-tailed t-test, unequal variance: $p = 0.019$). Although some *hslUV* populations had fitness values significantly < 1.0 (Table 3.1), *hslUV* populations had an average fitness decrease of 1% relative to the ancestor, which was not significantly different from 1.0 (one-tailed t-test: $p = 0.054$), suggesting less of a fitness tradeoff for *hslUV* populations compared to *rpoBC* populations.

DISCUSSION

We have performed experiments to characterize the genetic mutations that contribute to the phenomenon of rescue from otherwise lethal temperatures. Like Mongold et al. (1999), who previously studied the Lazarus effect, we find that rescue is infrequent, because it occurs in only 8.8% of experimental populations. It is nonetheless frequent enough that it has the potential to be a potent source of evolutionary innovation. Moreover, we have made a series of puzzling observations in the course of our experiment. The first is that there is a striking lack of independence among experiments, because we were more apt to find the same genetic mutations among populations within weeks, rather than between weeks. There were nonetheless notable parallels across weeks as well, suggesting that there is finite pool of potentially adaptive rescue mutations. The second is that there is a surprising amount of polymorphism within populations; even though there are clear 'driver' mutations that become fixed in Lazarus populations, we have detected additional polymorphic mutations at frequencies of >10% in most of our populations. Finally, there are clear patterns to the driver mutations, especially in relation to a previous thermal stress experiment at non-lethal temperatures (Tenaillon *et al.*, 2012). We have identified multiple driver mutations in two sets of genes: those encoding subunits of RNA polymerase

(*rpoB* and *rpoC*), and those encoding heat shock proteases (*hslU* and *hslV*). In the Discussion below, we consider these and other points further, culminating with a verbal model of the adaptive dynamics of Lazarus populations and potential mechanisms for population rescue.

Adaptation to Lethal and Nonlethal Temperature

One of the initial mysteries of the dynamics of population recovery was why populations adapted to high, but nonlethal temperatures did not concomitantly extend their upper thermal niche, thus enabling survival at even higher temperatures (Mongold *et al.*, 1999). Instead, lethal temperature conditions remained lethal for most populations, and recovery occurred primarily in populations already adapted to elevated temperature, suggesting that they were somehow pre-adapted to recover from high and otherwise lethal temperatures (Mongold *et al.*, 1999). Moreover, fitness measurements suggested that populations that recovered in lethal temperatures experienced tradeoffs at elevated, but nonlethal temperatures, implying that distinct sets of mutations are adaptive under nonlethal and lethal temperature regimes (Mongold *et al.*, 1999).

To determine if there is a distinction between the sets of mutations that arise under lethal and nonlethal conditions, we have compared the genetic data of our Lazarus experiments at 43.0°C to the results of a previous long-term evolution experiment conducted at a sustainable 42.2°C (Tenaillon *et al.*, 2012). These experiments can be compared directly because they were based on the same ancestor (REL1206), grown in the same media (DM25), and under the same conditions (i.e., 10 mL of media, with 120 RPM shaking). Although there are other differences (see below), the major difference is between a high but sustainable temperature (42.2°C) and a high but lethal temperature (43.0°C).

Among the parallel results between experiments, one was the recovery of point mutations in *rpoB* and *rpoC*. The *rpoB* gene is a common target for mutations across a wide

array of evolution experiments (Herring *et al.*, 2006; Conrad *et al.*, 2009; Charusanti *et al.*, 2010), and it was also the most mutated gene in the 42.2°C experiment (Tenaillon *et al.*, 2012). In that experiment, 76 of 115 clones contained an *rpoB* mutation, another 21 of 115 clones contained an *rpoC* mutation, and a total of 5 of 115 lines contained a mutation in both genes. At least some of the *rpoB* mutations were likely to have been drivers of adaptation, both because they fixed rapidly during the course of the 42.2°C experiment (Rodríguez-Verdugo *et al.*, 2013) and because they provided a ~22% fitness benefit, on average, at 42.2°C as single mutations in the REL1206 background (Rodríguez-Verdugo *et al.*, 2014). Our Lazarus experiments also suggest that *rpoB* and *rpoC* mutations are adaptive, because we have detected two populations with fixed (>85% estimated frequency) *rpoB* mutations and four populations with fixed *rpoC* mutations (Table S3.2). Of these six mutations, we can definitively identify at least one *rpoB* and one *rpoC* as driving mutations, because they were the lone high frequency variants in the recovered population (#3 and #26; Table 3.1).

The two experiments also identify mutations in the *hslUV* operon, but the frequency of mutations differs. In the 42.2°C experiment, only two of 115 (1.7%) clones contained a mutation in the *hslUV* region. Neither were obvious knockouts; one mutation was intergenic, and the other was a nonsynonymous replacement (I136N). In contrast, 16 of 26 (61%) of Lazarus populations contained a mutation in *hslUV* or their upstream region, several of which appear to be drivers because they are the lone fixed mutation in the population.

We can compare results between the two experiments more formally by considering the frequency of mutations within specific genes. For example, in the *hslU* gene, one mutation was observed out of 115 clones from the 42.2°C experiment, for a frequency of 0.009 (1/115). Assuming this is the probability of an adaptive mutation occurring in this gene in the Lazarus experiment, we can test whether our Lazarus observation of nine mutations events in 26

populations is expected. We find that the probability of observing nine or more *hslU* mutations in 26 observations is vanishingly small given a frequency of 0.009 (binomial: $p = 7.77 \times 10^{-13}$; Table 3.2). We note, however that this result changes somewhat if one assumes that the 26 populations are not independent (see below), and instead rely on the fact that *hslU* mutations occur in two of seven independent weekly trials (binomial: $p = 0.02$; not significant after Bonferroni correction; Table 3.2).

We find that many common mutations differ significantly between the two experiments (Table 3.2), most because they have occurred too infrequently in the Lazarus populations (e.g., *rpoB*, *ybaL*, *cls*, *rho*, *iclR*, and *rpoD* mutations), but others because they occur too often in Lazarus populations (e.g., *hslV*). Notably, *rpoB* and *ybaL*, the two most mutated genes within the 42.2°C experiment—and the *rho* and *cls* genes, which together define a second path of adaptation to 42.2°C adaptation—are significantly underrepresented in Lazarus populations. One important exception to this trend is *rpoC*, because we cannot reject the hypothesis that *rpoC* mutations have occurred with equal frequency in the two experiments (binomial: $p = 0.47$ for 26 populations, $p = 0.24$ for 7 weeks). There are caveats to using this binomial approach for comparing experiments: specifically, some mutations are not strictly independent because they occur together on haplotypes, and the test relies on a single point estimate of probability from the 42.2°C experiment. Nonetheless, the overarching impression is that the two experiments identify different sets of genes as loci for adaptive mutations, with the notable exception of *rpoC*.

Why might the sets of mutations differ between these two experiments? As we have noted, one difference between experiments is a modest but physiologically critical difference in temperature, but they also differ in at least three other features. The first is the method in which the experiments were started—single clones for each of the 115 populations at 42.2°C (Tenaillon *et al.*, 2012; Hug and Gaut, 2015) but a batch culture for each week of the Lazarus experiment.

These batch cultures likely introduced non-independence among replicates within weeks, but this non-independence is helpful for ultimately understanding the dynamics of rescue (see below).

A second difference is population size. Because REL1206 growth is sustained at 42.2°C, populations remained large (~10^6 individuals) for all phases of that experiment, whereas Lazarus populations initially crashed to undetectable or nearly undetectable levels (Figure 3.2). Within crashed populations, genetic drift may rapidly elevate the frequency of rare advantageous mutations, ultimately leading to an acceleration of the rate of adaptation (Lang *et al.*, 2013). We revisit this point further below.

Finally, the two experiments differ in duration; the 42.2°C experiment ran for a year, but the Lazarus populations were propagated for only five days. This time difference is likely important, because large-effect mutations tend to be fixed early in experimental populations (Kryazhimskiy *et al.*, 2014). After the initial fixation of large-effect mutations, the adaptive fixations that follow tend to have incrementally diminishing fitness benefits (Chou *et al.*, 2011; Khan *et al.*, 2011; Kryazhimskiy *et al.*, 2014). Given the short timespan of Lazarus experiments, we are likely uncovering only large-effect mutations. Unfortunately, we cannot compare them directly to early large-effect mutations in the 42.2°C experiment, because it is not yet known which mutations were fixed early in the that experiment (with the notable exception of *rpoB* mutations) (Rodríguez-Verdugo *et al.*, 2013). If we were to continue the Lazarus experiment, it is possible that the set of adaptive mutations separating lethal and nonlethal temperatures would diminish (Table 3.2). It is important to note, however, that time alone does not explain differences between the experiments, because mutations within *hslV* are common, fixed drivers in several Lazarus populations but were rare in the year-long 42.2°C experiment.

hslUV, *rpoBC*, and the Roles of Standing Variation and Genetic Drift

hslUV and *rpoBC* mutations drive adaptation and population recovery, but the evolutionary dynamics of mutations in these operons differ substantially for three reasons. First, *hslUV* and *rpoBC* mutations are not found together as fixed in the same population, which is statistically improbable given their respective frequencies across populations and given that at least one is fixed as a putative driver in each population with two or more fixed mutations ($p < 0.02$). The lack of co-occurring *rpoBC* and *hslUV* mutations suggests an initial canalization of the adaptive response, focused either on a pathway through heat shock proteins or through RNA polymerase, but not through both. Such a pattern is consistent with negative epistasis observed between the *rho* and *rpoB* pathways of the 42.2°C experiment (Tenaillon *et al.*, 2012). These results imply that epistasis shapes trajectories even at the earliest stages of the adaptive process. Second, as noted above, their patterns of parallelism differ. Specific *hslUV* mutations tend to repeat across populations within weeks, but *rpoBC* mutations repeat across weeks (Figure 3.4)—for example, the same nonsynonymous mutation in *rpoC* (W1020G) occurred in weeks three, six, and seven. Finally, populations with *rpoBC* vs. *hslUV* mutations had distinct fitness patterns. Based on cell densities and relative fitnesses at 42.2°C (Table 3.1), mutations in the *rpoBC* operon offer an advantage that is at least equivalent to, if not greater than, mutations in the *hslUV* operon. Despite this, *hslUV* mutations were more prevalent across all populations.

Together these observations suggest different evolutionary dynamics for mutations within *hslUV* and *rpoBC*. One possibility is that these two adaptive strategies are associated with different costs under non-stressful conditions. To assess this possibility, we competed Lazarus populations against the REL1207 ancestor at 37.0°C. Our results indicate that the *hslUV* populations do not differ statistically as a group in their fitness from the ancestor, suggesting that *hslUV* mutations are neutral (or only mildly deleterious) at 37.0°C. This observation may not be surprising, considering that the inactivation of heat shock genes is unlikely to be critical at

37.0°C. In contrast, *rpoBC* populations have a greater fitness tradeoff (significantly lower fitness) at 37.0°C than *hslUV* populations, suggesting that *rpoBC* mutations are deleterious at 37.0°C.

Together, these lines of evidence suggest a model for the establishment of recovery. This model relies critically on the fact that the replicates within a week were established from batch cultures at 37.0°C (Figure 3.1). Previous work in nonlethal experimental evolution systems has suggested that adaptation is not mutation-limited (Lang *et al.*, 2011), and we believe that to be the case here. Starting with our final cell densities from day zero (acclimation growth in DM25 at 37.0°C), using the expected 6.64 generations produced in DM25 (Lenski *et al.*, 1991), and assuming an *E. coli* mutation rate of 10^{-3} per genome per generation (Lee *et al.*, 2012), we can calculate the number of mutations expected at each nucleotide position during DM25 acclimation. Using a conservative six generations of binary fission, our populations would have sampled between 0.36 and 0.60 mutations per nucleotide, with an average of 0.47. It is important to note that these are only the mutations that could have arisen during growth in DM25; an additional eight hours of growth took place in LB prior to this (Figure 3.1).

Based on the length of the *hslUV* and *rpoBC* operons, we expect that mutations in *rpoBC* occur more often than in *hslUV*. However, several effects may have acted individually or together to generate a bias toward *hslUV* mutations. First, many *hslUV* mutations constituted frameshifts, suggesting that a simple loss of function of this operon was beneficial. Therefore, while *hslUV* might have been less likely to be mutated than *rpoBC* based on length, more mutations may result in a fitness advantage. In contrast, *rpoBC* mutations must retain function, suggesting there may be fewer potentially adaptive mutations from which to sample. Second, as noted above, populations with mutations in *hslUV* did not exhibit antagonistic pleiotropy to the same degree as populations with *rpoBC* mutations; they did not have an associated fitness cost at

the acclimation temperature of 37.0°C, suggesting *hslUV* mutations are more likely to exist in batch culture and may reach frequencies that enhance the possibility of sampling into separate populations, perhaps explaining the high degree of parallelism across populations within weeks. By contrast, because individuals with *rpoBC* mutations are less fit than the ancestor at 37.0°C, new *rpoBC* mutations in batch culture would tend to be lost rapidly, making them less likely to be sampled and distributed to individual Lazarus tubes. This argument presumes that much of the initial selection to high temperature acts on standing variation that was generated at 37.0°C. This is consistent with the work of Luria and Delbrück (1943) describing the appearance of mutations conferring virus resistance in *E. coli* populations, even in the absence of a viral selective pressure (Luria and Delbrück, 1943).

If selection acts on standing variation and *rpoBC* mutations are deleterious at 37.0°C, how do any *rpoBC* mutations survive? There are two possibilities. The first is that the *rpoBC* mutations occur *de novo* after transfer to lethal temperature. Under this hypothesis, such mutations would be expected to be rare, because the population size decreases immediately after introduction to high temperature. If this is true, however, one might predict that *rpoBC* populations recover later than other populations, and this is not the case. The second possibility is that the population crash increases genetic drift, occasionally establishing a rare, highly advantageous allele for subsequent fixation.

Mechanism

We have shown that population recovery occurs rarely and that it tends to be accompanied by mutations in *hslUV* or *rpoBC*, but not both. We believe that the pattern of parallelism within and between weeks is driven in part by the antagonistic pleiotropy of *rpoB* mutations and the relative lack thereof for *hslUV* mutations. But we have failed to address

another important question: what is the molecular mechanism by which these mutations permit population recovery?

There are already some insights into the potential benefits of *rpoB* mutations at high temperature, because similar *rpoB* mutations have been shown to have large effects on gene expression at high temperature (Rodríguez-Verdugo *et al.*, 2016). For example, single mutants with *rpoB* affect the expression of thousands of genes at high temperature, and these shifts tend to move gene expression from a stress pattern to a state more like that of the ancestor (Rodríguez-Verdugo *et al.*, 2016). Accompanying these wholesale shifts in gene expression is a trend toward increasing transcriptional efficiency, suggesting that at least some adaptive *rpoB* mutations slow the polymerase and enhance termination at high temperature (Rodríguez-Verdugo *et al.*, 2016). We have no evidence, however, that the distinct Lazarus mutations in *rpoB* and *rpoC* confer these same benefits.

How mutations in *hslUV* enable population recovery is more of a mystery. Given that most mutations in the operon are frameshifts, they likely trigger a loss of function of the heat shock protease system. In the short term, this poses a problem, because these genes are normally upregulated upon the onset of heat shock conditions, although only weakly (Chuang *et al.*, 1993). Over a longer period of thermal stress, they are downregulated to below pre-stress levels (Rodríguez-Verdugo *et al.*, 2016). If *hslUV* plays a roll in the initial heat shock response and is downregulated later, the benefit from knock out mutations is perplexing. Perhaps the function of the *hslUV*-encoded protease system—to degrade both misfolded and properly folded proteins (Miller *et al.*, 2013)—becomes detrimental to the cell under extreme stress. Three lines of evidence support this idea. First, the *hslUV* protease system is known to target σ^{32}, the major heat shock sigma factor of the cell, and therefore the absence of the *hslUV* proteases may enhance the production of other heat shock proteins (Kanemori *et al.*, 1997). Second, we

observed three mutants (#22, #23, and #25) with large duplications centered on the *groEL* and *groES* genes, which encode a chaperonin complex important to the heat shock response (Richter *et al.*, 2010), and which are among the heat shock proteins upregulated in *hslUV* deletion mutants (Kanemori *et al.*, 1997). Lastly, our third-most common fixed mutation impacted the downstream region of *clpA*, a subunit of the *clpAP* protease system that has similar functions to, and overlapping substrate specificities with, *hslUV* (Kwon *et al.*, 2004). It is conceivable that altering the downstream region of *clpA* could result in its downregulation or a change in its function, which would produce similar outcomes to knocking out *hslUV*. Ultimately, we cannot yet ascribe a mechanism, but our hypothesis—i.e., that *hslUV* knockouts may indirectly lead to upregulation of other heat shock proteins through an effect on σ^{32}—merits further attention.

FIGURES

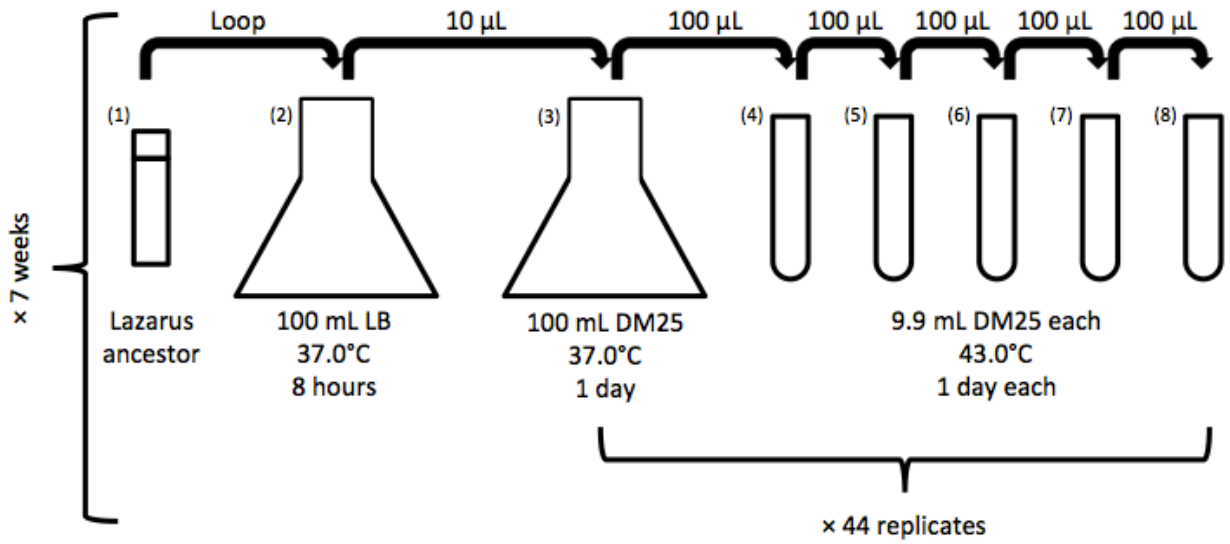

Figure 3.1: Experimental design for producing and observing Lazarus events. Bacteria were propagated from frozen through two flasks to acclimate them and to produce enough cells for experimental replication. Samples of flask culture were then propagated through culture tubes in 44 replicates for a total of five days. This procedure was repeated across seven different weeks.

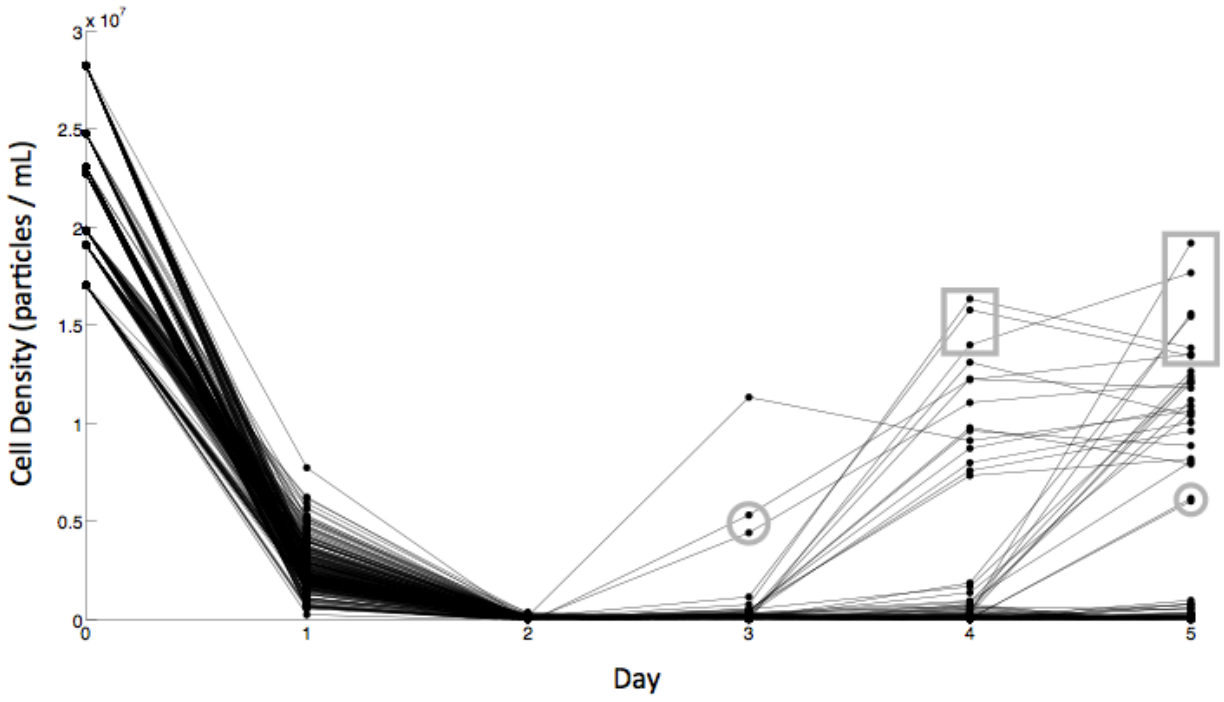

Figure 3.2: Population cell densities over time. Most populations went extinct over the course of five days. A total of 26 Lazarus events were observed across the third, fourth, and fifth days of growth. The timing of Lazarus events was determined by the day at which cell density increased by an order of magnitude over the previous day. Populations possessing *rpoBC* mutations are indicated by rectangles. Populations possessing duplications are circled.

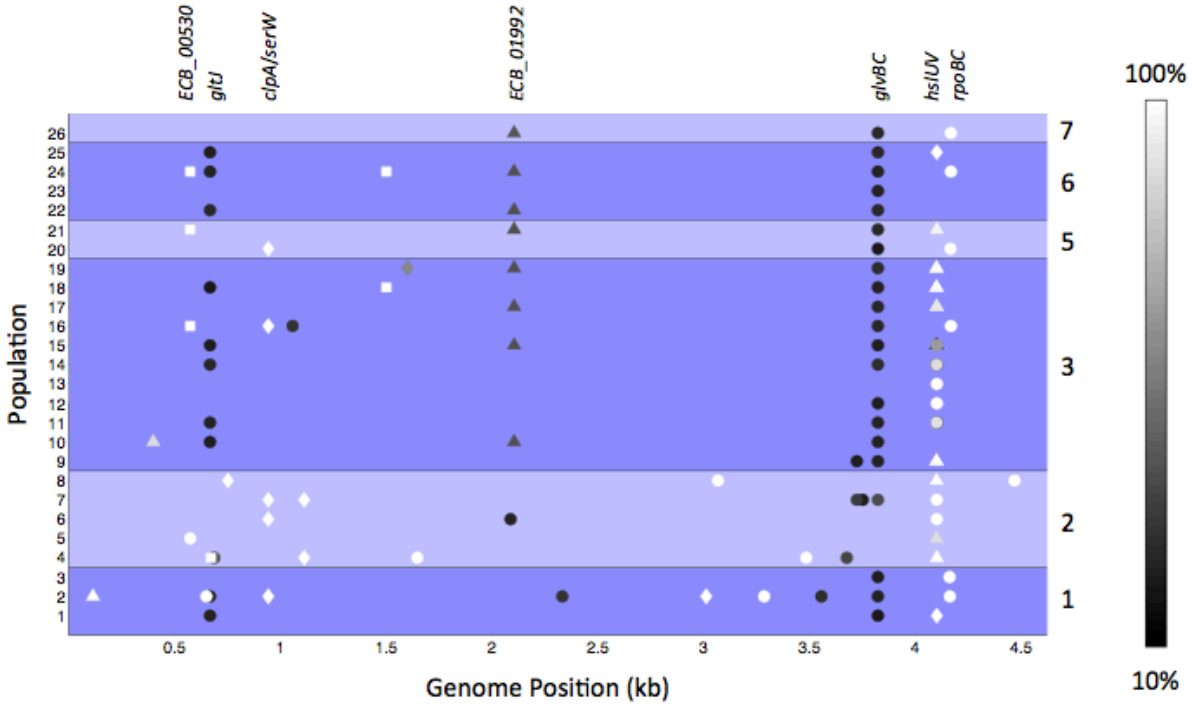

Figure 3.3: Genome-wide distribution of mutations in Lazarus populations. Different weeks are separated into groups and labeled at the right. Mutations are colored by their frequency in the population according to the scale at the right. Synonymous, nonsynonymous, indel, and intergenic mutations are represented by squares, circles, triangles, and diamonds, respectively. Only mutations at frequencies greater than 10% are shown. Mutations occurring in more than two populations are labeled at the top.

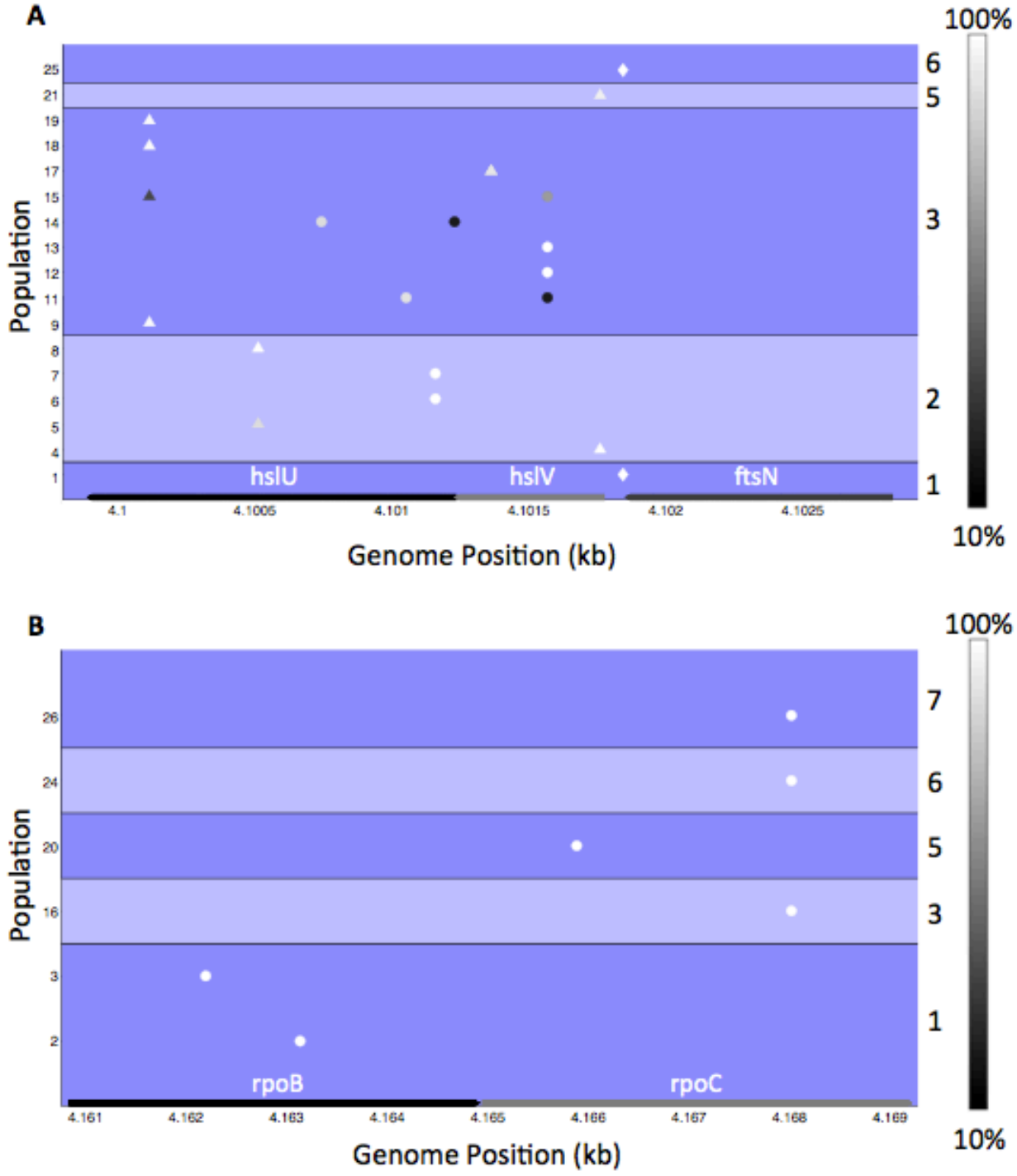

Figure 3.4: Distribution of mutations in the (A) *hslUV* and (B) *rpoBC* operons. Different weeks are separated into groups and labeled at the right. Mutations are colored by their

frequency in the population according to the scale at the right. Only mutations at frequencies greater than 10% are shown. Synonymous, nonsynonymous, indel, and intergenic mutations are represented by squares, circles, triangles, and diamonds, respectively, as in Figure 3.3.

TABLES

Table 3.1: Mutations present in populations at frequencies >10% and mean fitness values ($\bar{w}$) of populations relative to their ancestor at 42.2°C and 37.0°C.

Week	Population	Affected Region (Frequency in Population)	$\bar{w}_{42.2}$ (Standard Deviation)	*p*-value[a]	$\bar{w}_{37.0}$ (Standard Deviation)	*p*-value[a]
1	1	*hslV*/*ftsN* (1), *gltJ* (0.136), *glvBC* (0.108)	1.28 (0.06)	5.24E-05	0.96 (0.03)	1.58E-02
1	2	*mutT* (1), *mrdA* (1), *clpA*/*serW* (1), *ECB_02812*/*ECB_02813* (1), *arcB* (1), *rpoB* (1), *yfbM* (0.205), *yhhI* (0.178), *gltJ* (0.146), *glvBC* (0.136)	1.35 (0.13)	5.20E-04	0.95 (0.04)	1.56E-02
1	3	*rpoB* (1), *glvBC* (0.126)	1.33 (0.11)	3.71E-04	0.82 (0.02)	2.29E-06
2	4	*lnt* (1), *insE-1*/*serX* (1), *ynfL* (1), *rtcA* (1), *hslV* (1), *nagE* (0.352), *yiaN* (0.292)	1.12 (0.18)	7.35E-02	0.97 (0.03)	2.46E-02
2	5	*ECB_00530* (1), *hslU* (0.877)	1.34 (0.13)	6.97E-04	0.92 (0.06)	7.62E-03
2	6	*clpA*/*serW* (1), *hslU* (1), *yegM* (0.154)	1.33 (0.12)	6.06E-04	1.02 (0.04)	1.13E-01
2	7	*clpA*/*serW* (1), *insE-1*/*serX* (1), *hslU* (1), *glvBC* (0.309), *gpsA* (0.237), *dfp* (0.121)	1.13 (0.05)	4.52E-04	1.01 (0.02)	2.52E-01
2	8	*ybgG*/*cydA* (1), *yghS* (1), *hslU* (1), *pepA* (1)	1.05 (0.05)	2.96E-02	0.98 (0.05)	1.41E-01
3	9	*hslU* (0.986), *glvBC* (0.138), *gpsA* (0.117)	1.41 (0.12)	1.62E-04	1 (0.05)	4.09E-01
3	10	*secF* (0.872),	1.38 (0.24)	5.91E-03	0.95 (0.04)	1.78E-02

		ECB_01992 (0.332), *glvBC* (0.137), *gltJ* (0.116)				
3	11	*hslU* (0.893), *glvBC* (0.144), *gltJ* (0.136), *hslV* (0.106)	1.47 (0.19)	8.52E-04	1 (0.04)	4.69E-01
3	12	*hslV* (1), *glvBC* (0.122)	1.43 (0.06)	5.56E-06	0.97 (0.01)	6.53E-04
3	13	*hslV* (1)	1.38 (0.09)	8.31E-05	1.01 (0.03)	2.93E-01
3	14	*hslU* (0.874), *glvBC* (0.178), *gltJ* (0.165), *hslU* (0.111)	1.30 (0.04)	5.46E-06	1.03 (0.05)	7.82E-02
3	15	*hslV* (0.611), *ECB_01992* (0.326), *hslU* (0.286), *gltJ* (0.133), *glvBC* (0.133)	1.12 (0.10)	1.57E-02	1 (0.04)	3.84E-01
3	16	*ECB_00530* (1), *clpA*/*serW* (1), *rpoC* (1), *appB* (0.212), *glvBC* (0.157)	1.35 (0.28)	1.29E-02	0.96 (0.05)	5.28E-02
3	17	*hslV* (0.909), *ECB_01992* (0.329), *glvBC* (0.164)	1.17 (0.04)	1.20E-04	1.03 (0.03)	4.62E-02
3	18	*rhsE* (1), *hslU* (0.987), *glvBC* (0.135), *gltJ* (0.109)	1.20 (0.10)	2.49E-03	0.97 (0.04)	7.41E-02
3	19	*hslU* (0.989), *ydfJ*/*ydfK* (0.529), *ECB_01992* (0.318), *glvBC* (0.182)	1.08 (0.05)	3.85E-03	0.99 (0.02)	1.05E-01
5	20	*clpA*/*serW* (1), *rpoC* (1), *glvBC* (0.113)	1.24 (0.07)	1.79E-04	0.96 (0.05)	5.90E-02
5	21[b]	*ECB_00530* (1), *hslV* (0.956), *ECB_01992* (0.33), *glvBC* (0.171)	1.05 (0.06)	6.14E-02	0.95 (0.02)	8.25E-04

6	22[c]	*ECB_01992* (0.316), *gltJ* (0.173), *glvBC* (0.159)	1.12 (0.10)	1.90E-02	0.98 (0.03)	4.40E-02
6	23[c]	*glvBC* (0.137)	1.07 (0.09)	5.33E-02	0.93 (0.02)	1.89E-04
6	24	*ECB_00530* (1), *rhsE* (1), *rpoC* (1), *ECB_01992* (0.33), *glvBC* (0.144), *gltJ* (0.143)	1.14 (0.06)	1.51E-03	0.92 (0.02)	3.05E-04
6	25[c]	*hslV/ftsN* (1), *glvBC* (0.156), *gltJ* (0.121)	1.04 (0.07)	1.06E-01	0.96 (0.03)	1.04E-02
7	26	*rpoC* (1), *ECB_01992* (0.341), *glvBC* (0.155)	1.24 (0.11)	1.55E-03	0.92 (0.04)	1.36E-03

[a] Significance of relative fitness increases compared to a value of 1.0 was determined using one-tailed t-tests and six replicate fitness estimates per population.

[b] Population also contained a duplication of ~80 kb.

[c] Population also contained a duplication of ~20 kb.

Table 3.2: Mutations within specific genes that differ significantly in frequency between lethal and non-lethal high-temperature experiments.

Gene	Frequency in 115 Populations at 42.2°C	No. Observed across 26 Lazarus Populations	*p*-value[a]	No. Observed across 7 Lazarus Weeks	*p*-value[a]
rpoB	0.66	2	**7.92E-10**	1	**3.18E-11**
ybaL	0.56	0	**6.59E-10**	0	**6.59E-10**
cls	0.49	0	**2.91E-08**	0	**2.91E-08**
rho	0.39	0	**2.48E-06**	0	**2.48E-06**
iclR	0.32	0	**4.13E-05**	0	**4.13E-05**
rpoD	0.31	0	**5.76E-05**	0	**5.76E-05**
ECB_00503[b]	0.30	0	**7.98E-05**	0	**7.98E-05**
hslU	0.01	9	**7.77E-13**	5	**2.81E-06**
hslV	0.01	7	**2.14E-09**	2	0.0214

[a] Based on the cumulative binomial probability of observing the observed number of mutations and fewer (or greater, when appropriate) in Lazarus populations, given the frequency among 115 populations of the 42.2°C experiment. Bolded values are significant after Bonferroni correction for an experiment-wide significance value of 0.05.

[b] A recurring large deletion of ~76 kb

SUPPORTING INFORMATION

Table S3.1: Mutations present in Lazarus populations at >10% frequency.

week	sample	breseq evidence	genome position	mutation	population frequency	mutation type	gene(s)
1	1	RA	4101844	Δ1 bp	1	intergenic (-74/+19)	hslV/ftsN
1	1	RA	667294	T-->A	0.136	D203V (GAT-->GTT)	gltJ
1	1	RA	3823110	T-->A	0.108	I312F (ATT-->TTT)	glvBC
1	2	RA	114029	Δ1 bp	1	coding (182/390 nt)	mutT
1	2	RA	649901	T-->G	1	I301L (ATC-->CTC)	mrdA
1	2	RA	942604	A-->G	1	intergenic (+359/+339)	clpA/serW
1	2	RA	3012076	T-->G	1	intergenic (-516/-58)	ECB_02812/ECB_02813
1	2	RA	3285997	T-->G	1	E755A (GAA-->GCA)	arcB
1	2	RA	4163133	A-->C	1	N760H (AAC-->CAC)	rpoB
1	2	RA	2331921	T-->G	0.205	I122S (ATT-->AGT)	yfbM
1	2	RA	3555814	T-->G	0.178	F250C (TTC-->TGC)	yhhI
1	2	RA	667294	T-->A	0.146	D203V (GAT-->GTT)	gltJ
1	2	RA	3823110	T-->A	0.136	I312F (ATT-->TTT)	glvBC
1	3	RA	4162195	A-->T	1	H447L (CAC-->CTC)	rpoB
1	3	RA	3823110	T-->A	0.126	I312F (ATT-->TTT)	glvBC
2	4	RA	671609	A-->C	1	G456G (GGT-->GGG)	lnt
2	4	RA	1111967	A-->G	1	intergenic (-236/+166)	insE-1/serX
2	4	RA	1646703	T-->A	1	I29F (ATT-->TTT)	ynfL
2	4	RA	3484800	G-->A	1	S217F (TCC-->TTC)	rtcA
2	4	RA	4101760	+T	1	coding (11/531 nt)	hslV
2	4	RA	687648	T-->A	0.352	I537N (ATC-->AAC)	nagE
2	4	RA	3675965	A-->T	0.292	T236S (ACC-->TCC)	yiaN
2	5	RA	573229	T-->G	1	N87T (AAC-->ACC)	ECB_00530
2	5	RA	4100512	Δ1 bp	0.877	coding (719/1332 nt)	hslU
2	6	RA	942604	A-->G	1	intergenic (+359/+339)	clpA/serW
2	6	RA	4101159	C-->A	1	K24N (AAG--	hslU

						>AAT)	
2	6	RA	2087293	G-->C	0.154	G57A (GGC-->GCC)	yegM
2	7	RA	942604	A-->G	1	intergenic (+359/+339)	clpA/serW
2	7	RA	1111967	A-->G	1	intergenic (-236/+166)	insE-1/serX
2	7	RA	4101159	C-->A	1	K24N (AAG-->AAT)	hslU
2	7	RA	3823101	T-->G	0.309	T315P (ACC-->CCC)	glvBC
2	7	RA	3723692	G-->C	0.237	L246V (CTG-->GTG)	gpsA
2	7	RA	3751621	A-->T	0.121	H179L (CAT-->CTT)	dfp
2	8	RA	751779	T-->G	1	intergenic (+286/-561)	ybgG/cydA
2	8	RA	3067389	T-->G	1	H219P (CAC-->CCC)	yghS
2	8	RA	4100512	Δ1 bp	1	coding (719/1332 nt)	hslU
2	8	RA	4468692	T-->G	1	T163P (ACC-->CCC)	pepA
3	9	RA	4100115	+C	0.986	coding (1116/1332 nt)	hslU
3	9	RA	3823110	T-->A	0.138	I312F (ATT-->TTT)	glvBC
3	9	RA	3723691	A-->C	0.117	L246R (CTG-->CGG)	gpsA
3	10	JC	398683	+GGT	0.872	coding (756/972 nt)	secF
3	10	JC	2103887	+CAGC	0.332	coding (154/216 nt)	ECB_01992
3	10	RA	3823110	T-->A	0.137	I312F (ATT-->TTT)	glvBC
3	10	RA	667294	T-->A	0.116	D203V (GAT-->GTT)	gltJ
3	11	RA	4101052	C-->T	0.893	G60D (GGT-->GAT)	hslU
3	11	RA	3823110	T-->A	0.144	I312F (ATT-->TTT)	glvBC
3	11	RA	667294	T-->A	0.136	D203V (GAT-->GTT)	gltJ
3	11	RA	4101568	T-->G	0.106	H68P (CAT-->CCT)	hslV
3	12	RA	4101568	T-->G	1	H68P (CAT-->CCT)	hslV
3	12	RA	3823110	T-->A	0.122	I312F (ATT-->TTT)	glvBC
3	13	RA	4101568	T-->G	1	H68P (CAT-->CCT)	hslV
3	14	RA	4100743	A-->C	0.874	L163R (CTG-->CGG)	hslU
3	14	RA	3823110	T-->A	0.178	I312F (ATT-->TTT)	glvBC
3	14	RA	667294	T-->A	0.165	D203V (GAT-->GTT)	gltJ
3	14	RA	4101229	A-->C	0.111	M1R (ATG-->AGG) †	hslU
3	15	RA	4101568	T-->G	0.611	H68P (CAT-->CCT)	hslV

3	15	JC	2103887	+CAGC	0.326	coding (154/216 nt)	ECB_01992
3	15	RA	4100115	+C	0.286	coding (1116/1332 nt)	hslU
3	15	RA	667294	T-->A	0.133	D203V (GAT-->GTT)	gltJ
3	15	RA	3823110	T-->A	0.133	I312F (ATT-->TTT)	glvBC
3	16	RA	573246	G-->C	1	L81L (CTC-->CTG)	ECB_00530
3	16	RA	942604	A-->G	1	intergenic (+359/+339)	clpA/serW
3	16	RA	4168018	T-->G	1	W1020G (TGG-->GGG)	rpoC
3	16	RA	1057332	T-->C	0.212	S316P (TCA-->CCA)	appB
3	16	RA	3823110	T-->A	0.157	I312F (ATT-->TTT)	glvBC
3	17	RA	4101363	Δ1 bp	0.91	coding (408/531 nt)	hslV
3	17	RA	4101362	Δ1 bp	0.908	coding (409/531 nt)	hslV
3	17	JC	2103887	+CAGC	0.329	coding (154/216 nt)	ECB_01992
3	17	RA	3823110	T-->A	0.164	I312F (ATT-->TTT)	glvBC
3	18	RA	1500351	T-->G	1	G25G (GGT-->GGG)	rhsE
3	18	RA	4100115	+C	0.987	coding (1116/1332 nt)	hslU
3	18	RA	3823110	T-->A	0.135	I312F (ATT-->TTT)	glvBC
3	18	RA	667294	T-->A	0.109	D203V (GAT-->GTT)	gltJ
3	19	RA	4100115	+C	0.989	coding (1116/1332 nt)	hslU
3	19	RA	1600384	A-->T	0.529	intergenic (-332/-421)	ydfJ/ydfK
3	19	JC	2103887	+CAGC	0.318	coding (154/216 nt)	ECB_01992
3	19	RA	3823110	T-->A	0.182	I312F (ATT-->TTT)	glvBC
5	20	RA	942604	A-->G	1	intergenic (+359/+339)	clpA/serW
5	20	RA	4165883	A-->G	1	D308G (GAT-->GGT)	rpoC
5	20	RA	3823110	T-->A	0.113	I312F (ATT-->TTT)	glvBC
5	21	RA	573246	G-->C	1	L81L (CTC-->CTG)	ECB_00530
5	21	RA	4101760	Δ1 bp	0.956	coding (11/531 nt)	hslV
5	21	JC	2103887	+CAGC	0.33	coding (154/216 nt)	ECB_01992
5	21	RA	3823110	T-->A	0.171	I312F (ATT-->TTT)	glvBC
6	22	JC	2103887	+CAGC	0.316	coding (154/216 nt)	ECB_01992
6	22	RA	667294	T-->A	0.173	D203V (GAT-->GTT)	gltJ
6	22	RA	3823110	T-->A	0.159	I312F (ATT-->TTT)	glvBC

6	23	RA	3823110	T-->A	0.137	I312F (ATT-->TTT)	glvBC
6	24	RA	573246	G-->C	1	L81L (CTC-->CTG)	ECB_00530
6	24	RA	1500351	T-->G	1	G25G (GGT-->GGG)	rhsE
6	24	RA	4168018	T-->G	1	W1020G (TGG-->GGG)	rpoC
6	24	JC	2103887	+CAGC	0.33	coding (154/216 nt)	ECB_01992
6	24	RA	3823110	T-->A	0.144	I312F (ATT-->TTT)	glvBC
6	24	RA	667294	T-->A	0.143	D203V (GAT-->GTT)	gltJ
6	25	RA	4101844	Δ1 bp	1	intergenic (-74/+19)	hslV/ftsN
6	25	RA	3823110	T-->A	0.156	I312F (ATT-->TTT)	glvBC
6	25	RA	667294	T-->A	0.121	D203V (GAT-->GTT)	gltJ
7	26	RA	4168018	T-->G	1	W1020G (TGG-->GGG)	rpoC
7	26	JC	2103887	+CAGC	0.341	coding (154/216 nt)	ECB_01992
7	26	RA	3823110	T-->A	0.155	I312F (ATT-->TTT)	glvBC

Table S3.2: Cell densities over time for all 296 populations. t0 measurements were made from the acclimation flask. Negative values were indistinguishable from background particle counts. Lazarus populations are highlighted in yellow.

week	Lazarus population	t0 density (particles/mL)	t1 density (particles/mL)	t2 density (particles/mL)	t3 density (particles/mL)	t4 density (particles/mL)	t5 density (particles/mL)
7		28257000	2879225	-192065	-116375	548275	-152200
7		28257000	1562225	-120425	333325	663875	-112400
7		28257000	3624225	-128425	-64675	-140325	-9000
7		28257000	3241225	-40825	-132275	281575	-243800
7		28257000	1514225	-243805	-191975	-140325	-180100
7		28257000	3635225	-132325	-48775	-72625	-172100
7		28257000	2187225	-148325	-172075	-247725	-180100
7	26	28257000	2063225	-192065	-207935	297475	19190800
7		28257000	3249225	-160225	-164175	-263625	-184100
7		28257000	1864225	-215945	-56675	-231825	-204000
7		28257000	1769225	-192065	-68675	-223825	-200000
7		28257000	3480225	78575	-132275	-259725	-287560
7		28257000	2744225	-180125	-191975	-223825	-196000
7		28257000	7731225	-136325	-184075	-92525	349200
7		28257000	1900225	-227885	-72575	-239825	-192000
7		28257000	2839225	-211965	-195975	-196025	-247800
7		28257000	1940225	-196045	-223855	-299485	-164200
7		28257000	2509225	-172125	-136275	86575	-227900
7		28257000	1813225	-223905	-188075	-275605	-247800
7		28257000	3576225	-184105	-211915	-223825	-231800
7		28257000	3050225	-219925	-227835	-235825	-204000
7		28257000	2935225	-116425	-136275	-235825	-231800
7		28257000	2573225	-172125	3025	-124325	-148300
7		28257000	2875225	-172125	-104475	-251725	62700
7		28257000	2903225	-180125	-211915	-188025	-259700
7		28257000	3158225	-196045	-156175	-227825	-219900
7		28257000	1435225	-136325	-84575	-152225	-219900
7		28257000	2326225	-184105	-148275	-267625	-247800
7		28257000	2079225	-152225	-168175	-251725	-160200
7		28257000	1864225	-188085	-199975	-251725	-235800
7		28257000	2207225	-96525	-207935	-291525	-271640
7		28257000	1458225	-176125	-203955	420875	-247800
7		28257000	3210225	-160225	-120375	-227825	-208000
7		28257000	1371225	70575	-132275	-211925	-279600

7		28257000	2816225	-76625	-195975	-259725	-255700
7		28257000	1594225	-180125	-188075	-196025	-156200
7		28257000	4201225	-28925	-191975	-203925	-223900
7		28257000	2334225	66675	-203955	-148225	-259700
7		28257000	2883225	98475	-235795	-227825	-239800
7		28257000	3297225	-136325	-223855	38775	-235800
7		28257000	3174225	30775	-199975	-247725	-231800
7		28257000	3150225	-156225	-219875	-207925	-176100
7		28257000	2326225	-184105	-8975	-239825	-136300
7		28257000	3444225	-128425	-180075	-291525	-259700
6		24794900	2039930	-66675	-135300	-37775	-371125
6		24794900	2500930	84525	-75600	-41775	66675
6		24794900	5629930	-130355	-87500	-93535	-88525
6		24794900	2958930	235825	99500	-65675	774975
6		24794900	1748930	-110475	-119380	-101495	-311425
6		24794900	2282930	-86575	-7900	-53775	-291525
6		24794900	2083930	-66675	-115400	654725	-259725
6		24794900	3543930	-30875	-99480	-101495	-132325
6		24794900	2134930	-86575	-139280	-109455	-172125
6		24794900	5135930	-26875	-139280	-37775	-263625
6		24794900	2660930	-86575	-119380	-41775	-319385
6		24794900	3396930	-130355	-67600	-89555	-291525
6		24794900	1880930	-102475	-111420	-77575	-251725
6		24794900	4709930	40825	-51700	-89555	-307445
6		24794900	2588930	-74675	-123360	770125	-315405
6		24794900	2114930	-146275	-79600	-69675	-275625
6		24794900	2361930	-74675	-11900	2025	-200025
6		24794900	2727930	-106475	-107440	-141295	-327345
6		24794900	1888930	-118375	-139280	5925	-267625
6		24794900	2011930	-154235	-71600	-37775	-287525
6		24794900	2190930	-138315	-99480	-85575	-255725
6		24794900	2433930	-134335	-27800	25825	-343265
6		24794900	1267930	-134335	-103460	-93535	-267625
6		24794900	1896930	-114475	-119380	-121395	-247725
6		24794900	2114930	-130355	127400	-133335	-215925
6		24794900	2397930	-110475	-103460	-5975	-196025
6	22	24794900	2190930	-78575	4414000	11054925	12038975
6		24794900	2282930	-66675	-91500	5925	-156215
6		24794900	2230930	-138315	-103460	-109455	-295505
6	23	24794900	1903930	-138315	-151220	41825	6028975
6		24794900	1267930	-114475	-123360	-77575	-207925

6		24794900	2035930	-134335	-107440	-117415	-283525
6		24794900	2238930	-118375	-147240	-93535	-271625
6		24794900	2640930	-74675	-127340	-73675	-200025
6		24794900	2445930	-58675	35800	-89555	-203925
6	24	24794900	2898930	247725	15900	1852925	15448975
6		24794900	2624930	-58675	-95500	-109455	-80625
6		24794900	1919930	-162195	-43800	-49775	-327345
6		24794900	2071930	-70675	-87500	-117415	-311425
6		24794900	2298930	-94575	-135300	-49775	-307445
6		24794900	3082930	-54775	-103460	-113435	-319385
6		24794900	2457930	68625	-115400	-45775	-251725
6		24794900	2114930	-90575	143300	-105475	-287525
6	25	24794900	2524930	-158215	5317000	12214925	13518975
5	20	23059600	2029865	-7975	43750	377070	15591050
5		23059600	1886865	-79575	27850	10970	965050
5		23059600	1094865	-79575	-55750	-52730	-77650
5		23059600	1066865	-103475	-47750	-12930	-89550
5		23059600	1663865	-103475	-59650	-36830	-81550
5		23059600	1842865	-87575	-23850	146270	-109450
5		23059600	5241865	187025	-95550	54770	-125370
5		23059600	2029865	-99475	-23850	611970	503450
5		23059600	1973865	-119415	-39750	50770	-1950
5		23059600	1547865	-19875	-39750	26870	-109450
5		23059600	942865	-115435	43750	-68630	-61650
5		23059600	1925865	-131355	-95550	-84570	-41750
5		23059600	2188865	-99475	-79550	-12930	-73650
5		23059600	684565	-115435	-83550	-1030	-85550
5		23059600	569165	-95575	-19850	62670	567150
5		23059600	636765	-107475	95550	18870	193050
5		23059600	756165	-83575	262650	476570	-89550
5		23059600	5125865	31825	-107450	325370	37850
5		23059600	2200865	-95575	15950	62670	-61650
5		23059600	1512865	-99475	-83550	-44730	-113430
5		23059600	1834865	-15975	-95550	10970	-13950
5		23059600	2264865	-107475	-51750	-12930	21850
5		23059600	982865	-135335	-59650	114470	-101450
5		23059600	1878865	-27875	-67650	-8930	-141290
5		23059600	1810865	-119415	79650	-40830	-101450
5		23059600	2208865	-31875	-67650	50770	121350
5		23059600	1571865	-31875	87550	-28830	-97550
5		23059600	1818865	7925	-87550	22870	-113430

5		23059600	1058865	-11975	-3950	194070	770150
5		23059600	1488865	15925	-39750	-56730	-45750
5		23059600	1054865	258725	-39750	146270	320350
5	21	23059600	668665	-67675	-51750	29270	6167050
5		23059600	1953865	282525	-7950	-8930	9950
5		23059600	2296865	-59675	31850	6970	-41750
5		23059600	2534865	7925	-119350	-48730	272650
5		23059600	644765	-91575	-67650	-16930	-81550
5		23059600	1396865	-99475	-15950	-28830	49750
5		23059600	2196865	-39775	50	18870	-121390
5		23059600	2379865	-3975	-63650	22870	-73650
5		23059600	696465	-27875	322350	-4930	-117410
5		23059600	1858865	-43775	-119350	-1030	-125370
5		23059600	1826865	19925	-103450	6970	-133330
5		23059600	2574865	-75675	-15950	10970	-33850
5		23059600	1906865	83525	-63650	-44730	-41750
4		17037500	2362525	22900	82575	76650	294515
4		17037500	1586525	-64700	-36825	-134350	-51685
4		17037500	1315525	-104480	-60625	-70650	15915
4		17037500	2067525	-56700	126375	-110450	147315
4		17037500	1753525	-40800	-68625	-182090	-95485
4		17037500	2031525	22900	-12925	64650	250715
4		17037500	1785525	-36800	-84525	-106450	103515
4		17037500	1956525	162200	34875	351250	15915
4		17037500	1701525	-44800	-68625	-122350	31815
4		17037500	1741525	-1000	-92525	203950	135315
4		17037500	750525	-12900	-16925	-118450	-59685
4		17037500	1745525	-16900	-40725	-134350	-75585
4		17037500	1741525	-5000	-144245	-130350	131315
4		17037500	2015525	82600	-104425	20850	8015
4		17037500	5160525	86600	-88525	-38850	-23885
4		17037500	2998525	126400	18975	-30850	-75585
4		17037500	1641525	-60700	-72625	-10950	-63685
4		17037500	2127525	-68700	42775	20850	35815
4		17037500	1530525	-64700	6975	-90550	43815
4		17037500	1944525	7000	-116425	-66650	-95485
4		17037500	543325	14900	-56725	-110450	83615
4		17037500	1522525	-60700	62675	-78650	8015
4		17037500	1792525	166200	-116425	-118450	39815
4		17037500	1852525	3000	-124325	-78650	-59685
4		17037500	634825	-12900	-136285	60650	-39785

4		17037500	638825	-5000	158275	-34850	-39785
4		17037500	236825	-68700	-60625	-114450	-91485
4		17037500	1729525	-72600	-104425	-142250	-27885
4		17037500	2115525	-60700	-60625	-150250	79615
4		17037500	2099525	14900	-24825	-114450	-35785
4		17037500	658725	-48800	-48725	-178110	-43785
4		17037500	1570525	-5000	-64625	-18950	-35785
4		17037500	2242525	30800	-84525	-90550	-83585
3		19842800	5918660	192000	94520	86560	37800
3		19842800	4298660	-46800	-60680	54760	-133300
3		19842800	5699660	-26900	-24880	205960	153200
3	9	19842800	4039660	-10900	321420	937860	8041400
3		19842800	4684660	48800	126420	26860	-173100
3		19842800	4198660	32800	14920	209960	-101500
3	10	19842800	5305660	28900	118420	12257860	11789400
3		19842800	6156660	287600	-52720	114460	-21900
3	11	19842800	4763660	28900	245820	13107860	10389400
3		19842800	2873660	-18900	-52720	-32840	-6000
3		19842800	4927660	36800	-20880	-60740	-125400
3		19842800	3151660	-42800	-60680	-24840	-169200
3		19842800	3784660	-7000	-88540	-16940	-133300
3		19842800	2566660	100500	-88540	-32840	-165200
3	12	19842800	4433660	92500	464720	8718860	10879400
3		19842800	3661660	-14900	10920	-60740	-101500
3		19842800	4541660	5000	126420	201960	-109500
3		19842800	2753660	-30800	102520	18860	-201000
3		19842800	4166660	100500	26920	-64640	93500
3	13	19842800	4047660	104500	34820	492560	12089400
3		19842800	4095660	24900	-60680	-124380	-77600
3		19842800	3951660	-7000	237820	66660	-129400
3		19842800	3339660	12900	50720	-52740	-181100
3	14	19842800	3928660	44800	524420	7333860	8180400
3		19842800	2495660	92500	-48740	154260	-45800
3		19842800	2085660	80600	18920	-100500	-185080
3		19842800	2164660	124400	14920	-96520	-201000
3		19842800	3116660	-18900	90520	-96520	-33800
3		19842800	3999660	12900	22920	30860	-125400
3	15	19842800	4584660	28900	512420	1666860	12219400
3	19	19842800	5026660	16900	393020	9618860	8857400
3		19842800	3283660	52700	-76600	130360	9900
3		19842800	3585660	9000	-32780	46760	-177100

3	16	19842800	4246660	64700	751220	16347860	13839400
3		19842800	4771660	44800	54720	42760	121400
3		19842800	2415660	-50700	98520	-44740	-232840
3	17	19842800	3454660	-50700	416920	7584860	9601400
3		19842800	4974660	355200	114420	-16940	-145300
3		19842800	3207660	124400	3020	10960	595000
3		19842800	3991660	40800	46820	-12940	-208960
3		19842800	4636660	-22900	3020	134360	-153200
3	18	19842800	4326660	28900	245820	7986860	10039400
3		19842800	2113660	24900	-16880	86560	-49800
3		19842800	6232660	40800	58720	-68640	-165200
2	4	19101100	3529775	70660	153325	278625	12402630
2	5	19101100	3772775	154260	11326725	9109925	10622630
2		19101100	4027775	34860	266625	-11975	43730
2		19101100	4214775	66660	183125	-43775	131330
2		19101100	3617775	134360	107425	75625	27830
2		19101100	3653775	30860	226825	159225	55730
2		19101100	1926775	34860	51725	-55775	151230
2		19101100	2268775	78560	-39775	51725	91530
2		19101100	3446775	186060	163225	-55775	-11970
2		19101100	2901775	62660	-59695	-11975	59730
2		19101100	3959775	158160	274625	75625	91530
2		19101100	3243775	74660	11925	25	30
2		19101100	3983775	26860	119425	167125	199030
2		19101100	3466775	90560	-51735	-7975	79630
2		19101100	3617775	126360	35825	-15975	43730
2		19101100	2658775	66660	-63675	95525	-23870
2		19101100	2495775	22860	-15875	-11975	-7970
2		19101100	3064775	142260	83625	-43775	-23870
2		19101100	2960775	201960	-31875	-23875	155230
2		19101100	2901775	182060	143325	83525	31830
2		19101100	2089775	237760	-67655	25	63630
2		19101100	3709775	50760	226825	-47775	-39810
2	6	19101100	2730775	90560	107425	1360925	10522630
2		19101100	3311775	14960	19925	67625	115430
2		19101100	2654775	178060	-15875	-43775	-3970
2		19101100	3728775	205960	-63675	-103495	-31850
2		19101100	2710775	325360	-7975	107425	-43790
2		19101100	2089775	18860	-59695	-59675	-31850
2		19101100	2427775	142260	19925	119425	11930
2		19101100	3044775	110460	-31875	-19875	-55730

2		19101100	1782775	-1000	-67655	11925	-31850
2		19101100	2944775	38760	234825	226825	147230
2		19101100	2845775	62660	-67655	-71675	-51750
2		19101100	4417775	74660	-79595	131325	-27870
2		19101100	2383775	54760	143325	199025	167130
2		19101100	3028775	122360	-7975	-107475	39830
2		19101100	2053775	34860	-79595	-71675	-3970
2		19101100	1651775	-12940	-71635	-11975	-15970
2		19101100	3709775	50760	-67655	-95535	95530
2		19101100	2889775	42760	-11975	-83575	-43790
2		19101100	3943775	34860	155225	222825	23830
2		19101100	2785775	22860	-47755	63625	-7970
2	7	19101100	3406775	138260	238825	692525	11172630
2	8	19101100	3283775	106460	214925	871925	12642630
1		22748900	1880825	-35825	100525	1950	38775
1		22748900	1713825	63675	40825	85550	86575
1		22748900	2218825	-7925	-54675	-65670	10975
1		22748900	948825	147275	-126355	9950	-36825
1		22748900	2532825	55675	92525	41750	114375
1		22748900	1430825	131375	-42775	21850	2975
1		22748900	2011825	-25	-70675	-45770	-52745
1		22748900	1995825	155175	76625	-29850	22875
1		22748900	2803825	119375	259725	-33830	70675
1		22748900	1319825	-67665	-14875	53750	26875
1		22748900	1979825	-39825	60725	25850	142275
1		22748900	2095825	31875	263725	-41790	763175
1		22748900	865825	-23925	-74575	41750	-5025
1		22748900	718425	-59705	136325	-45770	90575
1		22748900	579125	-4025	-50775	-69650	-16925
1		22748900	1971825	-95525	-66675	-29850	10975
1		22748900	2206825	-79605	160225	-25850	-48765
1		22748900	1908825	119375	76625	25850	245775
1		22748900	1629825	-7925	-30875	-45770	18875
1		22748900	996825	119375	-10975	-1950	-16925
1		22748900	877825	71675	124425	-53730	14875
1		22748900	1677825	-25	108425	57750	6975
1		22748900	2055825	43775	164225	-61690	2975
1		22748900	1808825	-39825	-82575	5950	78575
1		22748900	2314825	67675	-78575	33850	134275
1	1	22748900	658725	83575	379125	9772650	7919675
1		22748900	782125	-43825	-138295	109450	122375

1		22748900	1880825	-23925	112425	-9950	-40805
1		22748900	1279825	-51725	-70675	-9950	-20925
1		22748900	2385825	115375	-118395	-69650	-84585
1		22748900	2119825	79575	-14875	29850	-36825
1		22748900	2660825	75575	-70675	-13950	30875
1		22748900	2903825	-79605	8925	53750	-28825
1		22748900	1076825	-83585	-106455	-13950	-64685
1		22748900	1991825	51775	116425	-49750	-24925
1	2	22748900	1665825	-7925	-50775	13990650	17673675
1	3	22748900	2166825	107475	1143025	15780650	13433675
1		22748900	2302825	-15925	-46775	141250	174075
1		22748900	1995825	-59705	-62675	21850	-1025
1		22748900	2218825	-67665	-26875	9950	50775
1		22748900	1876825	-71645	-58675	-53730	62675
1		22748900	1701825	-43825	128325	69650	-36825
1		22748900	2019825	-7925	-110435	-1950	-1025

Table S3.3: Colony counts for competitions to determine fitness of Lazarus populations relative to the ancestor (Ara+). Replicates for a given population are designated A-F. Competitions performed at 42.2°C are shown under the orange header. Competitions performed at 37.0°C are shown under the blue header. t0 and t1 represent the initial and final time points of the one-day competition, respectively.

42.2°C Competitions		t0			t1			
Lazarus Population		Ara+ Colonies	Lazarus Colonies		Ara+ Colonies	Lazarus Colonies		Fitness (w)
1A		121	27		92	66		1.269631607
1B		138	33		75	50		1.256615065
1C		120	28		82	53		1.241184655
1D		124	30		93	59		1.223283101
1E		139	24		76	64		1.396000931
1F		113	30		71	67		1.306295072
2A		115	64		68	215		1.42580579
2B		102	68		48	237		1.519895383
2C		96	48		92	109		1.189081771
2D		124	94		56	228		1.441174463
2E		98	84		70	170		1.2439746
2F		126	78		73	159		1.309906401
3A		108	76		84	157		1.224359039
3B		89	77		40	137		1.361572403
3C		100	88		44	168		1.387826155
3D		108	106		80	178		1.190113026
3E		120	93		54	157		1.347326303
3F		126	82		42	155		1.494877943
4A		126	49		104	28		0.916677653
4B		130	29		68	59		1.343244661
4C		107	30		63	69		1.3343421
4D		118	38		141	46		1.002713633
4E		127	57		74	40		1.045743659
4F		101	52		80	64		1.100806394

5A		137	34		70	75		1.371817636
5B		115	30		42	62		1.481724288
5C		109	32		92	52		1.147683631
5D		123	27		69	50		1.296557389
5E		120	28		52	70		1.464997334
5F		139	42		61	48		1.253104248
6A		115	54		64	82		1.249752175
6B		102	95		40	90		1.240394568
6C		124	46		65	103		1.366729151
6D		110	42		30	62		1.510830421
6E		123	77		31	72		1.406281005
6F		116	57		59	61		1.189323774
7A		111	34		76	47		1.166238021
7B		113	39		66	45		1.167386465
7C		136	35		75	34		1.141191709
7D		126	30		99	31		1.062775291
7E		149	27		142	38		1.085552816
7F		123	38		80	49		1.163925704
8A		127	36		80	34		1.097755512
8B		116	39		91	38		1.049686694
8C		113	48		84	38		1.014611749
8D		91	45		92	59		1.056312889
8E		119	40		106	32		0.976064263
8F		136	37		87	39		1.120091306
9A		116	19		70	40		1.304759145
9B		103	24		66	77		1.387208819
9C		147	27		55	80		1.571300036
9D		106	34		49	71		1.393352794
9E		101	24		37	60		1.533327027
9F		107	29		55	47		1.291482814
10A		123	66		57	175		1.454704758
10B		109	79		76	202		1.306141442
10C		87	77		79	107		1.094369273
10D		95	43		28	204		1.821218875
10E		128	73		77	147		1.294902149
10F		118	85		50	126		1.334255814

11A		136	44		43	92		1.546962759
11B		111	44		33	87		1.558566837
11C		102	44		44	78		1.375440243
11D		126	54		22	86		1.772956729
11E		119	46		60	63		1.254890952
11F		116	38		64	79		1.330776964
12A		130	52		52	115		1.463549748
12B		110	54		34	86		1.477837084
12C		112	43		47	107		1.476335629
12D		118	47		55	96		1.384597017
12E		99	54		53	107		1.328783969
12F		111	40		57	111		1.428347787
13A		121	38		47	93		1.50297652
13B		110	41		69	100		1.328107951
13C		131	32		70	89		1.414632593
13D		128	42		54	95		1.448743415
13E		84	46		65	102		1.242084095
13F		117	28		70	72		1.356383559
14A		136	70		63	102		1.298775363
14B		125	51		82	124		1.313141322
14C		117	39		69	99		1.358005604
14D		110	33		96	84		1.2395239
14E		117	42		77	83		1.262620393
14F		120	39		73	81		1.298899366
15A		130	48		78	35		1.047619997
15B		128	28		82	44		1.21570349
15C		102	36		154	59		1.016351657
15D		117	39		109	49		1.06595986
15E		140	34		224	91		1.101374923
15F		146	48		107	106		1.256850275
16A		118	56		65	162		1.413717788
16B		113	47		125	160		1.238860463
16C		131	74		128	207		1.229554789
16D		107	77		62	118		1.239581097
16E		132	59		26	163		1.886075937

16F		124	86		160	198		1.119139927
17A		99	37		82	58		1.144432642
17B		109	37		141	80		1.105642783
17C		129	36		155	93		1.159846011
17D		116	41		110	98		1.203095997
17E		142	27		123	66		1.23253484
17F		124	56		79	72		1.169015713
18A		104	57		167	92		1.001010952
18B		140	27		71	45		1.303037505
18C		134	26		105	55		1.22771151
18D		119	22		122	63		1.221853066
18E		126	25		155	95		1.234369162
18F		95	24		64	40		1.215149996
19A		113	34		154	56		1.038542543
19B		109	34		183	73		1.048008167
19C		123	37		54	29		1.153247711
19D		132	39		155	75		1.103509161
19E		111	50		110	60		1.041637594
19F		95	39		61	40		1.112518507
20A		142	79		83	104		1.199580341
20B		154	56		82	107		1.321441712
20C		138	59		86	110		1.265193454
20D		146	80		111	146		1.202179046
20E		161	70		88	124		1.293882457
20F		141	73		136	134		1.140835234
21A		119	46		156	52		0.969619945
21B		127	40		121	45		1.036468692
21C		135	47		101	34		0.992205634
21D		147	45		105	55		1.125832968
21E		138	37		109	36		1.047721266
21F		125	41		86	47		1.120661166
22A		117	43		118	73		1.112870537
22B		117	31		143	74		1.139290336
22C		126	43		105	57		1.104948812
22D		114	29		78	70		1.29833988

22E		103	28		151	40		0.994812079
22F		121	45		120	57		1.053229155
23A		166	40		73	42		1.230021378
23B		146	41		108	35		1.033285676
23C		119	42		95	31		0.982091968
23D		125	31		103	41		1.107256137
23E		111	32		110	44		1.071256525
23F		111	44		110	45		1.006858544
24A		160	73		106	139		1.251762361
24B		118	89		114	139		1.105087934
24C		109	64		111	133		1.154278465
24D		112	67		99	106		1.129886796
24E		116	63		90	85		1.127154103
24F		99	48		113	74		1.063451195
25A		96	42		132	85		1.078504759
25B		90	32		115	62		1.085824929
25C		119	41		174	52		0.97146294
25D		103	55		142	77		1.003120855
25E		102	54		114	52		0.968415297
25F		115	44		88	66		1.155169753
26A		139	53		82	139		1.365902544
26B		121	94		156	148		1.041128285
26C		96	66		55	103		1.247542203
26D		94	38		96	113		1.231019584
26E		118	60		99	144		1.237274749
26F		127	58		85	140		1.3051491
37.0°C Competitions		**t0**			**t1**			
Lazarus Population		**Ara+ Colonies**	**Lazarus Colonies**		**Ara+ Colonies**	**Lazarus Colonies**		**Fitness (w)**
1A		153	164		143	151		0.996695747
1B		148	181		136	117		0.922186278
1C		151	150		144	116		0.954016632
1D		121	152		113	113		0.94972412

1E		144	140		143	137		0.996804658
1F		154	160		140	127		0.969915574
2A		134	213		152	151		0.900648468
2B		122	123		141	122		0.967809456
2C		123	106		153	112		0.9661664
2D		135	154		142	125		0.94432848
2E		146	123		127	112		1.010241067
2F		146	177		126	98		0.900432273
3A		149	197		184	102		0.819520026
3B		119	164		186	97		0.807636901
3C		114	162		225	101		0.781956094
3D		119	151		136	71		0.812579176
3E		131	159		170	88		0.824864662
3F		132	154		171	95		0.847464537
4A		112	115		140	143		0.998916589
4B		108	123		105	121		1.002573098
4C		113	125		137	113		0.938821695
4D		114	124		112	102		0.961283855
4E		97	105		128	109		0.950858726
4F		125	108		172	136		0.981996189
5A		117	149		92	123		1.011139932
5B		106	182		102	117		0.911672296
5C		133	167		153	94		0.849368006
5D		125	148		128	119		0.947761671
5E		130	164		131	101		0.893252416
5F		126	172		143	116		0.89000499
6A		129	103		103	110		1.066399322
6B		120	115		147	136		0.992675276
6C		115	141		111	116		0.965038102
6D		132	110		139	125		1.016354722
6E		127	101		113	105		1.034676078
6F		135	103		126	123		1.054329499
7A		127	134		141	135		0.979375132
7B		117	103		136	130		1.017310881
7C		124	132		140	143		0.991258251

7D		111	108		143	152		1.01820215
7E		112	110		112	127		1.031205518
7F		124	121		135	130		0.997175086
8A		124	163		118	127		0.956105157
8B		117	132		129	159		1.018810846
8C		120	165		121	122		0.932757057
8D		132	132		138	157		1.027742496
8E		151	159		142	98		0.907017997
8F		136	146		133	149		1.009305412
9A		120	134		153	128		0.940439607
9B		110	123		123	148		1.015545025
9C		124	110		130	133		1.030654108
9D		128	105		127	109		1.00983848
9E		143	143		126	93		0.93219267
9F		144	127		128	137		1.043137993
10A		129	108		155	103		0.951759151
10B		164	121		167	111		0.977421369
10C		132	107		139	110		0.994841895
10D		125	114		125	75		0.909078206
10E		155	109		165	109		0.986605719
10F		125	100		159	80		0.904301089
11A		140	117		115	101		1.011263996
11B		132	114		109	98		1.009113198
11C		140	134		150	136		0.988409094
11D		127	102		148	117		0.996674356
11E		148	110		147	145		1.061550463
11F		119	120		124	95		0.940862352
12A		133	104		155	112		0.983403966
12B		116	104		172	129		0.964296835
12C		105	115		153	136		0.958095228
12D		109	126		142	147		0.977343647
12E		138	120		132	98		0.965340438
12F		141	127		168	118		0.947973414
13A		108	125		101	127		1.018263805
13B		147	108		133	98		1.000648092

13C		131	91		124	109		1.051734279
13D		111	120		129	118		0.964863633
13E		138	128		135	113		0.977599988
13F		134	132		120	138		1.034439587
14A		126	110		130	165		1.080711485
14B		125	135		121	119		0.979524312
14C		146	117		166	153		1.029551347
14D		136	134		99	143		1.089219314
14E		113	131		128	129		0.970394733
14F		131	110		136	149		1.057296943
15A		124	93		123	113		1.044133634
15B		128	104		108	93		1.013101258
15C		111	99		111	109		1.020895652
15D		131	118		122	86		0.945928236
15E		136	84		130	91		1.027447758
15F		138	119		145	112		0.97634523
16A		117	148		147	163		0.972747975
16B		126	117		154	132		0.983344703
16C		140	161		134	156		1.002686533
16D		119	176		143	111		0.865381173
16E		142	159		174	155		0.952435862
16F		138	150		126	128		0.985017661
17A		121	125		118	115		0.987276229
17B		155	104		105	85		1.044529943
17C		141	122		113	112		1.030989086
17D		140	103		129	133		1.074601986
17E		128	110		110	91		0.991451682
17F		122	130		111	143		1.042078146
18A		115	121		129	125		0.982551652
18B		112	125		138	111		0.93196023
18C		119	95		107	95		1.023626939
18D		117	111		116	104		0.987696183
18E		107	109		124	80		0.903890128
18F		122	122		113	107		0.987952183
19A		131	121		131	121		1

19B		133	130		147	133		0.98357819
19C		132	103		144	110		0.995469059
19D		121	133		116	103		0.953227877
19E		116	115		108	95		0.973620668
19F		131	112		160	148		1.016385954
20A		147	148		131	124		0.986259167
20B		152	181		148	116		0.908651651
20C		119	153		150	119		0.90017408
20D		158	129		136	134		1.042190311
20E		135	146		157	139		0.957927297
20F		137	140		120	104		0.96316251
21A		143	104		175	107		0.963906931
21B		135	143		170	120		0.916066534
21C		137	122		144	104		0.955002706
21D		120	130		141	112		0.934898198
21E		159	127		162	114		0.972602824
21F		136	131		166	115		0.931398088
22A		137	116		130	119		1.017128159
22B		144	110		153	99		0.96442511
22C		118	120		133	116		0.96749811
22D		115	104		115	80		0.943028324
22E		129	109		128	92		0.96481107
22F		140	136		130	127		1.001244779
23A		123	84		123	63		0.937530632
23B		104	114		149	102		0.905174811
23C		111	86		156	96		0.953427339
23D		114	98		150	90		0.926306637
23E		108	110		151	104		0.920807119
23F		119	94		148	94		0.954783937
24A		124	136		145	125		0.949430403
24B		116	122		149	125		0.953442035
24C		113	149		155	142		0.926002602
24D		99	128		156	120		0.897374646
24E		120	150		162	122		0.896699193
24F		113	156		158	144		0.915947733

25A		113	95		141	99		0.96267947
25B		102	99		119	101		0.971813203
25C		122	121		148	112		0.943630152
25D		107	114		134	94		0.913478477
25E		121	111		134	104		0.964482523
25F		132	108		121	100		1.002224432
26A		123	148		170	116		0.884913574
26B		111	140		145	119		0.911804288
26C		107	129		147	146		0.960630015
26D		104	148		145	115		0.881595657
26E		110	160		131	123		0.908427458
26F		109	133		136	140		0.964774782

REFERENCES

Barrick, J. E. et al., 2009 Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature **461**: 1243-1247.

Bennett, A. F., and R. E. Lenski, 1993 Evolutionary adaptation to temperature II. Thermal niches of experimental lines of *Escherichia coli*. Evolution 1-12.

Bochtler, M. et al., 2000 The structures of HslU and the ATP-dependent protease HslUHslV. Nature **403**: 800-805.

Charusanti, P. et al., 2010 Genetic basis of growth adaptation of *Escherichia coli* after deletion of pgi, a major metabolic gene. PLoS Genet **6**: e1001186.

Chevin, L. M., 2011 On measuring selection in experimental evolution. Biol Lett **7**: 210-213.

Chou, H. H., H. C. Chiu, N. F. Delaney, D. Segrè, and C. J. Marx, 2011 Diminishing returns epistasis among beneficial mutations decelerates adaptation. Science **332**: 1190-1192.

Chuang, S. E., V. Burland, G. Plunkett, D. L. Daniels, and F. R. Blattner, 1993 Sequence analysis of four new heat-shock genes constituting the hslTS/ibpAB and hslVU operons in *Escherichia coli*. Gene **134**: 1-6.

Conrad, T. M. et al., 2009 Whole-genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. Genome Biol **10**: R118.

Cooper, V. S., and R. E. Lenski, 2000 The population genetics of ecological specialization in evolving *Escherichia coli* populations. Nature **407**: 736-739.

Deatherage, D. E., and J. E. Barrick, 2014 Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq, pp. 165-188 in *Engineering and Analyzing Multicellular Systems*, edited by Springer,

Fayet, O., T. Ziegelhoffer, and C. Georgopoulos, 1989 The groES and groEL heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. Journal of bacteriology **171**: 1379-1385.

Hayer-Hartl, M., A. Bracher, and F. U. Hartl, 2016 The GroEL–GroES chaperonin machine: a nano-cage for protein folding. Trends in biochemical sciences **41**: 62-76.

Herring, C. D. et al., 2006 Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. Nat Genet **38**: 1406-1412.

Hug, S. M., and B. S. Gaut, 2015 The phenotypic signature of adaptation to thermal stress in *Escherichia coli*. BMC Evol Biol **15**: 177.

Jablonski, D., 1986 Causes and consequences of mass extinctions: a comparative approach.

Dynamics of extinction 183-229.

Kanemori, M., K. Nishihara, H. Yanagi, and T. Yura, 1997 Synergistic roles of HslVU and other ATP-dependent proteases in controlling in vivo turnover of sigma32 and abnormal proteins in *Escherichia coli*. J Bacteriol **179**: 7219-7225.

Khan, A. I., D. M. Dinh, D. Schneider, R. E. Lenski, and T. F. Cooper, 2011 Negative epistasis between beneficial mutations in an evolving bacterial population. Science **332**: 1193-1196.

Kryazhimskiy, S., D. P. Rice, E. R. Jerison, and M. M. Desai, 2014 Global epistasis makes adaptation predictable despite sequence-level stochasticity. Science **344**: 1519-1522.

Kwon, A. R., C. B. Trame, and D. B. Mckay, 2004 Kinetics of protein substrate degradation by HslUV. J Struct Biol **146**: 141-147.

Lang, G. I., D. Botstein, and M. M. Desai, 2011 Genetic variation and the fate of beneficial mutations in asexual populations. Genetics **188**: 647-661.

Lang, G. I. et al., 2013 Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. Nature **500**: 571-574.

Lee, H., E. Popodi, H. Tang, and P. L. Foster, 2012 Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. Proc Natl Acad Sci U S A **109**: E2774-83.

Lenski, R. E., and M. Travisano, 1994 Dynamics of adaptation and diversification: a 10,000 generation experiment with bacterial populations. Proc Natl Acad Sci U S A **91**: 6808-6814.

Luria, S. E., and M. Delbrück, 1943 Mutations of Bacteria from Virus Sensitivity to Virus Resistance. Genetics **28**: 491-511.

Lenski, R. E., M. R. Rose, S. C. Simpson, and S. C. Tadler, 1991 Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. American naturalist 1315-1341.

Maclean, R. C., G. Bell, and P. B. Rainey, 2004 The evolution of a pleiotropic fitness tradeoff in Pseudomonas fluorescens. Proc Natl Acad Sci U S A **101**: 8072-8077.

Miller, J. M., J. Lin, T. Li, and A. L. Lucius, 2013 *E. coli* ClpA catalyzed polypeptide translocation is allosterically controlled by the protease ClpP. J Mol Biol **425**: 2795-2812.

Missiakas, D., F. Schwager, J. M. Betton, C. Georgopoulos, and S. Raina, 1996 Identification and characterization of HsIV HsIU (ClpQ ClpY) proteins involved in overall proteolysis of misfolded proteins in *Escherichia coli*. The EMBO journal **15**: 6899.

Mongold, J. A., A. F. Bennett, and R. E. Lenski, 1999 Evolutionary adaptation to temperature. VII. Extension of the upper thermal limit of *Escherichia coli*. Evolution 386-394.

Richter, K., M. Haslbeck, and J. Buchner, 2010 The heat shock response: life on the verge of death. Mol Cell **40**: 253-266.

Rodríguez-Verdugo, A., B. S. Gaut, and O. Tenaillon, 2013 Evolution of *Escherichia coli* rifampicin resistance in an antibiotic-free environment during thermal stress. BMC Evol Biol **13**: 50.

Rodríguez-Verdugo, A., O. Tenaillon, and B. S. Gaut, 2016 First-Step Mutations during Adaptation Restore the Expression of Hundreds of Genes. Mol Biol Evol **33**: 25-39.

Rodríguez-Verdugo, A., D. Carrillo-Cisneros, A. González-González, B. S. Gaut, and A. F. Bennett, 2014 Different tradeoffs result from alternate genetic adaptations to a common environment. Proceedings of the National Academy of Sciences **111**: 12121-12126.

Rosenzweig, F., and G. Sherlock, 2014 Editorial: Experimental Evolution: Prospects and Challenges. Genomics **104**: v.

Tenaillon, O. et al., 2012 The molecular diversity of adaptive convergence. Science **335**: 457-461.

Wignall, P. B., and M. J. Benton, 1999 Lazarus taxa and fossil abundance at times of biotic crisis. Journal of the Geological Society **156**: 453-456.

Williams, G. C., 1957 Pleiotropy, Natural Selection, and the Evolution of Senescence. Evolution 398-411.

Woods, R., D. Schneider, C. L. Winkworth, M. A. Riley, and R. E. Lenski, 2006 Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. Proc Natl Acad Sci U S A **103**: 9107-9112.

CONCLUSIONS

Charles Darwin famously described the finches of the Galápagos Islands, using them as examples to support and explain his theory of evolution by natural selection. In the spring of 1977, a drought plagued the islet of Daphne Major. Parched for water, plants produced fewer seeds, a major food source for the islet's finch species. Having consumed most of the small, easy-to-open varieties, finches were forced to compete for larger seeds or go hungry, and their populations started to shrink. However, a handful of finches had beaks of sufficient size to open the large, hard seeds that remained. The birds with the larger beaks were the future of the island, and they were the ones who ensured that the finch species would survive to see another spring (Boag and Grant, 1981).

In this light, adaptation seems simple and easy, but not all natural populations afford such clean examples of evolution in action. As scientists, we must be able to observe adaptation, measure it precisely, and understand the mechanisms by which it works. Real environmental stressors are complex and confounding, and as a result the genetic underpinnings of adaptation in the wild are often numerous and tangled. We can use laboratory models to simplify these problems. In my dissertation research, I made use of new and existing (Tenaillon *et al.*, 2012) microbial evolution experiments to better understand the genotypic and phenotypic dynamics of adaptation.

In my first chapter, I assessed the extent to which adaptation restored phenotypes from a stress state to a pre-stress state. To do this, I quantified 94 phenotypes of 115 independent clones of *E. coli* evolved at high temperature (42.2°C) for 2,000 generations, as well as their ancestor grown under stress (42.2°C) and pre-stress (37.0°C) conditions. I found that most (58%) evolved phenotypes represented restorative shifts, supporting the idea that a major outcome of adaptation

is restoration (Fong *et al.*, 2005; Carroll and Marx, 2013; Sandberg *et al.*, 2014). Novelty was a less common outcome of adaptation, comprising between 7% and 20% of evolved phenotypes, consistent with the gene expression findings of other studies (Carroll and Marx, 2013; Sandberg *et al.*, 2014).

More than this, I was able to link specific genetic changes with phenotypic variation. Of particular interest was the strong association of a 71 kb deletion (ECB_00503_large) present in 35 of 115 clones and six of the nine significant principal component axes of phenotype space. Additionally, we were able to distinguish the phenotypes of clones possessing mutations in either *rho* or *rpoB*, which typify two genetic pathways to high-temperature adaptation (Tenaillon *et al.*, 2012). Interestingly, despite being distinguishable, *rho* and *rpoB* mutations were both associated with restorative shifts in phenotype, although to different degrees.

This work highlights an important aspect of adaptive evolution that is often overlooked: the acclimation response. Because of specialized genetic programs, phenotypic plasticity, or both, the phenotypes of an organism can be altered when exposed to the stress imposed by a selective pressure. This shift with stressful conditions constitutes an acclimation response, and this is the phenotypic state upon which natural selection acts. When this state is ignored, it may be possible to misjudge what adaptation actually does to a population of organisms. For example, comparing evolved phenotypes to unstressed ancestral phenotypes alone may suggest that adaptive evolution acts through novel changes. In reality, the unstressed phenotypes were never under selection; a separate set of acclimated phenotypes was visible to the environment. My findings extend the notion that adaptive evolution is predominantly restorative and that, rather than producing new phenotypes to meet new environmental challenges, adaptation tends to alter populations in a way that regenerates preexisting phenotypes. As such, future work must be done to characterize the acclimation response, its duration, its dynamics, and its associated

phenotypes. Furthermore, evolutionary studies must consider this state when drawing conclusions about the adaptive process.

In my second chapter, I compared the evolutionary dynamics of populations traversing either an *rpoB* or *rho* adaptive pathway in the context of pleiotropy and compensation. Using genome sequence data from four *rpoB* and four *rho* populations at 11 different time points saved over the course of 2,000 generations of evolution, I reconstructed the mutational trajectories of these populations and determined several parameters for each. Mutations in both *rpoB* and *rho* were among the earliest to sweep through populations and fix, and they were highly beneficial relative to all other mutations that occurred. Therefore, these two adaptive pathways are generally determined very early on in the course of evolution, canalizing subsequent adaptation. Relationships among mutational parameters, as well as data on the total numbers of mutations that occurred in each population, revealed that *rpoB* and *rho* populations differ in their compensatory evolution. Populations following an *rpoB* pathway have more compensatory mutations available to them, resulting in greater numbers of mutations appearing and accumulating during adaptation, as well as greater clonal interference among them.

These findings point to the importance of first-step adaptive mutations in determining not only which mutations might confer compensatory benefits, but also the overall dynamics of compensatory evolution. The pleiotropic effects of the first beneficial mutation to sweep through a population can have effects on mutation accumulation and competition that echo through the rest of the adaptive process.

Because only eight populations of 115 were sampled for sequencing, the addition of more samples would greatly enhance my findings. Additionally, this data could be coupled to fitness measurements over the course of evolution. Given differential compensation and the observation that clonal interference slows down adaptation (Lang *et al.*, 2011), we might expect to see the

fitness of *rpoB* populations take longer to plateau than *rho* populations. This would be an ideal chance to utilize more modern, sequencing-based methods for determining population fitness, such as FREQ-Seq (Chubiz *et al.*, 2012).

In my third chapter, I described the frequency of population recoveries in *E. coli* grown under lethal temperature conditions (43.0°C and 44.0°C), as well as the genetic drivers underlying such so-called Lazarus effects. In total, I evolved approximately 400 independent populations at lethally high temperature for five days and saved populations whose cell densities showed evidence of recovery after initially crashing. DNA from these Lazarus populations was then extracted and sequenced.

Lazarus effects were observed to occur in approximately 9% of populations grown at 43.0°C, making them infrequent, but common enough to have important evolutionary consequences. Population recoveries were driven predominantly by mutations in either *hslUV*, an operon encoding a heat shock protease system, or *rpoBC*, an operon encoding a portion of the RNA polymerase complex. Interestingly, populations never possessed fixed mutations in both operons, suggesting a negative epistatic relationship between them. Moreover, these two operons exhibited very different mutational and fitness properties. Mutations in *hslUV* were most common, exhibited a high degree of within-week parallelism at the nucleotide level, and were predominantly frameshifts. Additionally, populations with these mutations had no major fitness costs at the pre-stress condition of 37.0°C. On the other hand, *rpoB* mutations were less common, exhibited no within-week parallelism, and were solely nonsynonymous in their effects. Populations with these mutations were at a significant disadvantage when competed against the ancestor at 37.0°C, evidence of antagonistic pleiotropy. These differences suggest a model in which the mechanism of population recovery is determined to some extent by the accumulation of neutral variation during the brief acclimation phase prior to exposure to lethal heat stress.

This speaks once again to the importance of the acclimation response to the process of adaptation, as well as to the role that standing genetic variation can play in microbial evolution, a point most famously made by Luria and Delbrück (1943) in their classic experiment (Luria and Delbrück, 1943).

Overall, the work I performed for my dissertation highlights two important aspects of adaptation. First, there is not necessarily a singular solution to an evolutionary problem. Multiple adaptive pathways mediated by different genetic changes can lead to positive outcomes for a population. This was the case for both the *rpoB* and *rho* pathways studied in Chapters 1 and 2, as well as the *hslUV* and *rpoBC* pathways discovered in Chapter 3. Second, certain properties of a population's evolutionary trajectory can be determined very early on in the adaptive process. Epistasis and clonal interference can occur within as few as ~30 generations of adaptation (Chapter 3). Moreover, different routes to adaptation can have very different side effects and dynamics over the course of evolution. Different adaptive pathways can have distinct phenotypic consequences and restorative effects (Chapter 1), as well as compensatory landscapes (Chapter 2) and fitness consequences (Chapter 3), depending on the extent of their pleiotropy. Thus, not only do first-step adaptive mutations generate contingency for later mutations, but also for later mutational dynamics as a whole.

The question is occasionally raised as to whether or not laboratory evolution experiments tell us anything useful about evolution in nature. For example, mutations in RNA polymerase, such as those we observed in our evolution experiment, are actually rare in nature, but are a common route to adaptation in controlled conditions (Long *et al.*, 2015). Additionally, the selective pressures imposed in the laboratory are often strong, discrete, and of a single type, contrasting starkly with the multiple varying pressures provided by nature.

Laboratory evolution is a model, and no model completely recapitulates the dynamics of

nature. Instead, models are meant to simplify nature with the goal of explaining as much of it as possible with as few preconditions as possible. Evolution experiments, such as those I utilized and carried out for my dissertation, reveal how populations are capable of changing over time, given the manipulation of just a few important variables. Even with all we know about the model organisms used in these experiments, their genetics, and the environmental conditions governing their evolution, understanding the processes and outcomes of adaptation is still very difficult. As such, if we wish to understand evolution as it occurs in nature, we must continue to simplify the process and pull apart its pieces in the laboratory. Until we have a firm grasp of how evolution occurs under controlled conditions, even if these conditions do not always reflect the true complexity of natural conditions, we will not be able to understand how or why evolution works the way it does, nor the full capabilities, strengths, and weaknesses of real evolving populations. This understanding will be vital as we work toward finding solutions to pressing global problems, such as climate change, disease, and biological conservation.

REFERENCES

Boag, P., and P. Grant, 1981 Intense Natural Selection in a Population of Darwin's Finches (Geospizinae) in the Galápagos. Science **214**: 82-85.

Carroll, S. M., and C. J. Marx, 2013 Evolution after introduction of a novel metabolic pathway consistently leads to restoration of wild-type physiology. PLoS genetics **9**: e1003427.

Chubiz, L. M., M. C. Lee, N. F. Delaney, and C. J. Marx, 2012 FREQ-Seq: a rapid, cost effective, sequencing-based method to determine allele frequencies directly from mixed populations. PLoS One **7**: e47959.

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

Lang, G. I., D. Botstein, and M. M. Desai, 2011 Genetic variation and the fate of beneficial mutations in asexual populations. Genetics **188**: 647-661.

Long, A., G. Liti, A. Luptak, and O. Tenaillon, 2015 Elucidating the molecular architecture of adaptation via evolve and resequence experiments. Nat Rev Genet **16**: 567-582.

Luria, S. E., and M. Delbrück, 1943 Mutations of Bacteria from Virus Sensitivity to Virus Resistance. Genetics **28**: 491-511.

Sandberg, T. E. et al., 2014 Evolution of *Escherichia coli* to 42 degrees C and Subsequent Genetic Engineering Reveals Adaptive Mechanisms and Novel Mutations. Mol Biol Evol

Tenaillon, O. et al., 2012 The molecular diversity of adaptive convergence. Science **335**: 457 461.