DECONSTRUCTING THE CORRELATED NATURE OF ANCIENT AND EMERGENT TRAITS: AN EVOLUTIONARY INVESTIGATION OF METABOLISM, MORPHOLOGY, AND MORTALITY

By

Nkrumah Alions Grant

A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

Microbiology and Molecular Genetics—Doctor of Philosophy Ecology, Evolutionary Biology, and Behavior—Dual Major

ABSTRACT

DECONSTRUCTING THE CORRELATED NATURE OF ANCIENT AND EMERGENT TRAITS: AN EVOLUTIONARY INVESTIGATION OF METABOLISM, MORPHOLOGY, AND MORTALITY

By

Nkrumah Alions Grant

Phenotypic correlations are products of genetic and environmental interactions, yet the nature of these correlations is obscured by the multitude of genes organisms possess. My dissertation work focused on using 12 populations of Escherichia coli from Richard Lenski's long-term evolution experiment (LTEE) to understand how genetic correlations facilitate or impede an organism's evolution. In chapter 1, I describe how ancient correlations between aerobic and anaerobic metabolism have maintained – and even improved – the capacity of *E. coli* to grow in an anoxic environment despite 50,000 generations of relaxed selection for anaerobic growth. I present genomic evidence illustrating substantially more mutations have accumulated in anaerobic-specific genes and show parallel evolution at two genetic loci whose protein products regulate the aerobic-to-anaerobic metabolic switch. My findings reject the "if you don't use it, you lose it" notion underpinning relaxed selection and show modules with deep evolutionary roots can overlap more, hence making them harder to break. In chapter 2, I revisit previous work in the LTEE showing that the fitness increases measured for the 12 populations positively correlated with an increase in cell size. This finding was contrary to theory predicting smaller cells should have evolved. Sixty thousand generations have surpassed since that initial study, and new fitness data collected for the 12 populations show fitness has continued to increase over this period. Here, I asked whether cell size also continued to increase. To this end, I measured the size of cells for each of the 12

populations spanning 50,000 generations of evolution using a particle counter, microscopy, and machine learning. I show cell size has continued to increase and that it remains positively correlated with fitness. I also present several other observations including heterogeneity in cell shape and size, parallel mutations in cell-shape determining genes, and elevated cell death in the single LTEE population that evolved a novel metabolism – namely the ability to grow aerobically on citrate. This last observation formed the basis of my chapter 3 research where my collaborators and I fully examine the cell death finding and the associated genotypic and phenotypic consequences of the citrate metabolic innovation. Copyright by NKRUMAH ALIONS GRANT 2020 This dissertation is dedicated to the ancestors of my grandparents, Claudia (Petrie) and Stephenson Grant, and their descendants – including my three children – Genesis, Umoja, and Imani Grant

ACKNOWLEDGEMENTS

I can describe several times in my life where I slipped almost entirely through the cracks of the cobbled path leading to this submission. Thus, my dissertation does not come without my sincerest appreciation for those who have helped guide and encourage me on my pursuit toward this degree.

Firstly, I have been graced with the honor of being mentored by Dr. Richard Eimer Lenski. Notwithstanding his astounding intelligence and notoriety, Rich is extremely humble, approachable, and places the physical and mental well-being of his trainees above their lab work. I would like to thank Rich for taking a major chance on me. In my first email to Rich, I mistakenly addressed him by the name of a second individual I was interested in working with at Michigan State University (to which he didn't reply, oops!). Realizing my mistake, I sent a second email a few days later with a request that we meet in person. Rich scheduled a meeting without hesitation (he then asked that I send my unofficial transcript, statement of interest, and GRE scores by the end of that day which scared the *E. coli* out of me). Upon visiting, I was immediately impressed with the amount of respect Rich showed me. Indeed, as a black male wannabe-scientist at that time, he made me feel that I mattered. And Rich did so continuously throughout my training. Exemplifying this point, I began a two-year long divorce that culminated in my becoming a single father of two young children at the front-end of graduate school. This was an extremely difficult time for me – I wanted to end my graduate training, and sadly, life altogether. If it wasn't for Rich's understanding and advice, I wouldn't have had, or made the necessary time, to seek the help I desperately needed; I am living my life now in part because of the care Rich showed me. I would also like to thank Rich for his flexibility, which allowed me to do the science

vi

most interesting to me while also allowing me to be a father to my three children during their young – and arguably the most critical – period of their lives. These competing interests sometimes meant skipping lab meetings or working from home for weeks on in. Rich never complained.

I have also had the pleasure of a wonderful dissertation guidance committee, whose constructive feedback challenged me to approach my research questions critically and with rigor. That committee included: Dr. Terence Marsh, Dr. Charles Ofria, Dr. Gemma Reguera, and Dr. Christopher Waters. Thank you all for recognizing my promise, sharing your knowledge, and providing me with good counsel. To this list I would also like to thank Dr. Yann Dufour, who I recognize as an unofficial committee member. Thank you for opening your door to me and for providing a supplemental example of the type of mentor I wish to be for my own students.

I would also like to thank the Microbiology and Molecular Genetics (MMG) Department support staff. The administrative efforts of Roseann Bills, Rachael Stohlin, Katrina Conley, and Christine Vandeuren removed from my shoulders the burden of navigating university bureaucracy. Thank you all – your efforts are not in vain. While here, I too would like to thank the MMG department chair, Dr. Victor DiRita. Vic's enthusiasm is infectious, his knowledge broad, and his heart pure. He brought with him to the department a vibrancy that encouraged me to interact with my colleagues more frequently. Moreover, Vic provided me with several opportunities that helped increase my visibility. Collectively, his actions impressed upon me that I belonged in the department, which translated more broadly to a heightened and sustainable self-confidence for networking in our field at large. For that I am extremely grateful.

vii

It would be remiss of me if I did not acknowledge the people with whom I worked so closely with during my graduate studies, members of the Lenski lab. First and foremost, I was fortunate to have had the opportunity to mentor two very bright and determined undergraduate students, Joe Warren and Ali Abdel-Magid. Thank you both for helping me complete experiments, providing a medium for communicating ideas, and for offering your inputs as OUR science developed. In addition, thank you both for listening to my constructive criticisms and for holding me accountable. Indeed, my growth as a mentor was fostered by frequent reflections on conversations we had and on how to approach instructing the both of you given our individual differences. Therefore, know that mentorship was reciprocal, that is, I learned just as much from the both of you, as I hope you have learned from me. I have no doubt that the two of you will find success in your respective careers.

I would also like to thank Jay Bundy, Dr. Zachary Blount, Dr. Kyle Card, Josh Franklin, Dr. Rohan Maddamsetti, and Brian Wade, for cultivating an atmosphere that promoted my growth as a scholar. Each of you have at some point vetted my technical writings, respectfully challenged my ideas, have helped me when I was stuck on some jarring task, and were constant advocates for my success. I will cherish our conversations and the time we have spent with one another. Thank you all for your collegiality and friendship. I hope we stay in contact for many more years to come.

My success was contingent upon those who invested in me from a very long time ago. That being said, I must acknowledge my family. First, I would like to thank my grandmother, Claudia Grant, who immigrated to the United States in 1972. Her fearlessness paved a path to America for 10 children from the bushes of Guyana, South America, which

viii

undoubtedly provided opportunities for our family that would have not been available to us had she stayed.

Secondly, I would like to my mother, Dolorese Grant-Fall, who too has made significant sacrifices. My mother relocated from Harlem, NY to Saginaw, MI as a single woman in 1994 with four children. She often worked multiple jobs, while also cooking and selling meals, sewing and selling clothing, and doing whatever else she needed to do in order to make ends meet. My mother was also very involved in our education. She would frequently drop into school unannounced, regularly attended conferences, constantly supplemented assigned homework with her own material, and was not afraid to remove us from schools that were not meeting her expectations (hence the reason I attended on average one school per year during my K-12 education). More importantly, my mother never gave up on me. She loved me during the darkest times of my life and protected me from falling victim to the streets. This meant making tough decisions, like placing me in a translational living center for runaways (Innerlink – Saginaw, MI) and informing my probation officer that I wasn't abiding to court orders leading to my turning 16 years old while in juvenile detention. Her love, while sometimes tough, instilled within me a strong sense of morality and personal accountability.

I would also like to thank my stepfather, Amadou Fall, who took on the daunting responsibility of raising four children that were not his own, before and after, the birth of his first two biological children. It is from him that I learned the value of hard work and the importance of faith. Indeed, my DAD taught me how to be a nurturing man, an honorable spouse, and a doting father. I would also like to thank my siblings, Prince Robertson, Pierre

ix

Grant, Courtisha Grant, Khadim Fall, and Khoudia Fall for their love and support. Let us all keep loving, building, and making waves. #WeAllWeGot.

No one bears the burden of earning a Ph.D. more than the immediate family of an individual pursuing one. For they too often feel the energy of a failed experiment, writers block stress, and the pressure of a deadline placed on a task asked to be completed last minute. They too must be adaptable, willing to accept changes to plans made weeks and sometimes months ahead. And they must do all of this while also taking care of their own individual commitments. That said, I would like to thank my wife, Fatimata Ndiaye, whose unparalleled selflessness helped me reach my goal. Thank you for taking care of our home and our children in my absence; Thank you for taking care of me. I love you. I would also like to thank my three children Genesis, Umoja, and Imani Grant. Fathering the three of you motivates me to put my best foot forward in everything that I do. Indeed, being your father contextualizes life at large. Thank you all for accepting my repetitive "give me 5 more minutes" or "Sorry, we will have to do that tomorrows." Thank you for accompanying me to the lab and classroom when you could have been at home in front of the television. Thank you all for reminding me to eat and for stressing the importance of rest, too. None of my success would be possible without the support of my family. This is our Ph.D.!!! We made it my loves.

Х

PREFACE

Humans and other organisms alike are evolutionarily primed to make associations. Hence, with extreme humidity, grey skies, and floral scents that are stronger than usual, we might expect rain. Such *correlations* offer comfort in allowing one to reasonably predict outcomes within the boundaries of some set of conditions. But our inclination to assign order to a disordered nature often yields spurious associations. Exemplifying this point, violent crime and murder are known to increase with ice cream sales. Of course, ice cream in itself does not cause people to be more violent. Instead, the correlation arises because there are more opportunities for negative interactions between people to occur during the warmer months when ice cream sales are highest. Despite our apprehension of falsely making meaningless associations between objects and phenomena, correlations serve as catalysts for inquiry into why they exist in the first place.

Organisms often have features, or *phenotypes*, that correlate with other phenotypes. The nature of these correlations has long perplexed naturalists. For example, animal breeders during the 1800's and to this day have pondered why artificially selecting for tameness in domesticated breeds also lead to seemingly unrelated attributes including floppy ears and curly tails. We know now that phenotypic correlations arise in part due to interactions between genes and proteins underlying the phenotypes and their interplay with the environment. Nonetheless, our efforts toward systematically resolving such correlations are thwarted by the fact that organisms—even bacteria—contain thousands of genes. One cannot go back in time and watch how specific genotype-phenotype correlations arose. However, one can monitor how such correlations change over time in response to different perturbations and selection pressures, given that one knows the

xi

starting genotype and can track changes to that genotype and associated phenotypes over time. The research framework of experimental evolution permits this approach.

Experimental evolution, whereby organisms evolve under a set of conditions defined by an investigator, has emerged as a powerful tool for studying evolutionary dynamics. Such studies are often performed using single-cell organisms owing to their relatively small genomes, amenability to preserving intermediate genotypes, and the ability to track their phenotypic changes. My dissertation research used 12 replicate populations of *Escherichia coli* from Dr. Richard Lenski's >70,000 generation long-term evolution experiment (LTEE) to investigate how correlations between the genetic encoding of organismal traits facilitate or impede an organism's evolution.

I approach this conundrum in Chapter One with a simple motivating question: Are genes and associated phenotypes under relaxed selection less likely to be lost when they are strongly integrated with those under direct, positive selection? Of course, *E. coli* populations in nature do not live in constant environments like the 12 populations have experienced during their evolution in the LTEE, which has placed many of their phenotypes under relaxed selection. I address this question from the perspective of *E. coli's* facultative metabolism. Indeed, anaerobic metabolism in the 12 populations has been under relaxed selection because of the well-aerated flasks in which they live. Because evolution acts on those genes being expressed in a given environment, such that deleterious mutations are selectively purged from the genome, we might reasonably expect genes involved only in anaerobic metabolism to accumulate substantially more mutations. Furthermore, the associated phenotypic performance of the strains in an environment without oxygen (anoxic environment) should be reduced when compared to the LTEE ancestral strain. My

xii

research findings support the expectation of elevated mutations in anaerobic genes. However, I also show that the phenotypic performance of the LTEE populations has been maintained and even slightly increased in an anoxic environment, contrary to theoretical expectations. I provide evidence for abundant, complex interactions between aerobic- and anaerobic-specific genes, which I conclude has allowed selection to act on both sets of metabolic traits simultaneously, thereby maintaining the ancient and essential anaerobic metabolism of the LTEE populations.

In Chapter Two, I expand upon previous work in this system which found that fitness and cell size were positively correlated during the first 10,000 generations of evolution in the LTEE. Similar to the example of ice cream and crime, it has been suggested that the correlation between fitness and cell size reported earlier is spurious. This suggestion rests, in part, on theory that predicts evolution should favor smaller cells in a resource-limited environment. Given that the average fitness of the evolving populations has continued to increase in the LTEE, the motivating question for Chapter Two was this: Does the average cell size of the evolving LTEE populations also continue to increase and correlate with fitness? I addressed this question by thoroughly measuring changes in cell size and shape over 50,000 generations, and then by integrating my measurements with fitness estimates measured over the same period. My data show cell size has continued to increase and that it remains significantly correlated with fitness. In addition, I detail several other observations including: (i) changes in cell aspect and surface area-to-volume ratios; (ii) variability among the LTEE populations in cell size and shape, including the evolution of almost spherical cells in one population; (iii) mutations in genes known to maintain rodshaped cells in almost all of the LTEE populations: and (iv) evidence for numerous dead

xiii

cells in one LTEE population that was unique in another important respect. That peculiar phenotypic correlation then motivated the research that I present in my final chapter.

In Chapter Three, I examine in depth my serendipitous discovery that the single *E. coli* population in the LTEE that evolved a novel, and extremely beneficial, metabolism namely, the ability to grow on citrate—suffers the unexpected expense of elevated cell mortality. This finding provides a window for understanding how evolution proceeds when organisms evolve beneficial traits that nonetheless have correlated maladaptive effects. Moreover, this finding makes an interesting contrast with my work in Chapter One, where I show that the ancient correlation between aerobic and anaerobic metabolism is robust. In contrast, the fragility of the bacteria that evolved the ability to grow on citrate provides a model with which to investigate the physiological consequences associated with evolving novel metabolisms that must successfully integrate within existing metabolic networks. Indeed, this system offers an exciting research avenue for thinking about evolution over deep time, because the sequential addition of novel genes, metabolic pathways, and other innovations undoubtedly generated maladaptations that had to be overcome in order for life to flourish.

Charles Darwin discussed the nature of phenotypic correlations in the opening chapter of "On the Origin of Species," where he called them "mysterious." You will find that my research on the correlated responses described herein has answered many questions, but it has also yielded new mysteries. Resolving the mechanistic underpinnings of these mysterious correlations may one day allow improved predictions about how organisms will respond to changing selection pressures. This understanding could improve conservation measures aimed at maintaining species diversity. Another benefit of

xiv

understanding the genetic architecture underlying phenotypic correlations is that we may integrate synthetic genes, pathways, and traits into organisms that evolve in ways that align with, and not against, the organism's overall performance and fitness.

Sadly, my time as a student at Michigan State University has come to an end. My research focus will switch gears as I pursue new research interests at the University of Idaho. I hope that my graduate work inspires and enlightens my readers, as it has done for me the past five years of my life.

LIST OF TABLES	xviii
LIST OF FIGURES	xix
CHAPTER 1. MAINTENANCE OF METABOLIC PLASTICITY DESPITE RELAXED SEI	ECTION
IN A LONG-TERM EVOLUTION EXPERIMENT WITH ESCHERICHIA COLI	1
Abstract	
Introduction	2 3
Materials and Methods	7
Long-Term Evolution Experiment	7
Culture Conditions	
Competition Assays	8
Genomic and Statistical Analyses	
Aerobic and Anaerobic Network Connectivity	
Results	12
Signatures of Selection on Aerobic- and Anaerobic-specific Genes in th	e LTEE 12
Mutations in Genes that Regulate Metabolic Plasticity in Response to C)xvgen.17
Fitness of Evolved Bacteria Under Oxic and Anoxic Conditions	
Heterogeneity of Responses Among Replicate Populations	
Discussion	
Acknowledgments	
APPENDIX	
LITERATURE CITED	54
CHAPTER 2: CHANGES IN CELL SIZE AND SHAPE DURING 50,000 GENERATIONS	OF
EXPERIMENTAL EVOLUTION WITH ESCHERICHIA COLI	63
Abstract	64
Introduction	65
Materials and Methods	70
Strains	70
Culture conditions	70
Volumetric and shape measurements	70
Analysis of cell mortality in population Ara-3	71
Genomic and fitness data	72
Statistical analyses	73
Results	73
Cell volumes measured by two methods	73
Temporal trends in cell size in evolved clones	74
Monotonic cell size trends among whole populations	76
Differences in cell size between exponential and stationary phases	77
Changes in cell shape	79
Analysis of changes in the SA/V ratio	

TABLE OF CONTENTS

Nearly spherical cells in one LTEE population	83
Cell volume and fitness have remained highly correlated in the LTEE	84
Elevated cell mortality in the population that evolved to grow on citrate	85
Discussion	87
Acknowledgments	92
APPENDIX	
LITERATURE CITED	118
CHAPTER 3: GENOMIC AND PHENOTYPIC EVOLUTION OF ESCHERICHIA COLI IN A NO	OVEL
CITRATE-ONLY RESOURCE ENVIRONMENT	126
Abstract	127
Introduction	128
Materials and Methods	132
Evolution experiment	132
Isolation of evolved clones	133
Fitness assays	133
Growth curves	134
Microscopy and cell viability analyses	135
Genomic analysis and copy-number variation	137
Statistical test for selection on parallel IS150 insertions	138
RNA-Seq and transcriptome analysis	139
Construction of maeA plasmid	140
Competition experiments to assess fitness effects of maeA	140
Results	141
Experimental design and phylogenetic analysis of sequenced strains	141
Genome evolution is faster in the citrate-only environment than in the con	trol
environment	143
Fitness changes after 2500 generations in DM0 and DM25 environments	145
Changes in growth parameters after 2500 generations in DM0 and DM25	
environments	146
Evidence of cell death in clones isolated from both DM0 and DM25	
environments	148
Specificity of genome evolution in the DM0 and DM25 environments	150
Contribution of transposable insertion elements to parallel evolution	153
Parallel amplification mutations in the DM0- and DM25-evolved populatio	ns154
Increased MaeA expression is highly beneficial in the citrate-only environr	nent
	156
Transcriptomic analysis of DM0-evolved clones	157
Discussion	159
Acknowledgements	165
APPENDIX	166
LITERATURE CITED	191

LIST OF TABLES

Table 1.1: ANOVAs of relative fitness for clones sampled from the LTEE populations at threetimepoints and measured in either the oxic or anoxic environment
Table 1.2: List of <i>E. coli</i> clones used in study
Table 1.3: ANOVAs of relative fitness for clones sampled from non-mutator populations atthree timepoints and measured in the oxic or anoxic environment53
Table 2.1: List of <i>E. coli</i> clones used in study116
Table 2.2: List of <i>E. coli</i> whole-population samples used in study
Table 3.1: Copy number of amplified <i>citT</i> genes in sequenced clones. 189

Table 3.2: Copy number of amplified *maeA* and *dctA* genes in sequenced clones from populations that evolved for 2500 generations in either DM0 or DM25 environments.190

LIST OF FIGURES

Figure 1.5: Evolved clones are better adapted to the aerobic environment in which they evolved than to the novel, but otherwise identical, anoxic environment. Each point is the mean fitness for an evolved clone sampled from the indicated population at three different generations, measured relative to the LTEE ancestor with the opposite marker state. Orange and teal points indicate that the corresponding population had or had not evolved

Figure 1.7: Among-population variance component for fitness in the oxic and anoxic environments. Each point is the square root of the variance component estimated from a corresponding random-effects ANOVA. Error bars are approximate 95% confidence intervals obtained using the Moriguti-Bulmer procedure (Sokal and Rohlf 1995). Negative values of estimates and confidence limits are truncated at 0, the lower theoretical bound.40

Figure 1.10: Cumulative number of mutations in aerobic- and anaerobic-specific genes in the LTEE whole-population samples. Each panel shows the number of synonymous mutations in aerobic- (black) and anaerobic-specific (red) genes in the indicated population. For comparison, random sets of genes of equal cardinality to the aerobic- (227) or anaerobic-specific (345) gene sets were sampled 1,000 times, and the cumulative number of synonymous mutations was calculated to generate a null distribution of the expected number of mutations for each population. The gray and pink points show 95% of these null distributions (excluding 2.5% in each tail) for aerobic and anaerobic comparisons, respectively.

Figure 1.12: Relative fitness trajectories for the LTEE populations shown by their mutator status. Each point is the grand-mean fitness value of a set of evolved clones sampled at 2,000, 10,000, and 50,000 generations relative to the ancestors. Non-mutators and mutators are shown in teal and orange, respectively. The upper and lower trajectories were measured in the oxic and anoxic environments, respectively. Error bars are 95% confidence intervals.48

Figure 2.7: Average rate of cell volume increase. Slopes were calculated for each population over each of three intervals. Each point shows the grand mean for the 12 populations. Error bars are 95% confidence intervals, and brackets show the statistical significance (*p* value) based on one-tailed Wilcoxon tests, which account for the paired nature of the samples. 101

Figure 2.13: Average surface area-to-volume ratio (SA/V) of ancestral and evolved clones. The surface area and volume of individual cells were calculated from microscopic images, as

Figure 2.17: Correlation between mean fitness relative to the LTEE ancestor and grand median cell volumes, both based on whole-population samples. Four points (Ara+6 at 10,000 generations; Ara–2, Ara–3, and Ara+6 at 50,000 generations) are absent due to missing fitness values reported by Wiser et al. (2013). Kendall's τ = 0.6066, *N* = 34, *p* < 0.0001....111

Figure 2.20: Difference in cell size between exponential and stationary phases. We calculated the grand median cell volume of each 50,000-generation clone grown for 2 h (exponential) or 24 h (stationary), and then computed the grand mean of the 12 populations.

Figure 2.21: Isometric analysis of the surface area-to-volume ratio (SA/V) to assess the effect of the change in aspect ratio (length/width) from 10,000 to 50,000 generations. For each population, we calculated a hypothetical mean SA/V using its average aspect ratio at 10,000 generations (no shape change), and we compared it to a hypothetical mean SA/V calculated using its average aspect ratio at 50,000 generations (with shape change). This latter value differs from that shown in Figure 2.14, because that figure is based on direct measurements of individual cells, followed by averaging the cells in a single assay, averaging the replicate assays for each population, and calculating the grand mean of the 12 populations. By contrast, calculating a hypothetical mean SA/V for a population using its aspect ratio from a different generation can only use average values of the relevant parameters. Therefore, we applied the same population-level averages to compare ratios with and without shape changes here. Error bars are 95% confidence intervals, and the bracket shows the statistical significance (*p* value) based on a one-tailed paired *t*-test.....115

Figure 3.5: Schematic of the log-slope method to calculate growth rates. We logetransformed optical densities, and used the slope of the curve in the interval OD_{420} nm = [0.01, 0.02] to calculate the exponential growth rate on glucose (h⁻¹), r_{glucose}. We used the slope of the curve in the interval OD_{420} nm = [0.05, 0.1] to calculate the exponential growth rate on citrate (h⁻¹) r_{citrate}. In making this interpretation, we assumed a diauxic shift between growth on glucose and citrate, rather than simultaneous growth on both substrates. In any case, growth rates during these intervals are relevant phenotypes even without assuming diauxie. We estimated lag time (τ) as the time (h) until OD_{420} nm = 0.01 was reached.171

Figure 3.7: Growth parameters for whole-population samples that evolved in DM0 and their Cit⁺ ancestors. (A) Estimates of various growth parameters for the ancestral strains and DM0-evolved populations at 2500 generations, using the log-slope method. Ancestral strain CZB151 and its descendants are shown in black, CZB152 and its descendants are in orange, and CZB154 and its descendants are in blue. Units for growth rates *r* are h⁻¹, and units for lag times are h. Bias-corrected and accelerated (BCa) bootstrap 95% confidence intervals around parameter estimates were calculated using 10,000 bootstraps. (B) Estimates of log₂-transformed ratios of growth parameters for the evolved populations and their ancestors. The growth curves we used to estimate these parameters are shown in Figures 3.8 and 3.9.

Figure 3.10: Growth parameters for clones from populations that evolved in DM0 and their Cit⁺ ancestors. (A) Estimates of growth parameters for the ancestral strains and DM0-

Figure 3.13: Growth parameters of the 12 DM25-evolved clones and their 3 Cit⁺ ancestors. (A) Estimates of growth parameters for each ancestral and DM25-evolved clone, using the log-slope method (Figure 3.2). Estimates for ancestral strain CZB151 and its descendants are shown in black, estimates for CZB152 and its descendants are in orange, and estimates for CZB154 and its descendants are in blue. Units for growth rates *r* are h⁻¹, and units for lag times are h. Bias-corrected and accelerated (BCa) bootstrap 95% confidence intervals around parameter estimates were calculated using 10,000 bootstraps; no confidence interval is shown if a parameter could not be estimated accurately from the available data. Aberrant estimates that fall outside of these ranges are not shown. (B) Estimates of log₂-transformed ratios of growth parameters for the evolved clones and their ancestors. The growth curves used to estimate these parameters are shown in Figures 3.14 and 3.15.....179

Figure 3.16: Correlations between estimated growth rates across substrates and media for DM0-evolved clones and populations. All tests are two-tailed, because growth rates across substrates and media might, in principle, exhibit tradeoffs. (A) Correlations between rglucose and rcitrate in DM25 are not significant (Pearson's r = 0.4788, d.f. = 12, p = 0.0833 for clones; r = -0.0392, d.f. = 13, p = 0.8897 for populations). (B) Correlations between rcitrate in DM25 are highly significant (r = 0.7513, d.f. = 12, p = 0.0020 for clones; r = 0.8041, d.f. = 13, p = 0.0003 for populations). Circles and triangles indicate

CHAPTER 1: MAINTENANCE OF METABOLIC PLASTICITY DESPITE RELAXED SELECTION IN A LONG-TERM EVOLUTION EXPERIMENT WITH *ESCHERICHIA COLI*

Authors: Nkrumah A. Grant, Rohan Maddamsetti, and Richard E. Lenski

Abstract

Traits that are unused in a given environment are subject to processes that tend to erode them, leading to reduced fitness in other environments. Although this general tendency is clear, we know much less about why some traits are lost while others are retained, and about the roles of mutation and selection in generating different responses. We addressed these issues by examining populations of a facultative anaerobe, Escherichia coli, that have evolved for >30 years in the presence of oxygen, with relaxed selection for anaerobic growth and the associated metabolic plasticity. We asked whether evolution led to the loss, improvement, or maintenance of anaerobic growth, and we analyzed gene expression and mutational datasets to understand the outcomes. We identified genomic signatures of both positive and purifying selection on aerobic-specific genes, while anaerobic-specific genes showed clear evidence of relaxed selection. We also found parallel evolution at two interacting loci that regulate anaerobic growth. We competed the ancestor and evolved clones from each population in an anoxic environment, and we found that anaerobic fitness had not decayed, despite relaxed selection. In summary, relaxed section does not necessarily reduce an organism's fitness in other environments. Instead, the genetic architecture of the traits under relaxed selection and their correlations with traits under positive and purifying selection may sometimes determine evolutionary outcomes.

Introduction

"[I]f man goes on selecting, and thus augmenting, any peculiarity, he will almost certainly unconsciously modify other parts of the structure, owing to the mysterious laws of the correlation of growth." — Charles Darwin, *On the Origin of Species*, 1859

Organisms seldom experience static conditions. Instead, they typically experience fluctuations in both their external environments and internal states. Organisms have adapted to these fluctuations by evolving a variety of mechanisms to maintain homeostasis and survive, that is to be *phenotypically robust*, in the face of environmental and genetic perturbations (Lenski et al. 2006; Frankel et al. 2010; Fraser and Schadt 2010; Siegal and Leu 2014). One mechanism to maintain homeostasis is metabolic plasticity, by which we mean the innate capacity to change metabolic fluxes in response to changes in the environment (Jia et al. 2019). Metabolic plasticity is controlled by genes, the expression of which is coupled to one or more environmental signals (Paudel and Quaranta 2019). Bacillus subtilis, for example, produces metabolically dormant endospores when cells are starved for nutrients (Setlow 2006). Metabolic plasticity can sometimes go awry, such as when cancer cells perform glycolysis instead of oxidative phosphorylation (Warburg 1956) to generate ATP and synthesize biomass, promoting tumorigenesis and metastatic potential (Payen et al. 2016). More generally, metabolic plasticity determines the environmental conditions in which an organism can survive and grow. However, much remains unknown about the mechanisms underlying metabolic plasticity and the resulting phenotypic robustness, and how these mechanisms and robustness have evolved and continue to evolve (Siegal and Leu 2014; Nijhout et al. 2017).

Experimental evolution with microorganisms has proven to be a powerful way to study the evolutionary process (Elena and Lenski 2003; Lenski 2017; Van den Bergh et al. 2018). These experiments typically maintain relatively simple and constant conditions, which places many traits, including metabolic plasticity, under *relaxed selection*. Accordingly, many such studies have shown losses of functions owing to antagonistic pleiotropy (including the cost of expressing unneeded traits), mutation accumulation in unused genes, or both (Cooper and Lenski 2000; Cooper et al. 2001b; Maughan et al. 2009; Leiby and Marx 2014; Lamrabet et al. 2019). However, unneeded traits may sometimes be maintained and even improved if the underlying genes serve multiple purposes, such that the unneeded trait is genetically correlated with a trait under positive selection (Bennett et al. 1990). The latter scenario has been termed *buttressing pleiotropy* (Lahti et al. 2009). Computational models and experiments with artificial organisms suggest that buttressing pleiotropy readily occurs when new functions evolve by building upon existing functions (Wagner and Mezey 2000; Lenski et al. 2003; Ostrowski et al. 2015). However, the extent of buttressing pleiotropy in biological systems remains unclear and has been little studied.

To study how metabolic plasticity evolves under relaxed selection, we analyzed both phenotypic performance and genetic changes in *Escherichia coli* populations from the longterm evolution experiment (LTEE). The 12 LTEE populations have been evolving independently in a glucose-limited minimal medium with constant aeration for more than 70,000 generations (Lenski et al. 1991; Good et al. 2017). Samples are frozen every 500 generations, generating a "frozen fossil record" from which bacteria can be revived for genetic and phenotypic analyses, including measuring their fitness in the LTEE environment and other environments that differ in various respects. By 50,000 generations

the populations were, on average, about 70% more fit than their common ancestor in the LTEE environment (Wiser et al. 2013). Most of that improvement occurred in the first 10,000 generations, but the populations have continued to improve at slower rates throughout the experiment (Lenski et al. 2015; Lenski 2017). The bacteria experienced relaxed selection for anaerobic growth during the long duration of their evolution in the strictly oxic LTEE environment. To date, several hundred clones and more than 1400 mixed population samples have been sequenced (Tenaillon et al. 2016; Good et al. 2017), providing material for examining the coupling between aerobic and anaerobic fitness and the underlying genetic changes. Six LTEE populations evolved hypermutability during the experiment (Tenaillon et al. 2016; Good et al. 2017), which should increase the rate at which unused genes accumulate mutations and unused functions decay over time (Cooper and Lenski 2000; Leiby and Marx 2014). The hypermutable lineages exhibited ~100-fold increases in their point-mutation rate (Sniegowski et al. 1997; Wielgoss et al. 2013), but the rate of fitness gain in these lineages increased by only a few percent relative to other populations (Wiser et al. 2013; Lenski et al. 2015). This difference facilitates disentangling the effects of antagonistic pleiotropy from those of mutation accumulation on the phenotypic and genomic consequences of relaxed selection.

In addition to the strength of the LTEE model system for asking evolutionary questions, *E. coli* is an excellent model for studying the evolution of metabolic plasticity. As a facultative anaerobe, *E. coli* is able to survive, grow, and reproduce in both oxic and anoxic environments (Unden et al. 1994). This plasticity allows *E. coli* to inhabit diverse environments that vary in oxygen availability, from the gastrointestinal tracts of mammals (and some birds and reptiles) to freshwater and soil environments (Gordon and Cowling

2003). The molecular control of this plasticity is well-understood, and two global regulatory systems play critical roles (Gunsalus and Park 1994; Unden and Bongaerts 1997). The Fumarate and Nitrate Reductase protein (FNR), encoded by the *fnr* gene, is a transcription factor that directly senses oxygen (Gunsalus and Park 1994; Kang et al. 2005). The Anoxic Respiratory Control system, encoded by *arcA* and *arcB*, is a canonical two-component regulatory system that responds to oxygen and the redox status of the cell, as shown in Figure 1.1 (luchi and Lin 1988; Gunsalus and Park 1994; Unden and Bongaerts 1997). The transcriptional control conferred by the *fnr* and *arcAB* regulons allows *E. coli* cells to commit physiologically to either aerobic or anaerobic metabolism, depending on oxygen availability. Mutations in either regulon may disrupt this control (Melville and Gunsalus 1990). In particular, some mutations in *arcAB* have been shown to alter metabolism by causing the constitutive expression of genes that would normally be responsive to oxygen concentration and internal redox balance (luchi and Lin 1988; Saxer et al. 2014).

In this study, we sought to determine whether relaxed selection during 50,000 generations of strictly aerobic growth led to the loss of anaerobic performance and the associated metabolic plasticity. Alternatively, genetic correlations between aerobic and anaerobic physiology might have favored the maintenance or even improvement of anaerobic metabolism during evolution in the oxic environment of the LTEE. This alternative, if observed, might reflect the fact that anaerobic metabolism evolved more than 2 billion years before the origin of aerobic metabolism (Müller 1977; Soo et al. 2017), such that the genetic and biochemical networks underpinning metabolism in these conditions might be tightly coupled. It could also indicate the biochemical promiscuity of many

proteins involved in metabolism (Nam et al. 2012). In any case, we tested the phenotypic correlation in performance by measuring the fitness of the evolved LTEE clones against a marked ancestor in oxic and anoxic environments. In fact, anaerobic growth capacity was not only maintained under relaxed selection, but in some cases actually improved. Fitness gains were seen even in some populations that evolved high mutation rates—a change that promoted mutation accumulation in anaerobic-specific genes—and despite mutations affecting the ArcAB regulon.

Our results highlight the importance of understanding how a trait is encoded in a genetic network, and how that encoding affects its evolutionary fate in the absence of direct selection. A better understanding of the mechanisms that maintain anaerobic growth may also help synthetic biologists design more robust systems by exploiting pleiotropy (Stirling et al. 2017; Blazejewski et al. 2019; Geng et al. 2019). In addition, this work may provide a framework for better predicting how the genetic encoding of traits affects an organism's evolutionary potential, including in response to ecological challenges such as those caused by climate change.

Materials and Methods

Long-Term Evolution Experiment

The LTEE consists of 12 *E. coli* populations derived from a common ancestral strain, REL606 (Lenski et al. 1991). Six populations descend directly from REL606. The other six descend from REL607, which differs from REL606 by two mutations that are selectively neutral under LTEE conditions (Tenaillon et al. 2016). One is a point mutation in the *araA* gene that allows REL607 to utilize arabinose, and the second is an inadvertent secondary

mutation of no consequence. The mutation in *araA* provides a phenotypic marker that can be readily scored in the competition assays used to measure relative fitness. When plated on tetrazolium arabinose (TA) indicator agar plates, REL606 and its direct descendants form red colonies, whereas REL607 and its descendants form white colonies. The ability to freeze and revive viable strains has allowed the establishment of the frozen fossil record that includes samples from all 12 populations at 500-generation intervals. This record allows both genotypic and phenotypic changes to be quantified retrospectively. In this study, we examined ancestral and evolved clones that were frozen at generations 2,000, 10,000 and 50,000 (Table 1.2).

Culture Conditions

Unless noted otherwise, we grew strains in oxic and anoxic environments in 10 mL of Davis Mingioli minimal salts medium supplemented with 25 µg/mL glucose (DM25). We prepared anaerobic media by boiling 500-mL batches for 25 min while sparging in nitrogen gas using a hypodermic needle inserted through a butyl rubber stopper. Cultures were incubated at 37°C in 50-mL Erlenmeyer flasks, with orbital shaking at 120 rpm in the oxic but not the anoxic environment. These conditions are the same as those used during the LTEE, except for the absence of oxygen and shaking during anaerobic growth.

Competition Assays

We revived frozen clones isolated from each LTEE population at generations 2,000, 10,000 and 50,000. We used clones for which whole-genome sequences are available (Tenaillon et al. 2016). The notation Ara–1 to Ara–6 denotes clones descended from the ancestral strain
REL606, while Ara+1 to Ara+6 are derived from REL607. We excluded from these assays the 50,000-generation clones from three populations (Ara-2, Ara-3, Ara+6), as their evolved phenotypes make the assays unreliable (Wiser et al. 2013). We also excluded the clones sampled from Ara+6 at both earlier time-points because their growth was erratic in the anoxic environment. Clones were revived by inoculating 15 µL of thawed frozen stock into 10 mL of Luria-Bertani (LB) broth, and they were grown at 37°C in an orbital shaker under atmospheric conditions for 24 h. We then diluted each competitor 1:10,000 into DM25 medium for a preconditioning step that depended on whether the assay would be performed in the oxic or anoxic environment. For the former, the competitors were preconditioned in DM25 under the standard LTEE conditions. For the latter, the competitors were preconditioned in anaerobic DM25 in an anaerobic chamber under a 95%-N₂:5%-H₂ atmosphere. Each preconditioned competitor was then diluted 1:200 into a flask containing the relevant medium under the appropriate atmosphere, and the competition ran for one day, during which time the combined population grew 100-fold. The one-day competition assays encompassed the same lag, exponential growth, and stationary phases as populations experienced during the LTEE (Lenski et al. 1991; Vasi et al. 1994). We competed the Ara⁻ evolved clones against REL607 and the Ara⁺ evolved clones against REL606. Competitions were replicated five-fold for all clones across all generations and in both environments. Owing to technical errors, only four replicates yielded data for: Ara-1 at 2,000 generations, and Ara+1 and Ara+4 at 50,000 generations, in the oxic environment; and Ara–2 and Ara+2 at 2,000 generations, Ara+1 at 10,000 generations, and Ara-1 at 50,000 generations in the anoxic environment. We calculated the

fitness of an evolved clone relative to the ancestral competitor as the ratio of their growth rates realized during the competition assay (Lenski et al. 1991; Wiser et al. 2013).

Genomic and Statistical Analyses

The genomes of the clones used in this study were previously sequenced (Tenaillon et al. 2016). We used an online tool (http://barricklab.org/shiny/LTEE-Ecoli/) to identify all of the mutations discovered specifically in the *fnr* and *arcAB* genes. To identify genes regulated in response to oxygen levels, Salmon et al. (2005) performed a Bayesian analysis of gene-expression data obtained for *E. coli* K12 in oxic and anoxic environments. We then used the set of genes from that study that showed differential expression between oxic and anoxic conditions at a posterior probability greater than 99%. Those genes were mapped onto the REL606 reference genome using the OMA (Altenhoff et al. 2018) and EcoCyc (Keseler et al. 2017) databases. We call genes that are upregulated under oxic conditions "anaerobic-specific genes." We performed binomial tests to compare the numbers of mutations in the LTEE-derived genomes in aerobic- and anaerobic-specific genes to a null expectation based on the summed length of genes in the two gene sets. That analysis was conducted using an R script called aerobic-anaerobic-genomics.R.

In addition, the LTEE metagenomics dataset includes mutations found by sequencing whole-population samples for all 12 populations through 60,000 generations. These mutations were downloaded from https://github.com/benjaminhgood/LTEE-metagenomic/. The data were reformatted (*.csv) and analyzed using an R script called aerobic-anaerobic-metagenomics.R. In brief, the cumulative number of mutations observed

in each population was plotted, after normalizing by gene length, for various categories of mutations. In particular, we examined subsets of these data based on mutation type (nonsynonymous, synonymous, and all others including indels, nonsense, and structural variants) and by function (occurring in the aerobic- or anaerobic-specific genes). To generate a null expectation, we chose 10,000 random sets of genes (with the same cardinality as the aerobic- and anaerobic-specific genes in the specific comparison), and the cumulative number of mutations in each set was calculated. The proportion of replicates in which the cumulative number of mutations in the random set was larger than the corresponding number in the aerobic- or anaerobic-specific gene set was used as an empirical *p*-value for testing statistical significance. For visualization, the relevant figures show only the middle 95% of the cumulative mutations for 1,000 (rather than 10,000) random sets. Statistical analyses were performed in R (version 3.5.0; 2018-04-23). Datasets and R analysis scripts are available on the Dryad Digital Repository (DOI pending publication).

Aerobic and Anaerobic Network Connectivity

We used the Cytoscape platform (Shannon et al. 2003) to visualize the biomolecular interaction network between aerobic- and anaerobic-specific genes. In brief, we imported our gene sets into Cytoscape and then used the software to query the STRING database (Szklarczyk et al. 2017), which includes empirically known and computationally predicted protein-protein interactions. Interactions are evaluated using seven lines of evidence, and a score is assigned to each (Szklarczyk et al. 2017). The STRING software then computes a combined "confidence score" for each interaction, which is effectively the likelihood that

the interaction truly exists given the evidence. We constructed our network of the aerobicand anaerobic-specific genes showing only those interactions with confidence scores greater than 70%, a threshold considered "high" by the database curators.

Results

Signatures of Selection on Aerobic- and Anaerobic-specific Genes in the LTEE

We hypothesized that a subset of the aerobic-specific genes experienced positive selection to acquire mutations that better adapt the bacteria to the LTEE environment. By contrast, we expect that many anaerobic-specific genes were under relaxed selection in the LTEE. We tested these predictions using the genomic and metagenomic datasets spanning 50,000 and 60,000 generations, respectively (Tenaillon et al. 2016; Good et al. 2017). At various times, 6 of the 12 LTEE populations evolved roughly 100-fold higher point-mutation rates than the ancestral strain (Sniegowski et al. 1997; Tenaillon et al. 2016; Good et al. 2017). These mutator populations gained fitness slightly faster than the non-mutator populations (Wiser et al. 2013; Lenski et al. 2015). However, genomic evolution in these populations was dominated by the accumulation of random mutations (Tenaillon et al. 2016; Couce et al. 2017; Maddamsetti et al. 2017). For these reasons, we made and tested separate predictions for the mutator and non-mutator populations. For the non-mutator populations, where previous studies found compelling evidence for positive selection (Woods et al. 2006; Tenaillon et al. 2016; Good et al. 2017), we predicted that aerobicspecific genes would have more mutations than anaerobic-specific genes. By contrast, in the mutator populations, previous studies indicated that random mutations (neutral or nearly neutral) accumulated in those genes under relaxed selection, given the high

mutation pressure. Of course, some sites even within aerobic-specific genes could have accumulated neutral or nearly neutral mutations in the oxic LTEE environment. However, purifying selection should lead to fewer mutations in aerobic- than in anaerobic-specific genes in the mutator populations.

To test these predictions, we examined the sets of 345 and 227 anaerobic and aerobic-specific genes, respectively, as described in the Materials and Methods. Given 4,143 protein-coding genes in the genome of the LTEE ancestor (Jeong et al. 2009), the anaerobicand aerobic-specific genes constitute 8.3% and 5.5% of that total, respectively. We asked whether these two sets accumulated different numbers of mutations in clones sampled from the LTEE populations at 50,000 generations. We controlled for differences in mutational target size by summing over the length of the genes in each set. In clones from the six non-mutator lineages, the aerobic-specific genes had 38 mutations, whereas the anaerobic-specific genes had 21 mutations (two-tailed binomial test: $p < 10^{-6}$). By contrast, the hypermutator clones had 836 mutations in anaerobic-specific genes and 333 mutations in aerobic-specific genes (two-tailed binomial test: p = 0.0040). The first result is consistent with stronger positive selection for beneficial mutations in aerobic- than anaerobic-specific genes. The second result is consistent with relaxed selection on anaerobic-specific genes, which could also be described as stronger purifying selection on aerobic-specific genes.

Next, we examined these dynamics using whole-population metagenomic data that includes mutations that fixed as well as those that reached a frequency above $\sim 5\%$ in a population through $\sim 60,000$ generations (Good et al. 2017). Rather than reanalyzing these data from scratch, we analyzed the dataset previously generated by Good et al. (2017). We first visualized the evolutionary dynamics for all mutations in aerobic- and anaerobic-

specific genes in each population (Figure 1.9). For the non-mutator populations (panels A– F in Figure 1.9), mutations in aerobic-specific genes were more common than those in anaerobic-specific genes, especially during the first 10,000 generations, indicating that many mutations in aerobic-specific genes experienced positive selection. In the populations that evolved hypermutator phenotypes (panels G–L in Figure 1.9), however, it is difficult to tell by eye whether the rate of molecular evolution differs between these two sets of genes. Therefore, we counted the number of observed mutations in aerobic and anaerobic-specific genes in each population over time, normalized by gene length.

When we examine the occurrence of nonsynonymous mutations, all six populations that were never mutators, along with two others (Ara-1, Ara-3) before they became hypermutable, have substantially more mutations in aerobic-specific genes than expected under the null distribution calculated by resampling random sets of 227 genes (the cardinality of the aerobic-specific genes) (Figure 1.2). The probability of this directional outcome occurring by chance under a one-tailed binomial expectation is $(1/2)^8 = 1/256 \approx$ 0.004, which is thus very significant. The number of nonsynonymous mutations in anaerobic-specific genes in those same populations, by contrast, is much lower in every case. For the mutator populations, including the two (Ara-1, Ara-3) that evolved hypermutability fairly late in the LTEE, the rates of mutations in anaerobic- and aerobicspecific genes track one another more closely. In at least three of these populations (Ara-2, Ara+3, Ara+6), the rates of mutation accumulation decreased later in the LTEE. These decelerations correspond to reversions or compensatory alleles that arose in those populations and caused their mutation rates to decline (Tenaillon et al. 2016; Good et al. 2017). Some of these populations also suggest a slower rate of mutation accumulation in

aerobic- than in anaerobic-specific genes, in particular near the end of the time course. Indeed, two mutator populations, Ara+3 and Ara+6, accumulated significantly fewer nonsynonymous mutations in aerobic genes than expected under the null distribution (non-parametric bootstrap with 10,000 replicates, p < 0.0001 for each). The slower mutation accumulation in aerobic-specific genes suggests purifying selection, as expected because the mutator phenotype increases the rate of deleterious mutations. It might also reflect, in part, saturation of possible beneficial mutations in aerobic-specific genes, in accord with a "coupon-collecting" model of molecular evolution (Good et al. 2017). Population Ara–4 is an outlier, however, in that its rate of mutation accumulation in aerobic-specific genes slightly exceeded the rate observed in anaerobic-specific genes for most of its history, despite its mutator phenotype (Figure 1.2).

To look more deeply into the role of purifying selection, we examined the accumulation of insertions and deletions (indels), structural variants (including those generated by transposable elements), and nonsense mutations in protein-coding genes in all 12 populations. These types of mutations typically destroy protein function. Although such knockout mutations are sometimes beneficial in evolution experiments (e.g., Cooper et al. 2001b), they would be highly deleterious in conserved genes under purifying selection as well as in genes under positive selection to fine-tune protein function (Maddamsetti et al. 2017). If aerobic-specific genes faced strong purifying selection in the mutator populations, we reasoned that indels, structural variants, and nonsense mutations would be underrepresented in them (Figure 1.3). Indeed, that was the case in four of the six hypermutator populations (nonparametric bootstrap with 10,000 replicates: p = 0.0003 for Ara-3; p = 0.0014 for Ara-4; p < 0.0001 for Ara+3; p = 0.0003 for Ara+6). On balance, these

observations indicate stronger purifying selection on aerobic- than on anaerobic-specific genes in the LTEE, especially in the populations that evolved hypermutable phenotypes. This finding is consistent with the later evolution of anti-mutator alleles that reduced or reverted mutation rates to the ancestral level in most of the populations that evolved hypermutability. The fact that mutations accumulated more slowly in anaerobic-specific genes than expected under the null distribution in two mutator populations (nonparametric bootstrap with 10,000 replicates: p < 0.0001 for Ara+3; p = 0.0035 for Ara+6) suggests that some anaerobic-specific genes might also have experienced purifying selection, indicating functionality even during aerobic growth.

For synonymous mutations, which are effectively neutral in the vast majority of cases, we expect to see many more of them in the mutator populations, and indeed that is the case. We also do not expect to see any systematic association with aerobic- or anaerobic-specific genes. That expectation is also fulfilled: three of the six mutator populations had more synonymous mutations in aerobic- than in anaerobic-specific genes, and the other three show the opposite trend (Figure 1.10). It is a bit puzzling that the difference between the two gene sets is so noticeable in some cases in one direction or the other. These differences might reflect hitchhiking, whereby several synonymous mutations affecting one or the other gene set, all on the same background, were pushed to high frequency (or pulled to extinction) in a particular population (Maddamsetti et al. 2015). In any case, there is no overall pattern across the six mutator populations for synonymous mutations (Figure 1.10).

Mutations in Genes that Regulate Metabolic Plasticity in Response to Oxygen

After establishing the genome-wide signatures of selection on aerobic- and anaerobicspecific genes, we now turn our attention to three particular genes known to regulate metabolic plasticity in response to oxygen availability: *fnr, arcA*, and *arcB* (Figure 1.1). Only two LTEE populations, both mutators (Ara+3, Ara+6), have nonsynonymous mutations in the *fnr* gene; in both populations, the mutations arose well after the populations had evolved hypermutability. By contrast, 11 of the 12 populations have nonsynonymous mutations in *arcA*, *arcB*, or both (Figure 1.4A). The other population (Ara–6) has a 9-bp deletion in *arcA*. In another population, Ara–2, two lineages designated S and L have coexisted since about generation 6,000 (Rozen et al. 2005), and only the S lineage has a mutation in *arcA*. Many of the mutations in *arcA* and *arcB* were already present in the clones sequenced at 10,000 generations (Figure 1.4A), and *arcA* was previously identified as showing a signature of strong positive selection in the LTEE (Tenaillon et al. 2016). We mapped the *arcA* and *arcB* mutations in the 50,000-generation clones onto the encoded protein structures. Mutations in *arcA* impact both the response regulator and DNA binding domains of the protein (Figure 1.4B), and mutations in *arcB* map to several protein domains including the histidine kinase and histidine kinase receptor (Figure 1.4C). There are several ways that these mutations might affect metabolic plasticity in the evolved bacteria. Mutations in either gene could affect the stability of the proteins or their activities, thereby (i) altering the capacity of ArcB to sense redox changes through the quinone pool; (ii) altering the rate or efficiency of phosphoryl transfer between the ArcB protein domains; (iii) decreasing the extent of ArcA phosphorylation; (iv) increasing the rate of ArcA dephosphorylation; (v) decreasing the extent of ArcA phosphorylation-dependent

oligomerization; or (vi) altering the DNA binding efficiency of ArcA. In any case, these mutations may impact ArcAB signaling and might thus affect the ability of the evolved strains to grow in an anoxic environment.

Fitness of Evolved Bacteria Under Oxic and Anoxic Conditions

Half of the LTEE populations were founded by *E. coli* B strain REL606 and half by REL607, an *araA* mutant of REL606. This mutation is selectively neutral in the LTEE environment (Lenski et al. 1991; Wiser et al. 2013), and it provides a readily scored marker for distinguishing competitors in assays of relative fitness. However, it was unknown whether the *araA* mutation is also neutral under anoxic conditions. To that end, we competed REL606 and REL607 in oxic and anoxic environments, with other conditions the same as those used in the LTEE. We saw no significant differences in relative fitness in either the oxic (t = 0.6972, d.f. = 4, two-tailed p = 0.5241) or anoxic (t = 1.1433, d.f. = 4, two-tailed p = 0.3167) environment. Thus, the *araA* mutation can serve as a useful marker for assaying the relative fitness of evolved and ancestral clones in both the anoxic and oxic environments.

We expected that the evolved bacteria would be better adapted to the oxic environment, where they evolved, than to the anoxic environment. To test this hypothesis, we competed clones sampled at 2,000, 10,000, and 50,000 generations against the reciprocally marked ancestral strains. As explained in the Materials and Methods section, we excluded one population (Ara+6) at all three time points, and two others (Ara–2, Ara–3) at the last time point, because of technical difficulties associated with enumerating these competitors. We calculated fitness as the ratio of the realized growth rate of the

evolved clone relative to that of the ancestor during a competition. This metric integrates the effects of differences in lag, growth, and stationary phases (Lenski et al. 1991; Vasi et al. 1994). Figure 1.5 shows the results obtained for each evolved clone, with standard errors based on replicate competition assays for that clone. Note that all points lie below the isocline corresponding to equal fitness in the anoxic and oxic environments, consistent with our hypothesis. As an overall assessment, we performed paired comparisons of fitness in the two environments at each time point, and in all cases the difference was highly significant (paired *t*-tests; generation 2,000: *t* = 8.7849, df = 10, one-tailed *p* < 0.0001; generation 10,000: *t* = 8.7986, df = 10, one-tailed *p* < 0.0001; generation 50,000: *t* = 5.9556, df = 8, one-tailed *p* = 0.0002). Indeed, all 31 clones tested had higher estimated fitness values in the oxic environment than in the anoxic one (sign test, *p* << 0.0001). Figure 1.6 shows the grand mean fitness values in the two environments over time, with confidence limits based on the replicate populations.

The fitness of the LTEE populations has increased monotonically in the oxic environment throughout the experiment (Lenski et al. 2015), and our data recapitulate this behavior (Figures 1.6, 1.11). Although we expected, and confirmed (Figures. 1.5, 1.6), that fitness relative to the ancestor would be higher in the oxic environment than in the anoxic environment, we did not have a clear expectation for the trajectory of fitness relative to the ancestor in the novel anoxic environment. On the one hand, the anoxic environment shares most aspects of the oxic environment including the limiting resource (glucose), temperature (37°C), and absence of any predators. On the other hand, some aspects of performance might tradeoff between the two conditions. Also, as we saw at the genomic level (Figures 1.2, 1.3), anaerobic-specific genes could decay by mutation accumulation,

especially in the mutator populations. Given these opposing expectations, one might expect anaerobic fitness to follow a quasi-random walk (Freckleton and Harvey 2006). The fitness trajectories measured in the anoxic environment show some apparent changes in direction, but the differences between consecutive time points are generally within the margin of error (Figure 1.11). In fact, none of the 31 comparisons between sequential points are significant at p < 0.05 after performing a Bonferroni correction. In any case, fitness tended to increase even in the anoxic environment in most populations (Figure 1.11). Across all of the clones tested, 23 of 31 had point estimates of their fitness relative to the ancestor in the anoxic environment greater than unity (two-tailed sign test, p = 0.0107). However, the grand mean fitness of the evolved bacteria in that environment was significantly greater than unity only at the 10,000-generation time point (Figure 1.6).

Heterogeneity of Responses Among Replicate Populations

Each population acquired a unique set of mutations over the course of the LTEE. However, adaptation to the common environment contributed to strong parallelism at the level of genes, especially in the non-mutator populations (Woods et al. 2006; Tenaillon et al. 2016; Good et al. 2017). For example, just 57 genes that make up only $\sim 2\%$ of the coding genome had $\sim 50\%$ of the nonsynonymous mutations that accumulated in the non-mutator populations through 50,000 generations (Tenaillon et al. 2016). The trajectories for fitness also showed strong parallelism (Lenski and Travisano 1994; Wiser et al. 2013; Lenski et al. 2015). For example, the square root of the among-population variance for fitness was only $\sim 5\%$ after 50,000 generations (Lenski et al. 2015), when the grand mean fitness itself had increased by $\sim 70\%$ (Wiser et al. 2013). By contrast, with relaxed selection on anaerobic-

specific genes, the accumulation of different sets of mutations should contribute to the populations having greater heterogeneity in fitness when assayed in the anoxic environment than in the oxic environment. We first examined these predictions by performing six one-way ANOVAs (two environments and three time points) to test whether the among-lineage variation was significant (Table 1.1). At 2,000 generations, we saw significant fitness heterogeneity in the anoxic environment, consistent with our expectation. At 10,000 generations, by contrast, there was no significant heterogeneity in either environment. Finally, we saw significant fitness variation in both test environments after 50,000 generations.

Of course, statistical significance, or the lack thereof, is a crude criterion by which to compare the among-lineage heterogeneity in fitness between the two environments. One can visualize the magnitude of the heterogeneity by estimating the variance attributable to lineages that is greater than expected from the measurement error across replicate assays. However, the statistical uncertainty in estimating variance components is often quite large, and indeed that was the case in our analyses, which showed no significant difference in fitness heterogeneity between the oxic and anoxic environments at any of the generations tested (Figure 1.7). We also considered the possibility that these analyses were unduly influenced by the subset of populations that evolved hypermutability, which might obscure differences in the among-lineage variation between the two environments. To that end, we repeated the ANOVAs (Table 1.3) and the estimation of variance components using only those lineages that did not evolve hypermutability. However, the results of these analyses were not appreciably different (Figure 1.13). In short, contrary to our expectation, we found no compelling evidence of greater among-lineage heterogeneity when fitness was

measured in the anoxic than in the oxic environment.

Discussion

If a trait or function is no longer useful to an organism because of a change in its environment, then it may be lost over time. It is unclear, however, what factors determine whether and how quickly traits under relaxed selection will be lost. In this study, we investigated the consequences of relaxed selection for metabolic plasticity by examining the maintenance of anaerobic metabolism—an ancient and core function—using *E. coli* strains that have been evolving in the laboratory under strictly oxic conditions for more than 70,000 generations. On the one hand, one can really imagine that anaerobic metabolism would decay because it has been unused during that time. On the other hand, one can imagine that anaerobic metabolism would be maintained even under relaxed selection owing to the ancient, tightly intertwined physiological and genetic networks that govern aerobic and anaerobic metabolism.

Two mechanisms have been proposed to explain trait loss under relaxed selection (Fong et al. 1995; Cooper and Lenski 2000): antagonistic pleiotropy (AP), and mutational degradation (MD). AP is a selection-driven process whereby improvements to traits that are beneficial in one environment trade off with traits that are useful in other conditions (Williams 1957). An example of AP seen in the LTEE is the loss of the ability to grow on maltose, which has occurred in most populations by mutations in a gene that encodes a transcriptional activator of other genes that encode proteins used to transport and metabolize that sugar (Pelosi et al. 2006; Leiby and Marx 2014). In the absence of maltose, the loss of expression of those genes confers a demonstrable competitive advantage,

indicative of AP. A familiar example in the realm of multicellular eukaryotes is senescence, whereby traits that increase reproductive potential early in life reduce survival late in life (Rodríguez et al. 2017). By contrast, MD occurs by a neutral process. In this scenario, traits under relaxed selection accumulate degradative mutations that are neutral in an organism's current environment, but which reduce the organism's fitness in environments where those traits are useful. AP and MD are not mutually exclusive, so both can act together to degrade unused functions. In those LTEE populations that evolved greatly elevated mutation rates, one would expect MD to cause much greater decay of unused genes (Cooper and Lenski 2000; Couce et al. 2017). The hypermutable lineages have also increased their fitness in the LTEE environment more than the populations that retained the low ancestral mutation rate, although the difference is small (Wiser et al. 2013; Lenski et al. 2015). Whether the LTEE populations would, on balance, lose fitness under anoxic conditions, and whether those lineages that evolved hypermutability would lose fitness to a greater extent, depends on the form and strength of the regulatory and physiological couplings between aerobic and anaerobic metabolism.

We examined genomic and metagenomic sequence data to detect possible signatures of different modes of selection acting on aerobic- and anaerobic-specific genes. Six of the 12 LTEE populations retained the low ancestral point-mutation rate throughout their history (Tenaillon et al. 2016; Good et al. 2017). All six accumulated many more nonsynonymous mutations in aerobic-specific genes than in anaerobic-specific genes (Figure 1.2). Moreover, these mutations were concentrated in a subset of genes, with that genetic parallelism indicative of adaptive evolution (Tenaillon et al. 2016; Good et al. 2017). In a number of cases, the inferred benefit of parallel changes was confirmed by

measuring the relative fitness of otherwise isogenic strains (Barrick et al. 2009). By contrast, the six populations that evolved hypermutability (Sniegowski et al. 1997; Wielgoss et al. 2013; Tenaillon et al. 2016) accumulated more nonsynonymous mutations in anaerobic- than in aerobic-specific genes (Figure 1.2). Consistent with purifying selection against mutations in aerobic-specific genes, presumptive knockout mutations (insertions, deletions, nonsense mutations, and structural variants) were underrepresented in those genes in most of the hypermutable populations (Figure 1.3). More generally, our results agree with several studies that have found a higher frequency of mutations in genes underlying traits under relaxed selection (Shabalina et al. 1997; Cooper and Lenski 2000; Funchain et al. 2000; Maughan et al. 2007; Shewaramani et al. 2017; Cui et al. 2019; Harrison et al. 2019). For example, Maughan et al. (2007) propagated *Bacillus subtilis* vegetatively (i.e., without sporulation) for 6,000 generations. They found that the evolved strains' inability to generate spores was driven largely by the accumulation of mutations in genes required for sporulation, but not for vegetative growth.

We also identified mutations in *arcA* and *arcB* in almost all of the evolved lines (Figure 1.4). These genes encode the two-component system responsible for regulating the switch between aerobic and anaerobic metabolism (Figure 1.1). Previous work found that *arcA* is among the top 15 genes in the LTEE in terms of nonsynonymous substitutions among non-hypermutable lineages, and this parallelism implies positive selection (Tenaillon et al. 2016). Moreover, *arcB* mutations have been implicated in improving growth on acetate—a waste product of glucose metabolism—in some LTEE populations (Plucain et al. 2014; Quandt et al. 2015; Leon et al. 2018).

Given the high rate of mutation accumulation in anaerobic-specific genes in the mutator populations (Figure 1.2), along with numerous mutations in *arcAB* in both mutator and non-mutator populations (Figure 1.4), one might reasonably expect that anaerobic metabolism would be degraded, or perhaps even completely lost, in the LTEE populations. However, that was not the case. Our results indicate a more nuanced outcome. On average, the grand mean fitness measured under anoxic conditions tended to increase over the first 10,000 generations of the LTEE, although to a much lesser extent than fitness measured under oxic conditions (Figure 1.6). Between 10,000 and 50,000 generations, the grand mean fitness measured in the anoxic environment showed no clear trend, even as fitness under the oxic conditions of the LTEE continued to increase, albeit at a slower rate (Figure 1.6). One might also expect to see greater variation in fitness among the replicate lines when measured in the novel anoxic environment than in the oxic environment where selection was in force during the LTEE. Substantially increased among-population variation in fitness has been reported, for example, when growth substrates (Travisano and Lenski 1996) and temperature were changed (Cooper et al. 2001a). However, we found no meaningful differences in the among-population variance in fitness at any of the time points tested (Figure 1.7). We also saw no consistent difference in fitness in the anoxic environment between the hypermutable and non-mutator populations (Figure 1.12), despite the compelling evidence for relaxed selection on anaerobic-specific genes in the hypermutable populations (Figure 1.2). These results broadly suggest that anaerobic metabolism, taken as a whole, did not decay appreciably under relaxed selection, perhaps owing to conserved underlying correlations with traits that contribute to aerobic

performance and that experienced a mixture of positive and purifying selection during the LTEE.

Similar correlated responses have been measured for other phenotypic traits in the LTEE populations. For example, the fitness gains made on glucose during the first 2,000 generations led to correlated improvements on lactose (Travisano and Lenski 1996). Leiby and Marx (2014) measured the growth of clones isolated at generations 20,000 and 50,000 on a large array of substrates. They found correlated improvements on some substrates that the populations had not seen for decades, including a few that the ancestors could not even use. Cooper (2002) competed evolved and ancestral strains in four different media, including two with altered glucose concentrations (DM2.5 and DM250), the base medium (DM25) with added bile salts, and a dilute version of the nutritionally complex LB medium. The relative fitness of the evolved lines had, on average, increased in all of these foreign environments. In another study, Meyer et al. (2010) showed that most late-generation LTEE lines had evolved resistance to phage lambda, despite never being exposed to lambda or any other phage during the experiment. They further showed that this positive correlated response was itself associated with a negative correlated response, namely the loss of the capacity to grow on maltose. Lambda uses a maltose transporter to infect cells, and reduced expression of that protein promotes fitness on glucose, compromises growth on maltose, and confers resistance to lambda (Meyer et al. 2010).

What might contribute to the maintenance and even slight improvement of anaerobic performance, despite relaxed selection? The evolved populations showed modest improvement, on average, in the anoxic environment at 10,000 generations (Figure 1.6). One possibility is that some traits and pathways, such as faster glucose transport and

glycolysis, are beneficial in both oxic and anoxic environments. As a consequence, some of the mutated genes that evolved in parallel early in the LTEE might be beneficial in the anoxic environment as well. Only 4 of the 15 genes (*pykF*, *hslU*, *infB*, and *rplF*) with the strongest signatures of parallel evolution (Tenaillon et al. 2016) are in the aerobic-specific gene set, and even some of them might sometimes contribute to anaerobic metabolism. This overlap may explain the correlated improvement in anaerobic fitness early in the LTEE. However, it is unclear whether it is sufficient to explain the maintenance of anaerobic growth capacity after 50,000 generations of relaxed selection, especially in those lines that were hypermutable for much of that time, because knocking out even a single function that is truly required for anaerobic growth should disrupt that metabolic plasticity. In any case, future work might examine the fitness effects of specific mutations that arose in the LTEE, including those that fixed early in the LTEE, on anaerobic fitness to determine if they do, in fact, provide correlated advantages.

Another possible explanation for the maintenance of anaerobic performance hinges on a connection between biochemistry and metabolism. The formulation of the DM25 medium does not provide an alternative terminal electron acceptor that would permit anaerobic respiration. Thus, the bacteria are presumably fermenting glucose in the anoxic environment. Even in the oxic environment of the LTEE, however, the cells might be fermenting rather than respiring glucose, using a process called overflow metabolism (Basan et al. 2015; Swain and Fagan 2019). The use of overflow metabolism would favor the maintenance and even improvement of genes involved in fermentative metabolism. Acetate is a major byproduct of fermentative metabolism, and the fact that several LTEE populations evolved frequency-dependent interactions mediated by cross-feeding on

acetate (Elena and Lenski 1997; Rozen et al. 2005; Großkopf et al. 2016; Leon et al. 2018) provides support for this hypothesis. A prediction of this hypothesis is that the bacteria in the LTEE are not respiring, and therefore not using dissolved oxygen, at least during some phase of their population growth. Future studies can test this prediction, and they could also use alternative terminal electron acceptors to measure the respiratory capacity of the evolved lines under anoxic conditions to determine whether their anaerobic growth has become restricted to fermentative metabolism.

A third explanation for the maintenance and even slight improvement in anaerobic fitness of the LTEE populations involves the ArcAB system (Figure 1.1). These two proteins work in concert to regulate the expression of genes relevant to both aerobic and anaerobic growth (Iuchi and Lin 1988; Gunsalus and Park 1994; Unden and Bongaerts 1997). In particular, ArcA and ArcB together repress the genes that encode TCA-cycle enzymes under anoxic conditions. In their seminal paper describing this system, luchi and Lin (1988) isolated *arcA* mutants that produce abnormally high levels of enzymes that are normally repressed under anoxic conditions. These mutants were pleiotropic, so that several aerobic-specific enzymes had increased expression under anoxic conditions including those involved in the TCA cycle and in fatty acid degradation as well as some flavoprotein dehydrogenases and a ubiquinone oxidase. Saxer et al. (2014) also reported extensive metabolic changes caused by *arcA* mutations in short-term evolution experiments with both *E. coli* and *Citrobacter freundii*. They performed proteomic analyses that recapitulated the changes in gene expression reported by luchi and Lin (1988), and they also saw increased expression of genes involved in amino-acid metabolism. Saxer et al. (2014)

concluded that mutations in global regulators like the ArcAB system could, in one step, expand the niche of an organism by substantially remodeling its cellular metabolism.

All of these potential physiological and genetic explanations for the unexpectedly strong performance of the evolved LTEE bacteria when tested under anoxic conditions might fit under a common theme, which has been called "buttressing pleiotropy" (Lahti et al. 2009). In contrast to the tradeoffs (negative correlations) generated by antagonistic pleiotropy, the idea of buttressing pleiotropy is that "the function of the correlated trait is buttressing or maintaining values of the focal trait" (Lahti et al. 2009). The relevance of buttressing pleiotropy to our results would be strengthened if the genetic architecture underlying aerobic and anaerobic metabolism had many connections. To that end, we used empirical and computationally derived information on protein-protein interactions to infer and visualize the topology of the aerobic- and anaerobic-specific gene sets in our study (Shannon et al. 2003; Szklarczyk et al. 2017). One immediately sees many connections that demonstrate these two sets are far from independent (Figure 1.8). It is quite possible, therefore, that selection for improved aerobic growth would buttress anaerobic performance.

A related idea, also relevant to our system, is robustness. *E. coli* is often described as a facultative anaerobe because it can grow not only in the well-oxygenated conditions widely used in most laboratories, but also when deprived of oxygen in special growth chambers. However, *E. coli* might be better described as a facultative aerobe, because its natural home is the anoxic mammalian colon. To be sure, many *E. coli* cells periodically exit their hosts, and the ability to survive in the presence of oxygen is essential for colonizing new hosts. However, we would suggest that most of the 100-million year or so history of

this species has been spent living under anerobic conditions—even if there are as many *E. coli* cells outside as inside mammalian hosts at any given time, those that are outside are much more likely to be evolutionary dead-ends. If so, then evolution might have favored a more robust anaerobic metabolism, one that could not easily be disrupted by short-sighted selection for improved fitness in the oxic environment. A test of this anaerobic-robustness hypothesis would be to perform an experiment identical to the LTEE, except in a strictly anoxic environment. If populations that evolved for 50,000 generations in the absence of oxygen lost the ability to grow in its presence, then that would indicate that aerobic metabolism is less robust that anaerobic metabolism.

In a different vein, it is known that some enzymes exhibit substrate promiscuity (Nam et al. 2012). Changes to global regulators—including the ArcAB system and others that have changed during the LTEE (Cooper et al. 2003, 2008; Philippe et al. 2007; Crozat et al. 2011)—might increase expression of some promiscuous enzymes, and those latent functions might yield benefits in multiple environments. Future work might compare the transcriptomes of ancestral and evolved bacteria in oxic and anoxic environments, as well as examine the effects of specific mutations to ArcAB and other global regulators on the transcriptomes. Another area for future work would examine the speed with which the ancestral and evolved lines can respond to sudden changes in the environment from oxic to anoxic and vice versa. Yet another interesting direction might use a combination of genetic engineering and experimental evolution to generate obligate aerobic and obligate anaerobic strains, which could then be studied to better understand the constraints on each type of metabolism.

All in all, our results show that anaerobic metabolism was surprisingly robust during 50,000 generations of adaptation to a strictly oxic environment. The capacity for anaerobic growth persisted even in those lineages that evolved hypermutability, despite genetic signatures that showed increased mutation accumulation in anaerobic-specific genes in those lines. Thus, relaxed selection on a functional trait does not always result in its loss. We suggest that one must know how the genes and proteins responsible for traits that experience relaxed selection interact with the rest of an organism's genome and physiology, in order to understand the potential for loss, maintenance, or even correlated improvement of such traits. More generally, investigating the role of genetic architecture in the evolutionary process might help us better predict how organisms will respond to novel selection pressures and environmental perturbations, including climate change. Such understanding may also suggest new ways of promoting or constraining the evolutionary trajectories of organisms and their traits, which could be used for synthetic biology, on the one hand, and to limit the evolution of pathogens, on the other hand.

Acknowledgments

We thank Terence Marsh, Charles Ofria, Gemma Reguera, and Chris Waters for feedback as this research progressed; Zachary Blount for helpful comments on the manuscript; and members of the Lenski lab for valuable discussions. We thank Terence Marsh for providing access to an anaerobic chamber, and the MSU Department of Microbiology and Molecular Genetics for related supplies. This work was supported in part by a grant from the National Science Foundation (currently DEB-1951307), the BEACON Center for the Study of Evolution in Action (DBI-0939454), and the USDA National Institute of Food and

Agriculture (MICL02253). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funders. APPENDIX



Figure 1.1: Schematic representation of ArcAB-dependent gene regulation. In an anoxic environment, the transmembrane sensor kinase ArcB undergoes autophosphorylation. This reaction is enhanced by fermentation metabolites such as _D-lactate, pyruvate and acetate that act as effectors. Three conserved residues (His₂₉₂, Asp₅₇₆, His₇₁₇) in ArcB sequentially transfer the phosphoryl group onto ArcA, the response regulator, at a conserved Asp₅₄ residue. Phosphorylated ArcA (ArcA-P), in turn, represses the transcription of many operons involved in respiratory metabolism, while activating those encoding proteins involved in fermentative metabolism. In an oxic environment, ArcB autophosphorylation ceases and ArcA is dephosphorylated by reversing the phosphorelay, leading to the release of inorganic phosphate into the cytoplasm. Figure adapted from Kwon et al. (2003).



Figure 1.2: Cumulative numbers of mutations in aerobic- and anaerobic-specific genes in the LTEE whole-population samples. Each panel shows the number of nonsynonymous mutations in aerobic- (black) and anaerobic-specific (red) genes in the indicated population through 60,000 generations. For comparison, random sets of genes of equal cardinality to the aerobic- (227) or anaerobic-specific (345) gene sets were sampled 1,000 times, and the cumulative number of nonsynonymous mutations was calculated to generate a null distribution of the expected number of mutations for each population. The gray and pink points show 95% of these null distributions (excluding 2.5% in each tail) for aerobic and anaerobic comparisons, respectively.



Figure 1.3: Cumulative numbers of indel, nonsense, and structural mutations in proteincoding genes in the LTEE whole-population samples. The black and red points are mutations in aerobic- and anaerobic-specific genes, respectively. The gray and pink points show the corresponding null distributions based on randomized gene sets. See Figure 1.2 for additional details.



Figure 1.4: Mutations in *arcA* and *arcB* in clones from the LTEE. (A) Shading indicates that the clone has a mutation in *arcA* or *arcB*, which encode proteins ArcA and ArcB, respectively. Clones from two coexisting lineages, labeled S and L, are shown for population Ara–2. Sequence data are from Tenaillon et al. (2016), with one exception: Plucain et al. (2014) show that the *arcA* mutation had already fixed in the Ara–2 S lineage by 10,000 generations. Mutations present in the 50,000-generation clones are mapped onto the functional domains of the (B) ArcA and (C) ArcB proteins. The red arrow marks a deletion of 3 amino acids; all other mutations are point mutations resulting in amino-acid substitutions.



Figure 1.5: Evolved clones are better adapted to the aerobic environment in which they evolved than to the novel, but otherwise identical, anoxic environment. Each point is the mean fitness for an evolved clone sampled from the indicated population at three different generations, measured relative to the LTEE ancestor with the opposite marker state. Orange and teal points indicate that the corresponding population had or had not evolved hypermutability, respectively. Error bars show the standard errors based on replicated fitness assays in each environment for each clone.



Figure 1.6: Relative fitness trajectories during 50,000 generations of evolution in an oxic environment. Each point is the grand mean fitness of evolved clones sampled at 2,000, 10,000, and 50,000 generations relative to the LTEE ancestors. Black and red symbols correspond to fitnesses measured in the oxic and anoxic environments, respectively. Error bars show the 95% confidence intervals based on 11, 11, and 9 assayed populations at 2,000, 10,000, and 50,000 generations, respectively.



Figure 1.7: Among-population variance component for fitness in the oxic and anoxic environments. Each point is the square root of the variance component estimated from a corresponding random-effects ANOVA. Error bars are approximate 95% confidence intervals obtained using the Moriguti-Bulmer procedure (Sokal and Rohlf 1995). Negative values of estimates and confidence limits are truncated at 0, the lower theoretical bound.



Figure 1.8: Network topology of aerobic- and anaerobic-specific genes. We used Cytoscape (Shannon et al. 2003) to visualize protein-protein interactions between aerobic- (black) and anerobic-specific (red) genes predicted by STRING (Szklarczyk et al. 2017). Each edge reflects a confidence score of at least 0.7 (i.e., high) that an interaction exists, obtained using a maximum likelihood approach and seven lines of evidence. Figure 1.14 shows additional genes that are not integrated within the main network shown here.



Figure 1.9 (panels A-L): Evolutionary dynamics of mutations in aerobic-specific genes. Each panel (A–L) shows the allele-frequency trajectories for new mutations that arose in the indicated LTEE population and reached an observable frequency. Panels A–F show the six populations that never evolved hypermutability; panels G–L show the six populations that became mutators at various time points. In each panel, the top sub-panel shows in gray all observed mutations; the middle sub-panel colors only those mutations in aerobic-specific genes; and the bottom sub-panel colors only those mutations in anaerobic-specific genes. The underlying metagenomic data and statistical criteria are from Good et al. (2017). The stars in the middle and bottom sub-panels mark the "appearance time" of mutations, which Good et al. (2017) defined as 250 generations before the first sample in which a given mutation reached observable frequency (i.e., the midpoint between that sample and the preceding sample, given 500 generations between successive samples).

Figure 1.9 (cont'd)



Figure 1.9 (cont'd)


Figure 1.9 (cont'd)





Figure 1.10: Cumulative number of mutations in aerobic- and anaerobic-specific genes in the LTEE whole-population samples. Each panel shows the number of synonymous mutations in aerobic- (black) and anaerobic-specific (red) genes in the indicated population. For comparison, random sets of genes of equal cardinality to the aerobic- (227) or anaerobic-specific (345) gene sets were sampled 1,000 times, and the cumulative number of synonymous mutations was calculated to generate a null distribution of the expected number of mutations for each population. The gray and pink points show 95% of these null distributions (excluding 2.5% in each tail) for aerobic and anaerobic comparisons, respectively.



Figure 1.11: Relative fitness trajectories of individual LTEE populations in oxic and anoxic environments. Black and red points show competitions performed in oxic and anoxic environments, respectively. Each point is a replicate competition assay in which a clone from the indicated population was competed against the reciprocally marked ancestral strain. Two trajectories are truncated owing to technical difficulties (see Materials and Methods). Wide hash marks are means; error bars are 95% confidence intervals.



Figure 1.12: Relative fitness trajectories for the LTEE populations shown by their mutator status. Each point is the grand-mean fitness value of a set of evolved clones sampled at 2,000, 10,000, and 50,000 generations relative to the ancestors. Non-mutators and mutators are shown in teal and orange, respectively. The upper and lower trajectories were measured in the oxic and anoxic environments, respectively. Error bars are 95% confidence intervals.



Figure 1.13: Among-population variance component for fitness, excluding mutator populations. See Figure 1.7 for additional details. The among-population variance is similar in the two environments whether the mutator populations are included (Figure 1.7) or excluded (this figure).



Figure 1.14: Network topology aerobic- (black) and anerobic-specific (red) genes that are not connected to the large core network shown in Figure 1.8. See the legend to that figure for additional details.

		Oxic			Anoxic		
Generation	Source	df	F	р	df	F	р
2,000	Lineage	10	1.0441	0.4246	10	3.6004	0.0016
	Error	43			42		
10,000	Lineage	10	1 1091	0 3770	7	0 9220	0 5225
	Error	44	1.1071	0.3770	31	0.9220	0.5225
50,000	Lineage	8	3.9393	0.0022	5	2.7139	0.0194
	Error	34			24		

Table 1.1: ANOVAs of relative fitness for clones sampled from the LTEE populations at three timepoints and measured in either the oxic or anoxic environment.

Note: Eleven lineages were examined at 2,000 and 10,000 generations, but only 9 lineages at 50,000 generations. The fitness assays were replicated 5-fold for each clone, but technical errors resulted in a few missing values.

Clone ID	Population	Generation		
REL606	Ara [_] ancestor	0		
REL607	Ara ⁺ ancestor	0		
REL1158A	Ara+1	2,000		
REL1159A	Ara+2	2,000		
REL1160A	Ara+3	2,000		
REL1161A	Ara+4	2,000		
REL1162A	Ara+5	2,000		
REL1163A	Ara+6	2,000		
REL1164A	Ara-1	2,000		
REL1165A	Ara-2	2,000		
REL1166A	Ara-3	2,000		
REL1167A	Ara-4	2,000		
REL1168A	Ara-5	2,000		
REL1169A	Ara-6	2,000		
REL4530A	Ara+1	10,000		
REL4531A	Ara+2	10,000		
REL4532A	Ara+3	10,000		
REL4533A	Ara+4	10,000		
REL4534A	Ara+5	10,000		
REL4535A	Ara+6	10,000		
REL4536A	Ara-1	10,000		
REL4537A	Ara-2	10,000		
REL4538A	Ara-3	10,000		
REL4539A	Ara-4	10,000		
REL4540A	Ara–5	10,000		
REL4541A	Ara-6	10,000		
REL11392	Ara+1	50,000		
REL11342	Ara+2	50,000		
REL11345	Ara+3	50,000		
REL11348	Ara+4	50,000		
REL11367	Ara+5	50,000		
REL11370	Ara+6	50,000		
REL11330	Ara-1	50,000		
REL11333	Ara-2	50,000		
REL11364	Ara-3	50,000		
REL11336	Ara-4	50,000		
REL11339	Ara-5	50,000		
REL11389	Ara-6	50,000		

Table 1.2: List of *E. coli* clones used in study.

		Oxic			Anoxic		
Generation	Source	df	F	р	df	F	р
2,000	Lineage	10	1.0441	0.4246	10	3.6004	0.0016
	Error	43			42		
10,000	Lineage	7	0.5832	0.7644	7	0.7684	0.6180
	Error	32			31		
50,000	Lineage	5	3.2700	0.0234	5	2.6544	0.0478
	Error	22			24		

Table 1.3: ANOVAs of relative fitness for clones sampled from non-mutator populations at three timepoints and measured in the oxic or anoxic environment.

Note: Data were obtained for 11, 8, and 6 lineages at 2,000, 10,000, and 50,000 generations, respectively. See Table 1.1 for additional details.

LITERATURE CITED

LITERATURE CITED

- Altenhoff, A. M., N. M. Glover, C. M. Train, K. Kaleb, A. Warwick Vesztrocy, D. Dylus, T. M. De Farias, et al. 2018. The OMA orthology database in 2018: Retrieving evolutionary relationships among all domains of life through richer web and programmatic interfaces. Nucleic Acids Research 46(D1):D477–D485.
- Barrick, J. E., D. Yu, S. Yoon, H. Jeong, T. Oh, D. Schneider, R. E. Lenski, et al. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature 461:1243–1247.
- Basan, M., S. Hui, H. Okano, Z. Zhang, Y. Shen, J. R. Williamson, and T. Hwa. 2015. Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. Nature 528:99–104.
- Bennett, A. F., K. M. Dao, and R. E. Lenski. 1990. Rapid evolution in response to high-temperature selection. Nature 346:79–81.
- Blazejewski, T., H. I. Ho, and H. H. Wang. 2019. Synthetic sequence entanglement augments stability and containment of genetic information in cells. Science 365:595–598.
- Cooper, T. F., D. E. Rozen, and R. E. Lenski. 2003. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. Proceedings of the National Academy of Sciences of the USA 100:1072–1077.
- Cooper, T. F., S. K. Remold, R. E. Lenski, and D. Schneider. 2008. Expression profiles reveal parallel evolution of epistatic interactions involving the CRP regulon in *Escherichia coli*. PLoS Genetics 4:e35.
- Cooper, V. S. 2002. Long-term experimental evolution in *Escherichia coli*. X. Quantifying the fundamental and realized niche. BMC Evolutionary Biology 2:12.
- Cooper, V. S., and R. E. Lenski. 2000. The population genetics of ecological specialization in evolving *Escherichia coli* populations. Nature 407:736–739.
- Cooper, V. S., A. F. Bennett, and R. E. Lenski. 2001a. Evolution of thermal dependence of growth rate of *Escherichia coli* populations during 20,000 generations in a constant environment. Evolution 55:889–896.
- Cooper, V. S., D. Schneider, M. Blot, and R. E. Lenski. 2001b. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. Journal of Bacteriology 183:2834–2841.
- Couce, A., L. V. Caudwell, C. Feinauer, T. Hindré, J. P. Feugeas, M. Weigt, R. E. Lenski, et al. 2017. Mutator genomes decay, despite sustained fitness gains, in a long-term experiment

with bacteria. Proceedings of the National Academy of Sciences of the USA 114:E9026–E9035.

- Crozat, E., T. Hindré, L. Kühn, J. Garin, R. E. Lenski, and D. Schneider. 2011. Altered regulation of the OmpF porin by Fis in *Escherichia coli* during an evolution experiment and between B and K-12 strains. Journal of Bacteriology 193:429–440.
- Cui, R., T. Medeiros, D. Willemsen, L. N. M. Iasi, G. E. Collier, M. Graef, M. Reichard, et al. 2019. Relaxed selection limits lifespan by increasing mutation load. Cell 178:385–399.
- Darwin, C. 1859. On the Origin of Species. John Murray, London.
- Elena, S. F., and R. E. Lenski. 1997. Long-term experimental evolution in *Escherichia coli*. VII. mechanisms maintaining genetic variability within populations. Evolution 51:1058–1067.
- Elena, S. F., and R. E. Lenski. 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nature Reviews Genetics 4:457-469.
- Fong, D. W., T. C. Kane, and D. C. Culver. 1995. Vestigialization and loss of nonfunctional characters. Annual Review of Ecology and Systematics 26:249–268.
- Frankel, N., G. K. Davis, D. Vargas, S. Wang, F. Payre, and D. L. Stern. 2010. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature 466:490–493.
- Fraser, H. B., and E. E. Schadt. 2010. The quantitative genetics of phenotypic robustness. PLoS ONE 5:e8635.
- Freckleton, R. P., and P. H. Harvey. 2006. Detecting non-Brownian trait evolution in adaptive radiations. PLoS Biology 4:2104–2111.
- Funchain, P., A. Yeung, J. L. Stewart, R. Lin, M. M. Slupska, and J. H. Miller. 2000. The consequences of growth of a mutator strain of *Escherichia coli* as measured by loss of function among multiple gene targets and loss of fitness. Genetics 154:959–970.
- Geng, P., S. P. Leonard, D. M. Mishler, and J. E. Barrick. 2019. Synthetic genome defenses against selfish DNA elements stabilize engineered bacteria against evolutionary failure. ACS Synthetic Biology 8:521–531.
- Good, B. H., M. J. McDonald, J. E. Barrick, R. E. Lenski, and M. M. Desai. 2017. The dynamics of molecular evolution over 60,000 generations. Nature 551:45–50.
- Gordon, D. M., and A. Cowling. 2003. The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: Host and geographic effects. Microbiology 149:3575–3586.

- Großkopf, T., J. Consuegra, J. Gaffé, J. C. Willison, R. E. Lenski, O. S. Soyer, and D. Schneider. 2016. Metabolic modelling in a dynamic evolutionary framework predicts adaptive diversification of bacteria in a long-term evolution experiment. BMC Evolutionary Biology 16:163.
- Gunsalus, R. P., and S. J. Park. 1994. Aerobic-anaerobic gene regulation in *Escherichia coli*: Control by the ArcAB and Fnr regulons. Research in Microbiology 145:437–450.
- Harrison, M. C., E. B. Mallon, D. Twell, and R. L. Hammond. 2019. Deleterious mutation accumulation in *Arabidopsis thaliana* pollen genes: A role for a recent relaxation of selection. Genome Biology and Evolution 11:1939–1951.
- Iuchi, S., and E. C. Lin. 1988. arcA (dye), a global regulatory gene in *Escherichia coli* mediating repression of enzymes in aerobic pathways. Proceedings of the National Academy of Sciences of the USA 85:1888–1892.
- Jeong, H., V. Barbe, C. H. Lee, D. Vallenet, D. S. Yu, S.-H. Choi, A. Couloux, et al. 2009. Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). Journal of Molecular Biology 394:644–652.
- Jia, D., M. Lu, K. H. Jung, J. H. Park, L. Yu, J. N. Onuchic, B. A. Kaipparettu, et al. 2019. Elucidating cancer metabolic plasticity by coupling gene regulation with metabolic pathways. Proceedings of the National Academy of Sciences of the USA 116:3909–3918.
- Kang, Y., K. D. Weber, Y. Qiu, P. J. Kiley, and F. R. Blattner. 2005. Genome-wide expression analysis indicates that FNR *of Escherichia coli* K-12 regulates a large number of genes of unknown function. Journal of Bacteriology 187:1135–1160.
- Keseler, I. M., A. Mackie, A. Santos-Zavaleta, R. Billington, C. Bonavides-Martínez, R. Caspi, C. Fulcher, et al. 2017. The EcoCyc database: Reflecting new knowledge about *Escherichia coli* K-12. Nucleic Acids Research 45:D543–D550.
- Kwon, O., D. Georgellis, E.C. Lin. 2003. Rotational on-off switching of a hybrid membrane sensor kinase Tar-ArcB in *Escherichia coli*. Journal of Biological Chemistry 278:13192–13195.
- Lahti, D. C., N. A. Johnson, B. C. Ajie, S. P. Otto, A. P. Hendry, D. T. Blumstein, R. G. Coss, et al. 2009. Relaxed selection in the wild. Trends in Ecology and Evolution 24:487–496.
- Lamrabet, O., M. Martin, R. E. Lenski, and D. Schneider. 2019. Changes in intrinsic antibiotic susceptibility during a long-term evolution experiment with *Escherichia coli*. mBio 10:e00189–19.
- 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. 2017. Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. ISME Journal 11:2181–2194.
- Lenski, R. E., and M. Travisano. 1994. Dynamics of adaptation and diversification: a 10,000generation experiment with bacterial populations. Proceedings of the National Academy of Sciences of the USA 91:6808–6814.
- 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., C. Ofria, R. T. Pennock, and C. Adami. 2003. The evolutionary origin of complex features. Nature 423:139–144.
- Lenski, R. E., J. E. Barrick, and C. Ofria. 2006. Balancing robustness and evolvability. PLoS Biology 4:e428.
- Lenski, R. E., M. J. Wiser, N. Ribeck, Z. D. Blount, J. R. Nahum, J. J. Morris, L. Zaman, et al. 2015. Sustained fitness gains and variability in fitness trajectories in the long-term evolution experiment with *Escherichia coli*. Proceedings of the Royal Society B 282:20152292.
- Leon, D., S. D'Alton, E. M. Quandt, and J. E. Barrick. 2018. Innovation in an *E. coli* evolution experiment is contingent on maintaining adaptive potential until competition subsides. PLoS Genetics 14:e1007348.
- Maddamsetti, R., R. E. Lenski, and J. E. Barrick. 2015. Adaptation, clonal interference, and frequency-dependent interations in a long-term evolution experiment with *Escherichia coli*. Genetics 200:619–631.
- Maddamsetti, R., P. J. Hatcher, A. G. Green, B. L. Williams, D. S. Marks, and R. E. Lenski. 2017. Core genes evolve rapidly in the long-term evolution experiment with *Escherichia coli*. Genome Biology and Evolution 9:1072–1083.
- Maughan, H., J. Masel, C. W. Birky, and W. L. Nicholson. 2007. The roles of mutation accumulation and selection in loss of sporulation in experimental populations of *Bacillus subtilis*. Genetics 177: 937–948.
- Maughan, H., C. W. Birky, and W. L. Nicholson. 2009. Transcriptome divergence and the loss of plasticity in *Bacillus subtilis* after 6,000 generations of evolution under relaxed selection for sporulation. Journal of Bacteriology 191:428–433.
- Melville, S. B., and R. P. Gunsalus. 1990. Mutations in fnr that alter anaerobic regulation of electron transport-associated genes in *Escherichia coli*. Journal of Biological Chemistry 265:18733–18736.

- Meyer, J. R., A. A. Agrawal, R. T. Quick, D. T. Dobias, D. Schneider, and R. E. Lenski. 2010. Parallel changes in host resistance to viral infection during 45,000 generations of relaxed selection. Evolution 64:3024–3034.
- Müller, H. E. 1977. Age and evolution of bacteria. Experientia 33:979–984.
- Nam, H., N. E. Lewis, J. A. Lerman, D.-H. Lee, R. L. Chang, D. Kim, and B. O. Palsson. 2012. Network context and selection in the evolution to enzyme specificity. Science 337:1101– 1104.
- Nijhout, F. H., F. Sadre-Marandi, J. Best, and M. C. Reed. 2017. Systems biology of phenotypic robustness and plasticity. Integrative and Comparative Biology 57:171–184.
- Ostrowski, E. A., C. Ofria, and R. E. Lenski. 2015. Genetically integrated traits and rugged adaptive landscapes in digital organisms. BMC Evolutionary Biology 15:83.
- Paudel, B. B., and V. Quaranta. 2019. Metabolic plasticity meets gene regulation. Proceedings of the National Academy of Sciences of the USA 116:3370–3372.
- Payen, V. L., P. E. Porporato, B. Baselet, and P. Sonveaux. 2016. Metabolic changes associated with tumor metastasis, part 1: Tumor pH, glycolysis and the pentose phosphate pathway. Cellular and Molecular Life Sciences 73:1333–1348.
- Pelosi, L., L. Kühn, D. Guetta, J. Garin, J. Geiselmann, R. E. Lenski, and D. Schneider. 2006. Parallel changes in global protein profiles during long-term experimental evolution in *Escherichia coli*. Genetics 173:1851–1869.
- Philippe, N., E. Crozat, R. E. Lenski, and D. Schneider. 2007. Evolution of global regulatory networks during a long-term experiment with *Escherichia coli*. BioEssays 29:846–860.
- Plucain, J., T. Hindré, M. Le Gac, O. Tenaillon, S. Cruveiller, C. Médigue, N. Leiby, et al. 2014. Epistasis and allele specificity in the emergence of a stable polymorphism in *Escherichia coli*. Science 343:1366–1369.
- Quandt, E. M., J. Gollihar, Z. D. Blount, A. D. Ellington, G. Georgiou, and J. E. Barrick. 2015. Finetuning citrate synthase flux potentiates and refines metabolic innovation in the Lenski evolution experiment. eLife 4:e09696.
- Rodríguez, J. A., U. M. Marigorta, D. A. Hughes, N. Spataro, E. Bosch, and A. Navarro. 2017. Antagonistic pleiotropy and mutation accumulation influence human senescence and disease. Nature Ecology and Evolution 1:1–5.
- Rozen, D. E., D. Schneider, and R. E. Lenski. 2005. Long-term experimental evolution in *Escherichia coli*. XIII. Phylogenetic history of a balanced polymorphism. Journal of Molecular Evolution 61:171–180.

- Ruiz, J. A., R. O. Fernández, P. I. Nikel, B. S. Méndez, and M. J. Pettinari. 2006. Dye (Arc) mutants: Insights into an unexplained phenotype and its suppression by the synthesis of poly (3-hydroxybutyrate) in *Escherichia coli* recombinants. FEMS Microbiology Letters 258:55–60.
- Salmon, K. A., S. Hung, N. R. Steffen, R. Krupp, P. Baldi, G. W. Hatfield, and R. P. Gunsalus. 2005. Global gene expression profiling in *Escherichia coli* K12: effects of oxygen availability and ArcA. Journal of Biological Chemistry 280:15084-15096.
- Saxer, G., M. D. Krepps, E. D. Merkley, C. Ansong, B. L. Deatherage Kaiser, M. T. Valovska, N. Ristic, et al. 2014. Mutations in global regulators lead to metabolic selection during adaptation to complex environments. PLoS Genetics 10:e1004872.
- Setlow, P. 2006. Spores of *Bacillus subtilis*: Their resistance to and killing by radiation, heat and chemicals. Journal of Applied Microbiology 101:514–525.
- Shabalina, S. A., L. Y. Yampolsky, and A. S. Kondrashov. 1997. Rapid decline of fitness in panmictic populations of *Drosophila melanogaster* maintained under relaxed natural selection. Proceedings of the National Academy of Sciences of the USA 94:13034–13039.
- Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, et al. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Research 13:2498–2504.
- Shewaramani, S., T. J. Finn, S. C. Leahy, R. Kassen, P. B. Rainey, and C. D. Moon. 2017. anaerobically grown *Escherichia coli* has an enhanced mutation rate and distinct mutational spectra. PLoS Genetics 13:e1006570.
- Siegal, M. L., and J. Leu. 2014. On the nature and evolutionary impact of phenotypic robustness mechanisms. Annual Review of Ecology, Evolution, and Systematics 45:495–517.
- Sniegowski, P. D., P. J. Gerrish, and R. E. Lenski. 1997. Evolution of high mutation rates in experimental populations of *E. coli*. Nature 387:703–705.
- Sokal, R. R., Rohlf, F. J. 1995. Biometry: The principles and practice of statistics in biological research, 3rd Edition. W. H. Freeman, New York.
- Soo, R. M., J. Hemp, D. H. Parks, W. W. Fischer, and P. Hugenholtz. 2017. On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. Science 355:1436–1440.
- Stirling, F., L. Bitzan, S. O'Keefe, E. Redfield, J. W. K. Oliver, J. Way, and P. A. Silver. 2017. Rational design of evolutionarily stable microbial kill switches. Molecular Cell 68:686–697.

- Swain, A., and W. F. Fagan. 2019. A mathematical model of the Warburg Effect: Effects of cell size, shape and substrate availability on growth and metabolism in bacteria. Mathematical Biosciences and Engineering 16:168–186.
- Szklarczyk, D., J. H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, et al. 2017. The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Research 45:D362–D368.
- Tenaillon, O., J. E. Barrick, N. Ribeck, D. E. Deatherage, J. L. Blanchard, A. Dasgupta, G. C. Wu, et al. 2016. Tempo and mode of genome evolution in a 50,000-generation experiment. Nature 536:165–170.
- Travisano, M., and R. E. Lenski. 1996. Long-term experimental evolution in *Escherichia coli*. IV. Targets of selection and the specificity of adaptation. Genetics 143:15–26.
- Unden, G., and J. Bongaerts. 1997. Alternative respiratory pathways of *Escherichia coli*: energetics and transcriptional regulation in response to electron acceptors. Biochimica et Biophysica Acta 1320:217–234.
- Unden, G., S. Becker, J. Bongaerts, J. Schirawski, and S. Six. 1994. Oxygen regulated gene expression in facultatively anaerobic bacteria. Antonie van Leeuwenhoek 66:3–22.
- Van den Bergh, B., T. Swings, M. Fauvart, and J. Michiels. 2018. Experimental design, population dynamics, and diversity in microbial experimental evolution. Microbiology and Molecular Biology Reviews 82:e00008-18.
- Vasi, F., M. Travisano, and R. E. Lenski. 1994. Long-term experimental evolution in *Escherichia coli*. II. Changes in life-history traits during adaptation to a seasonal environment. American Naturalist 144:432–456.
- Wagner, G. P., and J. Mezey. 2000. Modeling the evolution of genetic architecture: A continuum of alleles model with pairwise AxA epistasis. Journal of Theoretical Biology 203:163–175.
- Warburg, O. 1956. On the origin of cancer cells. Science 123:309–314.
- Wielgoss, S., J. E. Barrick, O. Tenaillon, M. J. Wiser, W. J. Dittmar, S. Cruveiller, B. Chane-Woon-Ming, et al. 2013. Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. Proceedings of the National Academy of Sciences of the USA 110:222–227.
- Williams, G. C. 1957. Pleiotropy, natural selection, and the evolution of senescence. Evolution 11:398–411.
- Wiser, M. J., N. Ribeck, and R. E. Lenski. 2013. Long-term dynamics of adaptation in asexual populations. Science 342:1364–1367.

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*. Proceedings of the National Academy of Sciences of the USA 103:9107–9112.

CHAPTER 2: CHANGES IN CELL SIZE AND SHAPE DURING 50,000 GENERATIONS OF EXPERIMENTAL EVOLUTION WITH *ESCHERICHIA COLI*

Authors: Nkrumah A. Grant, Ali Abdel Magid, Joshua Franklin, Yann Dufour, and Richard E. Lenski

Abstract

Bacteria come in a wide variety of sizes and shapes, with many species exhibiting stereotypical morphologies. How morphology changes, and over what timescales, is less clear. Previous work on cell size and shape from an evolution experiment with *Escherichia coli* showed that both size and shape had changed substantially, with all 12 populations evolving larger cells and, in some cases, cells that were rounder (less rod-like). That experiment has now run for much longer, providing new material. In the meantime, genome sequencing has provided information on the genetic changes in these populations, and new computational methods enable high-throughput microscopic analyses. In this study, we measured cell volumes at stationary phase for the ancestor and all 12 populations at 2,000, 10,000, and 50,000 generations, along with measurements during exponential growth at the last timepoint. The samples were analyzed using both a Coulter counter, which measures the distribution of cell volumes, and microscopy, which provides data on shape as well as size. Our new datasets confirm the trend toward larger cells, while also revealing substantial variation in size and shape across the replicate populations. Most populations first evolved wider cells, but later reverted to the ancestral length-to-width ratio. All but one population acquired one or more mutations in genes involved in maintaining rod-shaped cells. We also observed many ghost-like (presumably dead) cells in the only population that evolved the novel ability to grow on citrate, supporting the hypothesis that this lineage struggles with balanced growth after cells shift from growing on glucose to this alternative substrate. Lastly, we show that cell size and fitness remain tightly correlated across 50,000 generations. Our results suggest that increased cell volume is beneficial in the experimental environment, while the later reversion toward rod-shaped

cells suggests partial compensation for the less favorable surface area-to-volume ratio in the larger evolved cells.

Introduction

For well over 100 years, cell biologists have wondered why cells adopt characteristic shapes (Young 2006) and sizes (Marshall et al. 2012). Cell size has been of particular interest owing to its importance for organismal fitness. For example, cell size influences a bacterium's susceptibility to predation by protists (Corno and Jürgens 2006; Batani et al. 2016) and phagocytosis by host immune cells (Champion et al. 2008; Doshi and Mitragotri 2010). Larger cell size has also been implicated in increasing susceptibility to bacteriophages (St-Pierre and Endy 2008; Choi et al. 2010) and reducing susceptibility to antibiotics (Koch 2003; Miller 2004; Nikolaidis et al. 2014). Cell size is generally tightly coupled to growth and division. Most eukaryotic cells follow a four-stage cycle in which they must reach a critical mass before partitioning into daughter cells (Cooper 2000). In contrast, the bacterial cell cycle involves less discrete periods due to the overlapping nature of cell growth, DNA replication, chromosome segregation and division (Chien et al. 2012). Bacterial cells are generally larger when they are growing faster (Schaechter et al. 1958; Bremer and Dennis 1987; Akerlund et al. 1995), in order to accommodate more genetic material (Wang and Levin 2009; Chien et al. 2012) and ribosomes (Valgepea et al. 2013). These facts suggest that cell size per se is a direct target of selection. However, it has also been suggested that cell size is a "spandrel" (Amir 2017), i.e., a phenotypic character that might appear to be the product of adaptive evolution, but is instead merely a byproduct of natural selection acting on some other trait (Gould and Lewontin 1979).

The distribution of cell size in prokaryotes spans many orders of magnitude (Heim et al. 2017). The smallest known bacteria occur in the genus *Palagibacterales*; they constitute 25% of all marine planktonic cells, and they have average volumes of only ~0.01 fL (1 fL = 1 μ m³) (Rappé et al. 2002; Giovannoni 2017). The largest heterotrophic bacteria, in the genus *Epulopiscium*, live in the intestines of surgeonfish; they have cytoplasmic volumes of ~2 x 10⁶ fL (Angert et al. 1993; Levin and Angert 2015). In contrast to these extremes, the average cell volumes of four widely studied bacteria—*Bacillus subtilis, Staphylococcus aureus, Escherichia coli, and Caulobacter crescentus*—range between about 0.4 – 3.0 fL (Levin and Angert 2015).

Large bacterial cells face significant challenges. Unlike multicellular eukaryotes that use elaborate vasculature or similar systems to transport nutrients and waste between cells, along with specialized cells to acquire nutrients from and dispose of wastes to the environment, bacteria rely on diffusion to grow and reproduce (Mika et al. 2010; Mika and Poolman 2011; Schavemaker et al. 2018). Diffusion must be considered from two perspectives. A cell must first acquire nutrients from the environment at the cell surface, and those nutrients must then diffuse internally to their sites of biochemical processing in a timely fashion. As cells grow, volume generally increases faster than surface area, such that the surface area-to-volume (SA/V) ratio decreases. The SA/V ratio might thus constrain viable cell sizes, as cells that are too large may be unable to obtain nutrients at a sufficient rate to service the demands of their biomass. However, bacteria have evolved a number of strategies that increase their rates of nutrient acquisition for a given cell volume. Rodshaped cells, for example, experience a smaller reduction in their SA/V ratio as they grow larger than do spherical cells. Other examples include various extracellular projections that

allow surface attachment while generating biomechanical motion to refresh the medium in the cell's immediate environment (Yang et al. 2016); chemotaxis, which allows cells to move along gradients of increasing nutrient concentration (Sourjik and Wingreen 2012); and invaginated cell envelopes, which increase the SA/V ratio (Tucker et al. 2010).

The surface area of a spherocylindrical (i.e., rod-shaped) cell is given by $S \approx 2\gamma V^{\frac{2}{3}}$, where $\gamma = \eta \pi \left(\frac{\eta \pi}{4} - \frac{\pi}{12}\right)^{-\frac{2}{3}}$ and η is the aspect ratio, i.e., the cell's length divided by its width (Ojkic et al. 2019). For rod-shaped species like *E. coli*, the SA/V ratio can be varied by changing either a cell's length or width, while holding the volume constant. Assuming that the rod shape is maintained, doubling a cell's width reduces its SA/V ratio by much more than doubling its length (Harris and Theriot 2018). If all else were equal, then SA/V considerations would predict relatively larger cells during nutritional upshifts (resources plentiful) and smaller cells during nutritional downshifts (resources scarce).

Now, suppose a bacterial population resides in a simple environment, one free of predators and stressors and with a predictable supply of carbon. As this population adapts to this environment by natural selection, the cells grow slightly faster. Do the cells also become larger, and if so, how much larger? Does cell size constrain the maximum growth rate that can be achieved, or is the causality in the opposite direction? If the cells evolve to become larger, are they larger while growing, in stationary phase when the limiting resource is depleted, or both? And what might change about the shapes of the cells including their aspect and SA/V ratios?

Experimental evolution has proven to be powerful for addressing such questions. This research framework provides the opportunity to study evolution in real time, both in

biological (Kawecki et al. 2012; Ratcliff et al. 2012; Graves et al. 2017) and digital (Lenski et al. 2003; LaBar and Adami 2017) systems. In the long-term evolution experiment (LTEE), 12 replicate populations of *E. coli* were started from a common ancestor and have been propagated by daily serial transfer in a minimal glucose-limited medium for more than 70,000 generations (32 years). Whole-population samples, and clones from each population, have been frozen every 500 generations, creating a frozen "fossil record" from which genotypic and phenotypic changes can be measured (Lenski et al. 1991; Lenski and Travisano 1994).

Evolution proceeded most rapidly early in the LTEE. By 2,000 generations, the populations were, on average, ~35% more fit than their ancestor. An increase in the exponential growth rate and a reduction in the duration of the lag phase prior to growth were major contributors to this improvement (Vasi et al. 1994). By 50,000 generations, the average population was ~70% more fit than the ancestor (Wiser et al. 2013; Lenski et al. 2015; Lenski 2017). The trajectory for fitness relative to the ancestor is well described a power-law function, which implies that fitness may continue to increase indefinitely, albeit at progressively slower rates of improvement (Wiser et al. 2013).

In the first 10,000 generations, cell size was found to have increased in all 12 LTEE populations and their trajectories were positively correlated with fitness (Lenski and Travisano 1994). The increase in cell volume was accompanied by a concomitant decrease in numerical yield, although the product of cell volume and number—the total biovolume yield—increased (Vasi et al. 1994). In the meantime, several populations were found to have diverged in shape, producing more spherical cells (Lenski and Mongold 2000;

Philippe et al. 2009); and fitness has continued to increase for at least 50,000 generations more (Wiser et al. 2013; Lenski et al. 2015).

In this study, we sought to determine if cell size has continued to increase over time, and whether it still tracked with fitness. To that end, we measured cell size in the ancestor and the evolving populations at 2,000, 10,000 and 50,000 generations. We used both a Coulter particle counter and microscopy to measure cell volumes, and microscopy to characterize changes in cell shape. All 12 populations evolved larger cells. As previously seen in the fitness trajectories, the rate of change in cell volumes was fastest early in the experiment, and the trend was monotonically increasing over time. By 50,000 generations, the average cell volume in most populations was well over twice that of the ancestor, both during exponential growth and in stationary phase. The evolved cells tended to increase more in width than in length during the first 10,000 generations, but they subsequently reverted to aspect ratios similar to the ancestral strain. However, there was considerable among-population variability in shape as well as size through the entire period. Analyses of genome sequence data also revealed mutations in cell-rod maintenance genes in almost every population. Lastly, we discovered greatly elevated cell mortality in the only population that evolved the novel ability to use citrate in the growth medium as a carbon source. Overall, our data suggest that cell size and shape are important targets of selection in the LTEE.

Materials and Methods

Strains

The *E. coli* LTEE is described in detail elsewhere (Lenski et al. 1991; Wiser et al. 2013; Lenski 2017). In short, 12 populations were derived from a common ancestral strain, REL606. Six populations descend directly from REL606. The other six descend from REL607, which differs from REL606 by two selectively neutral mutations (Tenaillon et al. 2016). Whole-population samples and clones from each population have been frozen at 500-generation intervals. These materials permit the retrospective analysis of genotypic and phenotypic evolution. In this study, we used both clones (Table 2.1) and wholepopulation samples (Table 2.2) from 2,000, 10,000 and 50,000 generations.

Culture conditions

Samples from the freezer were slightly thawed, inoculated into LB broth, and grown overnight at 37°C. These cultures were diluted 1:10,000 in 9.9 mL Davis Mingioli medium containing 25 ug mL⁻¹ glucose (DM25). Cultures were incubated at 37°C in 50-mL Erlenmeyer flasks, with orbital shaking at 120 rpm for 24 h. These conditions are the same as those used in LTEE. The following day, we diluted cultures 1:100 in fresh DM25 and grew them for 2 h or 24 h for exponential and stationary phase cell measurements, respectively.

Volumetric and shape measurements

Cell sizes were measured using two analytical approaches. In one, we used the Coulter Multisizer 4e (Beckman), an electronic device that measures cell volume following the Coulter principle (Don 2003). In this study, we used a 30- μ m aperture, and we measured particle sizes in the range from 2% to 60% of the aperture diameter, which corresponds to a volumetric range of 0.113 fL to 3,054 fL. However, we excluded any particles over 6 fL in our analyses. On several occasions we calibrated the aperture using 5.037- μ m diameter wide latex beads (Beckman). The measured variance in bead size was below the recommended threshold of 2.0% at each calibration.

In the second approach, we imaged cells using phase-contrast microscopy, and we processed the resulting micrographs using the *SuperSegger* package (Stylianidou et al. 2016). *SuperSegger* automatically identifies the boundary between cells and segments the individual cells on a micrograph. It returns measurements aligned to the midline of each cell for the long and short axes, which we used as length and width, respectively. The volume (in arbitrary units) of a cell is approximated by integrating over all segments within the cell's boundaries. Given the low density of cells in DM25 even at stationary phase, and to obtain sufficient numbers of cells for analysis in many visual fields, we concentrated most cultures 2-fold by centrifugation at 7,745 g for 2 min. Clones from two populations at generation 50,000 (Ara–1 and Ara–4) required 4-fold concentration. Samples from another population at generation 50,000 (Ara–3) were imaged without concentrating the medium. We then spotted 3-µl samples from each processed culture onto 1% agarose pads, and we imaged the cells using a Nikon Eclipse Ti-U inverted microscope.

Analysis of cell mortality in population Ara-3

We reanalyzed data on cell viability collected for two clones: the LTEE ancestor (REL606), and the 50,000-generation clone from population Ara–3 (REL11364) that evolved the

novel ability to use citrate as a source of carbon and energy (Cit⁺). We used the *Bac*Light viability kit for microscopy (ThermoFisher #L7007) following the manufacturer's directions for fluorescently labeling cells. In short, we mixed the provided components A and B in equal amounts, added 1 µl to 10-mL stationary-phase DM25 cultures of each clone, and incubated them for 20 min in the dark to prevent photobleaching. The two components contain two fluorescent dyes that differentially stain presumptively live and dead cells. For the Cit⁺ clone only, we also examined cells in DM0 medium, which contains the same concentration of citate as DM25, but no glucose. Full methods and additional results in the context of other work are reported in Blount et al. (2020).

Genomic and fitness data

We integrated our analyses of cell size and shape with previously published datasets on the fitness of the evolved bacteria relative to their ancestor, and on the mutations present in the various clones obtained by sequencing and comparing the evolved and ancestral genomes. The fitness data were previously collected by Wiser et al. (2013), who performed competition assays between evolved populations and reciprocally marked ancestors. We downloaded these data from the Dryad Digital Repository (accession https://doi.org/10.5061/dryad.0hc2m). The complete genomes of the ancestral strain and evolved clones used in our study were sequenced by Jeong et al. (2009) and Tenaillon et al. (2016), respectively. We used an online tool (https://barricklab.org/shiny/LTEE-Ecoli/) to identify all of the mutations that occurred in several genes (*mreB, mreC, mreD, mrdA*, and *mrdB*) known to be involved in maintaining rod-shaped cells in *E. coli*.

Statistical analyses

Statistical analyses were performed in R (Version 3.5.0, 2018-04-23). Our datasets and R analysis scripts will be made available on the Dryad Digital Repository (DOI pending publication).

Results

Our analyses and results are multi-faceted. They include: a comparison between two methods used to measure cell volumes; analyses of the evolutionary trends in cell size of both clones and whole-population samples; a comparison of sizes during exponential growth and at stationary phase; analyses of cell shape and the subsequent identification of mutations in genes known to affect cell shape; the correlation between cell size and relative fitness during the LTEE; and evidence for substantial cell mortality in a unique population.

Cell volumes measured by two methods

We first address whether the two approaches we used to estimate cell size provide comparable results. The Coulter-counter method directly estimates particle volumes, based on changes in conductance between two electrodes as cells suspended in an electrolyte solution are moved through a small aperture. The microscopy method involves obtaining cell images and processing them using software that defines the edges of objects, segments the objects into small pieces, and integrates the segments to estimate cell volumes. Figure 2.1 shows the highly significant correlation in the median cell volumes estimated using the two approaches for the two ancestors and 36 evolved clones from the 12 populations at three generations of the LTEE. All of these samples were grown in the same LTEE

conditions and measured in stationary phase at the 24-h mark (i.e., when they would be transferred to fresh medium under the LTEE protocol).

This concordance gives us confidence that we can use either approach when it is best suited to a given question. The Coulter counter method is especially well suited to efficient measurement of cell volumes for many cells from each of many samples. The microscopy and subsequent image processing, by contrast, is necessary to obtain information on changes in cell shape.

Temporal trends in cell size in evolved clones

It was previously reported that cell size increased in parallel across all 12 LTEE populations through 10,000 generations, and that the increase in cell volume was strongly correlated with the populations' improved fitness in the LTEE environment (Lenski and Travisano 1994). Subsequent papers reported continued fitness gains in the LTEE populations for an additional 40,000 generations (Wiser et al. 2013; Lenski et al. 2015), albeit at a declining rate of improvement. Here we ask whether cell size also continued to increase, focusing first on the clones isolated from each population at 2,000, 10,000 and 50,000 generations and measured during stationary phase.

Figure 2.2 shows that the evolved clones were all larger than their ancestors, although cell size did not always increase monotonically over the course of the LTEE. The median cell volumes of clones sampled from three populations (Ara–2, Ara–3, and Ara–6) were smaller at 10,000 generations than at 2,000 generations. Nonetheless, the median cell volumes of all 12 populations at 50,000 generations were greater than at 10,000 generations. However, the measurement noise associated with the rather small number of

biological replicates (i.e., independent cultures) for each clone, and the requirement to correct for multiple hypothesis tests, make it difficult to statistically ascertain the changes in cell volume between clones from successive generations. One possibility is that individual clones are not always representative of the populations from which they were sampled. If that were the case, then we would expect to see more consistent temporal trends in whole-population samples than in clones. We will address that issue in the next section. On balance, the median cell volumes of the evolved clones were on average 1.49, 1.68, and 2.55 times greater than the ancestor at 2,000, 10,000, and 50,000 generations, respectively (one-tailed paired *t*-tests: p = 0.0067, 0.0019, and 0.0006, respectively).

Besides the possible reversals in median cell size between 2,000 and 10,000 generations in a few populations, two other unusual cases are noteworthy. The 50,000-generation clone from population Ara–3 had by far the largest cells, with a median cell volume that was ~1.6 times greater than any other population at the same time point (Figure 2.2). That population is the only LTEE population that evolved the capacity to use the abundant citrate in the DM25 medium as an additional carbon source beyond the glucose that limits the other populations (Blount et al. 2008, 2012). The Cit+ phenotype is clearly advantageous, although it should also be noted that growth is slower on citrate than on glucose (Blount et al. 2020). Given that slower-growing *E. coli* cells tend to be smaller than faster-growing cells (Schaechter et al. 1958; Akerlund et al. 1995; Mongold and Lenski 1996; Bremer and Dennis 2008), and that this population's growth shifts in an apparent diauxic manner from glucose to citrate (Blount et al. 2020), it is surprising that this clone produces the largest stationary-phase cells of any clone we examined. Perhaps these cells are sequestering unused carbon, accounting for their large size; or perhaps the evident

stress they face during growth on citrate (Blount et al. 2020) leads to some decoupling of their growth and division. The other noteworthy population is Ara+1, which showed the smallest increase in cell volume (Figure 2.2). This population also achieved the smallest fitness gains of any of the LTEE populations (Wiser et al. 2013; Lenski et al. 2015). Given that growth rate is the main determinant of fitness in the LTEE (Vasi et al. 1994), it is therefore interesting (but not surprising) that Ara+1 is both the least fit and produces the smallest cells of any of the LTEE populations.

Figure 2.3 compares average cell volumes of the clones between the consecutive generations sampled. These analyses show that average cell size across the 12 LTEE lines increased significantly from the ancestor to generation 2,000, and between 10,000 and 50,000 generations; however, the increase between 2,000 and 10,000 generations was not significant.

Monotonic cell size trends among whole populations

We have so far established that the cell volume of clones usually, but not always, increased between the generations tested. However, the evolutionary changes in clones are not always representative of the populations from which they are sampled. For this reason, we measured the cell volumes of whole-population samples at the same three generations to see whether they might show more consistent temporal trends. Figure 2.4 shows the cell volume trajectories for these measurements. Indeed, the population samples showed more consistent trends toward larger cells than did the clones. The grand mean trend of the whole populations (Figure 2.5) closely mirrored the overall trend seen for clones (Figure 2.3). However, the correlation between cell volumes measured on clones and whole

populations, while highly significant overall, also showed considerable scatter (Figure 2.6), indicating that individual clones are not always representative of the populations from where they were sampled. One such difference was that the median volume in the 50,000-generation whole-population sample of Ara–3 was no longer an outlier when compared to the other populations (Figure 2.4), in contrast to the measurements on the individual clones (Figure 2.2). Another difference was the increase in median cell size from the ancestral state to generation 50,000 was much greater in the whole-population sample of Ara+1 than in the individual clone.

Overall, the temporal trend in cell volume does not appear to have reached any upper bound or asymptote, as each generation of whole-population samples that we tested had significantly larger cells than the preceding generation (Figure 2.5). However, the intervals between samples were also progressively longer. Therefore, we calculated the average rate of change in cell volume from the slopes calculated for each population between adjacent time points (Figure 2.7). The average rate of cell volume increase was ~ 0.17 fL per thousand generations in the first 2,000 generations but dropped to ~ 0.02 and ~ 0.007 fL per thousand generations in the following 8,000- and 40,000-generation intervals, respectively. In summary, these data show that cell size has continued to increase throughout the long duration of the LTEE, albeit at a decelerating pace and notwithstanding a few atypical evolved clones.

Differences in cell size between exponential and stationary phases

In the sections above, we established the following points: (i) there is good agreement between cell volumes estimated using the Coulter particle counter and by microscopy; (ii)

the evolved cells are generally much larger than their ancestors; (iii) there is a nearly monotonic trend over time toward larger cells, although at a declining rate and with a few clones as outliers; and (iv) the independently evolving populations show substantial variation in their average cell sizes after 50,000 generations. All of these conclusions were obtained using cells in stationary phase, and it is of interest to ask whether they also hold for exponentially growing cells. However, examining these issues with exponentially growing cells presents additional challenges. In particular, owing to evolved changes in growth rates and lag times (Vasi et al. 1994), cells from different generations and populations reach mid-exponential-phase growth at different times, complicating efforts to obtain consistent measurements. In addition, the DM25 medium in which the cells evolved is dilute: the stationary-phase population density of the ancestor is only $\sim 5 \ge 10^7$ cells per mL, and it is even lower for most evolved clones owing to their larger cells. Hence, cells in mid-exponential-phase growth are usually at densities less than 10⁷ cells per mL. For these reasons, and given the excellent correspondence between Coulter counter and microscopic data, we measured the distribution of cell volumes for exponentially growing cells using only the Coulter counter.

We measured cell volumes of the ancestors and 50,000-generation clones from all 12 LTEE populations 2 h and 24 h after they were transferred into fresh DM25 medium (Figure 2.8). At 2h, even the ancestors have begun growing exponentially (Vasi et al. 1994), and none of the evolved strains grow so fast that they would have depleted the limiting glucose by that time (Wiser et al. 2013). The 24-h time point corresponds to when the cells are transferred to fresh medium during the LTEE and hence leave stationary phase. This paired sampling strategy allows us to ask how predictive the stationary-phase cell volumes

are of exponentially growing cells. In fact, we found a strong positive correlation in cell volumes measured during exponential growth and stationary phase (Figure 2.9). The exponentially growing cells were consistently much larger than those in stationary phase for the ancestors as well as all of the 50,000-generation clones (Figure 2.8). For the evolved clones, the volumetric difference as a function of growth phase was \sim 2-fold, on average (Figure 2.20). It is well known that bacterial cells are larger during exponential growth, with each fast-growing cell typically having multiple copies of the chromosome and many ribosomes to support maximal protein synthesis. In the dilute glucose-limited DM25 minimal medium, cells hit stationary phase abruptly, with the last population doubling using up as much glucose as all the previous doublings combined. The \sim 2-fold volumetric difference between the exponentially growing cells and those measured many hours later in stationary phase implies that they typically undergo a reductive division, either as they enter or during stationary phase. At the same time, the range in size between the 12 independently evolved clones was also roughly 2-fold during both growth phases (Figure 2.8), which indicates that the striking morphological divergence extends across growth phases.

Changes in cell shape

Cell size has clearly increased during the LTEE. Has cell shape also changed? Cell shape has sometimes been regarded as invariant for a given species. For example, *E. coli* has rodshaped cells that typically maintain an aspect ratio (length-to-width) of ~4:1, independent of cell volume (Chang and Huang 2014; Harris and Theriot 2018). We examined and analyzed micrographs to see whether the larger cells that evolved in the LTEE maintained

their ancestral aspect ratio. Alternatively, larger volumes might have evolved by disproportionate increases in either the length or width of cells. Yet another possibility is that the lineages diverged in their aspect ratios not only from their common ancestor, but also from one another. Figure 2.10 shows representative micrographs of the ancestors and the 50,000-generation clones. It is readily apparent that the different lineages have evolved different aspect ratios. To investigate these differences more systematically, we processed multiple micrographs of the ancestors and clones from generations 2,000, 10,000, and 50,000 using the *SuperSegger* package (Stylianidou et al. 2016). Across all of the samples in total, we obtained lengths and widths (cross-sectional diameters) from >87,000 cells (see Methods). As a reminder, an increase in the aspect ratio relative to the ancestor implies a higher SA/V ratio for a given volume, whereas a decline in the aspect ratio indicates the opposite. Of course, having a larger cell alone also reduces the SA/V ratio, even without a change in the aspect ratio. One would typically expect a greater SA/V ratio to be beneficial in terms of resource acquisition, and therefore we might expect the evolved clones to have higher aspect ratios than the ancestral strains, especially given their increased volumes.

In fact, however, the opposite trend held, at least for the first 10,000 generations, as shown in Figure 2.11. Clones from 10 of the 12 populations, at both 2,000 and 10,000 generations, tended to produce relatively wider than longer cells in comparison to the ancestor (p = 0.0386 based on a two-tailed sign test at each time point). By 50,000 generations, the clones were split evenly: 5 had aspect ratios greater than the ancestor, 5 had aspect ratios lower than the ancestor, and 2 had aspect ratios nearly identical to the ancestor. Note that the 50,000-generation clone from population Ara–3 is an extreme outlier, with cells that are exceptionally long and very large. This population is the one that
evolved the novel ability to grow on citrate (Blount et al. 2008, 2012), and its unusual morphology is presumably related to its distinct metabolism.

Figure 2.12 shows the average length-to-width ratios and their associated 95% confidence intervals, excluding the Cit⁺ outlier at 50,000 generations. The ancestral cells had an average length-to-width ratio of 3.37. Recall that *E. coli* has been reported to typically maintain an average aspect ratio of about 4:1 (Chang and Huang 2014; Si et al. 2017; Harris and Theriot 2018). The aspect ratio we see is somewhat smaller. This difference might reflect variation between strains (the LTEE ancestor is a derivative of E. coli B, not K12), or some other factors. In any case, the mean aspect ratio across the evolved lines had declined to 2.90 and 2.87 at 2,000 and 10,000 generations, respectively, and then increased to 3.39 at generation 50,000, almost identical to the ancestral ratio. The early decline in the aspect ratio is significant, as is the subsequent reversal (Figure 2.12). This reversal would increase the SA/V ratio somewhat. However, it might not be sufficient to offset the reduction in the SA/V ratio associated with the much larger cell volumes at 50,000 generations. On balance, the LTEE lines evolved larger cell volumes by first increasing disproportionately in width, and later increasing their length, possibly to the benefit of a somewhat more favorable SA/V ratio.

Analysis of changes in the SA/V ratio

The reversion of the evolved clones to their ancestral aspect ratio (Figure 2.12), coupled with their overall increase in cell volume (Figure 2.3), raises the question of how much their SA/V ratios have changed. If selection to increase the diffusion of nutrients into cells is strong in the LTEE, then increasing cell length would be beneficial. However, the larger

cell volume would have the opposite effect. To examine the net result of these changes, we calculated the SA/V ratio of the evolved clones using the equations for spherocylindrical cells from Ojkic et al. (2019), which we presented in Introduction. We used the length and width values measured for clones using *SuperSegger* to compute for each cell γ , which depends on the aspect ratio, and from that the cell's surface area. We then divided that value by the cell's estimated volume to obtain its SA/V ratio. Given the early trend toward wider cells (lower aspect ratios) and the larger cell volumes at later generations, we expected lower SA/V ratios for the evolved clones relative to the ancestors, despite the later reversion toward the ancestral aspect ratio. Indeed, all 36 evolved clones had a SA/V ratio that was lower than the ancestors (Figure 2.13).

Figure 2.14 shows the average SA/V ratio and associated 95% confidence intervals over time. We included the 50,000-generation Ara–3 clone in this analysis because its SA/V ratio (Figure 2.13), unlike its aspect ratio (Figure 2.11), was not an extreme outlier; that is, its atypical aspect ratio was largely offset by its large average cell volume (Figure 2.2). The mean SA/V ratio declined monotonically and significantly from 0.461 in the ancestor to 0.430, 0.412, and 0.392 at 2,000, 10,000, and 50,000 generations, respectively. Even the reversion to the ancestral cell aspect ratio between 10,000 and 50,000 generations (Figure 2.12) was insufficient to offset the increase in cell volume over that same interval (Figure 2.3).

We also performed an isometric analysis to assess the extent to which the reversion to the ancestral aspect ratio between 10,000 and 50,000 generations changed the SA/V ratio. To do so, we used the cell aspect ratios measured at 10,000 generations and compared the average SA/V ratio at 50,000 generations to the hypothetical average using

the earlier aspect ratios. The average SA/V ratio at 50,000 generations was ~6% higher as a consequence of the change in cell aspect ratio (Figure 2.21), and this difference was significant (p = 0.0144). Even so, the mean SA/V ratio continued to decline (Figure 2.14) because the change in average cell aspect ratio over this period (Figure 2.12) was insufficient to offset the increase in average cell volume (Figure 2.3).

Nearly spherical cells in one LTEE population

While examining micrographs, we observed that cells from the Ara+5 population at 2,000 and 10,000 generations looked like stubby rods, many of which were almost spherical (Figure 2.15). By 50,000 generations, however, the cells were rod-shaped, suggesting that one or more mutations in morphogenic genes might contribute to this phenotype.

The typical rod-shaped cell morphology in *E. coli* is maintained by several proteins including MreB, MreC, MreD, MrdA (PBP2), and MrdB (RodA) (Kruse et al. 2005; Philippe et al. 2009). To this end, we examined published whole-genome sequence data (Tenaillon et al. 2016) for the clones in our study to identify any mutations in these genes. By 50,000 generations, all but one of the 12 lines (Ara–5) had nonsynonymous mutations in at least one of these five shape-maintaining genes (Figure 2.16). There were also a few synonymous changes, which were seen only in populations that had evolved point-mutation hypermutability, as well as one indel. However, the majority of mutations that arose and reached high frequency in these genes were nonsynonymous changes.

The 2000-generation clone from the Ara+5 population that produced the stubby cells had a single nonsynonymous mutation in *mreB*. This mutation was also present in the clones sampled from this population at 10,000 and 50,000 generations. There were no

other mutations in the other four rod-shape maintenance genes at any of the timepoints. *E.* coli cells have been shown to become spherical when MreB is depleted (Kruse et al. 2005), which strongly suggests that the *mreB* mutation is responsible for the stubby morphology observed in the early generations of this population. The fact that the Ara+5 cells were not stubby at 50,000 generations, despite the *mreB* mutation, suggests some compensatory change that did not involve the five morphogenic genes considered here. Four other populations also had nonsynonymous *mreB* mutations by generation 50,000 (Figure 2.16). Of these four, the clone from population Ara+1 also produced rather stubby cells (Figure 2.10), resulting in the lowest aspect ratio of any of the 50,000-generation clones (Figure 2.11). Whether the diverse effects of the *mreB* mutations on cell shape reflect the different mutations, the genetic backgrounds on which they arose, or both remains to be determined.

Cell volume and fitness have remained highly correlated in the LTEE

Cell size and relative fitness were previously shown to be strongly correlated during the first 10,000 generations of the LTEE (Lenski and Travisano 1994). The fitness of these populations has continued to increase throughout this experiment (Wiser et al. 2013; Lenski et al. 2015). In light of the continued increase in cell volumes reported in this work, we expected that cell size and fitness would continue to be correlated. To test this, we used the relative fitness data previously collected for the 12 LTEE populations through 50,000 generations (Wiser et al. 2013), and we asked how well those fitness values correlate with the cell volumes we measured for the ancestors and the whole-population samples from three later generations. Figure 2.17 shows that cell volume and relative fitness have

remained significantly correlated, although with substantial scatter. Some of this scatter reflects increased measurement noise when estimating relative fitness in later generations. These estimates are obtained by competing the evolved populations against a marked ancestor; as the relative fitness of the evolved bacteria increases, it becomes more difficult to enumerate accurately the relative performance of the two competitors.

Elevated cell mortality in the population that evolved to grow on citrate

We observed what we call "ghost" cells in micrographs of the 50,000-generation Cit⁺ clone from the Ara–3 population. These cells were quite distinct from the ancestral strain and evolved clones from all other populations (Figure 2.10). In terms of contrast with their background, the ancestor and other evolved clones had uniformly dark and opaque cells, in contrast to the lighter agar pad on which they were placed for imaging. Many of the Cit⁺ cells, by comparison, were translucent (Figure 2.10). Most translucent cells appeared intact, although we also saw some fragmented cells. We presume that the translucent cells that appear intact are nonetheless either dead or dying.

We also grew the Cit⁺ clone in DM0, which is the same medium as used in the LTEE and our other experiments, except DM0 contains only the citrate but no glucose. The proportion of ghost cells is even higher in this citrate-only medium (Figure 2.18). Some translucent cells had small punctations, or dots, within the cytoplasm, often at the cell poles (Figure 2.18). These dots are reminiscent of the polyhydroxyalkanoate storage granules that some bacterial species produce under conditions where their growth is unbalanced (Pötter and Steinbüchel 2006; Jendrossek 2009) or when cells are otherwise stressed

(Rowaihi et al. 2018; Obruca et al. 2020). It is also possible that these dots comprise the nucleoid or some other remnant of a leaky cytoplasm.

It is noteworthy that we observed these anomalous ghost cells at any appreciable frequency only in the unique Cit⁺ population (Blount et al. 2008, 2012). This observation of ghost cells, and the implication that many cells in this population are dead or dying, is supported by other observations that indicate the Cit⁺ cells struggle with maintaining balanced growth on citrate (Blount et al. 2020). To test whether the ghost cells are dead, dying, or at least physiologically incapacitated, we labeled stationary-phase cultures using a two-color live/dead stain. Our methods, full results, and in-depth analyses of these labelling experiments are presented elsewhere (Blount et al. 2020). Here we present a subset of the data, with an analysis that specifically compares the ancestor (REL606) and 50,000-generation Cit⁺ clone (REL11364). Figure 2.19A shows representative micrographs of the ancestral and evolved Cit⁺ cells grown to stationary phase in the standard DM25 medium that contains glucose as well as citrate. Figure 2.19B shows the estimated proportions of live (green) and dead (red) cells, obtained by pooling data from 5 independent cultures (i.e., biological replicates) for each clone. There was much more cell death in the cultures of the Cit⁺ clone when compared to the ancestor. On average, 43.6% of the Cit⁺ cells were scored as dead, based on greater intensity of the corresponding dye. By contrast, only 13.2% of the ancestral cells were scored as dead, and they exhibited much weaker intensity of that dye (Figure 2.19A). The difference in the proportion of dead cells between the ancestor and the Cit⁺ clone is highly significant (t = 2.9304, df = 8, one-tailed p = 0.0094). This result thus supports our hypothesis that the ghost cells seen in our original micrographs of the Cit⁺ clones were indeed dead or dying.

Discussion

During the first 10,000 generations of the LTEE, 12 populations of *E. coli* increased in fitness and cell size as they evolved in and adapted to their glucose-limited minimal medium (Lenski and Travisano 1994). The increase in cell size was unexpected, given the fact that larger cells have greater metabolic demands and have SA/V ratios that are less favorable for supporting those demands. In the >60,000 generations since that study, the populations have continued to adapt to the glucose media, and their fitness has continued to increase with trajectories that are well described by a simple power law (Wiser et al. 2013; Lenski et al. 2015). In this study, we sought to determine if cell size has continued to increase, and whether cell size still correlates with fitness. We measured changes in cell volume and shape for clones and whole-population samples. We used two methods: a Coulter counter that directly measures cell volume, and microscopy that allowed us to analyze both cell volume and shape using machine learning. The average cell volumes measured using the two methods were well correlated (Figure 2.1).

The average cell increased monotonically over time in the whole-population samples (Figures 2.4–2.5). Clones from three populations (Ara–2, Ara–3, Ara–6) deviated from this monotonic trend, producing smaller cells at 10,000 than at 2,000 generations (Figure 2.2). These idiosyncratic cases implies within-population heterogeneity. They might also be due, in part, to the clones being studied in an environmental context different from that in which they evolved. As an indication of the relevance of both of these explanations, two ecologically and genetically distinct lineages have coexisted in the Ara–2 population since ~6,000 generations, with coexistence mediated by differential growth on glucose and acetate, a metabolic byproduct (Rozen et al. 2005, Grosskopf et al. 2016). In

any case, our data show that average cell size and mean fitness have remained significantly correlated in the LTEE through 50,000 generations (Figure 2.17), despite variation within and between populations.

We obtained most of our data on average cell size with cells in stationary phase, at the end of the LTEE's standard 24-hour period prior to the transfer into fresh medium. We did so because analyzing exponentially growing cells presents additional challenges. In particular, the evolved cells reach exponential-phase growth faster than the ancestor, owing to changes in growth rates and lag times (Vasi et al. 1994). Also, cell densities are lower during exponential growth, especially given the low glucose concentration in the LTEE medium. Nonetheless, we performed a set of experiments to compare the average volumes of exponentially growing and stationary-phase populations (Figure 2.8). Exponentially growing cells were larger than stationary-phase cells, and this difference was observed using both the ancestor and evolved bacteria, Bacterial cells are larger during exponential growth to accommodate more ribosomes (Valgepea et al. 2013) and replicating chromosomes (Wang and Levin 2009; Chien et al. 2012). The approximately two-fold difference in average cell volume between exponential and stationary phases for the 50,000-generation clones (Figure 2.20) implies that these bacteria undergo a reductive division as they enter or during stationary phase.

The 12 LTEE populations have evolved shorter lag phases and faster maximal growth rates during their adaptation to the LTEE environment. Therefore, when compared to the ancestor, evolved cells spend more time in the stationary-phase period between transfers. In silico models of the daily transfer regime typical of experimental evolution systems, including the LTEE, have shown that virtual microbes can evolve to anticipate the

transfer interval (van Dijk et al 2019). A reductive division during stationary phase might prime the cells to grow faster when transferred into fresh medium by temporarily increasing their SA/V ratio, potentially reducing the duration of the lag phase. However, we note that the LTEE ancestors also undergo a similar reductive division, as indicated by smaller cells in stationary phase than during exponential growth (Figure 2.8). Thus, the reductive division per se does not account for the shortened lag phase in the evolved bacteria. In any case, future studies might examine when this reductive division occurs in the ancestral and evolved bacteria and, moreover, identify the metabolic cues and physiological processes involved.

We also observed substantial heterogeneity in the cell shape of the evolved lines (Figure 2.10). One population (Ara+5) evolved stubby, almost spherical, cells early in the LTEE (Figure 2.15A), evidently caused by a mutation in *mreB*, which encodes a protein involved in maintaining the rod shape that is typical of *E. coli*. This population later reevolved more rod-shaped cells (Figure 2.15B), although the genetic basis for that change is unclear. More generally, most populations evolved relatively wider cells during the first 10,000 generations (Figure 2.11), even though longer cells would have had a higher SA/V ratio (Harris and Theriot, 2018). This trend suggests that cell size evolution in the LTEE is not tightly constrained by the SA/V ratio. In later generations, the average cell aspect ratio (length/width) reverted to the ancestral ratio (Figure 2.12), but not enough to prevent a further decline in the average SA/V ratio (Figure 2.14), as the mean cell volume continued to increase (Figure 2.3).

For a given cell volume, wider cells have lower SA/V ratios than longer cells. From the standpoint of acquiring limited nutrients, wider cells would therefore seem

maladaptive, yet that is how the LTEE populations tended to evolve for the first 10,000 generations (Figure 2.11). Might wider cells have had some benefit that overcame their unfavorable SA/V ratios? As a bacterial cell grows in size, it simultaneously replicates multiple copies of its chromosome. These copies must then be fully segregated into the two daughter cells, which requires moving them away from the cell center before the division can be completed (Chien et al. 2012). Rod-shaped bacteria like *E. coli* typically divide at the middle of the cell; the midpoint is defined by the proteins MinCDE, which oscillate between the cell poles every 40-90 seconds while consuming ATP (Rasking and de Boer 1999a; Rasking and de Boer 1999b; Huang et al. 2003; Lutkenhaus 2007; Dajkovic et al. 2008; Arumugam et al. 2014). The number of MinCDE complexes doubles in cells longer than ~ 4 μm, while their oscillatory period remains constant (Rasking and de Boer 1999a). It has also been shown that MinCDE proteins do not oscillate at all in shorter cells, which have a reduced aspect ratio; instead, they exhibit stochastic switching between the two poles (Ramm et al. 2019). This stochastic switching reduces the rate at which these proteins use ATP (Fischer-Friedrich et al. 2010). Thus, one could imagine that evolving wider cells, which also have a reduced aspect ratio, would increase the ATP available for other metabolic processes. Future studies might study the oscillatory behavior and ATP consumption of these proteins in the LTEE lines.

Another potential advantage of wider cells is to minimize the macromolecular crowding that occurs within the highly concentrated cellular cytoplasm (Minton 2006). Gallet et al. (2017) suggested that the increased cell width in the LTEE lines might reduce the adverse effects of macromolecular crowding, but they did not directly test this hypothesis. However, they also proposed that the bacterial cells became larger in order to

become less densely packed, which would allow greater internal diffusion of resources and macromolecules. Gallet et al. (2017) found evidence in support of this second hypothesis in the one LTEE population they examined, where the cell density (dry mass-to-volume) declined over evolutionary time. If the rate of resource acquisition from the external environment does not limit growth, then increasing the rate of internal diffusion should increase the cell's metabolic rate and, at least potentially, lead to faster growth and higher fitness (Beveridge, 1988; Koch, 1996; Schulz and Jorgensen, 2001; Golding and Cox, 2006; Young, 2006; Beg et al., 2007; Ando and Skolnick, 2010; Dill et al., 2011). Therefore, it would be interesting to extend the analyses performed by Gallet et al. (2017) to all of the LTEE populations to assess the generality of their findings.

We also made the serendipitous discovery that one population, called Ara–3, evolved greatly elevated cell mortality (Figures 2.10 and 2.19). That population is the only one that evolved the ability to assimilate energy from citrate, which is in the LTEE medium as an iron chelator (Blount et al. 2008). We subsequently showed that this increased mortality has persisted in the population for almost 20,000 generations, and perhaps even longer (Blount et al. 2020). The persistence of this elevated death suggests some physiological constraint that is difficult to overcome, though this cost must be smaller than the benefit provided by the access to this additional resource. In any case, a 50,000generation clone that we analyzed from this population was also an outlier in other morphological respects, producing cells that are exceptionally large (Figure 2.2) and long (Figure 2.11). In addition to the many ghost-like cells that appear to be dead or dying (Figure 2.10), some of these translucent cells have inclusions within the cytoplasm. Future

studies may investigate the genetic and physiological bases of these unusual morphological traits and their relation to growth on citrate and cell death.

In summary, we have observed substantial changes in cell morphology, including shape as well as size, over the course of 50,000 generations of the *E. coli* LTEE. Some of the changes are highly repeatable including especially the parallel trend toward larger cells observed in all 12 independently evolving populations. At the same time, the replicate populations have evolved highly variable phenotypes, even under identical conditions, leading to approximately two-fold variation in their average cell volumes (Figure 2.8) as well as equally large differences in their aspect ratios (Figure 2.11). The consistent trend toward larger cells (Figure 2.4), the strong positive correlation of cell volume with fitness (Figure 2.17), and the parallel substitutions in genes involved in maintaining cell shape (Figure 2.16) all suggest that the evolution of cell morphology is not a mere spandrel, but instead reflects adaptation to the LTEE environment. The resulting among-population variation in size and shape, however, suggest that precise changes in cell morphology were not critical to performance, because most populations have improved in relative fitness to a similar degree (Lenski et al. 2015), despite different cell morphologies. Thus, the changes in cell size and shape during the LTEE reflect both natural selection and the idiosyncratic nature of the chance events, including mutations, particular to every evolving lineage.

Acknowledgments

We thank Terence Marsh, Charles Ofria, Gemma Reguera, and Chris Waters for feedback as this research progressed, and members of the Lenski lab for valuable discussions. We also thank Rohan Maddemsetti and Zachary Blount for their comments on the manuscript. This work was supported in part by a grant from the National Science Foundation (currently DEB-1951307), the BEACON Center for the Study of Evolution in Action (DBI-0939454), and the USDA National Institute of Food and Agriculture (MICL02253). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funders.

APPENDIX



Figure 2.1: Correlation between cell volume measurements obtained using microscopy and Coulter counter. Volumes obtained by microscopy are expressed in arbitrary units (a.u.) proportional to fL (i.e., μ m³); volumes obtained using the Coulter counter are expressed in fL. Each point shows the grand median of three assays for clones sampled from the 12 evolving populations or of six assays for the two ancestral strains. Kendall's coefficient τ = 0.5495, *N* = 38, *p* < 0.0001.



Figure 2.2: Cell size trajectories of clones obtained using Coulter counter. Each quantile (5th, 25th, 50th, 75th, and 95th) represents the median of the corresponding quantile from six replicates of each ancestor (REL607 for "Ara+" populations; REL606 for "Ara–" populations) and three replicates for clones sampled from each evolving population.



Figure 2.3: Tests of changes over time in average cell sizes of clones sampled from the 12 LTEE populations. Each point shows the grand mean across all populations of the median cell volume for each population, except the outlier clone from population Ara–3 at 50,000 generations (Figure 2.2) is excluded. Error bars are 95% confidence intervals, and brackets show the statistical significance (*p* value) based on one-tailed paired *t*-tests. The last comparison remains significant even if one includes the outlier clone (*p* = 0.0090).



Figure 2.4: Cell size trajectories for whole-population samples obtained using Coulter counter. Each quantile (5th, 25th, 50th, 75th, and 95th) represents the median of the corresponding quantile from six replicates of each ancestor (REL607 for "Ara+" populations; REL606 for "Ara–" populations) and three replicates for each evolved population.



Figure 2.5: Tests of changes over time in average cell size for whole-population samples. Each point shows the grand mean of the grand median cell volumes calculated for each population. Error bars are 95% confidence intervals, and brackets show the statistical significance (*p* value) based on one-tailed paired *t*-tests.



Figure 2.6: Correlation between cell volumes of clones and whole-population samples. The clone and population values are the medians from Figures 2 and 4, respectively. Kendall's coefficient $\tau = 0.4900$, N = 38, p < 0.0001.



Figure 2.7: Average rate of cell volume increase. Slopes were calculated for each population over each of three intervals. Each point shows the grand mean for the 12 populations. Error bars are 95% confidence intervals, and brackets show the statistical significance (*p* value) based on one-tailed Wilcoxon tests, which account for the paired nature of the samples.



Figure 2.8: Cell sizes measured during exponential and stationary phases of ancestral strains and 50,000-generation clones from all 12 populations. Each point represents the median cell volume for one assay at either 2 h (exponential growth) or 24 h (stationary phase) in DM25. Horizontal bars are the means of the 3 replicate assays for each strain. The points for some individual replicates are not visible because some values were almost identical.



Figure 2.9: Correlation between cell sizes during exponential growth and in stationary phase. Each point represents the average over 3 replicates of the median cell volume in each growth phase using the data shown in Figure 8. Kendall's coefficient $\tau = 0.7582$, N = 14, p << 0.0001.



Figure 2.10: Representative micrographs of ancestors (REL606 and REL607) and evolved clones from each population at 50,000 generations. Phase-contrast images were taken at 100 x magnification. Scale bars are 10 μ m.



Figure 2.11: Average cell aspect ratios (length/width) of ancestral and evolved clones. Each point shows the mean ratio for the indicated sample. The lines show deviations in the aspect ratio from the ancestral state. The mean aspect ratios were calculated from three replicate assays in all but 4 cases (Ara–4 at 10,000 generations; Ara–2, Ara–4, and Ara–5 at 50,000 generations), which had two replicates each.



Figure 2.12: Evolutionary reversal of cell aspect ratio. Each point is the grand mean of the cell aspect ratio (length/width) for the ancestors and evolved clones. N = 12, except at 50,000 generations, where N = 11 after excluding the outlier clone from the Ara–3 population. Errors bars are 95% confidence intervals, and brackets show the statistical significance (p value) based on two-tailed t-tests. The tests were paired for clones sampled from the same population at the consecutive time points, and the Ara–3 population was excluded from the final test.



Figure 2.13: Average surface area-to-volume ratio (SA/V) of ancestral and evolved clones. The surface area and volume of individual cells were calculated from microscopic images, as described in the text, and their ratio has arbitrary units (a.u.) proportional to μ m⁻¹. Each point shows the mean ratio for the indicated sample. The lines show deviations in the ratio from the ancestral state. The means were calculated from three replicate assays in all but 4 cases (Ara–4 at 10,000 generations; Ara–2, Ara–4, and Ara–5 at 50,000 generations), which had two replicates each.



Figure 2.14: Tests of changes over time in the average surface area-to-volume ratio (SA/V). Each point shows the grand mean of the average ratio calculated for the ancestor and evolved clones. Error bars are 95% confidence intervals, and brackets show the statistical significance (p value) based on one-tailed paired t-tests.



Figure 2.15: Representative micrographs of cells from (A) 2,000-generation and (B) 50,000generation clones of the Ara+5 population. Phase contrast images were taken on an inverted microscope at 100 x magnification. Scale bars are 10 μ m. Arrows point to examples of nearly spherical cells in the earlier sample, which are not seen in the later one.



Figure 2.16: Parallel mutations in genes known to be involved in the maintenance of rodshaped genes. Nonsynonymous mutations were found in all populations except Ara–5 by 50,000 generations. Populations Ara–2, Ara–4, Ara+3, and Ara+6 evolved hypermutable phenotypes between generations 2,000 and 10,000; populations Ara–1 and Ara–3 did so between 10,000 and 50,000 generations. Hence, all synonymous mutations were found in lineages with a history of elevated point-mutation rates.



Figure 2.17: Correlation between mean fitness relative to the LTEE ancestor and grand median cell volumes, both based on whole-population samples. Four points (Ara+6 at 10,000 generations; Ara–2, Ara–3, and Ara+6 at 50,000 generations) are absent due to missing fitness values reported by Wiser et al. (2013). Kendall's τ = 0.6066, *N* = 34, *p* < 0.0001.



Figure 2.18: Representative micrograph of 50,000-generation Cit⁺ clone from population Ara–3 grown in DM0. As shown in Figure 10, we observed translucent "ghost" cells in the only population that evolved the capacity to use citrate in the LTEE medium (DM25). This clone can also grow on citrate alone in the same medium except without glucose (DM0), which increased the proportion of presumably dead or dying ghost cells. Red arrows point to several ghost cells, some of which have darker punctate inclusions; white arrows point to several more typically opaque and presumably viable cells. Scale bar is 10 μ m.



Figure 2.19: Comparison of cell death in the ancestor and Cit⁺ clone. (A) Representative micrographs showing live-dead staining of the LTEE ancestor (REL606) and the 50,000-generation Cit⁺ clone from population Ara–3 (REL11364), both grown in DM25. Scale bars are 10 μ m. (B) Proportions of cells scored as alive (green) or dead (red), based on two-color stain assay. For each clone, we assayed cells from 5 biological replicates, which have been pooled in this figure.



Figure 2.20: Difference in cell size between exponential and stationary phases. We calculated the grand median cell volume of each 50,000-generation clone grown for 2 h (exponential) or 24 h (stationary), and then computed the grand mean of the 12 populations. Error bars are 95% confidence intervals, and the bracket shows the statistical significance (*p* value) based on a one-tailed Wilcoxon test, which accounts for the paired nature of the samples.



Figure 2.21: Isometric analysis of the surface area-to-volume ratio (SA/V) to assess the effect of the change in aspect ratio (length/width) from 10,000 to 50,000 generations. For each population, we calculated a hypothetical mean SA/V using its average aspect ratio at 10,000 generations (no shape change), and we compared it to a hypothetical mean SA/V calculated using its average aspect ratio at 50,000 generations (with shape change). This latter value differs from that shown in Figure 2.14, because that figure is based on direct measurements of individual cells, followed by averaging the cells in a single assay, averaging the replicate assays for each population, and calculating the grand mean of the 12 populations. By contrast, calculating a hypothetical mean SA/V for a population using its aspect ratio from a different generation can only use average values of the relevant parameters. Therefore, we applied the same population-level averages to compare ratios with and without shape changes here. Error bars are 95% confidence intervals, and the bracket shows the statistical significance (*p* value) based on a one-tailed paired *t*-test.

Clone ID	Population	Generation
REL606	Ara [_] ancestor	0
REL607	Ara ⁺ ancestor	0
REL1158A	Ara+1	2,000
REL1159A	Ara+2	2,000
REL1160A	Ara+3	2,000
REL1161A	Ara+4	2,000
REL1162A	Ara+5	2,000
REL1163A	Ara+6	2,000
REL1164A	Ara-1	2,000
REL1165A	Ara-2	2,000
REL1166A	Ara-3	2,000
REL1167A	Ara-4	2,000
REL1168A	Ara-5	2,000
REL1169A	Ara-6	2,000
REL4530A	Ara+1	10,000
REL4531A	Ara+2	10,000
REL4532A	Ara+3	10,000
REL4533A	Ara+4	10,000
REL4534A	Ara+5	10,000
REL4535A	Ara+6	10,000
REL4536A	Ara-1	10,000
REL4537A	Ara-2	10,000
REL4538A	Ara-3	10,000
REL4539A	Ara-4	10,000
REL4540A	Ara–5	10,000
REL4541A	Ara–6	10,000
REL11392	Ara+1	50,000
REL11342	Ara+2	50,000
REL11345	Ara+3	50,000
REL11348	Ara+4	50,000
REL11367	Ara+5	50,000
REL11370	Ara+6	50,000
REL11330	Ara-1	50,000
REL11333	Ara-2	50,000
REL11364	Ara-3	50,000
REL11336	Ara-4	50,000
REL11339	Ara–5	50,000
REL11389	Ara–6	50,000

 Table 2.1: List of E. coli clones used in study.
Sample ID	Population	Generation
REL1158	Ara+1	2,000
REL1159	Ara+2	2,000
REL1160	Ara+3	2,000
REL1161	Ara+4	2,000
REL1162	Ara+5	2,000
REL1163	Ara+6	2,000
REL1164	Ara-1	2,000
REL1165	Ara-2	2,000
REL1166	Ara-3	2,000
REL1167	Ara-4	2,000
REL1168	Ara-5	2,000
REL1169	Ara-6	2,000
REL4530	Ara+1	10,000
REL4531	Ara+2	10,000
REL4532	Ara+3	10,000
REL4533	Ara+4	10,000
REL4534	Ara+5	10,000
REL4535	Ara+6	10,000
REL4536	Ara-1	10,000
REL4537	Ara-2	10,000
REL4538	Ara-3	10,000
REL4539	Ara-4	10,000
REL4540	Ara-5	10,000
REL4541	Ara-6	10,000
REL11383	Ara+1	50,000
REL11325	Ara+2	50,000
REL11326	Ara+3	50,000
REL11327	Ara+4	50,000
REL11362	Ara+5	50,000
REL11363	Ara+6	50,000
REL11318	Ara-1	50,000
REL11319	Ara-2	50,000
REL11354	Ara-3	50,000
REL11321	Ara-4	50,000
REL11322	Ara-5	50,000
REL11382	Ara-6	50,000

Table 2.2: List of *E. coli* whole-population samples used in study.

LITERATURE CITED

LITERATURE CITED

- Akerlund, T., K. Nordstrom, and R. Bernander. 1995. Analysis of cell size and DNA content in exponentially growing and stationary-phase batch cultures of *Escherichia coli*. Journal of Bacteriology 177:6791–6797.
- Amir, A. 2017. Is cell size a spandrel? eLife 6:1–8.
- Ando, T., and J. Skolnick. 2010. Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion. Proceedings of the National Academy of Sciences of the United States of America 107:18457–18462.
- Angert, E. R., K. D. Clements, and N. R. Pace. 1993. The largest bacterium. Nature 362:239–241.
- Arumugam, S., Z. Petrašek, and P. Schwille. 2014. MinCDE exploits the dynamic nature of FtsZ filaments for its spatial regulation. Proceedings of the National Academy of Sciences of the United States of America 111:E1192-E1200.
- Batani, G., G. Pérez, G. Martínez de la Escalera, C. Piccini, and S. Fazi. 2016. Competition and protist predation are important regulators of riverine bacterial community composition and size distribution. Journal of Freshwater Ecology 31:609–623.
- Beg, Q. K., A. Vazquez, J. Ernst, M. A. de Menezes, Z. Bar-Joseph, A-L Barabási, and Z. N. Oltvai. 2007. Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. Proceedings of the National Academy of Sciences of the United States of America 104:12663–12668.
- Beveridge, T. J. 1988. The bacterial surface: general considerations towards design and function. Canadian Journal of Microbiology 34:363–372.
- 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*. Proceedings of the National Academy of Sciences of the United States of America 105:7899–7906.
- Blount, Z. D., J. E. Barrick, C. J. Davidson, and R. E. Lenski. 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. Nature 489:513–518.
- Blount, Z. D., R. Maddamsetti, N. A. Grant, S. T. Ahmed, T. Jagdish, J. A. Baxter, B. A. Sommerfeld, A. Tillman, J. Moore, J. L. Slonczewski, J. E. Barrick, and R. E. Lenski. 2020. Genomic and phenotypic evolution of *Escherichia coli* in a novel citrate-only resource environment. eLife 9:e55414.

- Bremer, H., and P. P. Dennis. 2008. Modulation of chemical composition and other parameters of the cell by growth rate. *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology 3:1–2.
- Champion, J. A., A. Walker, and S. Mitragotri. 2008. Role of particle size in phagocytosis of polymeric microspheres. Pharmaceutical Research 25:1815–1821.
- Chang, F., and K. C. Huang. 2014. How and why cells grow as rods. BMC Biology 12:54
- Chien, A. C., N. S. Hill, and P. A. Levin. 2012. Cell size control in bacteria. Current Biology 22:R340–R349.
- Choi, C., E. Kuatsjah, E. Wu, and S. Yuan. 2010. The effect of cell size on the burst size of T4 bacteriophage infections of *Escherichia coli* B23. Journal of Experimental Microbiology and Immunology 14:85–91.
- Cooper, G. M. 2000. The Cell: A Molecular Approach, 2nd edition. Sinauer, Sunderland, Mass.
- Corno, G., and K. Jürgens. 2006. Direct and indirect effects of protist predation on population size structure of a bacterial strain with high phenotypic plasticity. Applied and Environmental Microbiology 72:78–86.
- Dajkovic, A., G. Lan, S. X, Sun, D. Wirtz, and J. Lutkenhaus. 2008. MinC spatially controls bacterial cytokinesis by antagonizing the scaffolding function of FtsZ. Current Biology 18:235–244.
- Dill, K. A., K. Ghosh, J. D. Schmit. 2011. Physical limits of cells and proteomes. Proceedings of the National Academy of Sciences of the United States of America 108: 17876–17882.
- Don, M. 2003. The Coulter principle: Foundation of an industry. Journal of the Association for Laboratory Automation 8:72–81.
- Doshi, N., and S. Mitragotri. 2010. Macrophages recognize size and shape of their targets. PLoS ONE 5:e10051.
- Fischer-Friedrich, E., G. Meacci, J. Lutkenhaus, H. Chaté, and K. Kruse. 2010. Intra- and intercellular fluctuations in Min-protein dynamics decrease with cell length. Proceedings of the National Academy of Sciences of the United States of America 107:6134–6139.
- Gallet, R., C. Violle, N. Fromin, R. Jabbour-Zahab, B. J. Enquist, and R. Lenormand. 2017. The evolution of bacterial cell size: the internal diffusion-constraint hypothesis. ISME Journal 11:1559-1568.
- Giovannoni, S. J. 2017. SAR11 Bacteria: The most abundant plankton in the oceans. Annual Review of Marine Science 9:231–255.

- Golding, I., and E. C. Cox. 2006. Physical nature of bacterial cytoplasm. Physical Review Letters 96:98–102.
- Gould, S. J., and R. C. Lewontin. 1979. The spandrels of San Marco and the panglossian paradigm: a critique of the adaptationist programme. Proceedings of the Royal Society B 205:581–598.
- Graves, J. L., K. L. Hertweck, M. A. Phillips, M. V. Han, L. G. Cabral, T. T. Barter, L. F. Greer, M. K. Burke, L. D. Mueller, M. R. Rose. 2017. Genomics of parallel experimental evolution in *Drosophila*. Molecular Biology and Evolution 34:831–842.
- Grosskopf, T., J. Consuegra, J. Gaffé, J. Willison, R. E. Lenski, O. S. Soyer, and D. Schneider. 2016. Metabolic modelling in a dynamic evolutionary framework predicts adaptive diversification of bacteria in a long-term evolution experiment. BMC Evolutionary Biology 16:163.
- Harris, L. K., and J. A. Theriot. 2018. Surface area to volume ratio: A natural variable for bacterial morphogenesis. Trends in Microbiology 26:815–832.
- Heim, N. A., J. L. Payne, S. Finnegan, M. L. Knope, M. Kowalewski, S. K. Lyons, D. W. McShea, P. M. Novack-Gottshall, F. A. Smith, S. C. Wang. 2017. Hierarchical complexity and the size limits of life. Proceedings of the Royal Society B 284:20171039.
- Huang, K. C., Y. Meir, and N. S. Wingreen. 2003. Dynamic structures in *Escherichia coli*: Spontaneous formation of MinE rings and MinD polar zones. Proceedings of the National Academy of Sciences of the United States of America 100:12724–12728.
- Ishiguro, N., M. Sasatsu, T. K. Misra, and S. Silver. 1988. Promoters and transcription of the plasmid-mediated citrate utilization system in *Escherichia coli*. Gene 68:181–192.
- Jendrossek, D. 2009. Polyhydroxyalkanoate granules are complex subcellular organelles (carbonosomes). Journal of Bacteriology 191:3195–3202.
- Jeong, H., V. Barbe, C. H. Lee, D. Vallenet, D. S. Yu, S.-H. Choi, A. Couloux, et al. 2009. Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). Journal of Molecular Biology 394:644–652.
- Kawecki, T. J., R. E. Lenski, D. Ebert, B. Hollis, I. Olivieri, and M. C. Whitlock. 2012. Experimental evolution. Trends in Ecology & Evolution 27:547–560.
- Koch, A. L. 1996. What size should a bacterium be? A question of scale. Annual Reviews of Microbiology 50:317–348.
- Koch, A. L. 2003. Bacterial wall as target for attack: past, present, and future research. Clinical Microbiology Reviews 16:673–687.

- Kruse, T., J. Bork-Jensen, and K. Gerdes. 2005. The morphogenetic MreBCD proteins of *Escherichia coli* form an essential membrane-bound complex. Molecular Microbiology 55:78–89.
- LaBar, T., and C. Adami. 2017. Evolution of drift robustness in small populations. Nature Communications 8:1012.
- Lenski, R. E. 2017. Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. ISME Journal 11:2181–2194.
- Lenski, R. E., and J. A. Mongold. 2000. Cell size, shape, and fitness in evolving populations of bacteria. In: J. Brown and G. West (eds). Scaling in Biology. Oxford University Press, Oxford, pp 221–234.
- Lenski, R. E., C. Ofria, R. T. Pennock, and C. Adami. 2003. The evolutionary origin of complex features. Nature 423:139–144.
- 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 10,000generation experiment with bacterial populations. Proceedings of the National Academy of Sciences of the United States of America 91:6808–6814.
- Lenski, R. E., M. J. Wiser, N. Ribeck, Z. D. Blount, J. R. Nahum, J. J. Morris, L. Zaman, C. Turner, B. Wade, R. Maddamsetti, A. R. Burmeister, E. J. Baird, J. Bundy, N. A. Grant, K. J. Card, M. Rowles, K. Weatherspoon, S. E. Papoulis, R. Sullivan, C. Clark, J. S. Mulka, and N. Hajela. 2015. Sustained fitness gains and variability in fitness trajectories in the long-term evolution experiment with *Escherichia coli*. Proceedings of the Royal Society B 282:20152292.
- Levin, P. A., and E. R. Angert. 2015. Small but mighty: cell size and bacteria. Cold Spring Harbor Perspectives in Biology 7:a019216.
- Lutkenhaus, J. 2007. Assembly dynamics of the bacterial MinCDE system and spatial regulation of the Z ring. Annual Review of Biochemistry 76:539–562.
- Marshall, W. F., K. D. Young, M. Swaffer, E. Wood, P. Nurse, A. Kimura, J. Frankel, J. Wallingford, V. Walbot, X. Qu, and A. H. K. Roeder. 2012. What determines cell size? BMC Biology 10:101.
- Mika, J. T., and B. Poolman. 2011. Macromolecule diffusion and confinement in prokaryotic cells. Current Opinion in Biotechnology 22:117–126.

- Mika, J. T., G. van den Bogaart, L. Veenhoff, V. Krasnikov, and B. Poolman. 2010. Molecular sieving properties of the cytoplasm of *Escherichia coli* and consequences of osmotic stress. Molecular Microbiology 77:200–207.
- Miller, C. 2004. SOS response induction by beta-lactams and bacterial defense against antibiotic lethality. Science 305:1629–1631.
- Minton, A. P. 2006. How can biochemical reactions within cells differ from those in test tubes? Journal of Cell Science 119:2863–2869.
- Mongold, J. A., and R. E. Lenski. 1996. Experimental rejection of a nonadaptive explanation for increased cell size in *Escherichia coli*. Journal of Bacteriology 178:5333–5334.
- Nikolaidis, I., S. Favini-Stabile, and A. Dessen. 2014. Resistance to antibiotics targeted to the bacterial cell wall. Protein Science 23:243–259.
- Obruca, S., P. Sedlacek, E. Slaninova, I. Fritz, C. Daffert, K. Meixner, Z. Sedrlova, and M. Koller. 2020. Novel unexpected functions of PHA granules. Applied Microbiology and Biotechnology 104:4795-4810.
- Ojkic, N., D. Serbanescu, and S. Banerjee. 2019. Surface-to-volume scaling and aspect ratio preservation in rod-shaped bacteria. eLife 8:1–11.
- Parry, B. R., I. V. Surovtsev, M. T. Cabeen, C. S. O'Hern, E. R. Dufresne, and C. Jacobs-Wagner. 2014. The bacterial cytoplasm has glass-like properties and is fluidized by metabolic activity. Cell 156:183–194.
- Philippe, N., L. Pelosi, R. E. Lenski, and D. Schneider. 2009. Evolution of penicillin-binding protein 2 concentration and cell shape during a long-term experiment with *Escherichia coli*. Journal of Bacteriology 191:909–921.
- Pötter, M., and A. Steinbüchel. 2006. Biogenesis and structure of polyhydroxyalkanoate granules. In: J. M. Shively (ed), Inclusions in Prokaryotes. Springer, Berlin, pp 110–136.
- Ramm, B., T. Heermann, and P. Schwille. 2019. The *E. coli* MinCDE system in the regulation of protein patterns and gradients. Cellular and Molecular Life Sciences 76:4245–4273.
- Rappé, M. S., S. A. Connon, K. L. Vergin, and S. J. Giovannoni. 2002. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. Nature 418:630–633.
- Raskin, D. M., and P. A. de Boer. 1999a. Rapid pole-to-pole oscillation of a protein required for directing division to the middle of *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America 96:4971–4976.
- Raskin, D. M., and P. A. de Boer. 1999b. MinDE-dependent pole-to-pole oscillation of division inhibitor MinC in *Escherichia coli*. Journal of Bacteriology 181:6419–6424.

- Ratcliff, W. C., R. F. Denison, M. Borrello, and M. Travisano. 2012. Experimental evolution of multicellularity. Proceedings of the National Academy of Sciences of the United States of America 109:1595–1600.
- Rozen, D. E., D. Schneider, and R. E. Lenski. 2005. Long-term experimental evolution in *Escherichia coli*. XIII. Phylogenetic history of a balanced polymorphism. Journal of Molecular Evolution 61:171–180.
- Rowaihi, I. S. Al, A. Paillier, S. Rasul, R. Karan, S. W. Grötzinger, K. Takanabe, and J. Eppinger. 2018. Poly(3-hydroxybutyrate) production in an integrated electromicrobial setup: Investigation under stress-inducing conditions. PLoS ONE 13:e0196079.
- Schaechter, M., O. Maaloe, and N. O. Kjeldgaard. 1958. Dependency on medium and temperature of cell size and chemical composition during balanced growth of *Salmonella typhimurium*. Journal of General Microbiology 19:592–606.
- Schavemaker, P. E., A. J. Boersma, and B. Poolman. 2018. How important is protein diffusion in prokaryotes? Frontiers in Molecular Biosciences 5:1–16.
- Schulz, H., and B. Jorgensen. 2001. Big bacteria. Annual Review of Microbiology 55:105–137.
- Sherratt, D. J. Oscillation helps to get division right. 2016. Proceedings of the National Academy of Sciences of the United States of America. 113:2803–2805.
- Si, F., D. Li, S. E. Cox, J. T. Sauls, O. Azizi, C. Sou, A. B. Schwartz, M. J. Ericstad, Y. Jun, X. Li, and S. Jun. 2017. Invariance of initiation mass and predictability of cell size in *Escherichia coli*. Current Biology 27:1278–1287.
- Sourjik, V., and N. S. Wingreen. 2012. Responding to chemical gradients: bacterial chemotaxis. Current Opinion in Cell Biology 24:262–268.
- St-Pierre, F., and D. Endy. 2008. Determination of cell fate selection during phage lambda infection. Proceedings of the National Academy of Sciences of the United States of America 105:20705–20710.
- Stylianidou, S., C. Brennan, S. B. Nissen, N. J. Kuwada, and P. A. Wiggins. 2016. SuperSegger : robust image segmentation, analysis and lineage tracking of bacterial cells. Molecular Microbiology 102:690–700.
- Taheri-Araghi, S., S. Bradde, J. T. Sauls, N. S. Hill, P. A. Levin, J. Paulsson, M. Vergassola, and S. Jun. 2015. Cell-size control and homeostasis in bacteria. Current Biology 25:385–391.
- Tenaillon, O., J. E. Barrick, N. Ribeck, D. E. Deatherage, J. L. Blanchard, A. Dasgupta, G. C. Wu, S. Wielgoss, S. Cruveiller, C. Médigue, D. Schneider, and R. E. Lenski. 2016. Tempo and mode of genome evolution in a 50,000-generation experiment. Nature 536:165–170.

- Tucker, J. D., C. A. Siebert, M. Escalante, P. G. Adams, J. D. Olsen, C. Otto, D. L. Stokes, and C. N. Hunter. 2010. Membrane invagination in *Rhodobacter sphaeroides* is initiated at curved regions of the cytoplasmic membrane, then forms both budded and fully detached spherical vesicles. Molecular Microbiology 76:833–847.
- Valgepea, K., K. Adamberg, A. Seiman, and R. Vilu. 2013. *Escherichia coli* achieves faster growth by increasing catalytic and translation rates of proteins. Molecular BioSystems 9:2344–2358.
- van Dijk, B., J. Meijer, T. D. Cuypers, and P. Hogeweg. 2019. Trusting the hand that feeds: microbes evolve to anticipate a serial transfer protocol as individuals or collectives. BMC Evolutionary Biology 19:201.
- Varma, A., K. C. Huang, and K. D. Young. 2008. The Min system as a general cell geometry detection mechanism: branch lengths in Y-shaped *Escherichia coli* cells affect Min oscillation patters and division dynamics. Journal of Bacteriology 190:2106–2117.
- Vasi, F., M. Travisano, and R. E. Lenski. 1994. Long-term experimental evolution in *Escherichia coli*. II. Changes in life-history traits during adaptation to a seasonal environment. American Naturalist 144:432–456.
- Wang, J. D., and P. A. Levin. 2009. Metabolism, cell growth and the bacterial cell cycle. Nature Reviews Microbiology 7:822–827.
- Wiser, M. J., N. Ribeck, and R. E. Lenski. 2013. Long-term dynamics of adaptation in asexual populations. Science 342:1364–1367.
- Yang, D. C., K. M. Blair, and N. R. Salama. 2016. Staying in Shape: the impact of cell shape on bacterial survival in diverse environments. Microbiology and Molecular Biology Reviews 80:187–203.
- Young, K. D. 2006. The selective value of bacterial shape. Microbiology and Molecular Biology Reviews 70:660–670.

CHAPTER 3: GENOMIC AND PHENOTYPIC EVOLUTION OF *ESCHERICHIA COLI* IN A NOVEL CITRATE-ONLY RESOURCE ENVIRONMENT

Authors: Zachary D. Blount, Rohan Maddamsetti, Nkrumah A. Grant, Sumaya T. Ahmed, Tanush Jagdish, Jessica A. Baxter, Brooke A. Sommerfeld, Alice Tillman, Jeremy Moore, Joan L. Slonczewski, Jeffrey E. Barrick, Richard E. Lenski

Originally published in eLife 2020;9e55414

Abstract

Evolutionary innovations allow populations to colonize new ecological niches. We previously reported that aerobic growth on citrate (Cit⁺) evolved in an *Escherichia coli* population during adaptation to a minimal glucose medium containing citrate (DM25). Cit⁺ variants can also grow in citrate-only medium (DM0), a novel environment for *E. coli*. To study adaptation to this niche, we founded two sets of Cit⁺ populations and evolved them for 2500 generations in DM0 or DM25. The evolved lineages acquired numerous parallel mutations, many mediated by transposable elements. Several also evolved amplifications of regions containing the *maeA* gene. Unexpectedly, some evolved populations and clones show apparent declines in fitness. We also found evidence of substantial cell death in Cit⁺ clones. Our results thus demonstrate rapid trait refinement and adaptation to the new citrate niche, while also suggesting a recalcitrant mismatch between *E. coli* physiology and growth on citrate.

Introduction

Evolutionary novelties are qualitatively new traits that allow populations to invade previously inaccessible ecological niches (Simpson 1953; Mayr 1960). Novel traits are thus important drivers of speciation and adaptive radiations that promote biodiversity and ecological complexity. Indeed, many major transitions in evolution have been mediated by novel traits such as photosynthesis, multicellularity, endoskeletons, sociality, and cognition (Maynard Smith and Szathmary 1997, Lundgren et al. 2016, Erwin 2017; Erwin 2019).

We previously proposed a model in which novel traits can evolve in three distinct phases (Blount et al. 2012). In the potentiation phase, mutations accumulate in a lineage that make it possible to evolve the trait. In the actualization phase, a specific mutation produces the trait. Newly evolved traits are typically weak and ineffective. However, if the new trait confers even a slight advantage, it may spread throughout a population and, in the refinement phase, be improved by natural selection acting on subsequent mutations.

While potentiation and actualization enable the emergence of a novel trait, the capacity for refinement affects the trait's long-term persistence and potential to influence subsequent evolution (Quandt et al. 2015; Erwin 2015). Prospects for refinement depend on the capacity to generate heritable phenotypic variation that can improve the trait and integrate it with other aspects of organismal performance (Kirschner and Gerhart 1998; Pigliucci 2008), a facet of evolvability that we call 'refinement potential'.

Refinement potential is likely crucial for a population's long-term success in a new niche. A novel trait might allow a lineage to discover a new niche, but it does not guarantee long-term persistence. The new conditions may expose the population to selection pressures that differ in important respects from those of its ancestral niche, resulting in a

mismatch between the organism and its environment (Yeh 2004; Schluter and Conte 2009; Hu et al. 2017). Successful establishment can depend on ameliorating this mismatch (Chang et al. 2011; Turkarslan et al. 2011), and failure to further adapt may lead to invasion failures (Zenni and Nuñez 2013). Adaptation to a new niche therefore reflects a tension between evolvability and robustness (Lenski et al. 2006). The benefits of refining a novel trait must outweigh the costs (if any) of integrating that trait into organismal physiology.

Adaptation to novel niches has been widely studied in the context of invasive species that colonize and adapt to unfamiliar environments (Davis 2009; MacDougall et al. 2009; Logan et al. 2019). However, the ongoing refinement of traits that provide access to novel niches has received little attention, probably because most evolutionary novelties (and associated niche discoveries) occurred in the distant past and are therefore difficult to study. Experimental evolution allows researchers to overcome this challenge. It is possible to study evolutionary novelties that arise during experiments with microbial (Blount et al. 2008; Ratcliff et al. 2012; Barrick and Lenski 2013; Kassen 2019) and digital (Lenski et al. 2003) systems, in which evolution can be studied in real-time.

One such system is the Long-Term Evolution Experiment with *Escherichia coli* (LTEE), in which 12 bacterial populations founded from a common ancestral strain have been propagated for >70,000 generations in a glucose-limited minimal medium, DM25 (Lenski et al. 1991). DM25 also contains abundant citrate, which serves as an iron-chelating agent (Blount 2016). Many bacteria can grow aerobically on citrate, but most *E. coli* strains cannot because they are unable to transport citrate into the cell (Koser 1924; Hall 1982; Reynolds and Silver 1983; Pos et al. 1998).

Citrate was unexploited as a carbon and energy source in all of the LTEE populations until a Cit⁺ variant evolved in the population designated Ara-3 after ~31,000 generations (Blount et al. 2008). The Cit+ trait arose in one of three coexisting lineages in this population by a genetic duplication that activated a previously unexpressed di- and tricarboxylate transporter (Blount et al. 2012). The benefit of this duplication mutation was contingent, at least in part, on that lineage's prior evolution of an enhanced ability to use acetate excreted into the medium as a byproduct of glucose metabolism. That enhanced ability resulted from a mutation in citrate synthase that altered carbon flow into the tricarboxylic acid cycle in a manner that was pre-adaptive for growth on citrate (Quandt et al. 2015). Concurrently, the supply of competing beneficial mutations of large effect declined over time in the LTEE, allowing the Cit⁺ lineage to escape competitive exclusion (Leon et al. 2018). The Cit+ trait radically altered this population's ecology and subsequent evolution (Blount et al. 2008; Blount et al. 2012; Quandt et al. 2015; Quandt et al. 2014; Turner 2015; Turner et al. 2015). Access to the large citrate pool in the medium led to a several-fold increase in population size (Blount et al. 2008). Nonetheless, a Cit-lineage stably coexisted with the new Cit⁺ lineage for some 10,000 generations, before finally going extinct (Blount et al. 2008; Blount et al. 2012; Turner 2015; Turner et al. 2015). Even after 70,000 generations, none of the other 11 populations in the LTEE have evolved the ability to use the available citrate (Blount et al. 2018).

The emergence of Cit⁺ in the LTEE provides a powerful model system for studying the process of evolutionary innovation. Cit⁺ variants can grow not only in DM25, which contains both glucose and citrate, but also in DM0, a citrate-only medium in which *E. coli* normally cannot grow. How would the Cit⁺ trait be refined if these variants colonized and

adapted to this newly accessible citrate-only environment? To address this question, we founded 12 new, initially clonal Cit⁺ populations and allowed them to evolve in DM0 for 2500 generations. We also allowed a second set of 12 populations to evolve in the original DM25 medium for comparison. We sequenced the genomes of evolved clones sampled from all 24 populations to find parallel genetic changes that indicate likely targets of selection (Tenaillon et al. 2016; Deatherage et al. 2017). Among other parallel changes, we identified numerous IS element insertions and several large gene amplifications. Our results thus show that genomic structural variation involving transposable elements and amplifications can provide a rich source of plasticity and potential for novel trait refinement and adaptation to new niches.

We also compared the growth of the DM0-evolved clones to that of their ancestors in both DM0 and DM25, and we examined fitness changes at the level of both whole populations and individual clones. Although all populations show substantial adaptation reflected in their growth parameters, we also found evidence of persistent maladaptation, suggesting that this new function poses metabolic challenges that are difficult to overcome evolutionarily. Some individual evolved clones grow more poorly than their ancestors, even in the medium in which they had evolved. The fitness assays show atypically large variation across replicate assays of evolved populations and clones, as well as some paradoxical apparent declines in fitness despite 2500 generations of evolution. We also observed high levels of cell death in the ancestral and evolved Cit⁺ clones that we examined. This experimental system thus sheds light not only on how new traits are refined during adaptation to a novel niche, but also on how maladaptive phenotypes may persist for long periods in new environments.

Materials and Methods

Evolution experiment

We previously isolated three random Cit⁺ clones, designated as CZB151, CZB152, and CZB154, from the 33,000-generation sample of LTEE population Ara–3 (Blount et al. 2008). We also isolated spontaneous Ara+ revertants for each clone, designated as ZDB67, ZDB68, and ZDB69, respectively. For long-term preservation, we inoculated Luria Bertani (LB) broth with isolated colonies of each clone and its revertant, grew them overnight at 37°C with orbital shaking at 120 rpm, and froze samples of each at -80° C with glycerol as cryoprotectant. We revived the clones and revertants from the frozen stocks and grew them in LB overnight. We then diluted the LB cultures 10,000-fold into 9.9 mL of Davis Mingioli (DM) minimal medium supplemented with 25 mg/L glucose (DM25), and grew them at 37°C with orbital shaking at 120 rpm. After 24 hr, we diluted these cultures 100fold in 9.9 mL of fresh DM25 and grew them for another 24 hr. This preconditioning acclimated the bacteria to growing on citrate. The preconditioned cultures were then diluted 100-fold into 9.9 mL of base DM medium (DM0), which lacks any glucose but contains 1 g/L (1,700 mM) of citrate for carbon and energy. We started two replicate populations from each LTEE-derived clone and each revertant, for a total of 12 DM0 populations. At the same time, we inoculated 12 populations into DM25 (Figure 3.1). We maintained these DM25 populations at 37°C with orbital shaking, and transferred them by 100-fold dilution into fresh DM25 every 24 hr (i.e., the same conditions as in the LTEE) for 375 transfers and 2500 generations in total. The founding Cit⁺ clones grow poorly in the citrate-only resource environment. They were unable to reach stationary phase or, in some cases, exponential phase within 24 or even 48 hr. We therefore incubated the DM0

populations for 72 hr after their initial inoculation so they could reach stationary phase before transfer to fresh medium. We then diluted them 100-fold into 9.9 mL of DM0 every 48 hr for seven cycles (two weeks), and then subsequently every 24 hr for a total of 375 transfers and 2500 generations. Every 37 days (~250 generations) samples of each population were frozen with glycerol at -80° C.

Isolation of evolved clones

We revived each evolved population sample by inoculating 100 μ L of the stock frozen at generation 2500 into 9.9 mL of LB broth and incubating overnight at 37°C with orbital shaking. We then diluted the revived DM0- and DM25-evolved populations 10,000-fold in 9.9 mL of DM0 or DM25, respectively, and grew them for 24 hr at 37°C with orbital shaking, followed by 100-fold dilution into fresh DM0 or DM25 and another 24 hr period of growth at 37°C with orbital shaking. We then diluted each population 100,000-fold in 0.85% saline and spread 100 μ L on an LB agar plate marked with three dots on the bottom. We streaked the colony closest to each dot on an LB plate after 48 hr of incubation at 37°C, thereby providing three randomly chosen clones from each population. We then inoculated an isolated colony of each clone into LB broth, grew it overnight, and froze it as before.

Fitness assays

We measured fitness by performing competition experiments modified from those described by Lenski et al. 1991. We revived samples by inoculating 15 μ L (for clones) or 100 μ L (for whole populations) from a slightly thawed frozen stock into 10 mL of LB. These cultures then grew overnight at 37°C with 120 rpm orbital shaking, after which we diluted

each 10,000-fold into either DM25 or DM0 and preconditioned as described above. We inoculated 50 μ L of each competitor's preconditioned culture into 9.9 mL of the corresponding medium, vortexed to mix, and then we spread 100 μ L of 10–2 and 10–3 dilutions on Tetrazolium Arabinose (TA) indicator agar plates to estimate the competitors' initial densities. We estimated their densities again at the end of the assay by spreading 100 µL of 10–4 and 10–5 dilutions on TA plates. For whole populations, we assayed fitness with 3-fold replication in one-day competitions, in which final densities were estimated after 24 hr. For the evolved clones, we assayed fitness with 5-fold replication, and measured final densities after 3 days, with 100-fold serial transfers to fresh medium after 24 and 48 hr. The realized growth rates of the two competitors were determined from their starting and ending densities, accounting for the dilutions. We calculated the fitness of an evolved clone or population as its realized growth rate divided by that of the ancestral competitor. In the population fitness assays, ZDB67 was the common competitor for all Ara⁻ population samples, and CZB151 was the common competitor for all Ara⁺ population samples.

Growth curves

We chose one of the three evolved clones from generation 2500 from each DM0 or DM25 population, then revived and preconditioned it in DM0 or DM25 as described above. We diluted the cultures 100-fold into 9.9 mL of DM0 or DM25, vortexed, and dispensed six 200 μ L aliquots of each culture into wells in a 96-well plate. We randomized well assignments for the cultures to minimize position effects. We measured optical density (OD) at 420 nm

wavelength every 10 min for 48 hr using a Molecular Devices SpectraMax 384 automated plate reader. We discarded the measurements taken before 30 min from our analysis.

Microscopy and cell viability analyses

We performed microscopy and viability analyses on cells derived from five clones: the LTEE ancestor (REL606); one of the three Cit⁺ ancestors in our evolution experiment (CZB151); two of its descendants that evolved in DM0 and DM25 for 2500 generations (ZDBp871 and ZDBp910, respectively); and a Cit⁺ clone isolated at generation 50,000 of the LTEE (REL11364). We revived clones from the frozen stocks and preconditioned them as described above, except that the preconditioning steps in DM0 or DM25 were extended to four daily passages to ensure acclimation to these environments. We performed preparations for live/dead cell staining and microscopic analyses on the fifth day. In these preparations, we concentrated the cells in each culture by centrifugation at 7,745 g for 8 min and decanted the supernatant. We then resuspended the cell pellets in Corning tubes containing 10 mL of 0.85% saline, and incubated them at room temperature for 1 hr; we inverted the tubes every 15 min. We then centrifuged these cultures for an additional 8 min, decanted the supernatant, and resuspended the cell pellets in 0.85% saline. We adjusted the volume of saline based on variation in turbidity to ensure that we had sufficient cells in a typical field of view for microscopy. We examined 14–55 fields per replicate for each combination of strain and media treatment. Total cell counts ranged from approximately 15,000 to 60,000 for the various combinations of clones and culture media.

We used the LIVE/DEAD BacLight Viability Kit for microscopy (ThermoFisher #L7007), following the manufacturer's directions for fluorescently labeling cells. In short,

we mixed components A and B in equal amounts, added 1 µl to each culture containing resuspended cells, and incubated them for 20 min in the dark to prevent photobleaching. After labeling, we fixed 3 µL of each sample onto a 1% agarose pad and performed fluorescent microscopy using a Nikon Eclipse Ti inverted microscope. Phase-contrast images were taken using diascopic illumination with an exposure time of 100 ms. Fluorescence was measured with an exposure time of 200 ms at 25% power of the fluorescent light source using two filter sets, 49003-ET-EYFP and 49008-ET-mCherry Texas Red (Chroma), which correspond to the fluorescence spectra of 'live' and 'dead' cells, respectively. All images were taken at 100× magnification.

We analyzed micrographs using SuperSegger, an image-processing package (Stylianidou et al. 2016). We first filtered the data, keeping only those values for segmented regions in the micrograph that were scored by the neural-network classifier as having P(Cell = True) > 75%. (Region scores range between –50 and 50, so we used data only from regions with values between 25 and 50). We then used the fluorescence values from the SuperSegger output and scored individual cells as 'live' or 'dead' depending on whether the fluorescence signal on the green (YFP) channel was greater or lesser, respectively, than the signal on the red (RFP) channel. We calculated the proportion of dead cells across the many fields examined for each of the five replicate cultures that we analyzed for each combination of clone and growth medium, and we used these values in the statistical analyses.

Genomic analysis and copy-number variation

We thawed the 3 Cit⁺ founder strains (CZB151, CZB152, CZB154), their respective Araderivatives (ZDB67, ZDB68, ZDB69), and 25 evolved clones (one Cit⁺ clone from each DM0 and DM25 evolved population, plus the anomalous Cit⁻clone ZDBp874) and grew them overnight in LB broth. We isolated genomic DNA from each sample using the Qiagen Genomic-tip 100/G DNA extraction kit.

For genomes sequenced at UT Austin, we purified DNA from *E. coli* cultures using the PureLink Genomic DNA Mini Kit (Invitrogen). For each sample, we fragmented 1 µg of purified DNA using dsDNA Fragmentase (New England Biolabs). We then used the KAPA Low Throughput Library Preparation kit (Roche) to construct Illumina sequencing libraries according to the manufacturer's instructions with two exceptions. First, we reduced reaction volumes by half. Second, we designed DNA adapters that incorporate additional 6base sample-specific barcodes such that the barcodes are sequenced as the first bases of both read 1 and read 2. We performed paired-end sequencing with 300-base reads on an Illumina MiSeq at the University of Texas at Austin Genome Sequencing and Analysis Facility. Reads were demultiplexed using a custom python script. We trimmed barcodes and adapter sequences using Trimmomatic version 0.38 (Bolger et al. 2014).

When available, we combined short-read data from different platforms before mutation identification. We identified mutations using breseq version 0.33.2 (Deatherage and Barrick 2014). We used a bash script called 'generate-LCA.sh' to infer the last common ancestor (LCA) of all evolved strains by taking the intersection of mutations found in previously curated genomes for CZB152 and CZB154; those curated founder genomes (and others) are available at: https://github.com/barricklab/LTEE-Ecoli (Barrick, 2015). We

further analyzed the mutations called by breseq relative to the LCA using custom python and R scripts available and described at: https://github.com/rohanmaddamsetti/DM0evolution (Maddamsetti 2019).

We used the following algorithm to find copy-number variation in the genomes. The breseq pipeline models 1× copy number using a negative binomial distribution fit to coverage, truncating high and low coverage that might be caused by amplifications and deletions, respectively. We then identified all positions in the genome that rejected that negative binomial at an uncorrected p = 0.05. Finally, we calculated a Bonferroni-corrected p-value for contiguous stretches of the genome in which the 1× null model was rejected at each site. We examined coverage at sites separated by the maximum read length to ensure they were not spanned by a single read. For example, in the case of a region of elevated coverage that was 1000 bp in length, covered by 150-base Illumina sequencing reads, the value of P(coverage = min)6 would be calculated, where min is the minimum coverage in that region, P(coverage = min) is the probability of that minimum coverage under the negative binomial null model, and six represents the (integer) number of sites that are 150 bp apart in the 1000 bp stretch. The output was then filtered for regions longer than $2 \times$ 150 = 300 bp to remove potential false positives. The Bonferroni calculation included corrections for checking every site in the genome in addition to the number of sites that passed the initial 0.05 cutoff for deviations from the negative binomial expectation.

Statistical test for selection on parallel IS150 insertions

To test for positive selection on parallel IS150 insertions, we simulated a null model of insertion-site preferences based on the observed data. We conservatively assumed that

IS150 elements can only insert into the positions where we observed insertions in one or more sequenced genomes from either this experiment or the LTEE (Tenaillon et al. 2016). We also assumed that the probability of IS150 transposing into a given site is proportional to the observed number of IS150 insertions at that site across the sequenced genomes, as would be the case if mutational biases alone accounted for the parallel IS150 insertions. We then used the non-parametric bootstrap method (100,000 replicates) to calculate the probability that any particular site would be hit by so many IS150 elements among the DM0-evolved genomes, holding the number of IS insertions over that group fixed.

RNA-Seq and transcriptome analysis

We performed RNA-Seq on six clones: the three Cit⁺ clones from the LTEE used as ancestors in our evolution experiment (CZB151, CZB152, and CZB154) and three evolved descendants isolated after 2500 generations of adaptation to DM0 (ZDBp877, ZDBp883, and ZDBp889). We revived each clone from a frozen stock in LB as described above. For preconditioning to minimal medium, we diluted each culture 10,000-fold into DM25 with four-fold replication and allowed them to grow for 24 hr at 37°C with 120 rpm orbital shaking. We then diluted the 16 resulting cultures 100-fold in DM0 and grew them for 48 hr at 37°C with shaking, for preconditioning to the citrate-only medium. We diluted the mature cultures 100-fold again into fresh DM0, and grew them to OD600 0.2 – 0.3, corresponding to mid-log phase, at which point we extracted their RNA using the cold phenol-ethanol method (Bhagwat et al. 2003). We recovered RNA using a Qiagen RNeasy MiniKit (#74104), and removed DNA with a Qiagen RNase-free DNase set (#79254). RNA was diluted to 50 ng/mL with nuclease-free water and cDNA amplified by RT-PCR. Purified

cDNA was then sequenced by Admera Health (South Plainfield, NJ). We used kallisto version 0.44 (Bray et al. 2016) to quantify RNA transcripts and sleuth (Pimentel et al. 2017) to conduct differential-expression analysis and visualization.

Construction of maeA plasmid

We constructed a medium-copy-number plasmid based on the kanamycin resistance cassette-containing plasmid, pSB3K3, in which the *maeA* gene was placed under the control of a strong constitutive synthetic promoter and ribosome binding site, P089-R052, described by Kosuri et al. 2013. We used PCR to amplify the *maeA* gene from REL606 and the pSB3K3 plasmid. We ordered the P089-R052 promoter as an oligonucleotide. We assembled these components using circular polymerase cloning (Quan and Tian 2009) and Gibson assembly (Gibson 2011). We performed drop dialysis using Millipore membrane filters (VSWP01300) for 15 min to desalt the assembly reactions before electroporation. We isolated transformants on LB-Kanamycin plates and used PCR to find colonies that contained the P089-R052-*maeA* insert. We used Sanger-sequencing of plasmid inserts to verify that no unintended point mutations had occurred during construction. We designated the final plasmid containing the P089-R052-*maeA* insert in the pSB3K3 backbone RM4.6.2.

Competition experiments to assess fitness effects of maeA

We transformed the Cit⁺ ancestral clones CZB151 and CZB152 and their Ara+ revertants, ZDB67 and ZDB68, respectively, with the plasmid RM4.6.2. We also transformed the same

clones with the empty pSB3K3 vector. We froze stock cultures of each transformant at -80°C with glycerol as a cryoprotectant.

We competed each RM4.6.2 transformant against its cognate pSB3K3 transformant in the clone with the opposite Ara marker state. Briefly, we revived all eight transformants in LB supplemented with 50 µg/mL kanamycin and grew them overnight at 37°C with 120 rpm orbital shaking. We then diluted each overnight culture 10,000-fold in 9.9 mL DM0 and incubated for 48 hr at 37°C with orbital shaking, after which they were diluted 100-fold in fresh DM0 every 48 hr three times to acclimate cells to the citrate-only resource environment. We commenced the competition assays the next day by inoculating 9.9 mL DM0 with 50 µL each of an RM4.6.2 transformant and the oppositely marked pSB3K3 transformant, with fourfold replication for a total of 16 competitions. We ran three-day competitions to estimate fitness as described above.

Results

Experimental design and phylogenetic analysis of sequenced strains

We isolated three Cit⁺ clones (CZB151, CZB152, and CZB154) from the 33,000-generation sample of the Ara–3 population, and derived spontaneous Ara⁺ revertants of each clone (ZDB67, ZDB68, and ZDB69, respectively). We used each of the six clones to found two populations that evolved in the citrate-only medium (DM0) for 2500 generations and two populations that evolved for 2500 generations in the medium containing both glucose and citrate (DM25) as a control (Figure 3.1). Evolved clones were isolated from each of the 24 populations at the end of the experiment, and we sequenced their genomes along with

those of the six founding clones. We used these data to identify mutations that had accumulated during the evolution experiment.

We also used the genomic data to verify the presumed phylogenetic relationships among the ancestral (including the Ara⁺ revertants) and evolved clones, in the context of the Cit⁺ lineage of the Ara–3 population. This analysis showed that CZB154 is one mutation off the line of descent for the post-33,000 generation Cit⁺ lineage in the Ara–3 population, as it subsequently evolved in the LTEE (Blount et al. 2012). That mutation is a 1 bp deletion (GGGGGG \rightarrow GGGGG) in the promoter of the hypothetical protein-coding gene ECB 03525. CZB151 does not have that mutation, but it possesses all of the other mutations found in the CZB154 clone, as well as two additional mutations. One is a $C \rightarrow G$ transversion that causes a nonsynonymous E181K mutation in the *insD* transposase. The other is a deletion of a CGCGG repeat that restores both the reading frame and function to the pseudogene *dcuS* (Turner, 2015). The restored gene encodes a histidine kinase that regulates anaerobic fumarate respiration (The UniProt Consortium, 2017; Jeske et al. 2019). CZB152, by contrast, belongs to a lineage somewhat farther from the eventual line of Cit⁺ descent in the Ara–3 population, and it differs from CZB151 and CZB154 by several mutations (Blount et al. 2012). Genomic analysis also showed that the Ara+ revertant ZDB67 differs from its parent clone, CZB151, only in the expected restoration-of-function mutation in the araA gene. The Ara+ revertants ZDB68 and ZDB69 have secondary mutations relative to CZB152 and CZB154, respectively, in addition to the expected mutation in araA. ZDB68 has a $C \rightarrow G$ transversion that introduces a nonsynonymous T33I mutation in yfcC, which encodes a predicted inner-membrane protein; and ZDB69 has a 1 bp deletion in nplI, which encodes a hypothetical protein of unknown function. All 24 evolved clones evolved in DM0 or DM25

have the mutations that are unique to their ancestors. Therefore, no cross-contamination that would compromise the independence of the evolved lines took place during the experiment.

One evolved clone from population DM0–2, ZDBp874, lacks the *citT* amplification that confers the Cit⁺ trait (Blount et al. 2008; Blount et al. 2012). That clone also displays a negative reaction on Christensen's Citrate agar, confirming a Cit-phenotype. We therefore sequenced the genome of a second isolate, ZDBp875, from the same population. We verified that ZDBp875 has the Cit⁺ trait. The ZDBp874 and ZDBp875 genomes share only a single derived mutation, an IS150 insertion at the -35 position of the promoter of *yhiO*, which encodes universal stress protein B (UspB). The two clones thus appear to belong to coexisting lineages that diverged early during their evolution in DM0. We did not discover any additional Cit-variants in the DM0–2 population during a phenotypic screen of several hundred clones. Previous work has shown that the *citT* amplifications are prone to spontaneous collapse back to a single copy (Blount et al. 2012). This collapse, which presumably occurs by homologous recombination, eliminates CitT expression and causes reversion to the Cit⁻ phenotype. The ZDBp874 clone could be either a recent and fortuitously sampled 'amplification collapse' mutant or a representative of a rare and stably coexisting Cit-lineage in that population.

Genome evolution is faster in the citrate-only environment than in the control environment The populations that evolved in the citrate-only DM0 medium accumulated more mutations than those that evolved in DM25, which contains both glucose and citrate (Figure 3.2A vs. Figure 3.2B). The DM0-evolved genomes had an average of 19.5 mutations, whereas the DM25-evolved genomes had an average of 13.7 mutations (Mann-Whitney two-tailed test, *p* = 0.0116). The DM0 genomes had an average of 3.1 nonsynonymous SNPs in protein-coding genes, as compared to 1.1 on average for the DM25 genomes. The DM0 genomes also had more IS insertions on average than the DM25 genomes (10.3 vs. 6.6), driven largely by IS150 insertions (8.5 vs. 4.9). The evolved genomes from the DM0 and DM25 treatments had similarly low average numbers of synonymous mutations (0.1 vs. 0.3), deletions (3.8 vs. 3.8), non-IS insertions (0.8 vs. 0.8), consecutive bp-substitutions (0.1 vs. 0.1), and SNPs outside of coding regions (1.4 vs. 1.1). The nearly identical numbers of synonymous mutations and SNPs outside protein-coding genes that we see in genomes evolved in DM0 and DM25 imply that the disparities in nonsynonymous mutations and IS insertions between the two conditions were, at least in part, driven by stronger selection in DM0, as opposed to a higher mutation rate or differences in population dynamics caused by the more stressful DM0 environment (Frenoy and Bonhoeffer 2018).

In both resource environments, the spectrum of mutations identified in evolved clones was dominated by structural variation, including insertions, deletions, and mobile element transpositions (Figure 3.2A, B). This spectrum is quite different from that observed in the LTEE populations. Figure 3.2C shows the number and spectrum of mutations in clones isolated from 10 LTEE populations after 5000 generations (two other populations had evolved point-mutation hypermutability and are not shown). Despite having evolved for twice as many generations, the LTEE clones have roughly similar numbers of mutations as observed in our experiments. The mutational spectrum was dominated by nonsynonymous point mutations in all but one of the LTEE populations, Ara+1. The mutation spectrum in our study is similar to that particular population, which

evolved an elevated rate of IS150 transposition early in its history (Papadopoulos et al. 1999; Tenaillon et al. 2016). It is also similar to that of a sub-lineage within another LTEE population, Ara–5, which also evolved IS150-mediated hypermutability, but much later in that experiment (Tenaillon et al. 2016). Most clones from the DM0 and DM25 treatments also have more deletions than the LTEE clones, again despite having evolved for fewer generations. These differences suggest some genomic instability in our study populations, in addition to the high rates of IS150 transposition.

Fitness changes after 2500 generations in DM0 and DM25 environments

We conducted competition assays with evolved population samples to measure their fitness in DM0 and DM25 (Lenski et al. 1991). We had difficulty in obtaining neutral derivatives of ancestral clones with the opposite Ara marker state, possibly due to genomic instability. We therefore used CZB151 as a common competitor for the Ara⁺ populations and ZDB67 for the Ara⁻ populations. Regardless of the environment in which they evolved, the populations display high variance across replicates in DM0 (Figure 3.3A), and most exhibit high variance across replicates in DM25 (Figure 3.3B). Eight of the 12 DM0-evolved populations have average fitness values in DM0 higher than their respective ancestral controls, but owing to the high variances, only two cases (DM0–6, DM0+6) appear compelling. Even in DM25, where the variances are less extreme, only two DM25-evolved populations (DM25–5, DM25–6) appear a bit more fit than their ancestors, while some DM0-evolved populations (DM0–1, DM0–2, DM0–5) were clearly less fit in DM25.

We also examined fitness changes in evolved clones relative to their direct ancestors. We only tested clones from populations for which we were able to obtain neutral

ancestral variants with the opposite Ara marker state. We saw much lower variability across the clonal replicates than we did for the whole population samples. This reduced variance may reflect in part the higher replication and longer duration of the assays using clones; it might also be the case that the within-population genetic variation led to greater variation in the outcome of the whole-population competition assays. Nonetheless, we still saw inconsistent and paradoxical fitness changes in some clones. Two DM0-evolved clones (ZDBp880 and ZDBp886) were substantially less fit than their ancestors in DM0. All DM25evolved clones except ZDBp913 were also less fit in DM0 (Figure 3.4A). In DM25, one DM0evolved clone (ZDBp880) and one DM25-evolved clone (ZDBp915) were clearly less fit than their ancestors (Figure 3.4B).

Changes in growth parameters after 2500 generations in DM0 and DM25 environments To assess changes in growth parameters at the end of the evolution experiment, we compared the growth curves of evolved populations and clones to those of their respective ancestors (Figures 3.5–3.16). To quantify changes in growth parameters more precisely, we estimated the slope of the log-transformed growth curves over two separate intervals in DM25. In this medium, the Cit⁺ bacteria undergo an apparent diauxic shift from growth on glucose to growth on citrate. Therefore, we chose intervals of optical density (OD) in which the change in OD over time would correspond to the respective growth rates on those resources. We also estimated the duration of the lag prior to initial growth on glucose. A schematic of this method is shown in Figure 3.5. We calibrated the relevant intervals based on the growth kinetics of two Cit⁻strains in DM25: the founding LTEE strain, REL606 (Figure 3.6), and the anomalous evolved clone, ZDBp874 (Figures 3.11 and 3.12). In DM0,

we estimated the duration of the lag phase and the growth rate on citrate only. On balance, the populations evolved higher exponential growth rates and shorter lag phases in both DM0 and DM25. These demographic changes are consistent with those observed in the LTEE (Vasi et al. 1994). Indeed, all DM0-evolved populations show improvements in their growth on citrate in both DM0 and DM25 (Figure 3.7), whereas their growth on glucose in DM25 shows little or no change. In DM0, the populations also exhibit markedly reduced lags prior to commencing growth (Figures 3.8 and 3.9).

We observed substantially more variation in growth parameters among the evolved clones (Figures 3.10 and 3.13) than among the whole-population samples from which they were isolated. In fact, some evolved clones grow more poorly than their ancestors, as shown by non-overlapping confidence intervals on their growth parameter estimates in Figure 3.10A. Two CZB151-derived, DM0-evolved clones, ZDBp871 and ZDBp889, show little or no improvement in DM0, and they are markedly worse than CZB151 in DM25. Similarly, the anomalous Cit-clone, ZDBp874, is not only unable to grow in DM0, but also grows much more poorly than its ancestor in DM25 (Figures 3.11 and 3.12). All other DM0-evolved clones grow better than their ancestors in DM0, and most also grow about as well as their ancestors in DM25, with the additional exception of ZDBp901.

Our finding that some evolved clones are significantly less fit than their ancestor, along with the differences in the growth parameters of the evolved clones and the whole populations, implies that ecologically relevant genetic variation exists in both the DMO- and the DM25-evolved populations. We therefore considered the possibility that improved growth performance on citrate always comes at a cost of reduced growth on glucose. We used our separate estimates of growth rates on glucose and citrate in DM25 to determine if

growth on the two substrates was correlated. However, we found no significant correlation between growth rates on citrate and glucose for either the DMO-evolved clones or whole populations (Figure 3.16A). By contrast, the growth rates measured on citrate in the two media, DMO and DM25, are highly correlated for both clones and populations (Figure 3.16B).

Evidence of cell death in clones isolated from both DM0 and DM25 environments

The contribution of cell death to fitness in the LTEE is generally negligible compared to that of growth (Vasi et al. 1994). However, we serendipitously discovered evidence of substantial cell death in cultures of a Cit⁺ clone sampled from the Ara–3 population of the LTEE at 50,000 generations. This observation led us to examine the relationship between the Cit⁺ trait and cell death in more detail by using fluorescence microscopy (Figure 3.17). We analyzed five clones: the LTEE ancestor, REL606; the 33,000-generation Cit⁺ clone, CZB151; one of its DM0-evolved descendants, ZDBp871; one of its DM25-evolved descendants, ZDBp910; and the 50,000-generation Cit⁺ clone, REL11364. We labeled cells from 24 hr stationary-phase cultures (i.e., when they would be transferred to fresh medium in the evolution experiment) using two-color live/dead stains (Materials and methods). Proportions of dead cells were calculated for five independent cultures for each clone and medium combination (except ZDBp910, for which we had problems with growth in DM0 and so have only one replicate, and REL606, which cannot grow in DM0). Figure 3.17A shows representative fields for each clone in DM0 and DM25. Figure 3.17B shows the resulting estimates of the proportion of dead cells, along with 95% bias-corrected and accelerated (BCa) bootstrap confidence intervals (DiCiccio and Efron 1996) weighted by

the number of cells analyzed and scored in the replicate cultures. On average, 10.7% of the LTEE ancestral cells grown in DM25 were scored as dead in stationary phase. By contrast, when grown in the same DM25 medium, 29.6% and 39.9% of cells were scored as dead for the Cit⁺ clones isolated from LTEE population Ara–3 at 33,000 (CZB151) and 50,000 generations (REL11364), respectively. We observed similarly high proportions of dead cells for both clones in DM0 as well (33.1% and 44.2% for CZB151 and REL11364, respectively). These results indicate that the evolution of aerobic growth on citrate in the LTEE was associated with elevated mortality. Moreover, the increased mortality was not remedied even after almost 20,000 generations since the new trait arose in the Ara-3 population. The two evolved clones we examined from our evolution experiment, ZDBp871 and ZDBp910, show somewhat different patterns. Both show lower mortality in glucosecontaining DM25 (25.3% and 12.4% for ZDBp871 and ZDBp910, respectively) but higher mortality in citrate-only DM0 (53.0% and 51.6% for ZDBp871 and ZDBp910, respectively). The reduced mortality of ZDBp910 in DM25, in which it evolved for an additional 2500 generations, suggests that the apparent metabolic imbalance associated with growth on citrate may be reduced by evolving in a medium that also contains glucose. It is even more surprising, then, that we observed no comparable reduction in mortality in the 50,000generation Ara-3 clone, which might indicate that historically contingent ecological and genetic interactions are important for this trait. Moreover, the very high death rate of ZDBp871 in DM0, the medium in which it evolved, suggests that correcting the metabolic imbalance is even more difficult when citrate is the sole carbon and energy source.

Specificity of genome evolution in the DM0 and DM25 environments

We found evidence that the DM0 and DM25 environments selected for mutations in different genes. Following Deatherage et al. 2017, we compared the distribution of 'qualifying' mutations—nonsynonymous SNPs, deletions, duplications, and IS insertions that unambiguously affect single genes—that arose during evolution in each medium. We identified all genes in which we found at least two qualifying mutations across the 24 evolved Cit⁺ clones we sequenced. These genes are shown in Figure 3.19, where they are ranked by the absolute value of the difference in the number of qualifying mutations between the DM0 and DM25 conditions.

We then used the method of Deatherage et al. 2017 to quantify the extent of parallelism in genome evolution within and between the DM0 and DM25 treatments. We computed Dice's Coefficient of Similarity, S, for each pair of evolved clones, where $S=2||X\cap Y||/(||X||+||Y||)$. ||X|| and ||Y|| are the cardinalities of the sets of genes with qualifying mutations in two clones, and $||X\cap Y||$ is the cardinality of the set of genes with mutations in both clones. S thus ranges from 0, when the two clones have no qualifying mutations in common, to 1, when both clones have qualifying mutations in exactly the same set of genes. The grand mean similarity, S_m , is 0.135 across the 24 evolved clones. The mean within-treatment similarity, S_w , is 0.177, meaning that two clones that evolved independently in the same medium on average have 17.7% of mutated genes in common. By contrast, the mean between-treatment similarity, Sb, is 0.096, meaning that two clones that evolved in different media on average have only 9.6% of mutated genes in common. We evaluated the significance of the difference between S_w and S_b using a randomization test in which clones were permuted across samples 10,000 times, and the difference

between the two measures was calculated for each permutation. The observed difference between the DMO- and DM25-evolved clones was higher than in any of the permutations. The greater genomic parallelism within than between environments is therefore highly significant (p < 0.0001).

Five genes had significantly more parallel mutations in one environment than in the other. Eleven of the 12 DM0-evolved Cit⁺ clones had qualifying mutations associated with *yhiO*, encoding the universal stress protein UspB, compared to 4 of 12 clones that evolved in DM25 (Fisher's exact test: p = 0.0094). Similarly, we found qualifying mutations in *gltA*, which encodes citrate synthase, in 11 of the DMO-evolved Cit+ clones, whereas only 3 of the DM25-evolved clones had mutations in that gene (Fisher's exact test: p = 0.0028). The gene encoding isocitrate lyase, *aceB*, had only one qualifying mutation among the DM0-evolved genomes, but nine in the DM25-evolved genomes (Fisher's exact test: p = 0.0028). Among the DM0-evolved genomes, there are no qualifying mutations in *menC*, which encodes 0succinylbenzoate synthase, but 5 DM25-evolved genomes have mutations in that gene (Fisher's exact test: p = 0.0373). Six DM0-evolved genomes have qualifying mutations in the *fadL* gene, which encodes an outer membrane long-chain fatty acid channel, but none of the DM25-evolved genomes have mutations in this gene (Fisher's exact test: p = 0.0137). Moreover, we found nine additional qualifying mutations associated with four other genes (*fadA*, *fadE*, *fadD*, and *fadR*) in the fatty-acid degradation regulon among the DM0-evolved genomes, but none in the DM25-evolved genomes. Mutations in the fad regulon thus show a strong signature of adaptation specific to the DM0 medium. Thirteen of the 15 qualifying mutations in the fad regulon were mobile-element insertions.

The environment was much more important than the ancestral genotype in determining the genetic targets of selection. We found no difference in the total number of qualifying mutations between the 24 evolved Cit⁺ clones when grouped by ancestor (i.e., CZB151, CZB152, CZB154) (Kruskal-Wallis test, p = 0.8873). Moreover, by using the same randomization test described above to test the significance of the difference between S_w and S_b, we found no significant difference based on ancestral genotype (p = 0.5540, based on 10,000 replicates).

We also found five instances of parallel changes at the amino-acid level among the DM0-evolved genomes. Three of the five occurred in *gltA*, which encodes citrate synthase: M172I, A162T, I114F. All three of these substitutions are near the allosteric binding pocket for NADH (Figure 3.18). Quandt et al. 2015 reported an A162V substitution that likewise affects NADH binding, and which was previously shown to fine-tune carbon flux through citrate synthase (Maurus et al. 2003). These three *gltA* mutations presumably have similar effects. We also saw parallel I197L substitutions in *ygaF*, which encodes a protein that dehydrogenates L-2-hydroxyglutarate to alpha-ketoglutarate and replenishes the cell's reduction potential by feeding electrons from this reaction into the membrane quinone pool (Kalliri et al. 2008). There were parallel S351C substitutions in *atoS* in ZDBp871 and the anomalous Cit-clone ZDBp874. This gene encodes the sensor protein of a two-component regulatory system that stimulates short-chain fatty acid catabolism. Unlike some mutations that might reduce or destroy a protein's functionality, we expect that these parallel amino-acid substitutions fine-tune protein function (Maddamsetti et al. 2017).
Contribution of transposable insertion elements to parallel evolution

Notwithstanding the parallel amino-acid substitutions described above, most of the parallel genomic evolution reflects the activity of IS elements. In both environments, most new IS insertions are copies of IS150 elements (Figure 3.20A and B). We compared the number of IS150 insertions in clones evolved in the two media to the number that had accumulated through 50,000 generations in the Ara–3 population of the LTEE (Figure 3.20C). The rates of IS150 insertion accumulation in the Ara–3 population and the DM25-evolved Cit⁺ populations are comparable, but much lower than in the DM0-evolved populations. The difference between the DM0- and DM25-evolved genomes is significant (Mann–Whitney U test, two-tailed p = 0.0089), despite the high variability between genomes within each group (Figure 3.2A and B).

Insertions of IS150 into new sites were strongly parallel across the independently evolved populations within, but not between, the two environments (Figure 3.20A and B). These systematic differences led us to hypothesize that the parallel IS insertions reflect the influence of selection, rather than insertion-site biases (Figure 3.19, Figure 3.20B; Tenaillon et al. 2016). We evaluated this hypothesis by conducting a randomization test for selection-driven parallel IS150 insertions over and above a null model that assumes only insertion-site preferences (Materials and methods). The most extreme observed case of parallelism at the base-pair level was an IS150 insertion at the –35 position of the promoter for *yhiO*, which encodes the universal stress protein UspB, which happened in 9 of the 12 DM0 genomes (randomization test with 100,000 bootstraps: *p* = 0.014). This test is even more conservative because it excludes two other IS150 insertions affecting this same gene in the DM0 genomes: an IS150 insertion at the –36 position of the promoter and

another IS150 insertion in *yhiO* itself. Therefore, we can reject the null hypothesis that sitespecific insertion biases alone provide an adequate explanation for the distribution of IS150 insertions. Given the conservative nature of this test, it is quite possible that some other parallel IS-insertions also indicate positive selection.

Parallel amplification mutations in the DMO- and DM25-evolved populations

We detected tandem amplifications of large genomic regions, often to high copy-number, in many DMO- and DM25-evolved clones (Tables 3.1 and 3.2, Figure 3.21). All genomes include amplifications containing the novel genetic module that evolved during the LTEE, which places one or more copies of the *citT* gene under the control of the rnk promoter region, with the exception of the anomalous Cit-clone ZDBp874 (Table 3.1). This new rnk-citT module provides access to citrate, and mutations that increase its dosage improve growth on citrate (Blount et al. 2012; Van Hofwegen et al. 2016).

Other amplifications include the *dctA* gene (Table 3.2). DctA is a proton motive force-driven, generalized di- and tricarboxylic acid transporter. During growth on citrate, the CitT antiporter protein exports TCA cycle intermediates into the medium in exchange for citrate. DctA enables recovery of those intermediates. Two mechanisms of increasing *dctA* expression have been shown to improve growth on citrate. Quandt et al. 2014 identified mutations in the *dctA* promoter that cause high-level expression. Van Hofwegen et al. 2016 showed that increased copy number of *dctA* is likewise beneficial. We found evidence that these two mechanisms are anticorrelated. Two of the ancestral clones, CZB151 and CZB154, have a shared (identical by descent) mutation in the promoter sequence of *dctA*. The third ancestor, CZB152, lacks this mutation. Only one of the 16

evolved descendants of CZB151 or CZB154 has a *dctA* amplification, whereas five of CZB152's eight descendants have such an amplification (Fisher's exact test: p = 0.0069). Also supporting this anticorrelation, one of the three CZB152 descendants without a *dctA* amplification independently evolved a mutation affecting that gene's promoter.

We identified another set of parallel amplifications in six evolved genomes. These amplifications are large and highly variable in extent, but all include at least the *fdnI*, *yddM*, *adhP*, *maeA*, *rpsV*, and *bdm* genes. These amplifications were often present in high copy numbers. Three DM25-evolved genomes have 2–13 copies, and three that evolved in DM0 have 28–59 copies (Table 3.2). In one case, ZDBp889, the amount of DNA in the amplified region constitutes more than 15% of the total evolved genome (Figure 3.21). By contrast, the amplifications of *citT* and *dctA* contain an average of 4–5 and 2–3 copies, respectively (Tables 3.1 and 3.2).

These long, high-copy-number amplifications must exert a metabolic burden, due to the costs of additional DNA synthesis and increased gene expression (da Silva and Bailey 1986; Lenski and Nguyen 1988). The repeated evolution of amplifications of this genomic region suggests that they confer some selective benefit that outweighs their cost. We examined the genes shared among the amplifications to identify which might confer this benefit. The *rpsV* gene, which encodes the 30S ribosomal subunit protein D, appears to have been a minor target for adaptation to DM0 based on parallel mutations (Figure 3.19). The *maeA* gene encodes an NAD⁺-dependent oxaloacetate-decarboxylating malate dehydrogenase (EC 1.1.1.38) that catalyzes the decarboxylation of malate to pyruvate. This plausible connection to citrate metabolism led us to hypothesize that increased *maeA*

dosage and expression provides the benefit that overcomes the cost imposed by the amplifications.

Increased MaeA expression is highly beneficial in the citrate-only environment We tested our hypothesis that increased *maeA* dosage confers a fitness benefit by transforming the ancestral strains CZB151 and CZB152 with a low-copy plasmid, RM4.6.2, which contains a copy of *maeA* that is under the control of a strong constitutive synthetic promoter and ribosome-binding site. These Ara⁻ RM4.6.2 transformants were competed in DM0 against Ara+ mutants (ZDB67 and ZDB68, respectively) of the same clones transformed with the empty-plasmid control. The RM4.6.2 transformants had a fitness advantage of ~28% in both the CZB151 (n = 6; mean fitness = 1.2790, t-distributed 95% confidence interval: [1.2636, 1.2944]) and CZB152 (n = 6; mean fitness = 1.2778, tdistributed 95% confidence interval: [1.2597, 1.2959]) backgrounds relative to their otherwise isogenic competitors. Overexpression of *maeA* is therefore highly beneficial in the DM0 environment, and its benefit likely explains the high-copy-number amplifications containing *maeA* found in many evolved clones.

We used RNA-Seq to verify that clones with *maeA*-containing amplifications have elevated transcription of that gene. We compared the transcriptomes of two ancestral clones (CZB151 and CZB152) and two DM0-evolved clones with *maeA* amplifications (ZDBp883, ZDBp889). Both evolved clones do, indeed, have much higher levels of *maeA* expression than their respective ancestors (Figure 3.22).

Despite the large fitness advantage conferred by increased *maeA* dosage in the citrate-only DM0 environment, most of the evolved Cit⁺ genomes we examined do not have

maeA amplifications. Moreover, although three of the DM25-evolved genomes in this study have large *maeA* amplifications (Table 1), none have been found in the sequenced genomes of Cit⁺ clones isolated from the Ara–3 parent population in the LTEE itself. This discrepancy might be explained by the evolution of increased *maeA* expression via other mutations. To evaluate this possibility, we also used RNA-Seq to examine the transcriptome of ZDBp877, a DM0-evolved clone without a *maeA* amplification. In contrast to ZDBp883 and ZDBp889, ZDBp877 expresses *maeA* at a level similar to that of the ancestral clones (Figure 3.22). This finding means that at least some, and perhaps all, of the evolved clones without *maeA* amplifications lack mutations that boost its expression.

Transcriptomic analysis of DM0-evolved clones

We identified other potentially adaptive differences in transcription between the DM0-evolved clones during growth in the DM0 medium (Figure 3.22). The two evolved clones with *maeA* amplifications, ZDBp883 and ZDBp889, both show increased expression of the fad fatty acid β-oxidation regulon, whereas ZDBp877, which lacks a *maeA* amplification, does not. ZDBp877 and ZDBp889, but not ZDBp883, both downregulate the cytochrome bo3 terminal oxidase complex, *cyoABCD*. We also found three genes with more extreme differential expression than *maeA* between the clones with and without the *maeA* amplification. These genes are *dinI*, *gltS*, and ECB_03510, all three of which are strongly downregulated in ZDBp877 in comparison to both clones with the amplification. DinI is a DNA-damage inducible protein that regulates the SOS response. GltS is a glutamate/sodium symporter, while ECB_03510 encodes a protein of unknown function that lies immediately downstream of *gltS*.

These differences aside, we found largely similar changes in gene expression across the three DM0-evolved clones relative to their ancestors (Figure 3.22). All three display strong downregulation of the UspB stress protein encoded by *yhiO*, presumably caused by the parallel IS150 insertions into that gene's promoter. We also found extensive downregulation of genes encoding ribosomal proteins (including *rpsB*, *rpsU*, *rpsO*, *rpsT*, *rplE*, *rplJ*, *rplN*, and *rplX*); genes involved in RNA transcription (*rpoA*, *rpoB*, *rpoC*, *rpoS*, *rho*); and DNA-replication associated genes (*gyrA*). Other down-regulated genes in the evolved clones include the nuo operon, which encodes NADH dehydrogenase in the respiratory electron transport chain, and key TCA cycle genes including those encoding the 2oxoglutarate dehydrogenase complex (*sucB*, *sucC*, and *lpdA*). By contrast, we see strong upregulation of genes encoding certain prophage-associated proteins (ECB_00826 and ECB_00827); some toxin-antitoxin pairs (*chpA-chpR*); proteins involved in recombinational DNA repair (*recA* and *recN*); SOS response proteins (*dinD*, *sulA*, *umuC*, and *umuD*); proteins associated with stationary phase (*csiE* and *sbmC*); a biofilm-associated stress protein (bhsA); and others of unknown significance.

The downregulation of transcription, translation, and NADH dehydrogenase genes and the increased expression of stress-associated genes suggest that adaptation to DMO involved reducing growth rate (relative to the faster growth on glucose), presumably to achieve balanced growth on citrate alone. The upregulation of fatty-acid β-oxidation genes (albeit less so in the ZDBp877 clone without the *maeA* amplification) also indicates some remodeling of the connection between fatty acid and citrate metabolism that is mediated by acetyl-CoA. Other changes, including the upregulation of the fad operon encoding fatty acid degradation in ZDBp883 and ZDBp889, the anaerobic glycerol-3-phosphate

dehydrogenase operon (*glpABCD*) in ZDBp877 and ZDBp889, and the glycerol-3-phosphate transporter (*glpD* and *glpT*) in ZDBp883, suggest adaptation to scavenging on dead and dying cells in the DM0 populations.

Discussion

It is rarely feasible to examine evolution in action as organisms invade, colonize, and adapt to a new niche in nature, especially with independently evolving replicates and control populations. In this study, we investigated how *E. coli* variants with the new ability to grow aerobically on citrate adapted to a novel, citrate-only resource environment in the laboratory. We examined the genomic and phenotypic evolution of 12 initially clonal populations after 2500 generations in this new environment, along with 12 initially identical control populations maintained for the same time in the ancestral environment that contains glucose as well as citrate, to better understand their post-invasion potential, including refinement of the Cit⁺ trait.

The founding clones grew poorly in their new medium, exhibiting long lag phases, slow growth, and high variation in their growth kinetics. However, those founding clones had substantial potential to adapt to citrate as their sole carbon and energy source, as the experimental populations evolved shorter lag times and faster growth rates in DM0. The evolved populations also showed correlated improvements in the ancestral glucose-citrate medium, DM25. These changes are consistent with selection pressures typical for evolution experiments that use a serial batch-culture regime like that of the LTEE (Vasi et al. 1994), upon which our experiment was based. In contrast to the LTEE, but consistent with other lines of evidence that growth on citrate is stressful for *E. coli*, the growth-curve trajectories

and even stationary-phase optical densities exhibited substantial variability for replicate assays performed using the same population sample. Assays of competitive fitness over the same 24 hr transfer cycle also showed extreme variability, especially in DM0, making it difficult to reliably estimate overall fitness gains. Increasing the duration and replication of the fitness assays should help reduce this variation in future work. Nonetheless, our difficulty in measuring adaptation in this system is striking in contrast to the ease of doing so in the LTEE (Lenski et al. 1991; Wiser et al. 2013). It is also possible that other demographic components of fitness besides shorter lag times and faster growth rates are at play in this citrate-based system. Indeed, we observed extensive cell death in Cit+ clones, and the level of mortality varied considerably even between replicate assays for reasons that we do not understand.

The evolved clones exhibit even greater variation in their growth phenotypes. While most have faster growth rates and shorter lag times, similar to the improvements observed at the population level, some evolved clones grow only slightly better, or even worse, than their ancestors. Similarly, most evolved clones show no significant increase in fitness, and some are less fit than their ancestor, even in the environment where they evolved. In some cases, the fitness estimates are discordant with measured growth characteristics. For example, one clone from a population that evolved in DM25 (ZDBp917) has a lag time far longer than its ancestor, yet it has a marginally higher fitness.

Complex ecological interactions between genotypes might explain such discordant outcomes between competitive fitness assays and growth parameters estimated in pure culture. For example, non-transitive competitive interactions can give rise to evolved clones that are more fit than their immediate predecessors, but less fit than their earlier

ancestors (Paquin and Adams 1983; Buskirk et al. 2019). We cannot exclude the possibility of non-transitive dynamics in our system at this time, but we note that they have not been observed in the LTEE on which our experiments are based (de Visser and Lenski 2002; Lenski 2017). A more likely alternative is that the paradoxical changes in fitness and growth are caused, in part, by cross-feeding and similar negative frequency-dependent interactions. The ancestral clones, for example, might be better at invading some of the evolved communities than they are at growing alone in the same medium. Similarly, some evolved clones with paradoxical growth phenotypes might be specialized ecotypes that have adapted to unknown niches, such as scavenging on dead cells or cross-feeding on metabolites produced by coexisting lineages (Turner et al. 1996; Rozen et al. 2009; Velicer and Mendes-Soares 2009; Le Gac et al. 2012; Maddamsetti et al. 2015; Good et al. 2017). We will investigate the possibility of complex ecological interactions in future work.

Genomic plasticity reflecting copy-number variation and transposable-element activity played a key role in adaptation to the citrate niche. These findings support and extend previous work showing the importance of such plasticity in adaptation to other selective challenges (Chang et al. 2013; Vandecraen et al. 2017; Press et al. 2019; Lauer and Gresham 2019), including the rapid evolution of antibiotic heteroresistance in some pathogens (Nicoloff et al. 2019). Our work especially bolsters previous demonstrations of the evolutionary importance of dynamic gene amplifications, which increase the dosage of genes encoding specific products needed at higher levels (Patrick et al. 2007; Andersson and Hughes 2009). For example, Blank et al. 2014 found that *E. coli* strains with single-gene knockouts rapidly re-evolved the capacity to grow in minimal medium in part via amplifications that increased genome size by more than 20%. Such amplifications may also

impose substantial metabolic costs, and they are prone to recombination-mediated collapse, so they are readily lost when the relevant gene products are no longer needed at higher levels. Amplifications also increase the opportunity for further mutations that may provide a benefit for a single copy, thereby favoring subsequent collapse and elimination of the cost of multiple copies (Andersson and Hughes 2009; Brennan et al. 2015; Näsvall et al. 2012). In addition to their role as substrates for promoting amplifications, IS elements seem to have both inactivated and modulated the expression of various genes in our evolution experiment. The activity of some IS elements appears to be increased by stress, which cells may experience when they invade a new niche to which they are poorly adapted (Vandecraen et al. 2017). Of course, this plasticity is a double-edged sword: the genomic instability that transposable elements cause can also produce deleterious mutations, which could impede adaptation and might even lead to extinction, especially when small founding populations invade a new niche.

Altogether, our results have several nuanced implications for evolution following the innovation-driven discovery of new niches. They suggest that genomes often possess latent potential to refine novel traits and adapt to new niches. This potential can be fulfilled not only by point mutations, but also via larger mutations such as gene amplifications and transpositions that may allow more rapid adaptation after niche discovery. Despite such adaptation, however, suboptimal traits may persist long after the new niche has been successfully invaded. In our study, the evolved Cit⁺ bacteria's physiology shows an evolutionary mismatch with their growth on the newly accessible citrate, even after thousands of generations of adaptation to that new niche. Evidence for the mismatch includes erratic growth trajectories and fitness measures that suggest extreme sensitivity

to small differences in the environment, the identity of competitors, or both. It is also seen in the high levels of mortality during stationary phase in some Cit⁺ clones, as shown using live-dead staining of cells. Further evidence comes from analyses of transcriptomic data, which shows that some Cit⁺ lines evolved increased expression of stress-associated genes during exponential phase. Thus, while a nascent ecotype's latent potential for adaptation may allow its establishment in a new niche (especially in the absence of established competitors), it may nonetheless continue to experience stress and suffer from suboptimal phenotypes for a long time before becoming truly well suited to its new conditions.

Our findings also highlight that the fact that organisms are historically contingent patchworks of traits and functions constructed by evolutionary tinkering, and so they are rife with design compromises (Pittendrigh 1958; Tinbergen 1965; Jacob 1977). Natural selection integrates these complex assemblies into idiosyncratic, but usually robust and stable, biological systems. This stability can be disrupted, however, such as when a novel trait that is beneficial, on the whole, nonetheless generates new stresses. These secondary tradeoffs are also implicit in Fisher's geometric model of adaptation, in which mutations of large effect, including those that produce new functions, are especially likely to disrupt other phenotypes (Fisher 1930; Orr 2000). We have shown that such disruptions can persist for thousands of generations, which implies that re-evolving a stable, robust system in which a novel trait is fully integrated with the preexisting physiology can be difficult.

Organisms typically maintain their physiological systems at a dynamic steady state. This homeostasis implies that organisms have evolved to maintain physiological variables within an acceptable range in the face of perturbations (Albergante et al. 2014). Failure to maintain homeostasis may result in illness and even death. Viability is sometimes possible

outside the usual range of homeostasis, but often at the cost of stress and lasting damage to the organism. Our findings imply that the disruptions caused by evolutionary novelties can maladaptively change the system parameters that maintain homeostasis, thereby causing stress and increasing mortality. Many questions remain about the nature and consequences of these homeostatic disruptions, as well as how novel traits might eventually become well integrated with an organism's existing physiology to restore its prior homeostasis.

Our experimental system has the potential to address these and other questions about innovation, adaptation, and maladaptation that are relevant to both evolutionary biology, in general, and evolutionary medicine, in particular. In humans, for example, cultural innovations, including the agricultural revolution, have vastly reshaped our diets and thereby also changed our gut microbiota (McMichael et al. 2007; David et al. 2014). Contemporary high-calorie diets and sedentary lifestyles have led to an epidemic of associated illnesses, including hypertension, diabetes, and obesity. Might new traits typically exhibit more phenotypic variation, reflecting greater sensitivity to intrinsic stochasticity, lower robustness to environmental perturbations, or both? If growth on citrate by *E. coli* required major changes in physiology and metabolism, then that innovation may have increased fragility due to new difficulties in coordinating cell growth and division (Scott et al. 2014; Schaechter 2015). By disrupting existing physiological and metabolic processes, innovations can introduce new compromises and imbalances, the resolution of which requires novel variation. That new variation may, in turn, affect correlated traits and the organism's overall robustness. We conjecture that the evolutionary refinement of traits that open new niches may often promote evolvability at the expense of robustness and overall good health (Lenski et al. 2006).

Acknowledgements

We thank Joshua Franklin and Yann Dufour for helpful discussions and assistance with the microscopy work; Simon D'Alton for assistance with genome sequencing; Daniel Barich for help in handling transcriptomics data; Neerja Hajela and Devin Lake for assistance in the laboratory; Jean Vila, Erik Quandt, Daniel Deatherage, Dacia Leon, Debora Marks, David Ding, Yarden Katz, Helen Murphy, and Kyle Card for helpful discussions; and Sandeep Venkataram, Sébastien Wielgoss, and anonymous reviewers for constructive feedback on previous versions of the manuscript.

APPENDIX



Figure 3.1: Experimental design and sequenced clone derivations. We isolated three Cit⁺ clones (red hexagons) from generation 33,000 of LTEE population Ara–3. We then derived Ara+ mutants (white hexagons) from those three LTEE clones. We used these six clones to found 24 populations. Twelve populations evolved for 2500 generations in citrate-only medium, DM0 (cyan lines). The remaining 12 evolved for 2500 generations in glucose and citrate medium, DM25 (black lines). The evolved clones we isolated after 2500 generations for genomic and phenotypic analysis are shown for each population.



Figure 3.2: Numbers and types of mutations in evolved genomes. (A) Evolved genomes from the DM0 treatment after 2500 generations. (B) Evolved genomes from the DM25 treatment after 2500 generations. (C) Evolved genomes in the 10 non-hypermutable LTEE populations after 5000 generations. Mutations are color-coded according to the key: indel, insertions and deletions (excluding large duplications and amplifications); intergenic, intergenic point mutations; mobile-element transpositions; multiple-base substitution, consecutive point mutations (including adjacent to and in conjunction with indels); nonsense, nonsynonymous, and synonymous point mutations in protein-coding genes; pseudogene, mutations in pseudogenes.



Figure 3.3: Fitness of evolved populations and their Cit⁺ ancestors relative to Cit⁺ ancestral clones CZB151 and ZDB67 in DM0 and DM25.To show the difference in scale across panels, dashed gray lines are drawn at 1.0 (neutrality) and 1.5 on the y-axis. Ancestral strain CZB151 and its descendants are shown in black, CZB152 and its descendants are in orange, and CZB154 and its descendants are in blue. (A) Fitness of evolved and ancestral populations relative to CZB151 and ZDB67 in DM0, as measured in one-day competition assays. Some confidence limits extend beyond the range shown on the y-axis. (B) One-day fitness of evolved and ancestral populations relative to CZB151 and ZDB67 in DM0, as relative to CZB151 and ZDB67 in DM25, as measured in one-day competition assays. Error bars are 95% confidence intervals.



Figure 3.4: Fitness of select evolved clones against their direct ancestors in DM0 and DM25. The dashed grey line shows neutrality. Ancestral strain CZB151 and its descendants are shown in black, CZB152 and its descendants are in orange, and CZB154 and its descendants are in blue. (A) Fitness of evolved clones relative to their direct ancestors in DM0 in a three-day competition assay. (B) Fitness of evolved clones relative to their direct ancestors in DM25 in a three-day competition assay. Error bars are 95% confidence intervals. We selected clones for fitness assays based only on the availability of ancestral genotypes with confirmed, neutral, opposing Ara marker states.



Figure 3.5: Schematic of the log-slope method to calculate growth rates. We logetransformed optical densities, and used the slope of the curve in the interval OD_{420} nm = [0.01, 0.02] to calculate the exponential growth rate on glucose (h⁻¹), r_{glucose}. We used the slope of the curve in the interval OD_{420} nm = [0.05, 0.1] to calculate the exponential growth rate on citrate (h⁻¹) r_{citrate}. In making this interpretation, we assumed a diauxic shift between growth on glucose and citrate, rather than simultaneous growth on both substrates. In any case, growth rates during these intervals are relevant phenotypes even without assuming diauxie. We estimated lag time (τ) as the time (h) until OD₄₂₀ nm = 0.01 was reached.



Figure 3.6: Growth curves for REL606 in DM25. We used these data to choose the interval for estimating the exponential growth rate on glucose. (A) Replicated growth curves in DM25. (B) The same data as in panel A except log_e-transformed. Dashed black lines indicate the interval used to calculate growth rates on glucose; the dashed red line shows the lower bound of the interval in which the growth rate on citrate would be estimated.



Figure 3.7: Growth parameters for whole-population samples that evolved in DM0 and their Cit⁺ ancestors. (A) Estimates of various growth parameters for the ancestral strains and DM0-evolved populations at 2500 generations, using the log-slope method. Ancestral strain CZB151 and its descendants are shown in black, CZB152 and its descendants are in orange, and CZB154 and its descendants are in blue. Units for growth rates *r* are h⁻¹, and units for lag times are h. Bias-corrected and accelerated (BCa) bootstrap 95% confidence intervals around parameter estimates were calculated using 10,000 bootstraps. (B) Estimates of log₂-transformed ratios of growth parameters for the evolved populations and their ancestors. The growth curves we used to estimate these parameters are shown in Figures 3.8 and 3.9.



Figure 3.8: Growth curves of the 12 DM0-evolved whole-population samples, measured in DM0 and DM25.For comparison, growth curves of the evolved populations are paired with those of their ancestors: CZB151 (top row), CZB152 (middle row), and CZB154 (bottom row). The evolved and ancestral curves are shown in purple and gray, respectively.



Figure 3.9: Log_e-transformed growth curves of the 12 DM0-evolved whole-population samples, measured in DM0 and DM25. Dashed black and red lines indicate the intervals we used to calculate growth rates on glucose and citrate, respectively.



Figure 3.10: Growth parameters for clones from populations that evolved in DM0 and their Cit⁺ ancestors. (A) Estimates of growth parameters for the ancestral strains and DM0-evolved clones sampled at 2500 generations, using the log-slope method. CZB151 and its descendants are in black, CZB152 and its descendants are in orange, and CZB154 and its descendants are in blue. (B) Estimates of log₂-transformed ratios of growth parameters for the evolved clones and their ancestors. The growth curves we used to estimate parameters are shown in Figures 3.11 and 3.12. We excluded the anomalous evolved Cit⁻clone. See Figure 3.7 for additional details.



Figure 3.11: Growth curves of the 12 DM0-evolved clones, measured in DM0 and DM25. For comparison, growth curves of the evolved clones are paired with those of their ancestors: CZB151 (top row), CZB152 (middle row), and CZB154 (bottom row). The evolved and ancestral curves are shown in purple and gray, respectively, except the anomalous Citevolved clone shown in orange.



Figure 3.12: Log_e-transformed growth curves of the 12 DM0-evolved clones, measured in DM0 and DM25. Dashed black and red lines indicate the intervals we used to calculate growth rates on glucose and citrate, respectively.



Figure 3.13: Growth parameters of the 12 DM25-evolved clones and their 3 Cit⁺ ancestors. (A) Estimates of growth parameters for each ancestral and DM25-evolved clone, using the log-slope method (Figure 3.2). Estimates for ancestral strain CZB151 and its descendants are shown in black, estimates for CZB152 and its descendants are in orange, and estimates for CZB154 and its descendants are in blue. Units for growth rates *r* are h⁻¹, and units for lag times are h. Bias-corrected and accelerated (BCa) bootstrap 95% confidence intervals around parameter estimates were calculated using 10,000 bootstraps; no confidence interval is shown if a parameter could not be estimated accurately from the available data. Aberrant estimates that fall outside of these ranges are not shown. (B) Estimates of log₂-transformed ratios of growth parameters for the evolved clones and their ancestors. The growth curves used to estimate these parameters are shown in Figures 3.14 and 3.15.



Figure 3.14: Growth curves of the 12 DM25-evolved clones, measured in DM25 only.(Many DM25-evolved clones grew inconsistently in DM0.) For comparison, growth curves of the evolved clones are paired with those of their founders: CZB151 (top row), CZB152 (middle row), and CZB154 (bottom row). The evolved and ancestral curves are shown in purple and gray, respectively.



Figure 3.15: Log_e-transformed growth curves of the 12 DM25-evolved clones, measured in DM25.Dashed black and red lines indicate the intervals we used to calculate growth rates on glucose and citrate, respectively (Figure 2). See Figure 3.14 for additional details.



Figure 3.16: Correlations between estimated growth rates across substrates and media for DM0-evolved clones and populations. All tests are two-tailed, because growth rates across substrates and media might, in principle, exhibit tradeoffs. (A) Correlations between rglucose and rcitrate in DM25 are not significant (Pearson's r = 0.4788, d.f. = 12, p = 0.0833 for clones; r = -0.0392, d.f. = 13, p = 0.8897 for populations). (B) Correlations between rcitrate in DM0 and rcitrate in DM25 are highly significant (r = 0.7513, d.f. = 12, p = 0.0020 for clones; r = 0.8041, d.f. = 13, p = 0.0003 for populations). Circles and triangles indicate ancestral and evolved samples, respectively. Colors distinguish the different Cit⁺ ancestors and their evolved descendants.



Figure 3.17: Elevated mortality in Cit⁺ strains. The Cit⁺ strains exhibit substantially elevated mortality in the citrate-only DM0 medium; some also show high mortality in DM25 as well. REL606 is Cit⁻ and cannot grow in DM0. CZB151 was isolated from LTEE population Ara-3 at generation 33,000, and its descendants, ZDBp871 and ZDBp910, had evolved for 2500 generations in DM0 and DM25 media, respectively. REL11364 was isolated from LTEE population Ara-3 at generation 50,000. (A) Representative micrographs of the five clones in the two media. We stained cells using the BacLight Viability Kit, and we scored them as dead if their red fluorescence exceeded their green fluorescence (see Materials and methods). Scale bars (lower right corner) represent 5 μ m. (B) Proportion of dead cells in five replicate cultures of each strain grown in DM0 and DM25 medium each (except for ZDBp910, with only one replicate). The wider symbols show estimated overall proportions weighted by the number of cells analyzed in each replicate culture. We calculated bias-corrected and accelerated (BCa) bootstrap 95% confidence intervals using 10,000 bootstraps (except for ZDBp910), and we weighted by the number of cells analyzed in each replicate intervals using 10,000 bootstraps (except for ZDBp910), and we weighted by the number of cells analyzed in each replicate intervals using 10,000 bootstraps (except for ZDBp910), and we weighted by the number of cells analyzed in each replicate number of cells analyzed in each replicate intervals using 10,000 bootstraps (except for ZDBp910), and we weighted by the number of cells analyzed in each replicate.



Figure 3.18: Parallel substitutions at the amino-acid level in citrate synthase, GltA. All of the evolved substitutions occur at the allosteric protein-ligand interface with NADH. GltA is shown in its dimeric, NADH-bound conformation (1NXG crystal structure in the Protein DataBank). The M172I, A162T, I114F substitutions are shown in purple. NADH is shown in orange.



Figure 3.19: Parallel genetic evolution. Genes with mutations in two or more sequenced genomes from the DMO- and DM25-evolved populations, ranked by the absolute value of the difference in the number of qualifying mutations (see main text) between DM0 and DM25. Mutations in the same genes in the six non-mutator LTEE lineages and in a Cit⁺ clone from LTEE population Ara–3 (which evolved hypermutability), all at 50,000 generations, are shown for comparison. Yellow, violet, or red fill indicates the presence of one, two, or three qualifying mutations, respectively.



Figure 3.20: Parallel IS-element insertions. (A) Counts of parallel IS-element insertions in labeled genes (including promoter and coding regions) summed across sequenced DMO- and DM25-evolved genomes, and arranged by position on the *E. coli* chromosome, relative to the inferred last common ancestor of all strains (Materials and methods). IS1 insertions are shown in pink, IS150 in lavender, IS186 in red, IS3 in black, and ISRSO11 in green. Some genes contain multiple sites with parallel IS-element insertions. (B) Location of insertions, shown separately for the DMO- and DM25-evolved genomes. Colors are the same as in panel A. (C) Total number of IS150 insertions in the DMO- and DM25-evolved genomes after 2500 generations. The corresponding numbers of IS-element insertions in clones isolated from LTEE population Ara–3 at time points over 50,000 generations of evolution are shown for comparison. DM0 clones are labeled as brown circles, DM25 clones as pink triangles, and LTEE Ara–3 clones as tan squares.



Figure 3.21: Genetic amplifications in evolved clones. Genomic regions with significant amplifications in DM0- and DM25-evolved clones, arranged by chromosomal position. The evolved clones from DM0 (top half) and DM25 (bottom half) are indicated at the near left, with the total amplified length shown at the far left. Dashed vertical lines mark the *maeA* and *dctA* loci. The boundaries vary among the subset of genomes with amplifications that encompass these genes; by contrast, the *citT* locus is amplified in all of these genomes, and with nearly uniform boundaries. Colors denote amplification copy-number on a log2 scale from dark (low copy-number) to light (high copy-number).



Figure 3.22: Transcriptomic analysis of ancestral and evolved clones. Differential expression analysis comparing two ancestral (CZB151 and CZB152) and three evolved clones (ZDBp877, ZDBp883, ZDBp889), produced by sleuth (Pimentel et al. 2017). The colored bar (at right) shows the level of RNA expression based on estimated counts and transformed as log2(1 + est_counts). The differentially expressed genes discussed in the main text are shown here. The numeric labels after the strain identifiers indicate the two or four biological replicates for each clone (i.e., RNA samples prepared from independently revived cultures of that clone.
Genome	Medium	Mean copy	Minimum copy	Maximum
		number	number*	copy number*
CZB151	DM25	4.21	3.39	5.47
CZB152	DM25	8.23	5.17	11.46
CZB154	DM25	4.14	1.72	9.83
ZDBp871	DM0	2.82	1.70	4.26
ZDBp875	DM0	11.37	8.05	14.93
ZDBp877	DM0	7.68	3.66	11.33
ZDBp880	DM0	3.82	1.79	5.93
ZDBp883	DM0	5.08	1.88	12.81
ZDBp886	DM0	4.76	2.27	6.88
ZDBp889	DM0	4.66	2.90	7.11
ZDBp892	DM0	5.69	2.21	8.76
ZDBp895	DM0	5.30	2.13	8.51
ZDBp898	DM0	6.14	2.77	9.73
ZDBp901	DM0	3.91	1.83	5.64
ZDBp904	DM0	3.84	1.78	5.68
ZDBp910	DM25	4.71	2.40	6.47
ZDBp911	DM25	3.13	1.58	5.01
ZDBp912	DM25	8.93	4.51	13.41
ZDBp913	DM25	4.83	2.68	6.94
ZDBp914	DM25	4.11	2.03	6.17
ZDBp915	DM25	3.31	1.78	5.19
ZDBp916	DM25	4.87	2.67	6.95
ZDBp917	DM25	3.20	1.77	4.63
ZDBp918	DM25	3.66	1.88	5.18
ZDBp919	DM25	2.91	2.06	3.81
ZDBp920	DM25	3.92	2.08	5.49
ZDBp921	DM25	5.76	2.93	9.15

Table 3.1: Copy number of amplified *citT* genes in sequenced clones.

*These bounds indicate the ratio of the minimum and maximum sequencing coverage measured at the *citT* locus to the mean coverage over the genome. In all cases, the estimated copy number is significantly greater than one at p<0.0001, even after Bonferroni corrections for multiple tests of the same hypothesis.

Genome	Medium	Gene	Mean copy	Minimum	Maximum	Adjusted
			number	number*	number*	p-value†
ZDBp880	DM0	dctA	2.33	1.67	3.26	< 0.0001
ZDBp886	DM0	dctA	3.20	1.88	4.25	< 0.0001
ZDBp898	DM0	dctA‡	2.09	1.71	2.76	0.0023
ZDBp913	DM25	dctA	3.17	1.91	4.68	< 0.0001
ZDBp918	DM25	dctA	3.41	1.80	5.19	< 0.0001
ZDBp919	DM25	dctA	2.60	2.06	3.29	< 0.0001
ZDBp883	DM0	maeA	58.47	22.61	95.46	< 0.0001
ZDBp889	DM0	maeA	34.71	2.35	55.39	< 0.0001
ZDBp904	DM0	maeA	28.08	15.30	44.72	< 0.0001
ZDBp911	DM25	maeA	2.22	1.54	3.39	< 0.0001
ZDBp917	DM25	maeA	4.72	3.56	6.18	< 0.0001
ZDBp919	DM25	maeA	12.81	2.16	18.00	< 0.0001

Table 3.2: Copy number of amplified *maeA* and *dctA* genes in sequenced clones from populations that evolved for 2500 generations in either DM0 or DM25 environments.

*These bounds indicate the ratio of the minimum and maximum sequencing coverage measured at the indicated locus to the mean sequencing coverage over the genome.

+Significance levels are shown after Bonferroni corrections for multiple tests of the same hypothesis.

‡There may be two discontinuous amplifications of *dctA* in this genome, or there may be a single continuous amplification with a short region of low coverage within the gene. The second region of amplification has a similar copy number. We present data for only one region, which provides a conservative estimate of the overall statistical significance in this case.

LITERATURE CITED

LITERATURE CITED

- Albergante, L. J. J. Blow, and T. J. Newman. 2014. Buffered qualitative stability explains the robustness and evolvability of transcriptional networks. eLife 3:e02863.
- Andersson, D. I., and D. Hughes. 2009. Gene amplification and adaptive evolution in bacteria. Annual Reviews Genetics 43:167–195.

Barrick J. E. 2015. GitHub. LTEE-Ecoli, version v2.0.2.

- Barrick J. E., and R. E. Lenski. 2013. Genome dynamics during experimental evolution. Nature Reviews Genetics 14:827–839.
- Bhagwat A. A., R. P. Phadke, D. Wheeler, S. Kalantre, M. Gudipat, and M. Bhagwat. 2003. Computational methods and evaluation of RNA stabilization reagents for genome-wide expression studies. Journal of Microbiology Methods 55:399–40.
- Blank D., L. Wolf, M. Ackermann, and O. K. Silander. 2014. The predictability of molecular evolution during functional innovation. Proceedings of the National Academy of Sciences of the United States of America 111:3044–2049.
- Blount Z. D., J. E. Barrick, C. J. Davidson, and R. E. Lenski. 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. Nature 489:513–518.
- 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*. Proceedings of the National Academy of Sciences of the United States of America 105:7899–7906.
- Blount Z. D., R. E. Lenski, and J. B. Losos. 2018. Contingency and determinism in evolution: replaying life's tape. Science 362:eaam5979.
- Blount Z. D. 2016. A case study in evolutionary contingency. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences 58:82–92.
- Bolger A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.
- Bray N. L., H. Pimentel, P. Melsted, and L. Pachter. 2016. Near-optimal probabilistic RNA-seq quantification. Nature Biotechnology 34:525–527.
- Brennan G., J. O. Kitzman, J. Shendure, and A. P. Geballe. 2015. Experimental evolution identifies vaccinia virus mutations in A24R and A35R that antagonize the protein kinase R pathway and accompany collapse of an extragenic gene amplification. Journal of Virology 89:9986–9997.

- Buskirk S. W., A. B. Rokes, and G. I. Lang. 2019. Adaptive evolution of a rock-paper-scissors sequence along a direct line of descent. bioRxiv. doi: <u>https://doi.org/10.1101/700302</u>.
- Chang A. L., A. M. H. Blakeslee, A. W. Miller, and G. M. Ruiz. 2011. Establishment failure in biological invasions: a case history of *Littorina littorea* in California, USA. PLoS One 6:e16035.
- Chang S-L, H-Y Lai, S-Y Tung, and J-Y Leu. 2013. Dynamic large-scale chromosomal rearrangements fuel rapid adaptation in yeast populations. PLoS Genetics 9:e1003232.
- DaSilva N. A., and J. E. Bailey. 1986. Theoretical growth yield estimates for recombinant cell. Biotechnology and Bioengineering 28:741–746.
- David L. A., C. F. Maurice, R. N. Carmody, D. B. Gootenberg, J. E. Button, B. E. Wolfe, A. V. Ling, A Devlin Sloan, Y. Varma, M. A. Fischbach, S. B. Biddinger, R. J. Dutton, and P. J. Turnbaugh. 2014. Diet rapidly and reproducibly alters the human gut microbiome. Nature 505:559– 563.

Davis M. Z. Invasion Biology. 2009. Oxford, Oxford University Press.

- de Visser J. A. G. M., and R. E. Lenski. 2002. Long-term experimental evolution in *Escherichia coli*. XI. Rejection of non-transitive interactions as cause of declining rate of adaptation. BMC Evolutionary Biology 2:19.
- Deatherage D. E., and J. E. Barrick. 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Engineering and analyzing multicellular systems: Methods in Molecular Biology 1151:165–188.
- Deatherage D. E., J. L. Kepner, A. F. Bennett, R. E. Lenski, and J. E. Barrick. 2017. Specificity of genome evolution in experimental populations of *Escherichia coli* evolved at different temperatures. Proceedings of the National Academy of Sciences of the United States of America 114:E1904–E1912.
- DiCiccio T. J., and B. Efron. 1996. Bootstrap confidence intervals. Statistical Science 11:189–228.
- Erwin D. H. 2015. Novelty and innovation in the history of life. Current Biology 25:R930–940.
- Erwin D. H. 2019. Prospects for a general theory of evolutionary novelty. Journal of Computational Biology 26:735–744.
- Erwin D. H. 2017. The topology of evolutionary novelty and innovation in macroevolution. Philosophical Transactions of the Royal Society B 372:20160422.

Fisher R. A. 1930. The Genetical Theory of Natural Selection. Oxford, Clarendon Press.

- Frenoy A., and S. Bonhoeffer. 2018. Death and population dynamics affect mutation rate estimates and evolvability under stress in bacteria. PLOS Biology 16:e2005056.
- Gibson D. G. 2011. Enzymatic assembly of overlapping DNA fragments. Methods in Enzymology. 498:349–361.
- Good B. H., M. J. McDonald, J. E. Barrick, R. E. Lenski, and M. M. Desai. 2017. The dynamics of molecular evolution over 60,000 generations. Nature 551:45–50.
- Hall B. G. 1982. Chromosomal mutation for citrate utilization by *Escherichia coli* K-12. Journal of Bacteriology 151:269–273.
- Hu Y., Q. Wu, M. Shuai, M. Tianxiao, S. Lei, W. Xiao, N. Yonggang, N. Zemin, Y. Li, X. Yungang, and W. Fuwen. 2017. Comparative genomics reveals convergent evolution between the bamboo-eating giant and red pandas. Proceedings of the National Academy of Sciences of the United States of America 114:1081–1086.
- Jacob F. 1977. Evolution and tinkering. Science 196:1161–1166.
- Jeske L., S. Placzek, I. Schomburg, A. Chang, and A. Schomburg. BRENDA in 2019: a European ELIXIR core data resource. Nucleic Acids Research 47:D542–D549.
- Kalliri E., S. B. Mulrooney, R. Hausinger. 2008. Identification of *Escherichia coli* YgaF as an L-2-ydroxyglutarate Oxidase. Journal of Bacteriology 190:3793–3798.
- Kassen R. 2019. Experimental evolution of innovation and novelty. Trends in Ecology and Evolution 34:712–722.
- Kirschner M., J. Gerhart. 1998. Evolvability. Proceedings of the National Academy of Sciences of the United States of America 95:8420–8427.
- Koser S. A. 1924. Correlation of citrate utilization by members of the colon-aerogenes group with other differential characteristics and with habitat. Journal of Bacteriology 9:59–77.
- Kosuri S., D. B. Goodman, G. Cambray, V. K. Mutalik, Y. Gao, A. P. Arkin , D. Endy, and G. M. Church. 2013. Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America 110:14024–14029.
- Lauer S., and D. Gresham. 2019. An evolving view of copy number variants. Current Genetics 65:1287–1295.
- 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 of the United States of America 109:9487–9492.

- Lenski R. E., J. E. Barrick, and C. Ofria. 2006. Balancing robustness and evolvability. PLoS Biology 4:e428.
- Lenski R. E., and T. T. Nguyen. 1988. Stability of recombinant DNA and its effects on fitness. Trends in Ecology and Evolution 3:S18-S20.
- Lenski R. E., C. Ofria, R. T. Pennock, and C. Adami. 2003. The evolutionary origin of complex features. Nature 423:139–144.
- 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. The American Naturalist 138:1315–1341.
- Lenski R. E. 2017. Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. The ISME Journal 11:2181–2194.
- Leon D., S. D'Alton, E. M. Quandt, and J. E. Barrick. 2018. Innovation in an *E. coli* evolution experiment is contingent on maintaining adaptive potential until competition subsides. PLoS Genetics 14:e1007348.
- Logan M. L., I. A. Minnaar, K. M. Keegan, and S. Clusella-Trullas. 2019. The evolutionary potential of an insect invader under climate change. Evolution 74:132–144.
- Lundgren M. R., P-A. Christin, E. G. Escobar, B. S. Ripley, G. Besnard, C. M. Long, P. W. Hattersley, R. P. Ellis, R. C. Leegood, and C. P. Osborne. Evolutionary implications of C₃-C₄ intermediates in the grass *Alloteropsis semialanta*. Plant, Cell & Environment 39:1874–1885.
- MacDougall A. S., B. Gilbert, and J. M. Levine. 2009. Plant invasions and the niche. Journal of Ecology 97:609–615.
- Maddamsetti R. 2019. GitHub. DM0-evolution, version 04321e0.
- Maddamsetti R., P. J. Hatcher, A. G. Green, B. L. Williams, D. S. Marks, and R. E. Lenski. 2017. Core genes evolve rapidly in the long-term evolution experiment with *Escherichia coli*. Genome Biology and Evolution 9:1072–1083.
- Maddamsetti R., R. E. Lenski, and J. E. Barrick. 2015. Adaptation, clonal interference, and frequency-dependent interactions in a long-term evolution experiment with *Escherichia coli*. Genetics 200:619–631.
- Maurus R., N. T. Nguyen, D. J. Stokell, A. Ayed, P. G. Hultin, H. W. Duckworth, and G. D. Brayer. 2003. Insights into the evolution of allosteric properties. The NADH binding site of hexameric type II citrate synthases. Biochemistry 42:5555–5565.

- Maynard Smith J., and E. Szathmary. 1997. The Major Transitions in Evolution. Oxford, Oxford University Press.
- Mayr E. The emergence of evolutionary novelties. 1960. In Evolution after Darwin (ed. Tax S) Vol. 1:349–380. Chicago, University of Chicago Press.
- McMichael A. J., J. W. Powles, C. D. Butler, and R. Uauy. 2007. Food, livestock production, energy, climate change, and health. The Lancet 370:1253–1263.
- Näsvall J, L. Sun, J. R. Roth, D. I. Andersson. 2012. Real-time evolution of new genes by innovation, amplification, and divergence. Science 338:384–387.
- Nicoloff H, K. Hjort, B. R. Levin, D. I. Andersson. 2019. The high prevalence of antibiotic heteroresistance in pathogenic bacteria is mainly caused by gene amplification. Nature Microbiology 4:504–514.
- Orr H. A., 2000. Adaptation and the cost of complexity. Evolution 54:13–20.
- Papadopoulos D., D. Schneider, J. Meier-Eiss, W. Arber, R. E. Lenski, and M. Blot. 1999. Genomic evolution during a 10,000-generation experiment with bacteria. Proceedings of the National Academy of Sciences of the United States of America 96:3807–3812.
- Paquin C. E., and J. Adams. 1983. Relative fitness can decrease in evolving asexual populations of *S. cerevisiae*. Nature 306:368–371.
- Patrick W. M., E. M. Quandt, D. B. Swartzlander, and I. Matsumura. 2007. Multicopy suppression underpins metabolic evolvability. Molecular Biology and Evolution 24:2716–2722.
- Pigliucci M. 2008. Is evolvability evolvable? Nature Reviews Genetics 9:75–82.
- Pimentel H., N. L. Bray, S. Puente, P. Melsted, and L. Pachter. 2017. Differential analysis of RNA-seq incorporating quantification uncertainty. Nature Methods 14:687–690.
- Pittendrigh C. S. 1958. Adaptation, natural selection, and behavior. In Behavior and Evolution (eds Roe A, Simpson GG), pp. 390 416. New Haven, Yale University Press.
- Pos K. M., P. Dimroth, and M. Bott. 1998. The *Escherichia coli* citrate carrier CitT: A member of a novel eubacterial transporter family related to the 2-oxoglutarate-malate translocator from spinach chloroplasts. Journal of Bacteriology 180:4160–4165.
- Press M. O., A. N. Hall, E. A. Morton, and C. Queitsch. 2019. Substitutions are boring: some arguments about parallel mutations and high mutation rates. Trends in Genetics 35:253–264.

- Quan J, and J. Tian. 2009. Circular polymerase extension cloning of complex gene libraries and pathways. PloS One 4:e6441.
- Quandt E. M., D. E. Deatherage, A. D. Ellington, G. Georgiou, J. E. Barrick. 2014. Recursive genomewide recombination and sequencing reveals a key refinement step in the evolution of a metabolic innovation in *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America 111:2217–2222.
- Quandt E. M., J. Gollihar, Z. D. Blount, A. D. Ellington, G. Georgiou, J. E. Barrick. 2015. Finetuning citrate synthase flux potentiates and refines metabolic innovation in the Lenski evolution experiment. eLife 4:e09696.
- Ratcliff W.C., R. F. Denison, M. Borrello, and M. Travisano. 2012. Experimental evolution of multicellularity. Proceedings of the National Academy of Sciences of the United States of America 109:1595–1600.
- Reynolds C. H., and S. Silver. 1983. Citrate utilization by *Escherichia coli*: Plasmid- and chromosome-encoded systems. Journal of Bacteriology 156:1019–1024.
- Rozen D. E., N. Philippe, J. Arjan de Visser, R. E. Lenski, D. Schneider. 2009. Death and cannibalism in a seasonal environment facilitate bacterial coexistence. Ecology Letters 12:34–44.
- Schaechter M. 2015. A brief history of bacterial growth physiology. Frontiers in Microbiology 6:289.
- Schluter D., and G. L. Conte. 2009. Genetics and ecological speciation. Proceedings of the National Academy of Sciences of the United States of America 106:9955–9962.
- Scott M., S. Klumpp, E. M. Mateescu, and T. Hwa. 2014. Emergence of robust growth laws from optimal regulation of ribosome synthesis. Molecular Systems Biology 10:747.
- Simpson G. G. 1953. The Major Features of Evolution. New York, Columbia University Press.
- Stylianidou S., C. Brennan, S. B. Nissen, N. J. Kuwada, and P. A. Wiggins PA. SuperSegger: robust image segmentation, analysis and lineage tracking of bacterial cells. 2016. Molecular Microbiology 102:690–700.
- Tenaillon O., J. E. Barrick, N. Ribeck, D. E. Deatherage, J. L. Blanchard, A. Dasgupta, G. C. Wu, S. Wielgoss, S. Cruveiller, C. Médigue, D. Schneider, and R. E. Lenski. 2016. Tempo and mode of genome evolution in a 50,000-generation experiment. Nature 536:165–170.
- Tinbergen N. Behavior and natural selection. 1965. In: Ideas in Modern Biology (ed. Moore JA), pp. 519–542. New York, Natural History Press.

- Turkarslan S., D. J. Reiss, G. Gibbins, L. S. Wan, M. Pan, J. C. Bare, C. L. Plaisier, and N. S. Baliga. 2011. Niche adaptation by expansion and reprogramming of general transcription factors. Molecular Systems Biology 7:554.
- Turner C. B., Z. D. Blount, and R. E. Lenski. 2015. Replaying evolution to test the cause of extinction of one ecotype in an experimentally evolved population. PLoS One 10:e0142050.
- Turner C. B. 2015. Experimental evolution and ecological consequences: new niches and changing stoichiometry. Ph.D. Dissertation, Michigan State University.
- Turner P. E., V. Souza, and R. E. Lenski. 1996. Tests of ecological mechanisms promoting the stable coexistence of two bacterial genotypes. Ecology 77:2119–2129.
- UniProt Consortium. 2017. UniProt: the universal protein knowledgebase. Nucleic Acids Research 45:D158 D169.
- Van Hofwegen D. J., C. J. Hovde, and S. A. Minnich. 2016. Rapid evolution of citrate utilization by *Escherichia coli* by direct selection requires *citT* and *dctA*. Journal of Bacteriology 198:1022–1034.
- Vandecraen J., M. Chandler, A. Aertsen, and R. Van Houdt. 2017. The impact of insertion sequences on bacterial genome plasticity and adaptability. Critical Reviews in Microbiology 43:709–730.
- Vasi F., M. Travisano, and R. E. Lenski. 1994. Long-term experimental evolution in *Escherichia coli*. II. Changes in life-history traits during adaptation to a seasonal environment. The American Naturalist 144:432–456.
- Velicer G. J., and H. Mendes-Soares. 2009. Bacterial predators. Current Biology 19:R55–R66.
- Wiser M. J., N. Ribeck, and R. E. Lenski. 2013. Long-term dynamics of adaptation in asexual populations. Science 342:1364–1367.
- Yeh P. J. 2004. Rapid evolution of a sexually selected trait following population establishment in a novel habitat. Evolution 38:166–174.
- Zenni R. D., and M. A. Nuñez. 2013. The elephant in the room: the role of failed invasions in understanding invasion biology. Oikos 122:801–815.