

THE EVOLUTION OF A KEY INNOVATION IN AN EXPERIMENTAL POPULATION OF
ESCHERICHIA COLI: A TALE OF OPPORTUNITY, CONTINGENCY, AND CO-OPTION

By

Zachary David Blount

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Microbiology and Molecular Genetics

2011

ABSTRACT

THE EVOLUTION OF A KEY INNOVATION IN AN EXPERIMENTAL POPULATION OF *ESCHERICHIA COLI*: A TALE OF OPPORTUNITY, CONTINGENCY, AND COPTION

By

Zachary David Blount

The importance of historical contingency in evolution has been extensively debated over the last few decades, but direct empirical tests have been rare. Twelve initially identical populations of *E. coli* were founded in 1988 to investigate this issue. They have since evolved for more than 50,000 generations in a glucose-limited medium that also contains a citrate. However, the inability to use citrate as a carbon source under oxic conditions is a species-defining trait of *E. coli*. A weakly Cit⁺ variant capable of aerobic citrate utilization finally evolved in one population just prior to 31,500 generations. Shortly after 33,000 generations, the population experienced a several-fold expansion as strongly Cit⁺ variants rose to numerical dominance (but not fixation). The Cit⁺ trait was therefore a key innovation that increased both population size and diversity by opening a previously unexploited ecological opportunity.

The long-delayed and unique evolution of the Cit⁺ innovation might be explained by two possible hypotheses. First, evolution of the Cit⁺ function may have required an extremely rare mutation. Alternately, the evolution of Cit⁺ may have been contingent upon one or more earlier mutations that had accrued over the population's history. I tested these hypotheses in a series of experiments in which I "replayed" evolution from different points in the population's history. I observed no Cit⁺ mutants among 8.4×10^{12} ancestral cells, nor among 9×10^{12} cells from 60 clones sampled in the first 15,000 generations. However, I observed a significantly greater

tendency to evolve Cit⁺ among later clones. These results indicate that one or more earlier mutations potentiated the evolution of Cit⁺ by increasing the rate of mutation to Cit⁺ to an accessible, though still very low, level. The evolution of the Cit⁺ function was therefore contingent on the particular history of the population in which it occurred.

I investigated the Cit⁺ innovation's history and genetic basis by sequencing the genomes of 29 clones isolated from the population at various time points. Analysis of these genomes revealed that at least 3 distinct clades coexisted for more than 10,000 generations prior to the innovation's evolution. The Cit⁺ trait originated in one clade by a tandem duplication that produced a new regulatory module in which a silent citrate transporter was placed under the control of an aerobically-expressed promoter. Subsequent increases in the copy number of this new regulatory module refined the initially weak Cit⁺ phenotype, leading to the population expansion. The 3 clades varied in their propensity to evolve the novel Cit⁺ function, though genotypes able to do so existed in all 3, implying that potentiation involved multiple mutations.

My findings demonstrate that historical contingency can significantly impact evolution, even under the strictest of conditions. Moreover, they suggest that contingency plays an especially important role in the evolution of novel innovations that, like Cit⁺, require prior construction of a potentiating genetic background, and are thus not easily evolved by gradual, cumulative selection. Contingency may therefore have profoundly shaped life's evolution given the importance of evolutionary novelties in the history of life. Finally, the genetic basis of the Cit⁺ function illustrates the importance of promoter capture and altered gene regulation in mediating the exaptation events that often underlie evolutionary innovations.

Copyright by
ZACHARY DAVID BLOUNT
2011

This dissertation is dedicated to the memory of my great-grandmother, Geneva Mae Sikes (February 5, 1905 – January 27, 1997), my brother, Nicholas Andrew Blount (April 2, 1982 – April 17, 2008), and my aunt, Carmen Brightwell Poland (October 31, 1938 – March 18, 2011).

ACKNOWLEDGEMENTS

I think I first got the notion that I wanted to get a PhD when I was maybe six or seven. I don't quite recall at this point. I do remember it was during a time when I had gotten interested in chemistry, and decided I wanted to be a chemist when I grew up (and an astronaut). My mother told me around then that it took a lot of education, in fact a PhD to work in an area of science. So I decided I was going to get one, a choice that was reinforced by the discovery of the reruns of "Doctor Who" that GPTV showed on Saturday nights. I kept up with that decision over the years, though what I would become a doctor of changed. I lost interest in chemistry and got interested in physics, and maintained that interest until college, when I discovered that physics required an enormous amount of math that I found utterly baffling. I had developed a major interest in biology in 7th grade, when I first discovered evolution, and I had done very well in the subject all through school. When I took AP biology in high school, I can recall Ms. Pugh suggesting that I consider a career in biology, which I disregarded because I wanted to do physics. So, upon learning that physics wasn't going to work out, into biology I went. Here I am, almost thirty years after first deciding to get a PhD, finally getting one. Wow.

This dissertation marks two distinct milestones. First, it provides an interim (because it really isn't done) capstone to a research project that began when I first rotated through the Lenski lab back in 2004. Second, and in a larger sense, it marks the achievement of that childhood dream, and the end of a long formal education that began in 1982. I have a lot of people and institutions who have helped me along the way to this achievement, and to whom I owe a debt of gratitude that I would like to recognize. I will recognize those specific to the first milestone first,

and then move on to the longer list of those who have contributed to the whole long journey. I don't think there is a length restriction on this section, so I will ignore the music if it comes up.

I have great fortune to have had Dr. Richard Lenski as my mentor. He is not only an astonishingly brilliant scientist whose contributions will echo in evolutionary biology for generations to come, but also one of the kindest and most fundamentally decent human beings I have ever known. I want thank him for taking a chance on me, for his guidance, encouragement, and support. Moreover, as an advisor, he has that special talent of being able to grant enormous freedom to explore and strike out, and then knowing just when to provide just the right nudge to ensure one goes in the right direction. It is difficult to describe, but he has helped me to grow and develop confidence as a researcher and scientist in ways I hadn't imagined possible when I first entered the lab. I also want to thank him for being willing to accept, if not encourage little eccentric projects like draping the lab with plants, putting up signs in the windows, not to mention stockpiling used Petri plates for building tottering towers of infectious waste.

I have also had the benefit of a wonderful research advisory committee, which really helped me to clarify my ideas over the years. I would like to thank the members of that committee, Dr. Terrence Marsh, Dr. Robert Pennock, Dr. Charles Ofria, and Dr. Thomas Schmidt, for their wisdom and good counsel.

I had a wonderful group of collaborators without whom the research detailed in this dissertation could not have been done, and I thank them for all of their help and work. Dr. Chris Borland was overseeing the citrate project when I came into the lab as a rotation student. Her initial work on Cit⁺ laid a great foundation for what I ended up doing. She also trained me in the ways of the Lenski lab, got me going in my work, and taught me skills I still use. I wish I had gotten to work with her longer. Dr. Sean Sleight helped me greatly in the initial genetic analysis

of the Cit⁺ trait. Dr. Carla Davidson put an enormous amount of time and effort into some critical expression experiments that were important to telling the story found in Chapter 3. She also became a very dear friend helped me in many ways – I don’t know how I would have gotten through the last few years without her willingness to listen and reassure me. As far as the research goes, I owe the biggest thank you to Dr. Jeff Barrick. Jeff is really one of the most amazing brilliant people I have ever met, and his humility and self-effacement only barely makes tolerable the blindingly Apollonian incandescent shine of his intellect and talent. Jeff is the bioinformatics genius responsible for turning the jumbo-enormous amounts of genomics data that provides the core of Chapter 3 into something comprehensible by way of his custom-designed data mining Breseq platform (which I will forever simply refer to as simply “The Gnomes”). Were it not for him, I would still be trying to make sense out of all those letters. His work ethic and dedication to the pursuit of so many questions have also provided me with much inspiration. As an addendum to this section, I also want to thank Caroline Turner, who is delving into the ecological aspects of the citrate story, for future collaboration. I look forward to working with her on the next phase of work, and I love that her research means I get answers to big questions without my having to do the work! Finally, I want to thank Mark Kauth, an undergraduate who came into the lab initially as one of Jeff’s helpers, but who will be staying on as a graduate student. He has provided critical last-minute help on a number of aspects of the work covered in Chapter 3, and will likely be doing work on some aspect of the citrate story later. Thank you for your help so far; I look forward to working with you in the future.

In addition to research collaborators, my work would not have been possible without the efforts of the Lenski lab support staff. They are the ones who made sure that I had all the supplies that I needed, who cleaned up after me, and who were willing to pour incredible

numbers of plates, sometimes with little notice. They are the backbone of the lab, and without them, everything would screech to a halt. The core of this staff is Neerja Hajela, the lab's manager and resident goddess. Neerja oversees everything with calm, patience, care, and attention. My work would have been absolutely impossible had it not been for her. Moreover, during my time in the lab, she has become one of the best friends I have ever had, and really my second mother. She has always been there to encourage, to listen, to comfort, to counsel, and to just talk and joke around with. I love you, Neerja, and I am so grateful to have had your help, guidance, and friendship. Assisting Neerja in keeping the lab going have been a number truly wonderful undergraduate helpers, including Brian Chernoff, Florence Emanajo, Michelle Mize, and Camorrie Bradley. They are the ones who washed the dishes, ran errands, prepped most of the media, and generally kept things going. Y'all are awesome. I couldn't have done it without you. Finally, I want to single out one of those undergraduate assistances, Marwa Adawe, for special thanks. Marwa was helper during the largest and most frenetic of my experiments, including the big plating experiments discussed in Chapter 2. She did incredible amounts of work without complaint or delay. Amidst all that work, she showed a wonderful curiosity about everything in the lab, and was wonderfully patient for some pretty long-winded explanations. Over the years, she has become a dear and close friend who is very much the sister I never had. Marwa, you are an incredible person with a great mind and greater heart, and I know you will go far. I love you, and I am so grateful to have had your help, and to have even simply known you.

I want to thank our computer tech Brian Baer for keeping the computers running, answering stupid questions, and being patient with a functional Luddite, not to mention being willing to loan various books and videos for years on end. Brian has also been master of everything electronic, taken pictures, set up time-lapse videos, helped in setting up presentations,

and been an all around jack-of-all trades. He is also one of the kindest, most helpful people I have every known, in addition to being a good friend, and a great person to talk to about books, movie, and random nerdy topics. He's definitely made the lab a more fun place to work.

My work has been made possible also by the generous funding of a number of agencies and organizations. Through grants to Dr. Lenski, my work has been primarily funded by the National Science Foundation (NSF), including through the BEACON Center, and the Defense Advanced Research Projects Agency (DARPA). In my own name, I have been supported by an MMG travel grant (2005), two summer fellowships funded by the Ecology, Evolutionary Biology, and Behavior (EEBB) program (2006 and 2007), a Barnett Rosenberg Fellowship (2008 – 2009), an American Society for Microbiology Student Travel Award (2009), a Rudolf Hugh Fellowship (2008), a Duvall Family Award (2009), and a Ronald M. and Sharon Rogowski Fellowship (2011).

Beyond its boss and support staff, the Lenski lab has been a wonderful place to work. This is a place where a student can do really amazing science, and really feel like one is a part of something big. Not only is this home to the LTEE, a truly epic experiment, but it has been the nurturing ground of so many luminaries of experimental evolution. And it still nurtures a diverse array of future luminaries. Therefore, one of the great things about the lab is the opportunity to work with brilliant people who are examining fascinating questions. These undergraduate, graduate, and postdoctoral colleagues have provided wonderful resources of intellect and support, and who are always there to provide a different perspective, to read over or discuss a paper, or to answer a question or lend a hand. I have already mentioned a number of them who ended up playing a major role in my work, and I would like to additionally recognize all those others with whom I have had the honor of working: Dr. Kristina Hillesland, Dr. Christopher

Marx, Dr. Dusan Misevic, Dr. Elizabeth Ostrowski, Dr. Chris Strelhoff, Dr. Robert Woods, Dr. Gabe Yedid, James Dittmar, Devin Dobias, Nathan Johns, Rohan Maddamsetti, Daniel Mitchell, Dr. Jeff Morris, Christian Orlic, Mike Wiser, and Luis Zaman. I want to particularly thank Brian Wade and Justin Meyer for a good bit of discussion and friendship over the years.

I gave a poster at the 2005 ASM general conference in Atlanta. During my session, a man came up, looked over the poster, and listened to me give my spiel. When I was done, and had answered his questions, he asked me if I knew how lucky I was to be working in the Department of Microbiology and Molecular Genetics at MSU. He named a number of the prominent researchers in the department, and declared it the “Athens of Microbiology”. I have always been struck by that interaction, and, indeed, I have had amazing fortune as a doctoral student to have studied in MMG, where I have gotten a first rate graduate education and have been able to work with incredible people while doing cutting-edge science in a vibrant environment. Moreover, needed resources are always at hand. This is a remarkable place, and if it is the Athens of microbiology, it is Athens with the resources of Rome. I want to thank Dr. Walter Esselman, the chair of the department, and the entire department faculty for making this such a great place to learn and work.

I also doubt a debt of gratitude to a number of people who have generally helped or simply brightened my day during my time at MSU, including Stephanie Eichorst, Uri Levine, Fan Yang, Rhia Leveque, Jenifer Mayrberger, and Subrena Jones.

Finally, I want to thank some people I have never met, but whose work has greatly influenced me and my work. First, Aristotle invented biology and provides a model for clear thinking. Had it not been for him, I would likely have been a historian. Charles Darwin has provided a source of great intellectual delight, as well as being a model of not only how to be a

good and creative scientist, but also of how to be a good human being. It is remarkable just how wonderful and helpful his writings are a century and half on. And then there is Stephen Jay Gould. His writings and ideas have provided the impetus and intellectual foundation of much of my work, and I think it is correct to say that my research has been contingent upon his earlier efforts. Moreover, his extensive body of delightful, if often frustrating, writings have given me a great number of ideas, moments of puzzled silence, and hours of entertainment. I regret that I will never get to meet him and talk about contingency; he left the world far too soon.

On the broader journey that began long before I came to MSU, I have another myriad to thank. First and foremost, I must thank my family for supporting me in so many ways during my life. I am so grateful for their help, and I know how fortunate I have been to have had such a supportive family. My parents, Donna and David Blount, taught me to be resilient and independent, and always encouraged me to follow my dreams. Mom taught me to love learning and reading, and was a great emotional support through the years. Dad taught me a solid work ethic, the value of hard work, and the lesson that any job worth doing is worth doing well – no matter who might know. Working for Dad was not always pleasant or easy, but it gave me the foundation that has helped me succeed. My brothers, Jonathan and Nicholas, were good companions growing up (usually), and they gave me many hours of fun, adventure, and frustration. Nick passed away too soon, but he was a kind and gentle soul. Jonathan has been a model of dedication and discipline, and I am grateful to him for the advice and help he has given me over the years. My great-grandmother, Geneva Sikes, who I knew only as “Greatmommie”, was the light of my childhood – a kindly presence who was always there to encourage, to comfort, to listen, and to tell me stories of her life. I have missed her greatly since she left this world. I live every day hoping that I make her proud. I want to thank my maternal grandparents,

Frankie and John Davis, for their care, love, and support over the years. They have always been there for me, and I couldn't have asked for a better pair of grandparents. I thank my Aunt Carmen and Uncle Tom for being such an important and nurturing part of my childhood. As the dedication indicates, Aunt Carmen died recently, and I will miss her. She was one of the most magical people I have ever had the pleasure of knowing. And, come to think of it, she also introduced me to *Doctor Who*. And I also thank my nieces, Jordan and Taylor, and my nephews, Bailey, Christian, and Xavier for their entertainment value. I love you all.

I have had a small, but close circle of dear friends – Tim Clemons, Lee Goodson, Laurie Schwartz, Crystal Wenzler, Yasemin Tulu, and Matt and Karen Williams – who have been important parts of my life for years. Thanks for the encouragement, willingness to listen, and the years of friendship and fun. I hope I can one day repay you for it all.

I owe a deep debt of gratitude to my long-time significant other, Dr. Rae Braudaway, for her help, love, and support over the years. She has been integral to getting me through a difficult decade. I love you, Raeny.

I want to thank my cats, Aggie and Phrossy, for their companionship and what I choose to interpret as affection.

I have had many, many teachers over the years who helped to get me here. Among that great number, there is a select group of teachers who played a special role in my education and life, to whom I owe a special debt. Paula Gleeson, my kindergarten teacher, gave me a good start. Mary Ann Watson was my second grade teacher, and she did a lot to correct damage inflicted by a truly ghastly first grade experience. She gave me love a love of learning, introduced me to Mozart and Bach, and put me on a solid academic path. One day I will find you and let you know just how much good you did for that scared little boy. Lynn Pugh was my

high school chemistry and AP biology teacher. She was kind enough to take a particular interest in me, and, as I mentioned above, was the first to recommend that I consider becoming a biologist. She seems to have been correct in that recommendation. She was also the first teacher to formally introduce me to many of the fundamental concepts of evolution, and to devote significant class time to the subject. Dr. Thomas Tornabene was my introductory microbiology professor at Georgia Tech, and he introduced me to the wonders of the microbial world. He had an enthusiasm for the subject that was infectious, and that feeling really took me. And all these years later, I still believe in Santa Claustidium. Dr. Dennis Grogan was my advisor for my master's work at the University of Cincinnati, and he really showed me the ropes of formal scientific research. I was rather grateful for his patience. Finally, I want to thank Dr. James Miller for his Nature and Practice of Science course here at MSU. That class was hands down the most valuable I have ever taken during my graduate career. I do not exaggerate when I say it taught me how to think like a scientist. My work has been much better-structured and thought out than it would have been without Dr. Miller's help. Indeed, I chose to work with Chris Borland on the citrate project because the prospectus she put together of a possible project conformed to Platt's strong inference ideas, which he taught me to so respect. I honestly don't understand why that class, or at least something like it, isn't required of every graduate science program.

Finally, to She of the clear gray eyes, I give thanks for inspiration. May You forever guide my head and my heart.

In all things of Nature, there is something of the marvelous.

– Aristotle, *The Parts of Animals* (350 BCE?)

There is grandeur in this view of life, with its several powers, having been originally breathed into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved.

– Charles Darwin, *On the Origin of Species* (1859)

It is hard to realize that the living world as we know it is just one among many possibilities; that its actual structure results from the history of the earth. Yet living organisms are historical structures: literally creations of history. They represent, not a perfect product of engineering, but a patchwork of odd sets pieced together when and where opportunities arose. For the opportunism of natural selection is not simply a matter of indifference to the structure and operation of its products. It reflects the very nature of a historical process full of contingency.

– François Jacob, “Evolution and Tinkering” (1977)

TABLE OF CONTENTS

LIST OF TABLES.....	xix
LIST OF FIGURES.....	xx
KEY TO SYMBOLS AND ABBREVIATIONS.....	xxii
CHAPTER 1	
HISTORICAL CONTINGENCY AND EVOLUTION.....	1
A Brief History of the Neglect of History in Evolutionary Biology.....	1
Stephen Jay Gould and “Replaying the Tape of Life”	4
Difficulties in Gould’s Notions of Contingency.....	8
Reconsideration of Historical Contingency.....	12
Contingency and Human History.....	12
A historical vignette.....	12
Counterfactual analysis of history.....	13
Properties of a Historically Contingent System.....	16
Evolution and Contingency.....	22
Empirical Studies of Historical Contingency in Evolution.....	26
Macroevolutionary Studies.....	28
Evolution Experiments.....	35
The <i>E. coli</i> Long-term Evolution Experiment.....	37
An Open Ecological Opportunity.....	43
History of the Inclusion of Citrate in DM Medium.....	44
<i>E. coli</i> has the Potential to Evolve a Means of Exploitation the Citrate Opportunity.....	49
Cit ⁺ Evolves in the Ara–3 Population.....	50
Research Discussed in this Dissertation.....	53
Epilogue.....	55
Increased Concern for Historical Contingency Has Developed in Parallel Across Multiple Disciplines, and in the American Culture at Large.....	55
References.....	62
CHAPTER 2	
HISTORICAL CONTINGENCY AND THE EVOLUTION OF A KEY INNOVATION IN AN EXPERIMENTAL POPULATION OF <i>ESCHERICHIA COLI</i>	79
Abstract.....	80
Background and Introduction.....	82
Results.....	87
Evolution of Cit ⁺ Function in Population Ara–3.....	87

Historical Contingency in the Evolution of Cit ⁺	92
Statistical Analysis of the Replay Experiments.....	97
Rates of Mutation to Cit ⁺ and Quantification of the Potentiation Effect.....	99
Frequency-dependent Selection Maintains Ecological Diversity.....	103
Discussion and Future Directions.....	105
Methods and Materials.....	111
The Long-term Evolution Experiment.....	111
Media and Culture Conditions.....	111
Analysis of the Population Expansion.....	112
Search from Cit ⁺ Variants.....	112
Growth Curves.....	113
Replay Experiments.....	113
Confirmation of Cit ⁺ Variants.....	113
Test for Frequency-dependent Interaction.....	114
Acknowledgements.....	114
Supporting Information.....	114
SI Materials and Methods.....	114
Media and Culture Conditions.....	114
Analysis of the Population Expansion.....	115
First Replay Experiment.....	115
Second and Third Replay Experiments.....	116
Confirmation of Cit ⁺ Variants.....	118
Test for Frequency-dependent Interaction.....	118
References.....	120

CHAPTER 3

GENOMIC ANALYSIS OF A NOVEL INNOVATION IN AN EXPERIMENTAL POPULATION OF <i>ESCHERICHIA COLI</i>	126
Abstract.....	127
Introduction and Background.....	128
Genome Sequencing and Phylogenetic History.....	130
Evolution of the Cit ⁺ Function.....	134
Actualization of the Cit ⁺ Function.....	135
Refinement of the Cit ⁺ Function.....	143
Potentiation of Cit ⁺ Evolution.....	158
Perspective.....	172
Methods.....	175
Evolution Experiment.....	175
Genomic DNA Isolation.....	176
Whole-genome Shotgun Re-sequencing and Mutation Detection.....	176
Phylogenetic Reconstruction.....	177
PCR Screens of <i>cit</i> Amplifications.....	178
Expression Experiments.....	178

Isogenic Strain Construction.....	181
Growth Trajectories.....	182
Fitness Assays.....	182
Plasmid Construction.....	182
Identification of Cit ⁺ -conferring Mutations in Replays.....	183
Acknowledgements.....	186
Appendix.....	188
References.....	199

LIST OF TABLES

Table 1.1	Development of DM medium.....	46
Table 2.1	Summary of replay experiments.....	94
Table 2.2	Statistical analyses of three replay experiments.....	98
Table 3.1	Historical Ara-3 clones subjected to whole genome sequencing.....	131
Table 3.2	Results of PCR screens on whole-population samples for <i>cit</i> amplification.....	137
Table 3.3	Detection of <i>cit</i> amplification in Ara-3 clones.....	138
Table 3.4	Point mutations in early Cit ⁺ genomes.....	147
Table 3.5	IS insertions in early Cit ⁺ genomes.....	148
Table 3.6	Deletions in early Cit ⁺ genomes.....	149
Table 3.7	Amplifications in early Cit ⁺ genomes.....	150
Table 3.8	Estimated <i>citT</i> module copy number.....	157
Table 3.9	Phylogenetically informative mutations.....	159
Table 3.10	Mutations affecting <i>cit</i> region in Cit ⁺ replay mutants.....	166
Table 3.11	Annotated differences between the genomes of Cit ⁺ mutants and their Cit ⁻ parent clones.....	167
Table 3.12	Primer pairs used in this study.....	180
Table 3.13	Clones used in growth – trajectory experiments.....	185
Appendix Table A.1	Phylogenetic placement of Cit ⁻ replay clones.....	189

LIST OF FIGURES

Figure 2.1	Population expansion during evolution of the Cit ⁺ phenotype.....	89
Figure 2.2	Growth of Cit ⁻ and Cit ⁺ cells in DM25 medium.....	90
Figure 2.3	Alternative hypotheses for the origin of the Cit ⁺ function.....	93
Figure 2.4	Mutation rates from Ara ⁻ to Ara ⁺ and Cit ⁻ to Cit ⁺ of the ancestor and a set of potentiated clones.....	101
Figure 2.5	Frequency-dependent selection allows stable co-existence of Cit ⁻ and Cit ⁺ clones.....	104
Figure 3.1	Phylogeny of Ara-3 population.....	132
Figure 3.2	Total number of mutations relative to the ancestor.....	133
Figure 3.3	Tandem amplification found in Cit ⁺ genomes.....	136
Figure 3.4	Annotation of sequence adjacent to the boundary of the <i>cit</i> amplification found in Cit ⁺ genomes.....	140
Figure 3.5	Expression levels from native <i>citT</i> , native <i>rnk</i> , and evolved <i>rnk-citT</i> regulatory regions during aerobic metabolism.....	142
Figure 3.6	New <i>rnk-citT</i> module confers Cit ⁺ phenotype in potentiated background.....	144
Figure 3.7	Growth of early Cit ⁺ clones in DM25 medium.....	145
Figure 3.8	Refinement of Cit ⁺ phenotype by increased copy number of <i>rnk-citT</i> module.....	156
Figure 3.9	Key for placement of clones used in replay experiments that were sampled before generation 20,000.....	161
Figure 3.10	Key for placement of clones used in replay experiments that were sampled from generations 20,000 or later.....	162

Figure 3.11	Evidence of epistatic interactions in potentiation of Cit ⁺ phenotype.....	164
Figure 3.12	Mutations that produced Cit ⁺ phenotype in 14 replay experiments.....	173
Figure 3.13	Growth of Cit ⁺ mutants derived in replay experiments.....	174

KEY TO SYMBOLS AND ABBREVIATIONS

Ara ⁻	Unable to use arabinose
Ara ⁺	Capable of using arabinose
Ara ⁻ ##.....	Designation for one of six LTEE lines founded with an Ara ⁻ ancestor
Ara ⁺ ##.....	Designation for one of six LTEE lines founded with an Ara ⁺ ancestor
C1.....	Ara-3 clade one
C2.....	Ara-3 clade two
C3.....	Ara-3 clade three
CI.....	Confidence interval
<i>cit</i> amplification.....	Gene amplification overlapping the <i>cit</i> operon
Cit ⁺	Aerobic citrate-utilizing phenotype or clone or group with that phenotype
Cit ⁻	Unable to use citrate aerobically
CRP.....	Catabolite repressor protein
DM.....	Davis and Mingolini minimal salts base
DM25.....	DM-Base supplemented with 25 µg/mL glucose
DM0.....	DM-Base with no added sugar
DM500.....	DM-base with 500 µg/mL glucose
DM1000.....	DM-base with 1 mg/mL glucose
DNA.....	Deoxyribonucleic acid
FS.....	Frameshift

IS.....	Insertion sequence element
L.....	Liter
LB.....	Luria-Bertani broth
LTEE.....	Long-term evolution experiment
MC.....	Minimal citrate medium
mL.....	Mililiter
NC.....	Noncoding
NS.....	Nonsynonymous
OD ₄₂₀	Optical density at wavelength 420 nm
PCR.....	Polymerase chain reaction
P ₀	Fraction of replicates producing no mutants in a fluctuation test
SNP.....	Single nucleotide polymorphism
S.....	Synonymous
TA.....	Tetrazolium arabinose indicator medium
TCA.....	Of or pertaining to the TriCarboxylic Acid cycle
UC.....	Unsuccessful clade of Ara-3
μ.....	Mutation rate
μg.....	Microgram

CHAPTER 1

HISTORICAL CONTINGENCY AND EVOLUTION

Evolution involves the interplay of chance and deterministic components in a sequential process of descent with modification that occurs over time in organismal lineages that stretch back unbroken over billions of years. Evolution is therefore an intrinsically historical process (Darwin 1859, Mayr 1980, 1988, Bowler 2003, Gould 2002). Over the last few decades, the importance of the historical nature of evolution, and particularly its impact on the repeatability of evolutionary outcomes, has been among the most vibrant and far-reaching of debates within evolutionary biology.

A BRIEF HISTORY OF THE NEGLECT OF HISTORY IN EVOLUTIONARY BIOLOGY

The historical nature of evolution has been a core tenet of biology since Darwin. However, though Darwin certainly recognized its potential importance (Darwin 1866, Beatty

2006a, 2008), the role of history in evolution was long given little emphasis or consideration. This neglect is largely a reflection of how evolutionary biology developed after the publication of *On the Origin of Species*. The failure of Darwin to gain wide acceptance of natural selection led to the “Eclipse of Darwinism”. During this period much of evolutionary biology was concerned with struggles to develop satisfactory theories of the core mechanisms evolution, with neo-Lamarckism, orthogenesis, saltationism, and theistic evolution the dominant proposed alternatives to natural selection (Huxley 1942, Provine, 1971, Bowler 2003, Larson 2004). The conflict was exacerbated by early findings in genetics by Morgan, Pearson, de Vries, and Bateson in the wake of the re-discovery of Mendel’s work, which added the mutation theory of evolution to the other competing ideas. In the absence of a consensus on the fundamental mechanism of evolution, consideration of contributing factors such as history and chance was difficult during this period. Moreover, one of the motivating concerns for the search for alternatives to natural selection was the ongoing dominance of progressive, if not outright teleological, notions of evolution under which chance and history were *a priori* debarred from significant influence (Bowler 2003).

The conflict over the evolutionary mechanism was finally resolved during first half of the 20th century. Between 1918 and the mid-1930’s Fischer, Haldane, and Wright developed population genetics, which effectively integrated Mendelian genetics and natural selection (Provine 1971, Mayr and Provine 1980, Bowler 2003). Over the next decade, Mayr, Dobzhansky, Simpson, Huxley, and a number of other luminaries integrated population genetics with field biology and paleontology, resulting in the modern synthetic theory, which unified biology via the acceptance that, “... gradual evolution can be explained in terms of small genetic changes (“mutations”) and recombination, and the ordering of this genetic variation by natural

selection; and the observed evolutionary phenomena, particularly macroevolutionary processes and speciation, can be explained in a manner that is consistent with the known genetic mechanisms” (Mayr 1980).

Consideration of history’s role in evolution was not advanced by the synthesis. Due to its basis in mechanistic models of population genetics, the issue of history in evolution had been largely ignored [It must be acknowledged, though, that Wright’s shifting balance theory, formulated during this period, does involve a historical component due to its emphasis on the starting point of evolving populations, the chance role of genetic drift, and the effect of mass selection during environmental change on genetic composition (Wright 1932, 1977, Pigliucci 2009)]. Moreover, during its consolidation in the 1950’s and early 1960’s, the synthesis, or rather its interpretation, “hardened” around natural selection to the effective exclusion of other aspects of evolution. Drift, for instance, was noticeably downplayed (Gould 2002, Bowler 2003). One effect of this was an overemphasis on the capacity of natural selection to mold most any trait in most any way, which would mean that natural selection could effectively overwrite prior history, rendering it irrelevant (Gould and Lewontin 1979, Griffiths 1996, Gould 2002, Bowler 2003).

Beginning in the late 1960’s inklings of a renewed interest in the role of history and chance in evolution began to develop. Richard Lewontin, for instance, published on the idea of “historicity” in evolution, showing that the historical sequence of environmental fluctuations can have profound effects on allele frequencies, even under conditions in which selection is the only evolutionary force (1966). Soon afterward, Kimura introduced neutral theory that placed substantial emphasis on drift (1968). During this same period, a few paleontologists, most notably Norman Newell, began to reexamine the issue of mass extinctions and their

potential association with catastrophic events, which suggested a possible role for singular historical events in the evolution of life (Gould 2002, Bowler 2003, Allmon 2009). By the mid-1970s, a number of biologists began to call for re-evaluation of evolution to better explicitly consider the role history and chance (Beatty 2006).

Stephen Jay Gould and “Replaying the Tape of Life”

Among the biologists who began to this process of re-evaluation in the 1970's was one of Newell's former graduate students, Stephen Jay Gould. By the early 1980's, Gould had emerged as the foremost proponent of the view that history often swamps the deterministic power of natural selection to drive evolution to predictable ends, thereby rendering evolutionary outcomes fundamentally quirky and unpredictable. Gould best elaborated this view in his 1989 book, *Wonderful Life*, a prize-winning account of the Burgess Shale fossils written for the general public. The principle narrative focus of the book was the re-evaluation of the Burgess Shale fossils by Whittington, Briggs, and Conway Morris, which had shown that, contrary to the results of earlier, less intensive studies, the ~500 million year old fossil organisms display a far greater range of body plans than is currently seen across all extant phyla. Gould notes that each of the extinct body plans represented an alternate solution to basic biological challenges, and that diversification from each would have yielded descendents radically different from modern organisms that evolved with the body plans that did survive. The modern living world is therefore just one of many outcomes that could have resulted from a past that included the full myriad of Burgess shale organisms. Gould thought this realization suggested a general feature of evolution as a whole that he illustrated in a thought experiment that he called “replaying life's tape”:

You press the rewind button and, making sure you thoroughly erase everything that actually happened, go back to the any time and place in the past – say, to the seas of the Burgess Shale. Then let the tape run again and see if the repetition looks at all like the original. If each replay strongly resembles life’s actual pathway, then we must conclude that what really happened pretty much had to occur. But suppose that the experimental versions all yield sensible results strikingly different from the actual history of life? (Gould 1989)

In Gould’s view, the second possibility would prove correct, and “any replay of the tape would lead evolution down a pathway radically different from the road actually taken” (Gould 1989). This assertion was grounded in Gould’s view that the deterministic power of natural selection only constrains evolution within broad channels of possible outcomes. Within these broad channels are innumerable opportunities for possible evolutionary trajectories to outcomes that vary greatly in their details (Gould 1989, 2002, Beatty 1995). In any replay, these non-necessary details would be determined by chance historical events that would propagate down the ensuing causal chain of evolutionary history, potentially leading to vastly different outcomes. Therefore, because evolution involves an interaction of chance and necessity playing out over a continuous history, evolutionary outcomes are in some sense crucially dependent upon the precise, often chancy details of long, complex causal chains of antecedent states, making them the unpredictable and quirky products of historical contingency (Gould 1979, 1989, 1993, 2002, Allmon 2009). Gould thought that this role for historical contingency was the key feature of evolution, and that evolution could not be understood except in light of it (Gould 1989, 1993, 2002, Allmon 2009, Bambach 2009). Consequently, evolution is not like the simpler natural

phenomena studied by chemistry and physics that are historically insensitive, and can thus be explained by robust process explanations (Sterelny and Griffiths 1999). Instead, evolutionary outcomes can only be understood as historical events subject to actual sequence explanations:

Historical explanations take the form of narrative: E, the phenomenon to be explained, arose because D came before, preceded by C, B, and A. If any of these earlier stages had not occurred, or had transpired in a different way, then E would not exist (or would be present in a substantially altered form, E, requiring a different explanation). Thus, E makes sense and can be explained rigorously as the outcome of A through D ... A historical explanation does not rest on direct deduction from laws of nature, but on an unpredictable sequence of antecedent states, where any major change in any step of the sequence would have altered the final result (Gould 1989).

Needless to say, if correct, this perspective on evolutionary outcomes would have profound ramifications for how both evolution and the history of life may be understood (Gould 1985, Blaser 1999, Sterelny and Griffiths 1999, Beatty 2006, Gould 2002).

As Gould himself noted (1989), no evolutionary biologist dismisses chance history as irrelevant to evolution, but few are willing to grant it as much sway as he does. His views have thus engendered controversy and strong reactions. Richard Dawkins, coming from a position more tightly focused on adaptation, has argued that the deterministic and directional nature of natural selection is sufficient to drive evolutionary outcomes to similar highly adapted states (Dawkins 1996). Simon Conway Morris (1998, 2003, 2010) makes similar arguments, but emphasizes that the widespread occurrence of evolutionary convergence among lineages with different evolutionary histories suggests that the range of viable evolutionary endpoints are

limited. Because “the evolutionary routes are many, but the destinations are limited”, evolution is repeatable regardless of how contingent the routes taken may be on chance and history (Conway Morris 2003). With evolution so constrained, were one to replay the tape of life multiple times, the resulting outcomes would be markedly similar in their essential features, with contingency only resulting in relatively trivial differences (Van Valen 1991, Maynard Smith 1992, Dennett 1995, Dawkins 1996, Conway Morris 2003, 2010). Such repeatability would make evolution predictable and at least potentially subject to robust process explanations (Sterelny and Griffiths 1999).

There has been one history of life on Earth, and replaying the tape of that history on the planetary scale Gould envisioned is clearly not possible. This limitation has made resolution of the contingency debate difficult, and explains at least part of its contentiousness. Moreover, given the involvement of chance at multiple points in evolutionary histories, contingency no doubt plays at least some role in evolution. Indeed, the two hardened positions regarding contingency in evolution obscure that the real issue is more a matter of the relative importance of contingency in evolution, its effects, and the levels on which it works (Foote 1998). These are, of course, questions that can only be answered by empirical studies, and the two decades since the publication of *Wonderful Life* have seen numerous attempts to test Gould’s notions of contingency (Gould and Woodruff 1990, Travisano et al 1995, Losos et al 1998, Emerson 2001, de Queiroz and Rodriguez-Robles 2006, Vermeij 2006, Fukami et al 2007, Dick et al 2009). These studies have illuminated a number of aspects of contingency in evolution, as I will discuss later. Before doing so, however, there are some important conceptual issues that must be addressed.

Difficulties in Gould's Notions of Contingency

As empirical studies of evolutionary contingency have accumulated, there have been questions as to whether or not they actually engage with Gould's ideas (McRae, 1993, Beatty 2006b, Turner 2011). Such questions are likely typical of attempts to examine any difficult and controversial issues, but they seem to be exacerbated by the difficulty in determining precisely what Gould meant by contingency. Contingency was one of Gould's favorite themes, and one of the central foci of his later thinking on evolutionary theory. However, he never produced a formal treatment in which he explicitly developed and clarified his ideas on contingency and its sources (Gould 2002, Bambach 2009). Even in Gould and Woodruff (1990), his one technical, empirical examination of historical contingency in production of an evolutionary outcome, he does not formally define his concept. Moreover, most of Gould's considerations of contingency are presented informally and sometimes tangentially in non-technical writings published over many years. While these writings, particularly his essays in *Natural History*, are delightful and interesting, the descriptions of his concept of contingency they provide are imprecise, and lack the sharply drawn explication that would have made them more amenable to empirical tests. Adding to this difficulty, Gould's description of contingency varies across these informal treatments in potentially significant ways.

Inconsistency in Gould's description of his notions of contingency may even be found in *Wonderful Life*, which contains his most detailed consideration of the subject. Indeed, Beatty (2006b) points out that Gould actually gives two different versions of contingency in *Wonderful Life*. Gould presents the first when he asserts that replaying the "tape of life" would yield a different result each time. Beatty calls this version "contingency as unpredictability", and

identifies it with Gould's position of denying "that evolution by natural selection is sufficient to guarantee the same evolutionary outcome, even given initially indistinguishable ancestral lineages and indistinguishable environments..." This version shifted over time to suggest that the unpredictability it deals with exists on a continuum (Beatty 2006). In the initial version of the replay experiment in *Wonderful Life*, Gould suggests that a replay from *any* past time point would yield a different outcome. He later amended this to specify that the replay would need to be from a *distant* past time point (Gould 1990), a requirement that he later used to dismiss the macroevolutionary repeatability among *Anolis* lizards identified by Losos et al (1998) as having any real evidentiary relevance to contingency's role in evolution (Beatty 2006b). In the second, labeled "contingency as causal dependence", Gould emphasizes the dependence of outcomes on the precise chain of underlying antecedent events. In describing this version, Gould links different outcomes to differences in the antecedent states. However, as McRae (1993) has pointed out, there are inconsistencies in this version as well. Gould suggests at one point in *Wonderful Life* that *any* change in an antecedent event will have significant consequences to the evolutionary outcome, but later states that only a *major* change would do so. The two versions also intersect at times, with Gould suggesting at one point that replaying from identical starting points would yield different outcomes, but later specifying that that the starting point would have to be changed for a different outcome to occur.

Turner (2011) discerns a third possible version of contingency in Gould's concern for chance in macroevolution and certain inconsistencies in his consideration of stochastic forces. Gould had been part of a study group at the Marine Biological Lab at Woods Hole, Massachusetts in the 1970's that had also included David Raup, Thomas Schopf, Daniel Simberloff, and Jack Sepkoski, which produced a series of papers that demonstrated how

stochastic models could replicate observed phyletic patterns, and suggesting a role for stochastic processes in macroevolution (Raup et al 1973, Gould et al 1977). Gould often credited his experiences with this group as having a seminal effect on his subsequent ideas about the intrusion of random historical events into evolution (Gould 2002, Allmon 2009, Bambach 2009). Despite the group's emphasis on stochasticity, Gould later differentiated historical contingency from stochastic processes, among which he specifically cited drift¹ (Gould and Woodruff 1990, Beatty 2006b). However, in his final description of contingency, Gould specifically mentions stochasticity as a contributing factor:

...the tendency of complex systems with substantial *stochastic* components, and the intricate nonlinear interactions among components, to be unpredictable in principle from full knowledge of antecedent conditions, but fully explainable after time's actual unfoldings (Gould 2002).

Turner thinks that these inconsistencies suggest that Gould's actual view was that contingency was synonymous with macroevolutionary stochasticity introduced by random species sorting and the effects of mass extinction. Gould's dismissal of microevolutionary stochastic processes such as drift may then be related to a subsidiary notion of contingency at the microevolutionary level that Gould thinks may be interesting, but not important over the whole of evolutionary history. This comports well with Gould's view that macroevolutionary forces swamp patterns introduced by microevolution (Gould 1985, Allmon 2009).

These inconsistencies and general lack of clarity in Gould's conception of contingency present numerous difficulties. First and foremost, different studies intended to study

¹ Oddly, Gould explicitly describes bottleneck effects, mutations, and mutation order as contributing to contingency despite their stochasticity (Gould and Woodruff 1990).

contingency can engage with different versions of Gould's conception of contingency (Beatty 2006b). This ambiguity can complicate the evaluation of individual studies, as well as the comparison of multiple studies, leading researchers to talk past each other. Moreover, if Gould did actually conceive of contingency as impacting evolution principally at the macroevolutionary level, then his ideas should not actually be looked to for technical guidance or evaluation of microevolutionary studies. Therefore, while Gould deserves credit for bringing notions of historicity and contingency to the forefront of evolutionary thought and to the attention of empirical researchers, it would perhaps be best to re-think the notion of contingency in evolution independent of his speculations upon it. This divorce of the concept from Gould's writings would carry two benefits. First, it provide an opportunity to develop a clearer, sharper concept of contingency that could then be used to guide research and interpret empirical findings. Second, it would explicitly divide consideration of contingency's role in evolution from Gould's rather extreme advocated position that, like Conway Morris' position, ultimately comes down to a philosophical position (Dennet 1995, Sterelny and Griffin 1999). In the following section, I make a brief first attempt at this separation. I will begin by examining the general concept of historical contingency in human history. From there, I will attempt to first define the characteristics necessary for a system to be contingent, and then to derive a general definition of contingency from those features and considerations of philosophical treatments of the subject. I will thereafter examine what features of evolution render it subject to contingency based upon this definition, and then review a number of empirical treatments of contingency in evolution. This approach will allow me to set my empirical work with the evolution of aerobic citrate-using *E. coli* into

context prior to giving an extended background to lead into the empirical work presented in chapters 2 and 3.

RECONSIDERATION OF HISTORICAL CONTINGENCY

Contingency and Human History

For want of a nail the shoe was lost.

For want of a shoe the horse was lost.

For want of a horse the rider was lost.

For want of a rider the battle was lost.

For want of a battle the kingdom was lost.

And all for the want of a horseshoe nail.

- Traditional Proverb

Cleopatra's nose, had it been shorter, the whole face of the world would have been different.

- Blaise Pascal, *Pensées*

But fortune, which has great power in all matters ... causes great shifts in human affairs with just a little disturbance.

- Gaius Julius Caesar, *Commentaries on the Civil War*

A Historical Vignette

During the fall of 1861, the US Army began operations to secure the Mississippi River. Early in November, 5,000 Confederate soldiers crossed the river into Union-held Missouri, where they

made camp near a town named Belmont. To counter this incursion, Brigadier General Ulysses S.

Grant sailed from Cairo, Illinois with 3,100 men on November 6. On the following day, he landed his force north of the city, and led them in an assault on the Confederate beachhead in what would be his first combat leadership role of the war. After initially taking the Confederate camp, Grant's force came under heavy attack from Confederate reinforcements and an artillery concentration on the opposite side of the river. Despite withering fire, Grant was able to guide his men back to their transport ships for a retreat to the safety of Paducah, Kentucky. Shortly after boarding one of the riverboats, Grant entered the captain's quarters, and lay down on a sofa to rest. He rose only a few minutes later to return to the deck. As he stepped away from the sofa, a Confederate musket ball fired from shore pierced the wall of the cabin, and struck the sofa precisely where his head had been moments before (Grant 1885). Grant later went on to take Vicksburg by siege, opening the full length of the Mississippi to Union control, and to win the battle of Chattanooga. Based on these successes, Lincoln made Grant Commanding General of the United States Army. From that position, Grant awarded command of the Army of the Tennessee to William Tecumseh Sherman, and developed a unified strategy of coordinated attacks on Southern forces across multiple fronts. Sherman later took Atlanta, which contributed to Lincoln's reelection in 1864. Grant's strategy ended up winning the war in 1865, and he went on to be elected president in 1868 (McPherson 1988).

Counterfactual Analysis of History

One aspect of the role of contingency in history can be illustrated by use of counterfactuals. A counterfactual is a type of historiographic thought-experiment² that is used to

² The first known use of a counterfactual is in Thucydides' history of the Peloponnesian War (1972), but Thucydides did not originate the concept. Psychological research over the last thirty

analyze causation, and involves the development of a “what if” scenario in an event or historical circumstance is changed, and the consequences of that change on later events are examined (Lewis 2001, Ferguson 2000, Day 2008). In this case, we will pose a counterfactual of the above history in which Grant did not get up from that sofa after the battle of Belmont when he did, and look at the putative consequences of this slight change in a minor event on the later event of his election to the presidency. Assuming that he did not get up when he did, Grant would have remained on the couch until the bullet entered the cabin, been struck in the head by the bullet. It is impossible to assess the energy possessed by the bullet when it struck, but we will assume that the resulting wound would have been fatal. Note that this event belongs to a particular class of events with irreversible immediate consequences. Subsequent history would have had to have unfolded without Ulysses S. Grant. Obviously, Grant dying on November 7, 1861 would have precluded his being elected president in 1868. This is to say that his election is not resilient to the counterfactual, and it is thus possible to say that Grant’s election to the presidency, as well as any event in which he took part after November 7, 1861, was causally dependent up him having gotten up from that sofa before the bullet entered the cabin. Now consider a second counterfactual, this being that Grant was killed at Shiloh, as he almost was in the actual history (Flood 2006). Again, Grant’s election in 1868 would be precluded under the counterfactual situation. Indeed, innumerable events, both large and minute, antecedent to Grant’s election can

years has shown that humans are inherently prone to counterfactual thinking (Kahneman and Tversky 1982). Indeed, as decisions of future action often involve deciding between the potential consequences of multiple course of action, counterfactual thinking appears to play an important role in learning (Kahneman & Miller 1986, Coricelli and Rustichini 2010).

be identified for which counterfactuals may be constructed to which the election would not be resilient.

Now consider the effect of the first counterfactual on another event, this time Lincoln's reelection in 1864. In actual history, Grant played an important part in this event. Historians generally credit Sherman's capture of Atlanta with ensuring a change in public mood that contributed greatly to the reelection. Sherman had his position at the head of the Army of the Tennessee because Grant, who had been named Commanding General of the United States Army by Lincoln, put him there. However, unlike the case of Grant's presidential election, it is difficult to say what effect the counterfactual would have on the fact of Lincoln's election. There are many plausible alternate scenarios in which Lincoln was reelected without Grant's involvement, suggesting some degree of counterfactual resilience, so we cannot logically argue that Lincoln's reelection was causally dependent upon Grant getting up from the couch when he did. However, it is clear that the particular event of Lincoln's reelection as it occurred in history was causally dependent upon Grant getting up from the couch. In this case, we can say that there is necessary causal sensitivity in the *how* of the later event to the earlier event, but not in the *fact* of the later event to the earlier event. This situation could be called sensitive counterfactual resilience indicative of causal relationship, but not necessary causal dependence. Innumerable antecedent events, many of them tiny and of little or no obvious significance at the time, can be identified that would display this same relationship to Lincoln's reelection, just as there are also innumerable antecedent events that can be identified to which it would display an absolute lack of counterfactual resilience.

For virtually any other historical event, there is a similar range of antecedent events for which counterfactuals may be posited to which the later event would be either not resilient or

sensitively resilient. Now consider that each of those events occurs in the context of many other coincident events for which this is also true. Moreover, this historical context has emerged from earlier historical contexts with the same sort of complex causal structure. Consequently, historical events are crucially dependent upon the sequential, and coincident occurrence of many interacting and intersecting antecedent events to which they are causally connected to various degrees (Bloch 1953, Ferguson 2000, Lewis 2001). Historical causation is thus bewilderingly complex, with small, seemingly trivial, odd happenstance events often playing critical roles in the chain of events that leads to an outcome.

The complexity of historical causation makes prediction of historical outcomes notoriously difficult, as all the relevant causal factors and events cannot be accounted for from a given starting point. Moreover, the involvement of small, otherwise trivial events lends a sense that history could have been different had those tiny events occurred differently. There are two notions of contingency that come out of these points. First, there is the idea of contingency as an inability to understand historical outcomes from an initial historical point, provided sufficient temporal distance due to the inability to consider all of the causal factors linking the initial point to the outcome. The second, which I maintain is the relevant notion, is that there is true indeterminacy from a given historical starting point, with the consequence that history matters in determining what the outcome of that starting point is. With the possibility of different histories, historical outcomes are only explicable in terms of the actual sequence of events that lead to them, and the causal relationship between those events (Ben-Menahem 1997, Ferguson 2000). I maintain that this notion of historically sensitive outcomes is the core of historical contingency.

Properties of a Historical Contingent System

The above discussion indicates that, in order for a system to be subject to historical contingency, it must be a system in which history is possible, and in which that history plays a determinative role in outcome instantiation. These requirements suggest that, at heart, a historically contingent system must be one in which there are causal relationships between temporally ordered sequences of events, as well as to the possibility of alternate outcomes (Lamprecht 1971, McPherson 1988, Ferguson 2000). There therefore would seem to be at least five properties necessary for contingency to be a feature of any temporally ordered system, process, or phenomenon. First, there must be a necessary causal relationship between an outcome and at least some of its sequence of antecedent states. If an outcome is not necessarily causally dependent upon any antecedent states, then it cannot be said to be contingent upon them, and should be historically insensitive. Second, multiple outcomes must be possible from a given initial state. Third, there must be at least as many possible causal chains as there are possible outcomes such that each outcome may be reached by one or more chains, but no one particular chain can result in multiple outcomes. The second and third requirements flow from the first. A single possible outcome would require that any and all possible sequences of antecedent states result in that outcome³. In this case, the outcome would inevitably flow from the initial state no matter the intermediate states, and would only have a causal relationship with the initial state. (This situation would suggest a system that would display transient contingency and contingent intermediate outcomes despite an inevitable final outcome, as might well be the

³ To avoid future confusion, when I speak of deterministic processes, this is what I have in mind. If an outcome is guaranteed by a given initial state, then the process is deterministic in my parlance. Determinism, causality, and contingency are different, though related, things.

case with the universe as a whole.) Alternately, were there a single possible sequence of antecedent states, but multiple possible outcomes, the outcomes would be essentially random with respect to, and therefore not causally related to their antecedents. Fourth, the multiple causal chains cannot be fully reticulated, but must be at least partially independent. If they were fully reticulated, then all outcomes would be reachable via all causal chains. This scenario would effectively reduce down again to the outcome being random with respect to the underlying causal chain.

While a contingent system must be causal in the way outlined by the first four criteria, as a fifth criterion, it must also have a source of indeterminacy in order for alternate causal chains to be initiated and alternate end states achieved. Highly complex deterministic systems, for instance, are subject to chaotic effects such that even small chance variations in their initial states can result in wildly divergent and unpredictable outcomes (Ruelle 1991). In the case of historically contingent systems, however, there must be added inputs of indeterminacy beyond that of the initial state in order for history to matter to the outcome. If this were not so, then contingency would essentially be synonymous with highly complex deterministic causation⁴. In

⁴ There is a distinction here that I like to think of as the difference between Fate and Fortuna. In mythology, the Fates have a plan for all of history and everyone and everything within it. The Fates are implacable, and their plan will come to be. Once they decide upon it, it will occur. Also in mythology, there is Fortuna, the goddess of luck. She is not implacable, but fickle, swayable, and capricious. Fortuna can change the course of events at any time she pleases. It is possible that history may better be explained by the metaphor of Fate, where the outcomes of history, strange and improbable though they might seem, are and have been inevitable. I do not argue that this is not the case so. However, if this is so, then I don't think that contingency can

considering this issue, Lamprecht (1971) traced contingency to the fact that outcomes result from the non-necessary interaction of multiple independent causal chains that were causally unconnected prior to their convergence. He illustrated this point with a particularly vivid and informative example:

A former neighbor of mine went out into the woods on his farm the first morning of deer season a few years ago. He is an excellent shot with a rifle, as I had had occasion to observe frequently. He sighted a buck just after sunrise, aimed and shot his rifle with confidence, and then saw the buck gallop off quite unharmed. Puzzled by his failure to hit his quarry, he advanced slowly and carefully along the line of his shot. And he found a small twig of a shrub which had been freshly broken so that the upper part of the twig dangled loosely down. He concluded that his bullet had broken the twig as the twig swayed in the breeze, and that the bullet had been deflected by collision with the twig from the course it would have taken otherwise...An unexpected factor had intruded into the course of a carefully planned event. This factor, though having important causal consequences in the issues of the event, was not in any way causally related to other factors in the situation until it intruded with sudden efficacy. There was a causally continuous sequence in the hunter's intent, his sighting of the buck, his aiming and firing his rifle, and the swift flight of the bullet out of the muzzle of the rifle. There was also another causally continuous sequence in the movements of the buck through

be anything other than an illusory concept. Therefore, my discussion assumes that the Fortuna metaphor is more appropriate, and that there really is indeterminacy in history that permits for actual contingency. I may very well be quite wrong.

the woods. Those two sequences were brought together by the intent of the hunter to shoot the buck. But there was a third causally continuous sequence in the atmospheric conditions, the blowing of the breeze, and the swaying of the twig in the breeze. And this third sequence was quite independent of the other two sequences.

Note the similarity in the structure, though not the end result, of this story and the example from the Civil War that I discussed earlier. In history and any other contingent system, not all of the causal links leading to an outcome are necessarily related to or entailed by all the others. Only the outcome must be necessarily related to all the antecedent events in its underlying causal chain. This nonrelationship between causal links can result in the non-deterministic causal chains that underlie outcomes seen in history (Lamprecht 1971). This, of course assumes that the consequent interaction of the initially unrelated chains is not necessitated by the initial state. Alternatively, indeterminism may also arise from one or more causal links being either affected by or the product of chance or stochasticity (Day 2008). By this I mean that there must be some events in the causal chains such that either their occurrence or instantiated state is probabilistic rather than necessary given the initial state of the system.

These five criteria clearly suggest that the contingency of a system varies, and exists on a continuum between high contingency on one end, and complete determinism on the other (Desjardins 2011). Contingency should decline as the number of possible, or effectively reachable, outcomes declines, and increase as the number of possible outcomes increases. Contingency will also decline as the proportion of possible causal chains leading to a single outcome increases because the favored outcome will become increasingly counterfactually resilient (Ben-Menahem 1997). After all, a system in which there are twenty possible end states

and 500 possible causal chains will display less contingency if 481 of those causal chains lead to only one of the possible outcomes than if the number of causal chains were more evenly distributed. Of course, a corollary of this principle will be that, as the number of causal chains leading to a given outcome declines, that outcome becomes less counterfactually resilient, and thereby more contingent. Another complication is that not all causal chains will be equally likely, so that, even a large number of causal chains will not necessarily entail a highly contingent system if all but one of those chains are highly unlikely. Similarly, the contingency of a system should vary with the sources of indeterminacy, and the number of constituent causal links subject to them. The actual contingency of a system will therefore be determined by the number of possible end states, number of possible causal chains, the distribution of those chains among the outcomes, the distribution of probabilities among the different chains, the reticulation between the chains, and inputs of indeterminacy. Moreover, it is likely that contingency varies depending on the length of the causal chain, and the temporal proximity between the starting point and the end point. A final consideration is that the relevant contingency of a system will also vary depending on how significantly the alternate outcomes vary between each other.

Given the above described criteria, it is possible to give a general description of a historically contingent system. Historical contingency should arise from any system of temporally ordered events for which multiple outcomes are possible from a given initial state, with the occurrence of a particular outcome causally dependent upon the occurrence of one or more, but not all, series of events possible given the initial state, not all of which are necessarily related to each other and/or in which chance plays a part, but that do form a discrete, at least partially independent, causal chain necessarily related to the outcome. A historically contingent outcome is one that requires a particular causal chain or set of causal chains possible from an

initial state, but not necessarily entailed by that initial state. As there are multiple possible outcomes that are causally, but non-deterministically linked to the initial state, any prediction regarding outcomes within a contingent system is necessarily uncertain, though potentially probabilistic, depending on foreknowledge of the number of possible outcomes. Nonetheless, whichever outcome results, it must be in principle fully explicable in terms of the events in its causal chain, which is to say that precise recurrence of those events will result in the same outcome (Lamprecht 1971, Tucker 1999). It is important to note that this prior unpredictability and posterior explicability are consequences of contingency, and not contingency itself.

EVOLUTION AND CONTINGENCY

Evolution is a historical process of descent along continuous organismal lineages with hereditary modification over time. Perhaps unlike history⁵, evolution is subject to a strongly deterministic process. Natural selection works systematically to adapt populations to their prevailing environments, driving them toward local adaptive peaks (Wright 1932, 1988, Dawkins 1996b). However, chance inputs abound in evolution, and all hold the potential to contribute to evolutionary contingency. Natural selection acts on heritable variation generated by random mutation, the random reshuffling of recombination, or introduced by gene flow. Due to genetic drift, even beneficial mutations may be lost, while detrimental mutations may rise to high frequency. Multiple qualitatively similar beneficial mutations may arise, and the fixation of one as opposed to another may hinge on slight fitness differences and chance events including subsequent mutations. These alternative mutations may differ significantly in their pleiotropic effects and epistatic interactions, with significant consequences for possible future evolutionary

⁵ Hegel, Marx, and other proponents of strongly deterministic theories of history would, of course, disagree.

paths, including their reversibility and level of reticulation (Dollo 1893, Wright 1932, 1988, Jacob 1977, Cooper and Lenski 2000, Ostrowski et al 2005, 2008, Weinreich et al 2005, 2006, Cooper et al 2008, Bridgham et al 2009, Salverda et al 2011). Beneficial mutations will arise in some random order, with potentially significant consequences for the fitness and range of allowed future mutations (Mani and Clarke 1990, Lenski et al 1991, Weinreich et al 2005, Nosil and Flaxman 2011). The accumulation of neutral or even weakly detrimental mutations can likewise impact future evolutionary possibilities (Huynen et al 1996, Wilke 2001, Lenski et al 2006).

Any adaptation is also local, and potentially rapid and capricious biotic or abiotic environmental change can radically change fitness (Lewontin 1966, Gould 2002). The chance factors discussed above can drive populations to distinct adaptive peaks that, while similar in fitness in one environment, may have considerably different correlated fitnesses in other environments (Travisano et al 1995, Cooper and Lenski 2000). These differences can have potentially significant consequences for survival or extinction during environmental changes and the “different rules” that govern mass extinction events (Lewontin 1966, Gould 1985, Jablonski 1986, Raup 1991). Following mass extinctions, future evolutionary outcomes will be constrained to those reachable from the surviving lineages, which may be a subset of prior extant or reachable outcomes (Gould 2002, Sole et al 2002).

Is evolution a contingent process? Evolution is a temporally ordered process with causal relationships between outcomes and initial states. As discussed above, though causal chains in evolution are subject to the highly deterministic process of natural selection, chance is nonetheless an integral and inescapable part of them. Moreover, the large number of chance inputs necessarily entails multiple possible causal chains in evolution, though natural selection

and the de-reticulating effects epistasis clearly impact the likelihood or even possibility of the assembly of those chains. Evolution therefore fulfills four of the criteria necessary for a contingent process. Whether or not evolution is contingent in principle therefore rests on whether or not there are multiple possible evolutionary outcomes.

Whether or not there are multiple possible outcomes depends upon how evolutionary outcome is defined. The living world contains millions of diverse species and subspecies that occupy distinct niches, and are characterized by a diverse array of traits (Wilson 1993). If individual species may be considered to be different evolutionary outcomes, then the living world is testament to the existence of alternate evolutionary outcomes reachable from the same initial state that is the last universal common ancestor (Woese 1998). Evolutionary outcome may also refer specifically to adaptive traits and/or suites of traits evolved by particular lineages, and/or ecological interactions and ecological assemblages involving or composed of multiple lineages, in which case there are also clearly alternate outcomes in general (Foote 1998). On a smaller scale, using this definition, whether or not there are multiple possible evolutionary outcomes will vary depending on the environment. In general, for any environment, there should exist some set of possible allowed evolutionary outcomes that likely corresponds to the range of possible ecological opportunities that exist in that environment [Note that on a small scale what is mean by environment can get complicated because both biotic and abiotic aspects have to be taken into account, as both will impact the range of possible ecological opportunities actually available (Fukami et al 20007, Meyer and Kassen 2007)]. This set, which could be considered to be the evolutionary potential of the environment, will define the total number of possible evolutionary outcomes under those conditions. Given the aforementioned extreme diversity of life on Earth, this set is likely extremely large for most habitable environments (Wilson 1993).

On this definition, and on most scales, there will exist many possible evolutionary outcomes, making evolution at least potentially contingent.

However, whether or not there are multiple possible outcomes so defined will also depend on the lineages involved and their prior evolutionary histories. Future evolution must be based on the range of mutations possible and allowable from that ancestral state (Jacob 1977, Burch and Chao 1999, Weinreich et al 2006, Salverda et al 2011). Prior evolution may preclude some evolutionary outcomes by preventing access to their underlying causal chains due to limited reticulation between chains, and the irreversibility of much of evolutionary change (Dollo 1893, Muller 1939, Dawkins 1996a, Gould 2002, Yedid and Bell 2002, Weinreich 2006, Bridgham et al 2009). At the same time, prior evolution may also result in easier access to some outcomes than others. Moreover, new adaptations must be built on earlier structures, and prior adaptations may be incidentally co-opted for other functions, with qualitatively similar adaptations varying in their potential for cooption (Jacob 1977, 1982, Gould 1983, 1993, 2002, Lenski et al 2003a). Prior evolutionary history will therefore potentiate some evolutionary outcomes while constraining others (Jacob 1977, Gould 2002, Weinreich et al 2006, Salverda et al 2011). For this reason, the range of evolutionary outcomes permitted by the environment will almost certainly always be far larger than the number of outcomes possible from a given evolutionary starting point. This potential effect of past evolution has the interesting ramification that future evolution may be rendered highly repeatable due to prior evolution restricting outcomes, or else highly potentiating one particular outcome or set of outcomes (Yedid and Bell 2002). Paradoxically, this means that deterministically repeatable outcomes on this scale may be historically contingent, which can complicate interpretation.

An alternate view is of the highest scale, at which evolutionary outcome may be construed to refer to the biosphere as a whole, with its full complement of filled niches and adaptations. In this case, the relevant question of contingency is that of whether or not there existed alternatives to the whole assemblage of the living world. There are two problems with approaching the contingency of evolution from this perspective. First, given the possibility of contingent intermediate states in a system with a single final possible outcome, the question is likely unanswerable in principle while evolution is still going on. Second, there does not appear to be any way of assessing whether or not there are possible alternate outcomes at this level given the $n = 1$ problem. However, as a completely microbial biosphere does appear to be a viable possibility, there may be at least some possible scope for contingency even at this scale (Maynard Smith and Szathmary 1995, Gould 1996).

EMPIRICAL STUDIES OF HISTORICAL CONTINGENCY IN EVOLUTION

In my opinion, evaluating how contingent, complex, and chaotic history is can only be done empirically ... There is no philosophical superhighway to bypass the careful historiographical study of concrete historical processes and events.

– Aviezer Tucker (1999)

When one attempts to determine for a given trait whether it is the result of natural selection or of chance ... one is faced by an epistemological dilemma. Almost any change in the course of evolution might have resulted by chance. Can one ever prove this? Probably never.

- Ernst Mayr (1988)

The foregoing considerations strongly suggest that contingency is an intrinsic property of evolution. However, they do not address the issue of the actual impact of contingency on the evolutionary process, and the questions that pertain to it. What scope, for instance, does contingency have to impact or determine evolutionary outcomes? To what degree does cumulative selection along favorable paths canalize evolution toward particular outcomes? To what extent is the evolution of particular traits, combinations of traits, ecological assemblages, and ecological interactions contingent upon particular evolutionary or environmental histories? Are some outcomes more contingent than others? If so, are there common features among them? How much can the prior history of a lineage impact future evolution by restricting or potentiating access to various evolutionary paths (causal chains) to given outcomes? Is natural selection a robust enough search algorithm that it can always find the evolutionary path to a given outcome no matter prior history? Even if a particular evolutionary outcome is effectively unreachable by a particular lineage due to prior evolution history, does the extremely high number of lineages in the living world with distinct evolutionary histories guarantee that certain outcomes will always be found?

Addressing these fundamentally empirical questions requires investigation of three major impacts that contingency has been predicted to have on evolution. First, as Gould suggested, contingency should reduce evolutionary repeatability (Gould 1989, 2002). Second, evolutionary outcomes should be history-sensitive, meaning they should be impacted by differences in evolutionary and environmental history. Notably, variation in repeatability may arise from and be explicable in terms of the sensitivity to these differences. Of course, different outcomes may also arise from contingency inherent to microevolutionary processes that may leave no obvious

causal point of departure in the macroevolutionary record. Third, because contingency, while causal, is non-deterministic, it cannot accelerate the occurrence of particular evolutionary outcomes in the way that cumulative natural selection can (1996a,b). Consequently, the more contingent an adaptation is upon a particular history of multiple mutations from a given starting point, the longer the evolution of that adaptation may be expected to take (Foote 1998). While related to the issue of repeatability, this impact deals more properly with evolutionary timing.

Evaluation of these predictions is essential to understanding the actual importance of contingency in the evolutionary process. Over the past few decades, numerous empirical studies have begun to examine these predictions at both the macroevolutionary and microevolutionary levels, and using a variety of systems, including adaptive radiations on islands, and evolution experiments with bacteria and digital organisms. The findings as a whole suggest that the importance of contingency in evolution is not a simple question, likely reflective of the variation of contingency depending on the exact properties of the system under study. In the following, I review a number of these studies, focusing mainly on macroevolutionary studies. I will then focus on the use of evolution experiments with microorganisms in research into contingency, which will provide a basis for introducing the system and phenomenon addressed by the experimental work that is presented in chapters 2 and 3.

Macroevolutionary Studies

Macroevolutionary studies constitute the broadest array of work yet done on contingency in evolution, and have included investigation of all three predictions from contingency. Likely reflective of Gould's influence on the study of contingency, most macroevolutionary studies have involved examination of repeatability on either a grand or a small scale. Grand-scale

studies have primarily focused on examining to what degree adaptations, traits, and other biological properties have arisen independently over the course of evolution. Conway Morris (2003, 2010) has compiled a large number of instances of striking evolutionary convergences among a diverse array of lineages. Based on these findings, he has concluded that convergence on a limited number of viable outcomes is the principle feature of evolution, and that contingency therefore has little or no impact on the evolutionary process. However, Conway Morris' studies have been informal, have lacked a systematic and unbiased approach to examining the actual relative prevalence of convergence in evolution, and have not rigorously differentiated between convergent and parallel evolution. Consequently, while his findings certainly suggest that numerous biological properties can and do repeatedly evolve despite significant differences in evolutionary history, his broader conclusions are in doubt.

Vermeij (2006) conducted a more systematic grand-scale study of major innovations, which he defined as qualitatively new traits that either confer a fundamentally new function or improve a preexisting function. He examined 78 such innovations, their phylogenetic distributions, and their times or origin. Among these innovations, Vermeij identifies 55 that have demonstrably evolved multiple times, though he does so under loose definitions of phylogenetic independence that dismisses the possible effects of deep homology on repetitive evolution (Shubin et al 1997, 2009). The remaining 23 appear to have singular origins. Notably, the singular innovations appear to have originated at far earlier time points, on average, than the repeated innovations. Given that a significant proportion of the individual instantiations of repeated innovations occur in small clades for which information would rapidly decay over time due to extinction and the imperfection of the fossil record, Vermeij notes that ancient innovations may be more likely to appear singular due to long-branch attraction and information loss

obscuring multiple origins. From this consideration, he suggests that current data are insufficient to conclude that there are any truly singular innovations, though their existence cannot be excluded on the basis of the existing data. However, given that most innovations originate in the integration of pre-existing, repeatedly evolved components, their eventual production is likely over long time scales. Moreover, some innovations require longer to evolve simply because they are built on prior innovations. Vermeij therefore concludes that significant ecological, functional, and directional features of evolution are repeatable due to the iterative nature of evolution and the long periods of time involved, while contingency has only a subsidiary role that determines various details.

Dick and colleagues (2009) reached a similar conclusion in a study that addressed the prediction that adaptations contingent upon particular evolutionary histories will take longer to evolve. They examined the origin of two innovations in the cheilostome order of colonial bryozoans. Cheilostomata arose more than 150 million years ago during the late Jurassic. The two innovations, a calcified costal shield protecting the frontal membrane, and an underlying, water-filled ascus that corrects restrictions the shield places on function of the retractable feeding organ, or lophophore, required prior evolution and subsequent modification of articulated periopodial spines. After the evolution of these spines, the emergence of the two innovations showed long delays of 22 million years, and 33 million years, respectively. Nonetheless, the innovations evolved multiple times independently. The authors suggested that such traits that require particular histories or sets of histories to be actualized are ultimately very likely due to the long periods of time involved in evolution. They are thus highly repeatable and predictable features, with only the timing subject to historical contingency.

The most common approach to studying contingency's effect on evolutionary repeatability has been to examine small-scale, natural instantiations of Gould's "replaying the tape of life" thought experiment. These studies exploit situations in which evolution has played out independently under similar conditions, from similar starting points, or both. These studies have a number of benefits. First, they effectively permit tests of the repeatability prediction under natural conditions. Second, when differences between "replays" are identified, they provide instances of contingent outcomes that can be studied for better insight into the conditions under which contingency can drive divergence. Finally, they provide the prospect of at least potentially identifying possible historical or environmental differences to which the different outcomes may be causally attributed. The risk with such studies, of course, is that different outcomes may in fact result from different selection regimes imposed by subtle environmental differences between the "replicates".

In one such study, Knapp et al. (2006) studied mesic grassland ecosystems that have evolved independently in North America and South Africa. Both systems evolved under similar conditions subject to fire, grazing, and climatic variability, and had been previously shown to have significant structural similarities. Knapp and colleagues found that this structural convergence had been concurrent with convergence in the control of aboveground net primary productivity (ANPP) by between-year variation in precipitation. However, despite this overall functional convergence, the South African ecosystem showed greater ANPP sensitivity to variation in precipitation during the early part of the year that could have major consequences if climate change increases variability during these periods. The authors attribute this divergence to intra-annual differences in soil moisture availability between the two locations. This study shows that small differences in environment can lead to the evolution of slight divergences in

ecosystem function that can have potentially important long-term implications for a particular ecosystem's long-term stability and thereby not only its survival, but also that of its constituent organisms.

Despite the value of studies like that of Knapp et al., rare instances of replicated adaptive radiations on islands and in lakes are perhaps more useful and powerful study systems. First, they often permit examination of larger numbers of “replicates”. Second, the radiations often begin from very closely related ancestors, making it easier to examine the effect of contingency on divergence from similar starting points. Among island systems, the *Anolis* lizards radiation on Caribbean islands has been best studied in the context of contingency. Assemblages of multiple lizard species are found on the four islands of Cuba, Hispaniola, Jamaica, and Puerto Rico. The species in these assemblages cluster into seven distinct sets of highly morphologically similar ecomorphs that occupy similar ecological positions across the islands, with all but three ecomorphs found on all islands. Losos et al. (1998) demonstrated that each ecomorph evolved independently on each island, suggesting that natural selection and similar ecological opportunities were sufficient to drive repeated evolutionary outcomes among the replicates. At the same time, their analysis showed that ecomorphs rarely evolved multiple times on each island, suggesting that ecomorph evolution was contingent upon whether or not the ecological opportunity it suited had been taken. A similar pattern has been observed in the evolution of bacterial ecotypes, suggesting a general phenomenon (Fukami et al 2007). These findings suggested a constrained, incidental role for contingency in the radiations, with deterministic factors dominant. Interestingly, a second genus of lizards, *Phelsuma*, or day geckos, that has also experienced replicated adaptive radiations on islands in the Indian Ocean with highly similar environmental conditions and presumably similar ecological opportunities. However, the gecko

radiations show significant differences in the ecomorphs evolved, the range of morphospace they occupy, and how they partition resources relative *Anolis* (Harmon et al 2007, Losos 2010). These differences suggest that significant features of the two radiations have been contingent upon either subtle differences in the respective island environments, the prior evolutionary histories of the genera, or an interaction of both factors (Losos 2010).

Striking radiations of holoarctic fish have also been identified in hundreds of postglacial Canadian lakes colonized by marine ancestors within the last 12,000 years during periods of marine inundation, providing opportunities to investigate a range of important evolutionary issues (Schluter 1996). These radiations have involved a number of different species, but the most intensively studied have been those involving the three-spine stickleback, *Gasterosteus aculeatus*. The freshwater sticklebacks have shown substantial parallel evolution, including the accumulation of numerous similar characteristics associated with reduced predation pressure and exposure to new ecological opportunities (Schluter et al 2010). One commonly evolved trait is reduced armor plating, which has been found to be due to parallel fixation of an ancestral allele of *eda*, the gene for ectodysplasin (Colosimo et al 2005, Schluter et al 2010). These freshwater populations also often display two sympatric, reproductively isolated, ecomorphs, which occupy distinct benthic and limnetic niches (McPhail 1994, Schluter 1996). This degree of parallelism and repeated instances of evolution of the ecomorphs strongly suggests deterministic underlying factors, with ecological speciation driving the divergences. However, the lakes in which repeated stickleback divergences occurred are restricted to a small portion of the freshwater stickleback range, outside of which only single ecomorphic populations are seen. As these lakes are not environmentally different from those containing single ecomorphs, McPhail (1993) hypothesized that a historical event might have been responsible for the difference. He later

identified the lakes with sympatric pairs as those subjected to a second marine inundation ~1500-2000 years after the initial submergence. Later genetic analysis of these populations by Taylor and McPhail (2000) were consistent with a hypothesis that the sympatric pairs were deterministic outcomes contingent upon the environmental history of repeated marine submergence providing multiple waves of marine stickleback colonization. Indeed, the results of this study, those of the two lizard radiations, and the observation that adaptive radiation does not always occur under conditions in which it might be expected suggest that interplay between historical contingencies and deterministic factors determines important aspects of adaptive radiations, including their very occurrence (Taylor and McPhail 2000, Seehausen 2007, Losos 2010).

This interplay is also demonstrated by the findings of Emerson (2000), who investigated the sensitivity of later evolutionary outcomes on prior evolution in Southeast Asian fanged frogs. Her analysis of fifty monophyletic species strongly suggests that the evolution of small fang size later drove repeated evolution of small body size, terrestrial reproduction, and female-biased size sexual dimorphism. This finding supports the view that prior evolution can potentiate certain outcomes, either by constraining future evolution under certain conditions to a limited range of general outcomes, or by altering the morphological “cross-section” exposed to deterministic environmental and ecological forces.

Emerson’s findings, as well as those from the adaptive radiations support the conclusions that that environmental and evolutionary historical contingencies can actually promote repeated evolutionary outcomes in particular instances by providing a context that is effectively a part of the complement of deterministic forces in evolution. Supportive of this, Yedid and Bell (2002), demonstrated that digital organisms lacking a prior history were able to evolve to a wide variety of unrepeatable, well-adapted end states even in an extremely simple environment, but became

more constrained to particular end states as they accumulated history. Ultimately, these findings imply that contingency determines what is repeatedly evolved, which poses a problem for interpreting repetition as evidence of evolution as deterministic. Indeed, it would seem that this interpretation would be valid on a large scale only if one could demonstrate that the repetition was of all outcomes possible for the environment. If the repetition were of a constrained set of the full possible range, then it would actually support an interpretation of the importance of contingency, as the evolutionary constraint that underlay the repetition would have been due to prior history having reduced the range of probable outcomes. On a small scale, repetition would be supportive of the deterministic primacy of natural selection only if the repeated outcome were to correspond to the single most fit end state available in the environment. What these studies generally suggest, therefore, is that, beyond certain almost inevitable general features of living things, contingency and determinism have been tightly intertwined in evolution. History, by eliminating evolutionary totipotence, can produce starting points from which the range of outcomes is sufficiently constrained that certain outcomes are almost certain to result from them. It appears that robust process explanations can be used to predict and explain evolutionary outcomes given starting conditions, but those conditions are contingent, limiting predictability of evolution in the case of most particular instances.

Evolution Experiments

The rise in interest in the role of contingency in evolution fortuitously coincided with the development of evolution experiments with fast-growing microorganisms, including bacteria, algae, viruses, and unicellular fungi as an accepted means with which to address fundamental evolutionary questions (Elena and Lenski 2003). Indeed, these experimental systems have

proven to be among the best and most powerful systems with which to examine the issue of contingency in evolution. While lacking the complexity and temporal scope of macroevolutionary subjects, evolution experiments with microorganisms have numerous advantages that compensate for these deficits. Microorganisms are easy to handle, grow, and maintain in large populations under laboratory conditions, eliminating many of the logistical hurdles associated with field-based macroevolutionary work. This logistical simplicity also allows populations to be propagated over long periods of time. The value of this capacity is amplified by the rapid generation times of the organisms, which permits many generations of evolution over short time scales. A typical experiment can therefore be used to study hundreds to thousands of generations of evolution.

Most of the model microorganisms used are clonal, meaning that multiple populations can be founded from defined and isogenic points. Combined with the small space needed to maintain large microbial populations, this means that high levels of true evolutionary replication may be achieved. With microorganisms, significant stretches of the same tape of life really can be replayed multiple times, and done so simultaneously. Due to the defined starting point, selection in these replicates will act upon *de novo* variation introduced by mutation, rather than upon standing variation. Samples of evolving populations can be frozen to provide a complete fossil record of evolution. These frozen samples remain viable effectively indefinitely, allowing evolutionary changes to be identified and examined by direct comparison of ancestral, intermediate, and evolved organisms. Clonality also means that stable genetic markers may be used to differentiate organisms from different time points in an evolutionary history, which in turn permits fitness changes to be tracked by use of direct competition assays (Lenski et al 1991). Moreover, the wealth of genetic tools, including advanced sequencing technologies, permit ready

identification and manipulation of genetic changes, which may be directly linked to phenotypic changes (Herring et al 2003, Bentley 2006, Hegreness and Kishony 2007, Barrick et al 2009, Barrick and Lenski 2009). This capacity affords a far greater precision in examining and comparing evolutionary trajectories and outcomes than is possible in the natural world.

A final major advantage to these experiments is the high level of control that is possible. As already mentioned, unlike in nature, the genetic composition of founding populations can be controlled to eliminate confounding initial variation. Conversely, initial genetic states and prior evolutionary history can be manipulated to examine their effect on subsequent evolution (Travisano et al 1995). Similarly, important population genetic factors such as mutation supply, population size, and bottleneck stringency may be manipulated (Chao and Cox 1983, Burch and Chao 2000a, b, De Visser et al 1999, Elena et al 2001, Perfeito et al 2007). Moreover, both biotic and abiotic aspects of the environment of the experiment can be manipulated to examine their contingent effects on evolution (Levin and Lenski 1985, Bennett and Lenski 1993, Bohannan and Lenski 2000, Sleight et al 2006, Fukami et al 2007, Meyer and Kassen 2007). Consequently, extrinsic features of the environment may be held highly stable and constant, eliminating the effects of environmental variability that can confound the examination of other aspects of evolution.

The *E. coli* Long-Term Evolution Experiment

Among the microbial evolution experiments that have been used to examine contingency in evolution, the *E. coli* Long Term Evolution Experiment (LTEE) is the longest running and most intensively yet undertaken. Begun in 1988, the experiment was explicitly conceived and designed to examine not only evolutionary repeatability, but also the dynamics of evolutionary change and the relationship between phenotypic and genomic change (Lenski et al 1991, Lenski

2004). Twelve replicate populations were founded from the same *E. coli* B clone, REL606 (Lenski et al 1991). Save for a neutral single nucleotide polymorphism in half of the populations that confers the capacity to metabolize the sugar arabinose (Ara^+), all twelve populations were initially genetically identical. Every day, 1% of each population is diluted into 9.9 mL of DM25, a minimal salts medium amended with 139 mM glucose as a carbon and energy source (Davis and Mingioli 1950). Under these conditions, the populations each undergo approximately 6.64 doublings during a 24-hour period, resulting in a 100-fold expansion to a daily maximum population size that was initially $\sim 5 \times 10^8$ (Lenski et al 1991, Lenski 2004). Each population has experienced more than 50,000 generations of evolution over the 23 years of the experiment so far. Samples of each population have been frozen every 500 generations throughout the experiment.

The LTEE examines evolution under highly simplified conditions (Lenski et al 1991, Lenski and Travisano 1994, Lenski 2004). The founding clone is strictly asexual and lacks natural competence. Moreover, its genome lacks any plasmids or active prophages (Jeong et al 2009, Studier et al 2009). The populations are each maintained in separate culture vessels under strict isolation from each other and outside organisms, eliminating the possibility for gene flow. Evolution in the LTEE is therefore purely by the operation of genetic drift and natural selection upon variation generated de novo by mutation (Lenski et al 1991, Lenski 2004). Moreover, the populations are evolved under stable environmental conditions of 37°C with aeration. The only perturbations the populations experience are the highly regular “seasonality” caused by glucose exhaustion and bacterial alteration of the medium, and the brief period they spend at room temperature without aeration which undergoes the daily transfers. Consequently, the LTEE

permits the examination of the isolated effects of contingency inherent to the core evolutionary processes of mutation, drift, and natural selection.

Indicative of the power of natural selection to drive repeated evolutionary changes, the populations have shown significant and pervasive parallel evolution over the course of the experiment. All twelve populations have evolved significantly higher fitness relative to the ancestor under the conditions of the experiment. All showed a pattern of rapid fitness gain during the early generations that then decelerated markedly (Lenski and Travisano 1994, Lenski 2004). Other parallels have included reduced lag times upon transfer, increased maximal growth rates on glucose, increased average cell size, and decreased peak population density relative to the ancestor (Lenski and Travisano 1994, Vasi et al 1994, Lenski 2004, Philippe et al 2009). All populations have experienced metabolic specialization, as indicated by the loss of some or all of their capacity to grow on a number of substrates (Cooper and Lenski 2000). In the most striking case, all populations lost the capacity to metabolize ribose due to similar *IS150*-mediated deletion events (Cooper and Lenski 2000, Cooper et al 2001). The populations also have also experienced parallel changes in gene expression (Cooper et al 2003), global protein profiles (Pelosi et al 2006), global regulatory networks (Philippe et al 2007, Crozat et al 2011), epistatic interactions that affect the CRP regulon (Cooper et al 2008), and resistance to phages T6* and λ (Meyer et al 2010). Moreover, ten of the populations have evolved increased DNA supercoiling (Crozat et al 2005, 2010). At least three genes have experienced substitutions in all twelve populations (Woods et al 2006, Cooper et al 2001), and a number of others have experienced substitutions in multiple populations (Crozat et al 2005, Cooper et al 2003, Pelosi et al 2006, Woods et al 2006, Philippe et al 2009, Barrick et al 2009, Crozat et al 2010). Given that most genes across the populations have not accumulated any mutations (Lenski et al 2003), these

genetic parallelisms likely reflect the action of natural selection. Indeed, a number of these parallel mutations have been verified to confer increased fitness under the experimental conditions (Cooper et al 2001, Barrick et al 2009, Philippe et al 2009).

Amidst this remarkable parallelism, the LTEE populations have also diverged evolutionarily. While there is parallelism in a number of the adaptive genetic changes identified across the populations, there are still variations in the full complement of adaptive changes in each population (Woods et al 2006, Stanek 2009). Moreover, even when the same genes are mutated in multiple populations, the details and locations of the changes usually differ between the populations (Crozat et al 2005, Cooper et al 2003, Pelosi et al 2006, Woods et al 2006, Cooper et al 2001). Each population accumulated substantial and divergent gross genomic changes, including IS insertions, inversions, and deletions (Papadopoulos et al 1999). A number of the populations have evolved mutator phenotypes due to mutations in methyl-directed mismatch repair pathway and other genes, resulting in accelerated genomic divergence (Sniegowski et al 1997, Barrick and Lenski 2009, Barrick et al 2009). Significant between-population variation in relative fitness was detected after 2000 generations (Lenski et al 1991), and persisted even after 8,000 more generations of evolution (Lenski and Travisano 1994). A hundred-fold higher between-population variation in correlated fitness change on maltose and lactose than on glucose was also detected after 2000 generations (Travisano et al 1995b). After 20,000 generations, decay in unused catabolic functions also varied greatly between populations (Cooper and Lenski 2000).

Most of the divergence observed between the populations suggests that contingency has played a relatively constrained role in LTEE population's evolution. Indeed, natural selection appears to have repeatedly driven the populations to ascend adaptive peaks that are highly

similar in the stable, simple LTEE environment, if not in their correlated properties in other environments (Travisano 1995b, Cooper 2000). Of course, deterministic, repeated evolution is to be expected if either the environment affords few ecological opportunities, or if the prior evolutionary history of the starting population either restricts or precludes discovery of most of the available opportunities or highly predisposes evolution to exploit one opportunity. By design, the glucose resource environment of DM25 affords one major ecological opportunity. The large amount of parallelism with mostly subtle divergence in the LTEE suggests that the ancestor was predisposed to adapt to the exploitation of this opportunity, and to do so via a number of similar evolutionary paths. However, there are at least two populations that have shown much more significant divergence. Both have been based on finding and exploiting ecological opportunities in DM25 that have not been readily accessible via either the majority of available evolutionary paths, or by those most likely to be followed in the LTEE, suggesting that these niche discoveries have been historically contingent outcomes.

One population evolved to exploit an unforeseen ecological opportunity in the LTEE produced in part by an interaction of a sort that the design of the LTEE was intended to eliminate. The amount of glucose present in DM25 is much lower than is typically added to defined bacteriological media (Atlas 2004). The use of such a low amount is a byproduct of the experimental goal of examining evolutionary dynamics and the relationship between phenotypic and genotypic changes without unnecessary ecological complexities (Lenski et al 1991, Lenski 2004, Lenski personal comments). At the time the experiment was begun, and to some extent still today, the easiest way to assess these dynamics involved fitness assays in which evolved populations are competed against their ancestor (Lenski et al 1991). To permit detection of meaningful patterns of fitness change that could be related back to the sweep of individual

beneficial mutations through the populations using this method, it was necessary to simplify the evolutionary dynamics. Reducing the medium's glucose amendment was expected to achieve this simplification in two ways. First, by reducing the maximum size of the population, it was expected to reduce the possibility of clonal interference, a phenomenon in which multiple mutations arise simultaneously in different, competing lineages within the population (Gerrish and Lenski 1998, de Visser and Rozen 2006). Second, it was hoped that reduced glucose availability would restrict the opportunity for evolution of persistent ecological complexity based on cross-feeding or other bacterial niche construction interactions (Helling et al 1987, Tuner et al 1996, Pfeiffer and Bonhoeffer 2004). While low glucose amendment did reduce these complicating factors, neither was fully eliminated (Barrick and Lenski 2009).

The remaining potential for crossfeeding interactions, combined with the seasonality of the LTEE environment, afforded an open ecological opportunity that one population, Ara-2, has evolved to exploit. Ara-2 contains two monophyletic ecotypes that are referred as "S" and "L", based on their consistent production of relatively small and large colonies, respectively (Rozen and Lenski 2000, Rozen et al 2005). The two ecotypes originated prior to generation 6,000, and have stably coexisted since then due to negative frequency-dependent selection based on ecological differentiation between the two. In the mixed population, L ecotype cells display higher fitness during growth on glucose, but suffer higher mortality during the seasonal starvation phase that occurs following glucose exhaustion. S ecotype cells, by contrast, have lower fitness on glucose, but display increased fitness and decreased mortality during starvation in co-culture with L. This difference reflects S cells having improved capacity to scavenge resources from L cells that lyse during starvation (Rozen et al 2009). The scavenging capacity of S appears to have arisen in part, due to a loss of activity in RpoS, the stationary phase sigma

factor (Rozen et al 2009). Defects in RpoS have previously been observed in scavenging phenotypes that evolve as part of the GASP response, and are typically observed to confer improved substrate range, increased competitiveness at low substrate concentration, but reduced competitiveness at high concentration (Finkel 2006). The genetic basis of this activity change has yet to be determined, but it evolved very early in the S ecotype's history. Indeed, the existence of the Ara-2 balanced polymorphism and exploitation of the starvation season opportunity appears to have been contingent upon this loss of RpoS activity. Similar loss of RpoS activity has not been observed in any other LTEE populations. Its selective benefit, and thereby the S ecotype, therefore appear to be likewise contingent upon the L ecotype having improved growth on glucose via a mechanism that led to increased mortality during starvation (Rozen 2009). The evolutionary divergence of Ara-2 from the other populations therefore appears to have been an historically contingent product of multiple coincident events that opened an ecological opportunity based on niche construction and environmental seasonality, and then permitted exploitation of that opportunity.

This dissertation is concerned with the second population known to have evolved in a manner highly divergent from the pattern of the other LTEE populations. This population is known as Ara-3. Like Ara-2 it evolved to take advantage of an ecological opportunity not easily accessed. Unlike Ara-2, however, Ara-3's divergence was based on discovery of a pre-existing niche.

An Open Ecological Opportunity

From the beginning of the experiment, the LTEE has contained an open ecological opportunity. In addition to 139 μ M glucose, the DM25 medium in which the populations are propagated also contains 1700 μ M citrate. However, this abundant potential second carbon

source is inaccessible to the bacteria. Though *E. coli* has a complete TCA cycle that permits internal metabolism of citrate during aerobic growth on other substrates (Lara and Stokes 1952), it is incapable of transporting and using citrate as a carbon and energy source under oxic conditions. Moreover, this characteristic is highly stable. Spontaneous mutants of *E. coli* capable of aerobic growth on citrate (Cit⁺) are extraordinarily rare, with only one having ever been reported in the literature (Hall 1982). Consequently, the Cit⁻ phenotype has long been regarded as one of the key defining characteristics of *E. coli* as a species (Koser 1923, 1924, Scheutz and Strockbine 2005).

History of the Inclusion of Citrate in DM Medium

While citrate constitutes an open ecological opportunity in the LTEE, it was not included in DM25 for that purpose. Citrate has been a standard component of the DM basal salts solution since the medium was first developed and described (Davis and Mingioli 1950, Carlton and Brown 1981) (Table 1.1). DM developed out of a medium formulated by Bernard Davis for use in his experiments with the penicillin method of enriching for auxotrophic mutants of *E. coli* W (Davis 1949). This method exploits the fact that penicillin kills only actively dividing cells, permitting its addition to a culture growing in minimal medium to select for non-dividing auxotrophs. Davis sought a defined medium in which auxotrophs could not grow, cultures could be started from small inocula (10 to 10⁴ cells), and proliferating cells would be efficiently killed by penicillin. Davis notes that satisfaction of the second and third criteria required increasing the concentrations of potassium and magnesium above those found in previously established *E. coli* media such as F Medium (Cohen and Anderson 1946) and Medium 12 (Sahyun et al 1936). During this development, Davis also experimented with citrate amendment, and noted that, “A small amount of sodium citrate (0.01%) was also found to substitute for the high concentrations

of K^+ and Mg^{++} , and was sometimes added to the minimal medium since it augmented the bactericidal rate of penicillin.” The initial recipe was later refined, and a reduced concentration of citrate was reported as a standard ingredient in the DM recipe reported by Davis and Mingioli (1950). The penicillin method articulated by Davis (and Joshua Ledeborg, who developed his own version independently), soon came to enjoy wide use. As a consequence, DM came to be an accepted standard minimal medium amenable for use in a variety of experiments with a number of different organisms. This wide usage entailed occasional alterations (Vogel 1955). Indeed, the canonical version of DM currently used most widely is a variant developed for use with *Bacillus subtilis* that contains 33% less KH_2PO_4 (Carlton and Brown 1981).

The biological role of citrate in DM medium is related to iron availability. Under oxic conditions at neutral pH, iron predominantly occurs as an insoluble and biologically inaccessible ferric hydroxide polymer. Under conditions typical of bacteriological media, the equilibrium concentration of the soluble free ferric ion is $\sim 10^{-12}$ μM (Hartmann and Braun 1981). Citrate binds free ferric ions to form stable, soluble ferric dicitrate chelates. This chelation activity increases the amount of iron in solution and renders it accessible to *E. coli*, which has a dedicated, high affinity ferric dicitrate transport system (Cox et al 1970, Frost and

Table 1.1 Development of DM Medium					
	Proto-DM	DM	E Medium	DM for <i>B. subtilis</i>	DM25
Component	Davis 1949	Davis and Mingioli 1950	Vogel 1955 (E medium*)	Carlton and Brown 1981	Lenski et al 1991
K ₂ SO ₄	1 g	–	–	–	–
MgSO ₄ •7H ₂ O	0.025 g	0.1 g	0.2 g	0.1 g	0.1 g
(NH ₄)SO ₄	–	1 g	–	1 g	1 g
NaHPO ₄ •7H ₂ O	6 g	–	–	–	–
KCl	4 g	–	–	–	–
K ₂ HPO ₄	–	7 g	–	7 g	7 g
NaNH ₄ HPO ₄ •4H ₂ O	–	–	3.5 g	–	–
KH ₂ PO ₄	–	3 g	10 g	2 g	2 g
Thiamine	–	–	–	–	0.002 g
Sodium citrate	–	0.5 g	2 g	0.5 g	0.5 g
Adjusted pH	7.6	7.0	–	–	–
Ammonium lactate	2 g	–	–	–	–
Glucose	–	5.4 g	5 g	5 g	0.025 g
All quantities given are for 1 L of medium					
*Values given for 1x dilution. Standard recipe is for 50x stock solution.					

Rosenberg 1973, Hussein et al 1981, Härle et al 1995). *E. coli* can also acquire iron using an alternate ferric enterochelin uptake system (Cox et al 1970, Langman et al 1972), and is capable of growing quite well in citrate-free media such as M9. However, iron solubility is affected by Na^+ and K^+ concentrations, and a Na^+ to K^+ ratio of 3:1 or higher has been reported to be necessary to maintain the solubility of iron at a level that permits efficient iron acquisition by *E. coli* in the absence of citrate (Langman et al 1972). While M9's Na^+ to K^+ ratio of 3.2:1 fits this criterion, DM's is 1:18.6, suggesting that iron availability is likely low in the absence of citrate. *E. coli* can sometimes be cultured in DM medium without citrate, but anecdotal reports suggest that these cultures are prone to periodic instability and failure to propagate during serial transfer (Lenski personal communication, Blount personal observation). This instability could be explained by the cultures experiencing greater sensitivity to normal fluctuations in iron concentration in DM without citrate⁶.

⁶ It is interesting that iron levels in the water used by Davis for medium preparation were probably higher than they are in more modern laboratories. Davis developed the penicillin method during his time as the head of the US Public Health Service Tuberculosis Research Laboratory (Lochhead and Menninger 1949, Maas 1999). Davis' lab was located in the Kipps Bay Yorkville Health Center building on the campus of the Cornell University Medical Center. This building was constructed in 1937 with iron water pipes (Lewis Wetstein, former head of Cornell Medical Center Facilities Operations Department via personal communication with Elizabeth M. Shepard, Acting Head of Medical Center Archives of New York-Presbyterian Hospital/Weill Cornell). Copper pipes did not come into common use until the late 1940s, with modern plastic plumbing much later. It is unclear how much this technological shift has

While the need to ensure iron bioavailability explains the presence of citrate in DM25, there remains the question of its high concentration in the medium. This is a particularly important question, as it addresses the size of the ecological opportunity presented by the citrate in the medium used in the LTEE. Frost and Rosenberg (1973) determined that the concentration of dissolved ferric dicitrate rapidly declines as the ratio of citrate to total iron falls below 20:1, and that the *E. coli* ferric dicitrate transport system ceases to function at citrate concentrations below 10 μM . Moreover, they reported maximal growth at a citrate concentration of 1000 μM in a medium supplemented with 30 mM glucose. The level of citrate found in DM25 is therefore significantly higher than is necessary to prevent iron limitation even in a medium with far more glucose. This great excess of citrate in DM25 may be explained by three basic points. First, the formulation of DM was designed specifically for use in the penicillin method of obtaining auxotrophic mutants, and Davis makes clear that the addition of citrate was made purely to improve the killing efficiency of penicillin (1949). Given that *E. coli* was at the time thought to be metabolically inert toward exogenous citrate, it is unlikely that Davis was concerned with reducing citrate to the lowest level necessary to have the desired effect. Second, the DM minimal salt recipe provided by Davis and Mingioli (1950) was developed for supplementation with 11.1 mM glucose. At this level of glucose, the citrate concentration is slightly less than three times higher than strictly necessary. Third, the ferric dicitrate iron acquisition system was not identified until 1973 (Frost and Rosenberg), while Davis and Mingioli wrote in 1950. DM was developed for a particular experiment under particular conditions, and prior to the discovery of the biological role of citrate in the medium or of the concentrations strictly necessary for it to

mattered, but it does not appear to be often taken into account during preparation of growth media from recipes developed at much earlier points in time.

fulfill that role, and its formulation reflects this. Unless required for a particular experiment or other pragmatic reason, the recipes of growth media are rarely revised, and this is particularly true of minimal salts solutions. It is therefore unsurprising that the level of citrate in DM was neither changed after the medium came into common use in experiments beyond that for which it was originally intended, nor adjusted following the identification of the ferric dicitrate iron acquisition system. The high concentration of citrate in DM25, and the large ecological opportunity it has presented in the LTEE, is therefore a matter of historical accident and conservative lab practice⁷.

***E. coli* has the Potential to Evolve a Means of Exploiting the Citrate Opportunity**

The only known barrier to aerobic citrate utilization by *E. coli* is the inability to transport citrate into the cell under oxic conditions (Hall 1982, Reynolds and Silver 1983, Pos et al 1998). Indeed, numerous atypical Cit⁺ *E. coli* have been isolated from agricultural and clinical samples, all of which have been found to harbor plasmids that encode foreign citrate transporters (Ishiguro et al 1978, Ishiguro et al 1979). Hall (1982) reported the only known case of a spontaneous Cit⁺

⁷ Peter Reeves of the University of Sydney and John Roth of the University of California at Davis both offered an alternate historical origin to that described (personal communications). According to this explanation, DM was developed from a *Salmonella* medium that was originally kept in 50x stocks. At such a concentration, it was found that the amount of citrate had to be increased to keep MgSO₄ in solution. This explanation is likely incorrect, and appears to be based upon confusion of DM, which is specifically formulated for a 1x concentration, with the similarly constituted E medium (Table 1.1), which Vogel (1955) developed from DM following his use of it during a collaboration with Davis (1952).

mutant of *E. coli*. His findings suggested that either a complex mutation, or multiple mutations, activated one or more cryptic genes that expressed a citrate transporter, though he was unable to identify the genes involved. *E. coli* is also known to encode a dedicated citrate transporter, CitT, which permits fermentative anaerobic growth on citrate if a cosubstrate is available to provide reducing power (Lutgens and Gottschalk 1980). Pos et al. (1998) demonstrated that high-level constitutive expression of *citT* from a multi-copy plasmid was sufficient to confer a Cit⁺ phenotype in the genetic background of the K-12 strain they studied. However, in its native genomic context, *citT* is present in a single copy, and does not appear to be expressed under oxic conditions (Pos et al 1998).

These findings suggest that *E. coli* has the potential to evolve a Cit⁺ phenotype. Presumably, REL606, the ancestor of the LTEE populations shared this potential. However, the extreme rarity of Cit⁺ mutants of *E. coli* makes clear that this evolutionary potential is not easily actualized. Moreover, Hall's (1982) findings indicate that the evolution of Cit⁺ is complex, and requires multiple steps.

Cit⁺ Evolves in the Ara-3 Population

While the evolution of a Cit⁺ variant in one or more of the LTEE populations was considered to be a possibility during the experiment, no additional effort was made to test for that potential beyond seeing what happened in the LTEE itself. However, screens for Cit⁺ variants occur regularly as incidental byproducts of experimental procedures. For example, prior to the freezing of population samples every 500 generations, dilutions of each population are spread on minimal glucose (MG) and minimal arabinose (MA) agar plates. These media are

both derived from DM, and therefore contain substantial amounts of citrate. However, these platings are not especially effective as a means of screening for rare Cit^+ mutants, as all populations should show growth on MG, and half on MA. However, they do present an incidental means of screening for Cit^+ variants in the six Ara^- populations that should not normally produce colonies on MA. The best screen has been the use of DM25 itself. Due to the high concentration of citrate, the evolution of a Cit^+ variant that could effectively exploit the citrate opportunity would be signaled by a visually obvious increase in turbidity. Of course, as part of precautions against contamination, turbidity checks are made of each population at the time of transfer. If a population is found to display elevated turbidity, a sample is plated on TA, MA, and MG medium, and the resulting colonies are scrutinized for evidence of outside contamination. In those cases in which increased turbidity occurs and no evidence of outside contamination is detected, tests may then be done for Cit^+ variants.

On Saturday January 25, 2003, Tim Cooper performed LTEE transfer 4971. During the turbidity check, he found that population Ara-3 was markedly more turbid than the other populations. In accordance with protocol, he spread samples of Ara-3 and Ara-4 (for comparison) on TA plates to check for contaminants and to verify the elevated turbidity. Though significantly more colonies grew from the Ara-3 sample than from that taken from Ara-4, all displayed morphology, color, and odor consistent with *E. coli*. Given that all prior instances of elevated turbidity had been due to contamination, however, the population was terminated despite the lack of definitive evidence of contamination.

The Ara-3 population was restarted on March 21, 2003 from the sample frozen at generation 33,000. On April 10, 2003, Neerja Hajela, while performing LTEE transfer 5044,

reported that Ara-3 again showed increased turbidity. Refrigerated samples of the population from the April 8 and 9 transfers were plated on TA. Again, no obvious contaminant colonies were observed, though samples from the April 9 transfer showed a ~3.5-fold increase in colony number. Ten randomly chosen colonies were tested for resistance to phages T5 and T6, and all showed the T5^r, T6^s phenotype characteristic of REL606-derived clones. Similarly, Gram stains showed only Gram-negative coccal-rods consistent with *E. coli*. The population's turbidity continued to rise over subsequent transfers, peaking at approximately 10-fold higher than the other populations on April 13. At this point, the evolution of a Cit⁺ variant was considered, and samples from the April 11 transfers were plated on minimal citrate (MC) plates, and inoculated into DM0 (DM25 lacking glucose). The DM0 cultures showed turbidity consistent with growth after 48 hours of incubation, and well-formed colonies were observed on the MC plates at 72 hours, demonstrating that there were members of the population capable of using the citrate in the medium. To provide a further check, population samples from April 8 through 14 were plated on TA, and 20 to 400 typical colonies for each day were tested on TA, MC, and Simmon's citrate agar. While none of the clones checked for April 8 showed evidence of growth on citrate, 70% to 97% of clones tested from subsequent sample dates showed both growth on MC and a positive reaction on Simmon's citrate agar. DNA sequencing later showed that Cit⁺ clones isolated from the population have the same mutations in the *pykF* and *nadR* genes possessed by clones from earlier generations of the Ara-3 population, and which distinguish the population from the other LTEE populations (Woods et al 2006). These results established that a Cit⁺ variant of *E. coli* had indeed evolved in the Ara-3 population.

Research Discussed in this Dissertation

The following chapters of this dissertation detail the first stages of research into the origin and evolution of this novel metabolic innovation, which has been the most dramatic change observed over the course of the LTEE so far. In Chapter 2, the Cit^+ phenotype, the essentials of its invasion and rise to high frequency, and the initial impact of its numerical dominance on the population's structure and ecology are described in more detail. Notably, the rise of Cit^+ variants to high frequency did not result in the extinction of clones displaying the ancestral Cit^- phenotype. Instead, Cit^- variants persisted in the population until at least generation 40,000, indicating that the evolution of Cit^+ was also a cladogenic event.

The larger thrust of Chapter 2 deals with the question of the role of historical contingency in the evolution of the Cit^+ phenotype. Certainly, Cit^+ evolution fits two expected features of a historically contingent adaptation (Foote 1998). First, it evolved only once among the twelve populations. Second, its evolution was long-delayed, occurring 16 years into the experiment, and after more than 30,000 generations of evolution. In the core of Chapter 2, I describe a series of experiments in which I replayed the tape of Ara-3's evolution from various time points in its history. These replay experiments establish that the Cit^+ trait is indeed a historically contingent innovation that was causally dependent upon the prior occurrence of one or more mutations. The prior mutation(s) potentiated the actualization of the Cit^+ trait by increasing the rate of mutation to Cit^+ from the immeasurably low ancestral rate to a rate that, while still low, rendered the function mutationally accessible.

Cit^+ therefore offers a remarkable opportunity to study the evolution of a demonstrably historically contingent innovation in a context in which the details of the underlying history may be precisely determined. Evolution in the LTEE is restricted to the operation of only the core evolutionary processes of mutation, drift, and natural selection. In this context, evolutionary history is effectively the accumulation of mutations. Due to the advent of efficient, cost-effective whole-genome sequencing technologies, all mutations that have occurred along a line of descent may be identified (Bentley 2006, Hegreness and Kishony 2007, Barrick et al 2009, Barrick and Lenski 2009). Combined with the capacity to move mutations into relevant genetic backgrounds (Herring 2003), this technology holds the prospect of not only discovering the putative historical sequence of events underlying Cit^+ evolution, but also experimentally reconstructing that sequence to empirically identify the causal contributions of each event to the innovation's origin. While this approach has been done previously at the level of protein evolution (Weinreich et al 2006, Ortlund et al 2007, Bridgham et al 2009), it has not been done at the whole-genome level. Such experiments will not only help to better understand historical contingency, but also the origin of novel innovations from the modification of pre-existing structures, traits, and genetic information (Mayr 1960, Jacob 1977, 1982, Pigliucci 2008).

Chapter 3 describes the steps taken to realize this opportunity. The focus of the chapter are the findings of an extensive program of whole-genome sequencing of dozens of clones, including both Cit^+ and Cit^- , isolated from numerous points in the history of the Ara-3 population. The resulting data allowed reconstruction of the evolutionary history of this population. The phylogeny shows that the population was heterogeneous for most of its history. Three Cit^- clades had emerged by 20,000 generations that co-existed until after the evolution of

the Cit⁺ innovation. The Cit⁺ clones form a coherent, monophyletic clade within one of these three Cit⁻ clades. The data also led to the discovery that the Cit⁺ function's immediate cause was a tandem gene duplication event that produced a new regulatory module in which the citrate transporter gene, *citT*, is placed under the control of a promoter that directs expression during aerobic growth. Increased copy number of this regulatory module subsequently refined the initially weak Cit⁺ phenotype, leading to the rise of Cit⁺ variants to numerical dominance in the population. Further analysis of this mutation and the distribution of potentiated clones within the population phylogeny suggest that potentiation involved at least two mutations, one of which appears to have improved expression and thereby potentiated the evolution of Cit⁺ through functional epistatic interactions.

EPILOGUE

Increased Concern for Historical Contingency Has Developed in Parallel Across Multiple Disciplines, and in the American Culture at Large

The increased interest in historical contingency that arose in the later 20th century was not limited to biology, but instead appears to have been a general phenomenon across many fields of science, the humanities, and popular culture. Here I will give a quick overview of this more general phenomenon, which would seem to be ripe for further study by those who might be so inclined. I make no claim to understand the origins of this increased interest, though I suspect the rise of science fiction and the increased historical self-awareness of western culture in the

post-World War I era has something to do with it. Nonetheless, I find the phenomenon remarkable, and feel it proper to take note of it here. If nothing else, the pervasiveness of interest in contingency in academia and the broader culture will no doubt serve to spur further widespread work on the subject in a variety of areas.

During the period during which Gould was advocating for a better appreciation of contingency in evolution, similar interest in the concept arose in psychology. As previously suggested, because contingency underlie the occurrence of a single outcome from among multiple possible outcomes, it is related to counterfactual historical thinking, or the consideration of alternate historical realities that is the source of the power of what historian and writer Harry Turtledove (2001a) has called "... those two mournful little words ... What if ..." (Lewis 1986). In 1982, the cognitive psychologist and Nobel prize-winning economist Daniel Kahneman and Amos Tversky published a paper that suggested that counterfactual thinking is a core heuristic feature of human thought. Counterfactual thinking has since become the subject of considerable research (Mandel et al 2005). Counterfactual thinking appears to be particularly important in consideration of personal history, where it often carries significant emotional content due to linkage to regret and guilt (Ferguson 2000, Coricelli and Rustichini 2010). Moreover, as decisions regarding future actions often involve deciding between the potential consequences of multiple courses of action, counterfactual thinking appears to play an important role in learning (Kahneman and Miller 1986, and Morris and Moore 2000, Coricelli and Rustichini 2010).

As Beatty has noted (2006b), during the same period, sociology saw increased interest in history and historical approaches to their area of study. This manifested in part in attempts to discern boundaries between sociology and history due to the field's interest in the sociological dimension of history (Abrams 1983, Goldthorpe 1991, Bryant 1994). However, it has also

perhaps more importantly led to growth in the use of counterfactual approaches to causality in sociological phenomena (Sobel 1995, Harding 2003, Morgan 2007). Similar growth in counterfactual analysis has also occurred in economics over the same period (Cowan and Foray 2002, Cartwright 2007). Interestingly, the use of counterfactuals in economics goes back to the 1950's, during which interest in the role of history in biology first began to stir (Cartwright 2007).

Increased use of counterfactuals to trace causation, together with elevated appreciation for contingency has also been seen in professional historiography. This movement has been particularly noteworthy because, though Thucydides and Livy were the first known users of counterfactuals (Thucydides 1972, Livius 1982), historians have traditionally viewed counterfactuals with skepticism, if not disdain and outright hostility, and continue to do so (Thompson 1978, Bulhof 1999, Ferguson 2000). Ferguson (2000) traces this hostility to four factors. First, most historians have considered their jobs to chronicle and explicate what did occur, and not what might have occurred. Second, strongly deterministic, even teleological approaches to history have a long history within the field, as typified by the Hegelian and Marxist traditions born in the 19th century. Third, even among historians who do not ascribe to deterministic approaches, there is significant concern for the principle of causality in history that many see as threatened by consideration of contingency, which has often been conflated with a lack of causality (Ben-Menahem 1997). Finally, there is concern that counterfactuals encourage reductionistic historical explanations in which contributing causal factors in historical outcomes are mistaken for singular causes (see also Bulhof 1999, Fischer 2002).

Despite this traditional hostility, a few instances of the use of counterfactuals by academic historians may be found during the 20th century. The earliest notable example is the

collection of counterfactual historical essays, including one by Winston Churchill, *If It Had Happened Otherwise* (1931). Another, more serious effort is economic historian Robert Fogel's 1964 book, *Railroads and American Economic Growth*, in which the impact of railroads on America's economic expansion was assessed by positing a counterfactual in which canal development had continued in their stead. However, these early instances were sporadic at best, and inspired few followers. Then, in the 1980's and 90's, roughly coincident with Gould's push for reconsideration of contingency in evolution, professional historians began to show an increased willingness to use counterfactuals (Rosenfeld 1999, Bulhoff 1999, Fergusson 2000, Fischer 2002). This has been true of both works designed for the general public (Crowley and Ambrose 2000), as well as more academic works, in which counterfactuals have been increasingly considered for use as legitimate historiographic tools that help to provide better appreciation of the role of contingency in history (Reisch 1991, Ferguson 2000, 2007). It is unclear if this growth was in any way tied to Gould's work, though it is interesting that he specifically mentioned McPherson's explicit consideration of contingency's role in the Civil War in his magisterial *The Battle Cry of Freedom* (1988). Reciprocally, Fergusson has an extended discussion of Gould's ideas in the introduction of *Virtual History* (2000). While the interest in contingency in history and biology may have initially developed independently, they are clearly now intertwined to some degree⁸.

Perhaps the most obvious sign of the increased interest in the contingent nature of history during the late 20th century, however, is to be found in the enormous growth of "alternate history". Alternate history is a sub-genre of science fiction that originated in the 1930's and

⁸ I can assert this with certainty, at least on one side of the ledger, as I have routinely consulted counterfactual historical works during the course of the research I report in this dissertation.

focuses on exploration of various “what if” scenarios of worlds in which history played out differently than in reality (Rosenfeld 2002, see www.uchronia.net for an exhaustive listing). Examples include the persistence of the Roman empire (Silverberg 2003), complete destruction of the European population by the Black Death (Robinson 2002), avoidance of the American Revolution (Sobel 1973, Dreyfuss and Turtledove 1997), a Confederate victory in the American Civil War (Moore 1953, Turtledove 1997), and an Axis victory in World War II (Dick 1962, Harris 1992, Turtledove 2003a), with these last two being perhaps the most popular (Fergusson 2000, Rosenfeld 2002). Alternate histories commonly feature an explicit “point of divergence” from which the altered timeline under exploration diverged. In many cases, the point of divergence traces to a change in seemingly inconsequential event, in which case it is sometimes called a “Jonbar Hinge”⁹, or a horseshoe nail¹⁰. For instance, Harry Turtledove’s “Southern Victory” series (Turtledove 1997, 1998, 1999, 2000, 2001b, 2002, 2003b, 2004, 2005, 2006, 2007) describes an alternate world in which the Confederacy won the Civil War, and traces world history through to a Union victory in an alternate World War II. The Jonbar Hinge of this universe takes the form of a Southern soldier picking up a dropped copy of Lee’s Special Order 191 to the Army of Northern Virginia, when in reality this copy was found by Union soldiers, leading to the critical battle of Antietam (McPherson 2002). By way of another example, in Ray Bradbury’s highly influential short story, “A Sound of Thunder” (1952), the death of a single

⁹ The term Jonbar Hinge comes from Jack Williamson’s 1938 novella, “The Legion of Time”, in which history leads to either a utopian or dystopian future depending entirely upon whether a 12 year old boy named John Barr happens to pick up a magnet or a pebble while playing one day.

¹⁰ After the nursery rhyme quoted earlier in this chapter.

butterfly in the Jurassic has subtle, but significant effects on human history. While alternate history stories are principally intended for purposes of entertainment, more literary entries are concerned with exploration of various ideas, with contingency itself and the consequent fragility of history being common (Rosenfeld 2002). Though long a significant part of science fiction, alternate history experienced a flowering beginning in the late 1980's that accelerated in the 1990's, and continues in the early part of the 21st century. This growth is most evident from the increasing pace of publication of alternate history works that is catalogued at the Uchronia database (www.uchronia.net).

Increased interest in contingency has also manifested in other cultural media. On television, alteration of history has been a common plot point in numerous episodes of the various incarnations of *Star Trek*¹¹ (1966 – 1969, 1973 – 1974, 1987 – 1994, 1993 – 1999, 1995 – 2001, 2001 – 2005) and *The Twilight Zone* (1959 – 1964, 1985 – 1989, 2002 – 2003). Other shows had the alteration of history via contingency as part of the central premise of the story. *Quantum Leap* (1989 – 1993), for instance, involved the adventures of a physicist, Sam Beckett, who each week “leapt” into the body of a person from a different time period, with the charge to “put right what once went wrong”. The suggestion was generally that the small changes Sam made in each person's life was somehow having a larger, though hidden, impact on history in

¹¹ The most notable example is what is perhaps the most highly regarded episode of the original series, *City on the Edge of Forever* (1967), in which Dr. McCoy accidentally goes back in time to 1930's America. During his time there, he saves the life of Edith Keeler, with the consequence that World War II ends with an Axis victory. The remainder of the episodes revolves around the moral dilemma of having to rectify this alteration and set right the time line, which requires Keeler's death.

general. The later *Journeyman* (2007), featured a similar premise. Similarly, historical contingency and counterfactuals have been key concepts in a variety of movies, including, of course, *It's a Wonderful Life* (1946). As with alternate history titles, these films have been more frequent since the 1980's, and have included *Back to the Future* (1985) and its sequels (1989, 1990), *Groundhog Day* (1993), *Frequency* (2000), *The Butterfly Effect* (2004), *CSA: The Confederate States of America* (2004), *Inglourious Basterds* (2009), and *Never Let Me Go* (2010). On a smaller level of the contingencies that affect individual lives, *Sliding Doors* (1998), examines the consequences of a character catching or missing a subway train, while *The Family Man* (2000) traces the effects consequent to the alternative of a single decision in one man's life.

REFERENCES

REFERENCES

1. Abrams, P. (Ed) *Historical Sociology* (Cornell University Press, Ithaca, NY, 1983).
2. Allmon, W.D. The structure of Gould, In Allmon, W.D., Kelley, P.H. & Ross, R.M. (Eds.) *Stephen Jay Gould: Reflections on His View of Life* (Oxford University Press, New York, NY, 2009).
3. Atlas, R.M. *Handbook of Microbiological Media* (CRC Press, Boca Raton, FL, 2004).
4. Barrick, J.E. & Lenski, R.E. Genome-wide mutational diversity in an evolving population of *Escherichia coli*. *Cold Spring Harbor Symposia on Quantitative Biology* **74**, 119-129 (2009).
5. Barrick, J.E., Yu, D.S., Yoon, S.H., Jeong, H., Oh, T.K., Schneider, D., Lenski, R.E. & Kim, J.F. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* **461**, 1243–1247 (2009).
6. Barrick, J.E., Kauth, M.R., Strelisoff, C.C. & Lenski, R.E. *Escherichia coli rpoB* mutants have increased evolvability in proportion to their fitness defects. *Molecular and Biological Evolution* **27**, 1338–1347 (2010).
7. Beatty, J. The evolutionary contingency thesis, In Wolters, G. Lennox, J.G. & McLaughlin, P. *Concepts, Theories, and Rationality in the Biological Sciences* (University of Pittsburgh Press, Pittsburgh, PA, 1995).
8. Beatty, J. Chance variation: Darwin on orchids. *Philosophy of Science*, **73**, 629–641 (2006a).
9. Beatty, J. Replaying life's tape. *The Journal of Philosophy* **103**, 336–362 (2006b).
10. Beatty, J. Chance variation and evolutionary contingency: Darwin, Simpson, *The Simpsons*, and Gould. In Ruse, M. & Werkmeister L.T. (Eds). *The Oxford Handbook of Philosophy of Biology* (Oxford University Press, New York, NY, 2008).
11. Beatty, J. & Desjardins, E.C. Natural selection and history. *Biology and Philosophy* **24**, 231–246 (2009).
12. Ben-Menahem, Y. Historical contingency. *Ratio* **10**, 99–107 (1997).
13. Bennett, A.F. & Lenski, R.E. Evolutionary adaptation to temperature. II. Thermal niches of experimental lines of *Escherichia coli*. *Evolution* **47**, 1–12 (1993).
14. Bentley, D.R. Whole-genome resequencing. *Current Opinion in Genetics and Development* **16**, 545–552 (2006).

15. Blaser, K. The history of nature and the nature of history: Stephen Jay Gould on science, philosophy, and history. *The History Teacher*, **32**, 411–430 (1999).
16. Bloch, M. *The Historian's Craft*. (Vintage Books, New York, NY, 1953).
17. Bohannan, B.J.M. & Lenski, R.E. The relative importance of competition and predation varies with productivity in a model community. *American Naturalist* **156**, 329-340 (2000).
18. Bowler, P.J. *Evolution: The History of an Idea*. (University of California Press, Berkeley and Los Angeles, CA, 2003).
19. Bradbury, R. A sound of thunder. (1952). Reprinted in: *R is for Rocket* (Doubleday and Company, New York, NY, 1962).
20. Bridgham, J.T., Ortlund, E.A. & Thornton, J.W. An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* **461**, 515–519 (2009).
21. Bryant, J.M. Evidence and explanation in history and sociology: critical reflections on Goldthorpe's critique of historical sociology. *British Journal of Sociology* **65**, 3-19 (1994).
22. Bulhof, J. What if? Modality and history. *History and Theory* **38**, 145–168 (1999).
23. Burch, C. L. & Chao, L. Evolution by small steps and rugged landscapes in the RNA virus ϕ -6. *Genetics* **151**, 921–927 (2000a).
24. Burch, C.L. & Chao, L. Evolvability of an RNA virus is determined by its mutational neighborhood. *Nature* **406**, 625–628 (2000).
25. Carlton, B.C., and Brown, B.J. Gene mutation. In (Ed.) P. Gerhardt, *Manual of Methods for General Bacteriology*. (American Society for Microbiology, Washington, DC, 1981).
26. Cartwright, N.D. Counterfactuals in economics: a commentary. In Campbell, J.K, O'Rourke, Silverstein, H. (Eds). *Causation and Explanation* (MIT Press, Cambridge, MA, 2007).
27. Chao, L. & Cox, E.C. Competition between high and low mutation strains of *Escherichia coli*. *Evolution* **37**, 125–134 (1983).
28. Cohen, S.S. and Anderson, T.F. Chemical studies on host-virus interactions I: The effect of bacteriophage adsorption on the multiplication of its host, *Escherichia coli* B. *Journal of Experimental Medicine* **84**, 511–523 (1946).
29. Colosimo, P.F., Hosemann, K.E., Balabhadra, S., Villarrea, G. Jr., Dickson, M., Grimwood, J., Schmutz, J., Myers, R.M., Schluter, D. & Kinglsey, D.M. Widespread

- parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* **5717**, 1928–1933 (2005).
30. Conway Morris, S. *The Crucible of Creation: The Burgess Shale and the Rise of Animals* (Oxford University Press, Oxford, UK 1998).
 31. Conway Morris, S. *Life's Solution: Inevitable Humans in a Lonely Universe* (Cambridge University Press, Cambridge, UK, 2003).
 32. Conway Morris, S. Evolution: like any other science it is predictable. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **365**, 133–145 (2010).
 33. Cooper, V.S. & Lenski, R.E. The population genetics of ecological specialization in evolving *E. coli* populations. *Nature* **407**, 736–739 (2000).
 34. Cooper, V., Schneider, S., Blot, M. & Lenski, R. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *E. coli* B. *Journal of Bacteriology* **183**, 2834–2841 (2001).
 35. Cooper, T.F., Rozen, D.E. & Lenski, R.E. Parallel changes in gene expression after 20,000 generations of evolution in *E. coli*. *Proceedings of the National Academy of Sciences (USA)* **100**, 1072–1077 (2003).
 36. Cooper, V.S., Remold, S.K., Lenski, R.E., & Schneider, D. Expression profiles reveal parallel evolution of epistatic interactions involving the CRP regulon in *Escherichia coli*. *PLoS Genetics* **4**, e35 (2008).
 37. Coricelli, G. & Rustichini, A. Counterfactual thinking and emotions: regret and envy learning. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **365**, 241–247 (2010).
 38. Cowan, R. & Foray, D. Evolutionary economics and the counterfactual threat: on the nature and role of counterfactual history as an empirical tool in economics. *Journal of Evolutionary Economics* **12**, 539–562 (2002).
 39. Cox, G., Gibson, F., Luke, R., Newton, N., O'Brien, I., and Rosenberg, H. Mutations affecting iron transport in *Escherichia coli*. *Journal of Bacteriology* **104**, 219–226 (1970).
 40. Crowley, R. & Ambrose, S. (Eds). *What If? The world's foremost military historians imagine what might have been* (Berkeley Publishing Group, New York, NY 2000).
 41. Crozat, E, Philippe, N., Lenski, R.E., Geiselmann, J. & Schneider, D. Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. *Genetics* **169**, 523–532 (2005).

42. Crozat, E., Winkworth, C., Gaffe, J., Hallin, P.F. Riley, M.A, Lenski, R.E. & Schneider, D. Parallel genetic and phenotypic evolution of DNA superhelicity in experimental populations of *Escherichia coli*. *Molecular Biology and Evolution* **27**, 2113–2128 (2010).
43. Crozat, E., Hindre, T., Kuhn, L., Garin, J., Lenski, R.E. & Schneider, D. Altered regulation of the OmpF porin by Fis in *Escherichia coli* during an evolution experiment between B and K12 strains. *Journal of Bacteriology* **193**, 429–440 (2011).
44. Darwin, C.R. *On the Origin of Species* (John Murray, London, UK, 1859).
45. Davis, B.D. The isolation of biochemically deficient mutants of bacteria by means of penicillin. *Proceedings of the National Academy of Sciences (USA)* **35**, 1–10 (1949).
46. Davis, B.D. & Mingioli, E.S. Mutants of *Escherichia coli* requiring methionine or vitamin B12. *Journal of Bacteriology* **60**, 17–28 (1950).
47. Dawkins, R. *The Blind Watchmaker, Second Edition* (Norton, New York, NY, 1996a).
48. Dawkins, R. *Climbing Mount Improbable* (Norton, New York, NY, 1996b).
49. Day, M. *Philosophy of History: An Introduction* (Continuum, London, UK, 2008).
50. de Queiroz, A. & Rodriguez-Robles, J.A. Historical contingency and animal diets: The origins of egg eating in snakes. *The American Naturalist* **167**, 684–694 (2006).
51. De Visser, J.A.G.M., Zeyl, C.W., Gerrish, P.J., Blanchard, J.L. & Lenski, R.E. Diminishing returns from mutation supply rate in asexual populations. *Science* **283**, 404–406 (1999).
52. De Visser, J.A.G.M. & Rozen, D.E. Clonal interference and the periodic selection of new beneficial mutations in *Escherichia coli*. *Genetics* **172**, 2093–2100 (2006).
53. Dennett, D. *Darwin's Dangerous Idea* (Simon and Schuster, New York, NY, 1995).
54. Desjardins, E. Historicity and experimental evolution. *Biology and Philosophy, In press*.
55. Dick, M.H., Lidgard, S., Gordon, D.P. & Mawatari S.F. The origin of ascophoran bryozoans was historical contingent but likely. *Proceedings of the Royal Society B: Biological Sciences* **276**, 3141–3148 (2009).
56. Dick, P.K. *The Man in the High Castle* (Putnam, New York, NY, 1962).
57. Dreyfuss, R. & Turtledove, H.N. *The Two Georges* (Tom Doherty Associates, New York, NY, 1997).

58. Elena, S.F., Sanjuan, R., Borderia, A.V. & Turner, P.E. Transmission bottlenecks and the evolution of fitness in rapidly evolving RNA viruses. *Genetics and Evolution* **1**, 41-48 (2001).
59. Elena, S.F. & Lenski, R.E. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Reviews Genetics* **4**, 457-469 (2003).
60. Emerson, S.B. A macroevolutionary study of historical contingency in the fanged frogs of Southeast Asia. *Biological Journal of the Linnean Society* **73**, 139–151 (2001).
61. Enslin, W. The emperor and the imperial administration. In Baynes, N.H. and Moss, H.St.L.B. *Byzantium: An introduction to East Roman civilization*. (Clarendon Press, Oxford, UK, 1961).
62. Ferguson, N. Introduction. In Ferguson, N. (ed). *Virtual History: Alternatives and Counterfactuals* (Basic Books, New York, NY, 2000).
63. Ferguson, N. *The War of the World* (Penguin, New York, NY, 2007).
64. Finkel, S.E. Long-term survival during stationary phase: evolution and the GASP phenotype. *Nature Reviews Microbiology* **4**, 113–120 (2006).
65. Fischer, D.H. Editor's note In McPherson, J.M. *Crossroads of Freedom: Antietam* (Oxford University Press, New York, NY, 2002).
66. Flood, C.B. *Grant and Sherman* (Harper Perennial, New York, NY, 2006).
67. Fogel, R.W. *Railroads and American Economic Growth: Essays in Econometric History* (The Johns Hopkins University Press, Baltimore, MD, 1964).
68. Foote, M. Contingency and convergence. *Science* **280**, 2068–2069 (1998).
69. Frost, G., and Rosenberg, H. The inducible citrate-dependent iron transport system in *Escherichia coli* K-12. *Biochimical et Biophysica Acta* **330**, 90–101 (1973).
70. Fukami, T., Beaumont, H.J.E., Zhang, X-. X. & Rainey, P.B. Immigration history controls diversification in experimental adaptive radiation. *Nature* **446**, 436–439 (2007).
71. Gerrish, P.J. & Lenski, R.E. The fate of competing beneficial mutations in an asexual population. *Genetica* **102/103**, 127-144 (1998).
72. Goldthorpe, J.H. The uses of history in sociology: reflections on some recent tendencies. *British Journal of Sociology* **42**, 211 – 30 (1991).
73. Gould, S.J. *Hen's Teeth and Horse's Toes*. (Norton, New York, NY 1983).

74. Gould, S.J. The paradox of the first tier: an agenda for paleobiology. *Paleobiology* **11**, 2–12 (1985).
75. Gould, S.J. *Wonderful Life* (Norton, New York, NY, 1989).
76. Gould, S.J. *The Individual in Darwin's World: The Second Edinburgh Medal Address* (Edinburgh University Press, Edinburgh, UK, 1990).
77. Gould, S.J. *Eight Little Piggies* (W.W. Norton & Company, Inc., New York, NY, 1993).
78. Gould, S.J. *Full House* (Random House, New York, NY, 1996).
79. Gould, S.J. *The Structure of Evolutionary Theory* (The Belknap Press of Harvard University Press, Cambridge, MA, 2002).
80. Gould, S.J., Raup, D.M., Sepkoski, J.J. Jr., Schopf, T.J.M. & Simberloff, D.S. The shape of evolution: A comparison of real and random clades. *Paleobiology* **3**, 23–40 (1977).
81. Gould, S.J. & Lewontin, R.C. The spandrel of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London B: Biological Sciences* **205**, 581–598 (1979).
82. Gould, S.J. & Woodruff, D.S. History as a cause and area effects: An illustration from Cerion on Great Inagua, Bahamas. *British Journal of the Linnean Society*, **40**, 67–98 (1990).
83. Griffiths, P.E. The historical turn in the study of adaptation. *British Journal of the Philosophy of Science* **47**, 511–532 (1996).
84. Hall, B. 1982. Chromosomal mutation for citrate utilization by *Escherichia coli* K-12. *Journal of Bacteriology* **151**, 269–273 (1982).
85. Harding, D.J. Counterfactual models of neighborhood effects: the effect of neighborhood poverty on dropping out and teenage pregnancy. *American Journal of Sociology* **109**, 676 – 719 (2003).
86. Härle, c., Kim, I., Angerer, A., and Braun, V. Signal transfer through three compartments: transcription initiation of the *Escherichia coli* ferric citrate transport system from the cell surface. *The EMBO Journal* **14**, 1430–1438 (1995).
87. Harmon, L.J., Harmon, L.L. & Jones, C.G. Competition and community structure in diurnal arboreal geckos (genus *Phelsuma*). *Systematic Biology* **57**, 562–573 (2007).
88. Harris, R. *Fatherland* (Random House, New York, NY, 1992).

89. Hartmann, A. and Braun, V. Iron uptake and iron limited growth of *Escherichia coli* K-12. *Archives of Microbiology* **130**, 353–356 (1981).
90. Hegreness, M. & Kishony, R. Analysis of genetic systems using experimental evolution and whole-genome sequencing. *Genome Biology* **8**, 201 (2007).
91. Helling, R.B., Vargas, C.N. & Adams, J. Evolution of *Escherichia coli* during growth in a constant environment. *Genetics* **116**, 349–358 (1987).
92. Herring, C., Glasner, J., and Blatner, F. Gene replacement without selection: regulated suppression of amber mutations in *Escherichia coli*. *Gene* **331**, 151–163 (2003).
93. Hussein, S., Hantke, K., and Braun, V. Citrate-dependent iron transport system in *Escherichia coli* K-12. *European Journal of Biochemistry* **117**, 431–437 (1981).
94. Huxley, J.S. *Evolution: The Modern Synthesis* (Allen and Unwin, London, UK, 1942).
95. Huynen, M.A., Stadler, P.F. & Fontana, W.B. Smoothness within ruggedness: The role of neutrality in adaptation. *Proceedings of the National Academy of Sciences (USA)* **93**, 397–401 (1996).
96. Ishiguro, N., Oka, C. & Sato, G. Isolation of citrate positive variants of *Escherichia coli* from domestic pigeons, pigs, cattle, and horses. *Applied and Environmental Microbiology* **36**, 217–222 (1978).
97. Ishiguro, N., Oka, C., Hanazawa, Y. & Sato, G. Plasmids in *Escherichia coli* controlling citrate-utilizing ability. *Applied and Environmental Microbiology* **28**, 956–964 (1979).
98. Jacob, F. Evolution and tinkering. *Science* **196**, 1161–1166 (1977).
99. Jacob, F. *The Possible and the Actual* (University of Washington Press, Seattle, WA, 1982).
100. Jeong, H., Barbe, V., Lee, C.H., Vallenet, D., Yu, D.S., Choi, S.H., Couloux, A., Lee, S.W., Yoon, S.H., Cattolico, L., Hur, C.G, Park, H.S, Ségurens, B., Kim, S.C., Oh, T.K., Lenski, R.E, Studier, F.W., Daegelen, P. & Kim, J.F. Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *Journal of Molecular Biology* **394**, 644–653 (2009).
101. Kahneman, D. & Miller, D. Norm theory: Comparing reality to its alternatives. *Psychological Review* **93**, 136–153 (1986).
102. Kahneman, D. & Tversky, A. The simulation heuristic. In Kahneman, D., Slovic, P, & Tversky, A. (Eds). *Judgement under uncertainty: heuristics and biases*. (Cambridge University Press, New York, NY, 1982).

103. Keegan, J. *The First World War*. (Vintage, New York, NY, 2000).
104. Kimura, M. Evolutionary rate at the molecular level. *Nature* **217**, 624–626 (1968).
105. Knapp, A.K., Burns, C.E., Fynn, R.W.S., Kirkman, K.P., Morris, C.D. & Smith, M.D. Convergence and contingency in production-precipitation relationships in North American and South African C₄ grasslands. *Oecologia* **149**, 456–464 (2006).
106. Koser, S.A. Utilization of the salts of organic acids by colon-aerogenes group. *Journal of Bacteriology* **8**, 493–520 (1923).
107. Koser, S.A. Correlation of citrate-utilization by members of the colon-aerogenes group with other differential characteristic and with habitat. *Journal of Bacteriology* **9**, 59–77 (1924).
108. Lamprecht, S. Contingency in nature. *Philosophy and Phenomenological Research* **32**, 1–14 (1971).
109. Langman, L., Young, I.G., Frost, G.E., Rosenberg, H, and Gibson, F. Enterochelin system of iron transport in *Escherichia coli*: Mutations affecting ferric-enterochelin esterase. *Journal of Bacteriology* **112**, 1142–1149 (1972).
110. Lara, F. Stokes, J. Oxidation of citrate by *Escherichia coli*. *Journal of Bacteriology* **9**, 59–77 (1952).
111. Larson, E.J. *Evolution* (Modern Library, New York, NY, 2004).
112. Lenski, R. Phenotypic and genomic evolution during a 20,000-generation experiment with the bacterium *Escherichia coli*. *Plant Breeding Reviews* **24**, 225–265 (2004).
113. Lenski, R.E., Rose, M.R., Simpson, S.C. & Tadler, S.C. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *The American Naturalist* **138**, 1315–1341 (1991).
114. Lenski, R. & Travisano, M. Dynamics of adaptation and diversification: 10,000-generation experiment with bacterial populations. *Proceedings of the National Academy of Sciences (USA)* **91**, 6808–6814 (1994).
115. Lenski, R.E., Ofria, C., Pennock, R.T. & Adami, C. The evolutionary origin of complex features. *Nature* **423**, 139–144 (2003a).
116. Lenski, R.E., Winkworth, C.L. & Riley, M.A. Rates of DNA sequence evolution in experimental populations of *Escherichia coli* during 20,000 generations. *Journal of Molecular Evolution* **56**, 498–508 (2003b).

117. Lenski, R.E., Barrick, J.E. & Ofria, C. Balancing robustness and evolvability. *PLoS Biology* **4**, e428 (2006).
118. Levin, B.R. & Lenski, R.E. Bacteria and phage: a model system for the study of the ecology and co-evolution of hosts and parasites. In Rollinson, D. & Anderson, R.M. (Eds). *Ecology and the Genetics of Host-Parasite Interactions*. (Academic Press, London, UK, 1985).
119. Lewis, D.K. *Counterfactuals*. (Blackwell, Oxford, UK, 2001).
120. Lewontin, R.C. Is nature probably or capricious? *BioScience* **16**, 25–27 (1966).
121. Livius, T., Radice, B (Translator). *Rome and Italy: Books VI-X of The History of Rome from its Foundation* (Penguin Putnam Inc, New York, NY, 1982).
122. Lochhead, R. & Menninger, R. About the Department of Public Health and Preventative Medicine. *Cornell Medical Journal* **4**, 25–27 (1949).
123. Losos, J.B., Jackman, T.R., Larson, A., de Queiroz, K. & Rodriguez-Schettino, L. Contingency and determinism in replicated adaptive radiations of island lizards. *Science* **279**, 2115–2118 (1998).
124. Losos, J.B. Adaptive radiation, ecological opportunity, and evolutionary determinism. *The American Naturalist* **175**, 623–639 (2010).
125. Lutgens, M., and Gottschalk, G. Why a co-substrate is required for anaerobic growth of *Escherichia coli* on citrate. *Journal of General Microbiology* **199**, 63–70 (1980).
126. Maas, W.K. Bernard David Davis: 1916-1994. In *Biographical Memoirs, Volume 77* (National Academy Press, Washington, D.C., 1999).
127. Mandel, D.R., Hilton, D.J. & Catellani, P. (Eds). *The Psychology of Counterfactual Thinking*. (Routledge, New York, NY, 2005).
128. Mani, G.S. & Clarke, B.C. Mutational order: A major stochastic process in evolution. *Proceedings of the Royal Society of London B: Biological Sciences* **240**, 29–37 (1990).
129. Maynard Smith, J. Taking a chance on evolution. *New York Review of Books* May 14, 234–236 (1992).
130. Maynard Smith, J. & Szathmary, E. *The Major Transitions in Evolution* (Oxford University Press, Oxford, UK, 1995).
131. Mayr, E. The emergence of evolutionary novelties. In (Ed) Tax, S. *Evolution after Darwin* (University of Chicago Press, Chicago, 1960).

132. Mayr, E. Some thoughts on the history of the evolutionary synthesis. In (Ed.) Mayr, E. & Provine, W.B. *The Evolutionary Synthesis* (Harvard University Press, Cambridge, MA, 1980).
133. Mayr, E. *Toward a New Philosophy of Biology* (Harvard University Press, Cambridge, MA, 1988).
134. McPhail, J.D. Ecology and evolution of sympatric sticklebacks (*Gasterosteus*): origins of the species pairs. *Canadian Journal of Zoology* **71**, 515–523 (1993).
135. McPhail, J.D. Speciation and the evolution of reproductive isolation in the sticklebacks (*Gasterosteus*) of south-western British Columbia. In (Eds.) Bell, M.A. & Foster, S.A., *The Evolutionary Biology of the Threespine Stickleback* (Oxford Science Publications, Oxford, UK, 1994).
136. McPherson, J.M. *Battle Cry of Freedom: The Civil War Era* (Ballantine, New York, NY, 1988).
137. McPherson, J.M. *Crossroads of Freedom: Antietam* (Oxford University Press, New York, NY, 2002).
138. McRae, M.W. Stephen Jay Gould and the contingent nature of history. *Clio* **22**, 239–250 (1993).
139. Meyer, J.R. & Kassen, R. The effects of competition and predation on diversification in a model adaptive radiation. *Nature* **446**, 432–435 (2007).
140. Meyer, J.R., Agrawal, A.A., Quick, R.T., Dobias, D.T., Schneider, D. & Lenski, R.E. Parallel changes in host resistance to viral infection during 45,000 generations of relaxed selection. *Evolution* **64**, 3024–3034 (2010).
141. Moore, W. *Bring the Jubilee* (Random House, New York, NY, 1953).
142. Morgan, S.L. *Counterfactuals and Causal Inference: Methods and Principles for Social Research* (Cambridge University Press, New York, NY, 2007).
143. Morris, M.W. & Moore, P.C. The lessons we (don't) learn: counterfactual thinking and organizational accountability after a close call. *Administrative Science Quarterly* **45**, 737–765 (2000).
144. Muller, H.J. Reversibility in evolution considered from the standpoint of genetics. *Biological Reviews of the Cambridge Philosophical Society* **14**, 261–280 (1939).
145. Newell, N.D. Revolutions in the history of life. *Geological Society of America Special Paper* **89**, 62–91 (1967).

146. Nosil, P. & Flaxman, S.M. Conditions for mutation-order speciation. *Proceedings of the Royal Society of London B: Biological Sciences* **278**, 399–407 (2011).
147. Ortlund, E.A., Bridgham, J.T., Redinbo, MR. & Thornton, J.W. Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* **317**, 1544–1548 (2007).
148. Ostrowski, E.A., Rozen, D.E. & Lenski, R.E. Pleiotropic effects of beneficial mutations in *Escherichia coli*. *Evolution* **59**, 2343–2352 (2005).
149. Ostrowski, E.A., Woods, R.J. & Lenski, R.E. The genetic basis of parallel and divergent phenotypic responses in evolving populations of *Escherichia coli*. *Proceedings of the Royal Society of London B: Biological Sciences* **275**, 277–284 (2008).
150. Papadopoulos, D., Schneider, D., Meier-Eiss, J., Arber, W., Lenski, R.E. & Blot, M. Genomic evolution during a 10,000-generation experiment with bacteria. *Proceedings of the National Academy of Sciences (USA)* **96**, 3807–3812 (1999).
151. Pelosi, L., Kuhn, L., Guetta, D., Garin, J., Geiselmann, J., Lenski, R.E. & Schneider, D. (2006). Parallel changes in global protein profiles during long-term experimental evolution in *Escherichia coli*. *Genetics* **173**, 1851–1869 (2006).
152. Perfeito, L., Fernandes, L., Mota, C., & Gordo, I. Adaptive mutations in bacteria: high rate and small effects. *Science* **317**, 813–815 (2007).
153. Pfeiffer, T. & Bonhoeffer, S. Evolution of cross-feeding in microbial populations. *The American Naturalist* **163**, E126–135 (2004).
154. Philippe, N., Crozat, E., Lenski, R.E. & Schneider, D. (2007). Evolution of global regulatory networks during a long-term experiment with *Escherichia coli*. *BioEssays* **29**, 846–860 (2007).
155. Philippe, N., Pelosi, L., Lenski, R.E. & Schneider, D. Evolution of penicillin-binding protein 2 concentration and cell shape during a long-term experiment with *Escherichia coli*. *Journal of Bacteriology* **191**, 909–921 (2009).
156. Pigliucci, M. What, if anything, is an evolutionary novelty? *Philosophy of Science* **75**, 887–898 (2008).
157. Pigliucci, M. An extended synthesis for evolutionary biology. *Annals of the New York Academy of Sciences* **1168**, 218–228 (2009).
158. Pos, K., Dimroth, P., and Bott, M. 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 (1998).

- 159.Provine, W.B. *The Origins of Theoretical Population Genetics* (University of Chicago Press, Chicago, IL, 1971).
- 160.Raup, D.M, Gould, S.J., Schopf, T.J.M, & Simberloff, D.S. Stochastic models of phylogeny and the evolution of diversity. *Journal of Geology* **81**, 525–542 (1973).
- 161.Raup, D.M. *Extinction: Bad Genes or Bad Luck* (Norton, New York, NY, 1991).
- 162.Reisch, G. Chaos, history, and narrative. *History and Theory* **30**, 1–20 (1991).
- 163.Reynolds, C., and Silver, S. Citrate utilization by *Escherichia coli*: plasmid-and chromosome-encoded systems. *Journal of Bacteriology* **156**, 1019–1024 (1983).
- 164.Robinson, K.S. *The Years of Rice and Salt* (Bantam, New York, NY, 2002).
- 165.Rosenfeld, B. Why do we ask “what if?”: Reflections on the function of alternate history. *History and Theory* **41**, 90–103 (2002).
- 166.Rozen, D.E. & Lenski, R.E. Long-term experimental evolution in *Escherichia coli*. VIII. Dynamics of a balanced polymorphism. *The American Naturalist* **155**, 24–35 (2000).
- 167.Rozen, D.E., Schneider, D. & Lenski, R.E. Long-term experimental evolution in *Escherichia coli*. XIII. Phylogenetic history of a balanced polymorphism. *Journal of Molecular Evolution* **61**, 171–180 (2005).
- 168.Rozen, D.E., Philippe, N., de Visser, J.A. & Schneider, D. Death and cannibalism in a seasonal environment facilitate bacterial coexistence. *Ecology Letters* **12**, 34–44 (2009).
- 169.Ruelle, D. *Chance and Chaos*. (Princeton Science Library, Princeton, NJ, 1991).
- 170.Sahyun, M., Beard, P, Schultz, E.W., Snow, J., and Cross, E. Growth stimulating factors for microorganisms. *The Journal of Infectious Diseases* **58**, 28–44.
- 171.Salverda, M.L.M., Dellus, E., Gorter, F.A., Debets, A.J.M., van der Oost, J., Hoekstra, R.F., Tawfik, D.S. & de Visser, J.A.G.M. Initial mutations direct alternative pathways of protein evolution. *PloS Genetics* **7**, e1001321 (2011).
- 172.Scheutz, F., and Strockbine, N.A. Genus I. *Escherichia*, Castellani and Chalmers 1919. In (Ed.) Garrity, G.M., Brenner, D., Kreig, N.R., Staley, J.R. *Bergey’s Manual of Systematic Bacteriology, Volume Two: The Proteobacteria* (Springer, New York, NY, 2005).
- 173.Schluter, D. Ecological speciation in postglacial fishes. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **351**, 807–814 (1996).

174. Schluter, D., Marchinko, K.B., Barrett, R.D.H., & Rogers, S.M. Natural selection and the genetics of adaptation in threespine stickleback. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **365**, 2479–2486 (2010).
175. Seehausen, O. Chance, historical contingency, and ecological determinism jointly determine the rate of adaptive radiation. *Heredity* **99**, 361–363 (2007).
176. Shubin, N., Tabin, C. & Carroll, S. Fossils, genes, and the evolution of animal limbs. *Nature* **388**, 639–648 (1997).
177. Shubin, N., Tabin, C. & Carroll, S. Deep homology and the origins of evolutionary novelty. *Nature* **457**, 818–823 (2009).
178. Silverberg, R. *Roma Eterna* (Roc, New York, NY, 2003).
179. Sleight, S.C., Wigginton, N.S. & Lenski, R.E. Increased susceptibility to repeated freeze-thaw cycles in *Escherichia coli* following long-term evolution in a benign environment. *BMC Evolutionary Biology* **6**, 104–108 (2006).
180. Sniegowski, P., Gerrish, P. & Lenski, R. Evolution of high mutation rates: separating causes from consequences. *BioEssays* **22**, 1057–1066 (1997).
181. Sobel, M.E. Causal inference in the social and behavioral sciences. In Arminger, G., Clogg, C.C. & Sobel, M.E. (Eds). *Handbook of Statistical Modeling for the Social and Behavioral Sciences* (Plenum Press, New York, 1995).
182. Sobel, R. *For Want of a Nail* (MacMillan, New York, NY, 1973).
183. Sole, R.V., Montoya, J.M. & Erwin, D.H. Recovery after mass extinction: evolutionary assembly in large-scale biosphere dynamics. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **357**, 697–707 (2002).
184. Squire, J.C. (Ed) *If It Had Happened Otherwise: Lapses into Imaginary History* (Longmans, Green, London, UK, 1931).
185. Stanek, M.T., Cooper, T.F. & Lenski, R.E. Identification and dynamics of a beneficial mutation in a long-term evolution experiment with *Escherichia coli*. *BMC Evolutionary Biology* **9**, 302–315 (2009).
186. Sterelny, K., & Griffiths, P. *Sex and Death: An Introduction to the Philosophy of Biology* (University of Chicago Press, Chicago, IL, 1999).
187. Studier, F.W., Daegelen, P., Lenski, R.E., Maslov, S. & Kim, J.F. Understanding the differences between genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3) and comparison of the *E. coli* B and K-12 genomes. *Journal of Molecular Biology* **394**, 644–652 (2009).

188. Taylor, E.B. & McPhail, J.D. Historical contingency and ecological determinism interact to prime speciation in sticklebacks, *Gasterosteus*. *Proceedings of the Royal Society of London, Series B: Biological Sciences* **267**, 2375–2384 (2000)
189. Tetlock, P.E. & Belkin, Aaron (Eds.) *Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives*. (Princeton University Press, Princeton, NJ, 1996).
190. Thompson, E.P. *The Poverty of Theory and Other Essays* (Merlin, London, UK, 1978).
191. Thucydides, Warner, R. (Translator), Finley, M.I. (Ed). *The History of the Peloponnesian War: Revised Edition* (Penguin Putnam Inc, New York, NY, 1972).
192. Travisano, M., Mongold, J.A, Bennett, A.F. & Lenski R.E. Experimental tests of the roles of adaptation, chance, and history in evolution. *Science* **267**, 87–90 (1995a).
193. Travisano, M., Vasi, R. & Lenski, R. Long-term experimental evolution in *Escherichia coli*. III. Variation among replicate populations in correlated responses to novel environments. *Evolution* **49**, 189–200 (1995b).
194. Tucker, A. Historiographical counterfactuals and historical contingency. *History and Theory* **38**, 264–278 (1999).
195. Turner, D.D. Gould’s replay revisited. *Biology and Philosophy* **26**, 65–79 (2011).
196. Turner, P.E., Souza, V. & Lenski, R.E. Tests of ecological mechanism promoting the stable coexistence of two bacterial genotypes. *Ecology* **77**, 2119–2129 (1996).
197. Turtledove, H.N. *How Few Remain* (Ballantine, New York, NY, 1997).
198. Turtledove, H.N. *The Great War: American Front* (Ballantine, New York, NY, 1998).
199. Turtledove, H.N. *The Great War: Walk in Hell* (Ballantine, New York, NY, 1999).
200. Turtledove, H.N. *The Great War: Breakthroughs* (Ballantine, New York, NY, 2000).
201. Turtledove, H.N. Introduction. In (Ed.) Turtledove, H.N & Greenberg, M.H. *The Best Alternative History Stories of the 20th Century* (Del Rey, New York, NY, 2001a).
202. Turtledove, H.N. *American Empire: Blood and Iron* (Ballantine, New York, NY, 2001b).
203. Turtledove, H.N. *American Empire: The Center Cannot Hold* (Ballantine, New York, NY, 2002).

204. Turtledove, H.N. *In the Presence of Mine Enemies* (NAL, New York, NY, 2003a).
205. Turtledove, H.N. *American Empire: The Victorious Opposition* (Ballantine, New York, NY, 2003b).
206. Turtledove, H.N. *Settling Accounts: Return Engagement* (Ballantine, New York, NY, 2004).
207. Turtledove, H.N. *Settling Accounts: Drive to the East* (Ballantine, New York, NY, 2005).
208. Turtledove, H.N. *Settling Accounts: The Grapple* (Ballantine, New York, NY, 2006).
209. Turtledove, H.N. *Settling Accounts: In at the Death* (Ballantine, New York, NY, 2007).
210. www.uchronia.net, accessed February 18, 2011.
211. Vasi, F., Travisano, M. & Lenski, R. 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.
212. Vermeij, G. J. Historical contingency and the purported uniqueness of evolutionary innovations. *Proceedings of the National Academy of Sciences (USA)* **103**, 1804–1809 (2006).
213. Vogel, H.J., in McElroy, W.D., and Glass, B. In (Ed) McElroy, W.D. *A Symposium on Amino Acid Metabolism* (The Johns Hopkins Press, Baltimore, MD, 1955).
214. Vogel, H.J. and Davis, B.D. Glutamic gamma-semialdehyde and delta-1-pyrroline-5-carboxylic acid, intermediates in the biosynthesis of proline. *Journal of the American Chemical Society*, **74**, 109–112 (1952).
215. Weinreich, D.M., Watson, R.A. & Chao, L. Sign epistasis and genetic constrain on evolutionary trajectories. *Evolution* **59**, 1165–1174 (2005).
216. Weinreich, D.M., Delaney, N.F., DePristo, M.A. & Hartle, D.L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).
217. Wilke, C.O. Adaptive evolution on neutral networks. *Bulletin of Mathematical Biology* **63**, 715–730 (2001).
218. Williamson, J. *The Legion of Time* (Pyramid, New York, NY, 1967).
219. Wilson, E.O. *The Diversity of Life* (Norton, New York, NY, 1993).

220. Woese, C. The universal ancestor. *Proceedings of the National Academy of Sciences (USA)* **95**, 6854–6859 (1995).
221. Woods, R., Schneider, D., Winkworth, C.L., Riley, M.A. & Lenski, R.E. Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proceedings of the National Academy of Sciences (USA)* **103**, 9107–9122 (2006).
222. Wright, S. The roles of mutation, inbreeding, crossbreeding, and selection in evolution, in Jone, D.F. (Ed.) *Proceedings of the VI International Congress of Genetics*, (1932).
223. Wright, S. *Evolution and the Genetics of Populations, Volume 3. Experimental Results and Evolutionary Deductions*. (University of Chicago Press, Chicago, IL, 1977)
224. Wright, S. Surfaces of selective value revisited. *The American Naturalist* **131**, 115–123 (1988).
225. Yedid, G. & Bell, G. Macroeolution simulated with autonomously replicating computer programs. *Nature* **420**, 810–812 (2002).

CHAPTER 2

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

HISTORICAL CONTINGENCY AND THE EVOLUTION OF A KEY INNOVATION IN AN EXPERIMENTAL POPULATION OF *ESCHERICHIA COLI*

ABSTRACT

The role of historical contingency in evolution has been much debated, but rarely tested. Twelve initially identical populations of *E. coli* were founded in 1988 to investigate this issue. They have since evolved in a glucose-limited medium that also contains citrate, which *E. coli* cannot use as a carbon source under oxic conditions. No population evolved the capacity to exploit citrate for >30,000 generations, although each population tested billions of mutations. A citrate-using (Cit^+) variant finally evolved in one population by 31,500 generations, causing an increase in population size and diversity. The long-delayed and unique evolution of this function might indicate the involvement of some extremely rare mutation. Alternately, it may involve an ordinary mutation, but one whose physical occurrence or phenotypic expression is contingent on prior mutations in that population. We tested these hypotheses in experiments that “replayed” evolution from different points in that population’s history. We observed no Cit^+ mutants among 8.4×10^{12} ancestral cells, nor among 9×10^{12} cells from 60 clones sampled in the first 15,000 generations. However, we observed a significantly greater tendency for later clones to evolve Cit^+ , indicating that some potentiating mutation arose by 20,000 generations. This potentiating change increased the mutation rate to Cit^+ , but did not cause generalized hypermutability. Thus, the evolution of this phenotype was contingent on the particular history of that population. More generally, we suggest that historical contingency is especially

important when it allows key innovations that are not easily evolved by gradual, cumulative selection.

BACKGROUND AND INTRODUCTION

At its core, evolution involves a profound tension between random and deterministic processes. Natural selection works systematically to adapt populations to their current environments. However, selection requires heritable variation generated by random mutation, and even beneficial mutations may be lost by random drift. Moreover, random and deterministic processes become intertwined over time such that future alternatives may be contingent on the prior history of an evolving population. For example, multiple beneficial mutations will arise in some unpredictable order (Mani and Clarke 1990, Lenski et al 1991), and those that are substituted first may differ from others in their pleiotropic effects and epistatic interactions (Cooper and Lenski 2000, Cooper et al 2008), thus constraining some evolutionary paths while potentiating other outcomes (Jacob 1977, Wright 1988, Yedid and Bell 2002, Weinreich et al 2005, Weinreich et al 2006). These accidents of history may even determine the survival or extinction of entire lineages, given the capricious and sudden nature of some environmental changes (Lewontin 1966, Gould 1985).

Stephen Jay Gould maintained that these historical contingencies make evolution largely unpredictable. Although each change on an evolutionary path has some causal relation to the circumstances in which it arose, outcomes must eventually depend on the details of long chains of antecedent states, small changes in which may have enormous long-term repercussions (Gould 1989, 2002, Beatty 2006). Thus, Gould argued that contingency renders evolution fundamentally quirky and unpredictable, and he famously suggested that replaying the “tape of life” from some point in the distant past would yield a living world far different from the one we see today. Simon Conway Morris countered that natural selection constrains organisms to a

relatively few highly adaptive options, so that “the evolutionary routes are many, but the destinations are limited” (Conway Morris 2003). He and others point to numerous examples of convergent evolution as evidence that selection finds the same adaptations despite the vagaries of history. Evolution may thus be broadly repeatable, and multiple replays would reveal striking similarities in important features, with contingency mostly confined to minor details (Conway Morris 2003, Van Valen 1991, Dawkins 1996, Bermeij 2006).

Of course, replaying life’s tape on the planetary scale is impossible, but careful experiments can examine the role of contingency in evolution on a more modest scale (Beatty 2006, Travisano et al 1995). To address the repeatability of evolutionary trajectories and outcomes, the long-term evolution experiment (LTEE) with *E. coli* was started in 1988 with the founding of 12 populations from the same clone (Lenski et al 1991). These populations were initially identical except for a neutral marker that distinguished six lines from six others. They have since been propagated by daily 1:100 serial transfer in DM25, a minimal medium containing 25 mg/L glucose as the limiting resource (Lenski et al 1991, Lenski 2004). Environmental conditions have been controlled, constant, and identical for all 12 lines. To date, each population has evolved for over 44,000 generations, and samples have been frozen every 500 generations, providing a rich “fossil record” (Lenski and Travisano 1994). Moreover, these samples remain viable, allowing us to perform simultaneous measurements and other experiments with bacteria from different generations. The founding strain is strictly asexual, so populations have evolved by natural selection and genetic drift acting on variation generated by spontaneous mutations that have occurred during the experiment. Thus, the LTEE allows us to examine the effects of contingency that are inherent to the core evolutionary processes of mutation, selection, and drift.

Previous analyses of this experiment have shown numerous examples of parallel phenotypic and genetic evolution. All twelve populations underwent rapid improvement in fitness that decelerated over time (Lenski et al 1991, Lenski and Travisano 1994, Cooper and Lenski 2000, Lenski 2004). All evolved higher maximum growth rates on glucose, shorter lag phases upon transfer into fresh medium, reduced peak population densities, and larger average cell sizes relative to their ancestor (Lenski 2004, Lenski and Travisano 1994, Vasi et al 1994, Lenski and Mongold 2000, Novak et al 2006). Ten populations evolved increased DNA supercoiling (Croizat et al 2005), and those populations examined to date show parallel changes in global gene-expression profiles (Cooper et al 2003, Cooper et al 2008, Pelosi et al 2006). At least three genes have substitutions in all 12 populations (Woods et al 2006, Cooper et al 2001), while several others have substitutions in many populations (Croizat et al 2005, Cooper et al 2003, Pelosi et al 2006, Woods et al 2006), even though most loci harbor no substitutions in any of them (Lenski et al 2003). At the same time, there has also been some divergence between populations. Four have evolved defects in DNA repair, causing mutator phenotypes (Cooper and Lenski 2000, Sniegowski et al 1997). There is subtle, but significant, between-population variation in mean fitness in the glucose-limited medium in which they evolved (Lenski et al 1991, Lenski and Travisano 1994). In media containing other carbon sources, such as maltose or lactose, the variation in performance is much greater (Travisano et al 1995). And while the same genes often harbor substitutions, the precise location and details of the mutations almost always differ between the populations (Croizat et al 2005, Cooper et al 2003, Pelosi et al 2006, Woods et al 2006, Cooper et al 2001).

Throughout the duration of the LTEE, there has existed an ecological opportunity in the form of an abundant, but unused, resource. DM25 medium contains not only glucose, but also

citrate at a high concentration. The inability to use citrate as an energy source under oxic conditions has long been a defining characteristic of *E. coli* as a species (Koser 1924, Scheutz and Strockbridge 2005). Nevertheless, *E. coli* is not wholly indifferent to citrate. It uses a ferric dicitrate transport system for iron acquisition, though citrate does not enter the cell in this process (Frost and Rosenberg 1973). It also has a complete tricarboxylic acid cycle, and can thus metabolize citrate internally during aerobic growth on other substrates (Lara and Stokes 1952). *E. coli* is able to ferment citrate under anoxic conditions if a co-substrate is available for reducing power (Lutgens and Gottschalk 1980). The only known barrier to aerobic growth on citrate is its inability to transport citrate under oxic conditions (Hall 1982, Reynolds and Silver 1983, Pos et al 1998). Indeed, atypical *E. coli* that grow aerobically on citrate (Cit^+) have been isolated from agricultural and clinical settings, and were found to harbor plasmids, presumably acquired from other species, that encode citrate transporters (Ischiguro et al 1978, Ishiguro et al 1979).

Other findings suggest that *E. coli* has the potential to evolve a Cit^+ phenotype. Hall (Hall 1982) reported the only documented case of a spontaneous Cit^+ mutant in *E. coli*. He hypothesized that some complex mutation, or multiple mutations, activated cryptic genes that jointly expressed a citrate transporter, although the genes were not identified. Pos et al. (Pos et al 1998) identified an operon in *E. coli* K-12 that apparently allows anaerobic citrate fermentation, and which includes a gene, *citT*, encoding a citrate-succinate antiporter. High-level constitutive expression of this gene on a multi-copy plasmid allows aerobic growth on citrate, but the native operon has a single copy that is presumably induced only under anoxic conditions.

Despite this potential, none of the 12 LTEE populations evolved the capacity to use the citrate that was present in their environment for over 30,000 generations. During that time, each population experienced billions of mutations (Lenski 2004), far more than the number of possible point mutations in the ≈ 4.6 -million bp genome. This ratio implies, to a first approximation, that each population tried every typical one-step mutation many times. It must be difficult, therefore, to evolve the Cit^+ phenotype, despite the ecological opportunity. Here we report that a Cit^+ variant finally evolved in one population by 31,500 generations, and its descendants later rose to numerical dominance. The new Cit^+ function has been the most profound adaptation observed during the LTEE, and has had major consequences. As we will show, the population achieved a several-fold increase in size. Moreover, a stable polymorphism emerged, with a Cit^- minority coexisting with the new Cit^+ majority. Interestingly, the population that evolved the Cit^+ function is not one that had previously become hypermutable. It is also intriguing that this key innovation evolved so late in the experiment, given that the rate of fitness improvement had declined substantially in all the populations (Cooper and Lenski 2000, Lenski and Travisano 1994).

The long-delayed and unique evolution of the Cit^+ phenotype might indicate that it required some unusually rare mutation, such as a particular chromosomal inversion, that does not scale with typical mutation rates. Alternately, the occurrence or phenotypic expression of the mutation that generated the Cit^+ function might depend on one or more earlier mutations, such that its evolution was contingent upon the particular history of that population. Contingent adaptations should tend to be complex and require multiple steps, some of which might not be

beneficial, at least not uniquely so given other advantageous paths. Otherwise, cumulative selection would predictably favor the same steps, and the evolutionary path should be repeatable (Dawkins 1996). Contingent adaptations should thus display two characteristics. First, independent origins should be rare, because the same historical sequences would rarely recur (Vermeij 2006). Second, significant time-lags should occur between the presentation of ecological opportunities or challenges and the evolution of those traits that confer adaptation to those circumstances (Foote 1998).

Can the hypothesis of contingent adaptation be rigorously tested in the case of the evolution of the Cit⁺ function? And can this hypothesis be formally distinguished from the alternative explanation that this new function required an unusually rare mutation, but one that was not historically contingent? The answer to both questions is yes, owing to certain features of the LTEE in particular, and of bacteria in general, that allow us quite literally to replay the tape of life from various points in a population's history. We will first describe the emergence of this new function, and then present our experiments to distinguish between the hypotheses of mutational rarity and historical contingency in the origin of this key innovation.

RESULTS

Evolution of Cit⁺ Function in Population Ara-3.

The LTEE populations are transferred daily into fresh medium, and the turbidity is checked visually for each one at that time. Owing to the low concentration of glucose in DM25 medium, the cultures are only slightly turbid when transferred. Occasional contaminants that grow on citrate have been seen over the 20 years of this experiment, and the contaminated cultures reach much higher turbidity owing to the high concentration of citrate that allows the

contaminating bacteria to grow to high density. (When contamination occurs, the affected population is restarted from the latest frozen sample.) After ~33,127 generations, one population, designated Ara-3, had significantly elevated turbidity, which continued to rise for several days (Figure 2.1). A number of Cit⁺ clones were isolated from the population and checked for phenotypic markers characteristic of the ancestral *E. coli* strain used to start the LTEE: all were Ara⁻, T5-sensitive, and T6-resistant, as expected (Lenski et al 1991). DNA sequencing also shows that Cit⁺ clones have the same mutations in the *pykF* and *nadR* genes as do clones from earlier generations of the Ara-3 population, and each of these mutations distinguishes this population from all the others (Woods et al 2006). Therefore, the Cit⁺ variant arose within the LTEE and is not a contaminant.

The evolved Cit⁺ variant grows to high density in DM0 (a citrate-only medium), produces vigorous colonies on minimal citrate (MC) agar plates, and causes a positive color change on Simmon's citrate agar, all of which indicate that it can use citrate as a sole carbon source. In DM25, Cit⁺ cells undergo a period of rapid growth on glucose that is followed by slower growth on citrate (Figure 2.2). Also, growth on citrate is inhibited by the citrate analog 5-fluorocitrate (data not shown), as was observed for the one previously reported Cit⁺ mutant of *E. coli* (Reynolds and Silver 1983).

Cit⁺ clones could be readily isolated from the frozen sample of population Ara-3 taken at generation 33,000. To estimate the time of origin of the Cit⁺ trait, we screened 1280 clones randomly chosen from generations 30,000, 30,500, 31,000, 31,500, 32,000, 32,500 and 33,000

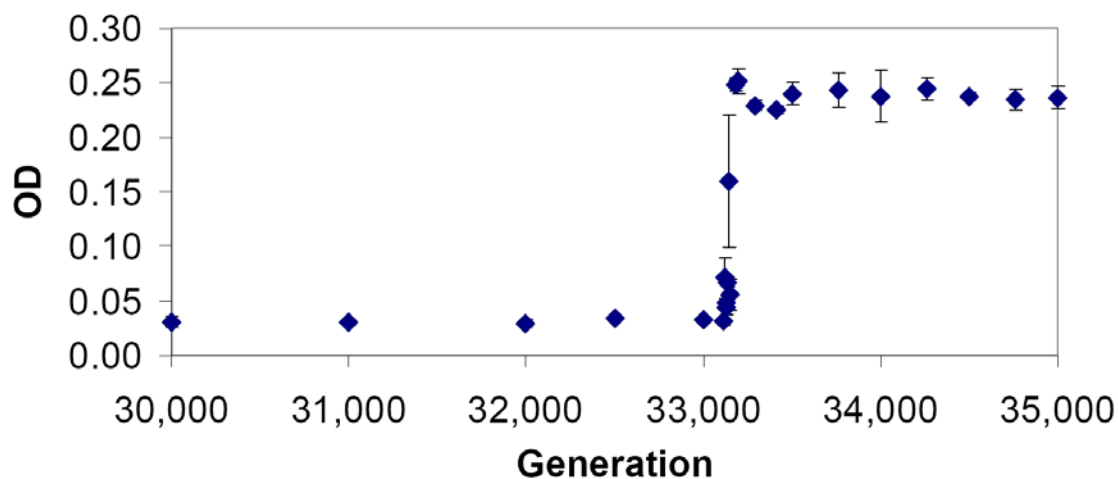


Figure 2.1. Population expansion during evolution of the Cit⁺ phenotype.

Samples frozen at various times in the history of population Ara-3 were revived, and three DM25 cultures established for each generation. Optical density (OD) was measured for each culture at 24 h. Error bars show the range of three values measured for each generation. For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

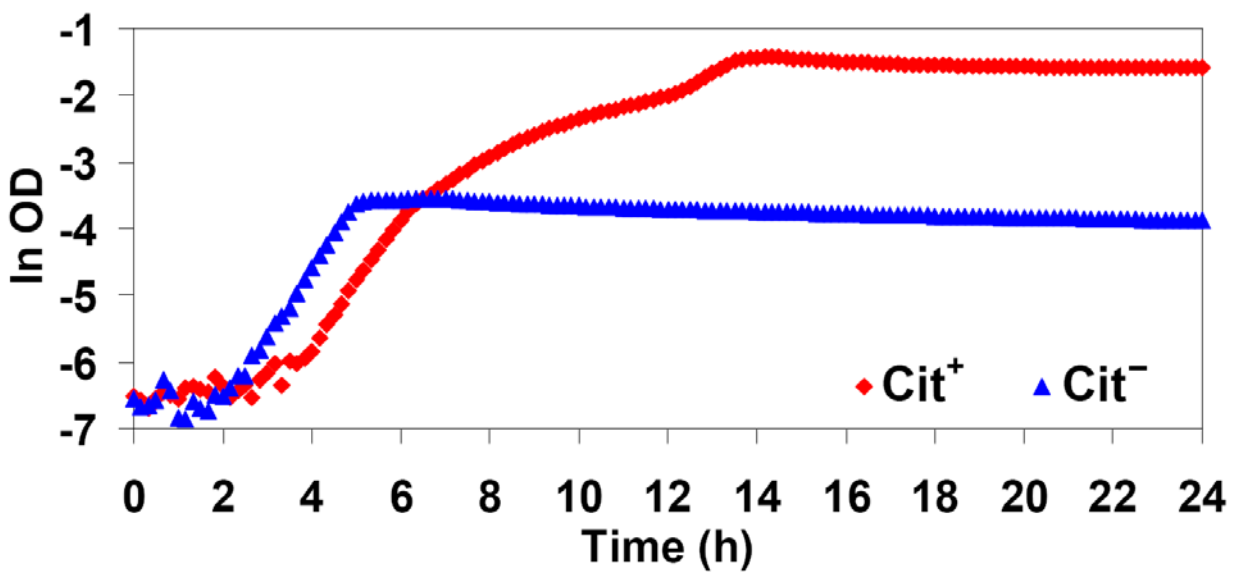


Figure 2.2. Growth of Cit^- (blue triangles) and Cit^+ (red diamonds) cells in DM25 medium.

Each trajectory shows the average OD for eight replicate mixtures of three clones, all from generation 33,000 of population Ara-3.

for the capacity to produce a positive reaction on Christensen's citrate agar, which provides a sensitive means to detect even weakly citrate-using cells. No Cit^+ cells were found in the samples taken at 30,000, 30,500 or 31,000 generations. Cit^+ cells constituted ~0.5% of the population at generation 31,500, then 15% and 19% in the next two samples, but only ~1.1% at generation 33,000. It appears that the first Cit^+ variant emerged between 31,000 and 31,500 generations, although we cannot exclude an earlier origin. The precipitous decline in the frequency of Cit^+ cells just prior to the massive population expansion suggests clonal interference (Gerrish and Lenski 1998), whereby the Cit^- subpopulation produced a beneficial mutant that out-competed the emerging Cit^+ subpopulation until the latter evolved some other beneficial mutation that finally assured its persistence. The hypothesis of clonal interference implies that the early Cit^+ cells were very poor at using citrate, such that a mutation that improved competition for glucose could have provided a greater advantage than did marginal exploitation of the unused citrate.

Indeed, the Cit^+ clones isolated from generations 32,500 and earlier grow much more slowly on citrate than those from 33,000 generations and later. After depleting the glucose in DM25, the earliest Cit^+ clones grow almost imperceptibly, if at all, for many hours before they begin efficiently using the citrate (data not shown), whereas later Cit^+ clones switch to growth on citrate almost immediately (Figure 2.2). Thus, the population expansion between generations 33,000 and 33,500 (Figure 2.1) was triggered by one or more mutations that improved citrate utilization, rather than by the growth of the original Cit^+ mutant. This finding also raises the

question of whether weak Cit^+ mutants might exist in any other LTEE population. We therefore screened the other 11 populations, in most cases using samples from generation 41,500. None of 220 cultures inoculated with heterogeneous population samples grew in glucose-free DM0, nor did any of >3,500 clones show a positive reaction on Christensen's citrate agar.

Historical Contingency in the Evolution of Cit^+ .

We performed three experiments to test whether the evolution of the Cit^+ function required an unusually rare mutation or, alternatively, was historically contingent and depended on the prior evolution of a certain genetic background. All three experiments used clones sampled from many generations of population Ara-3 to replay evolution starting from different genetic backgrounds. The two hypotheses make different predictions about the temporal pattern of Cit^+ evolution during these replay experiments (Figure 2.3). According to the rare-mutation hypothesis, Cit^+ variants should evolve at the same low rate regardless of the time of origin of the clone with which a replay started. By contrast, the historical-contingency hypothesis predicts that the mutation rate to Cit^+ should increase after some potentiating genetic background has evolved. Thus, Cit^+ variants should re-evolve more often in the replays using clones sampled from later generations of the Ara-3 population.

In our first experiment, we performed the replays under the same conditions as the LTEE. We isolated three random clones from each of twelve time points, from the ancestor to 32,500 generations (Table 2.1), and obtained neutral Ara^+ mutants of each clone to embed as protection against accidental cross-contamination during the experiment (Lenski et al 1991). In total, there

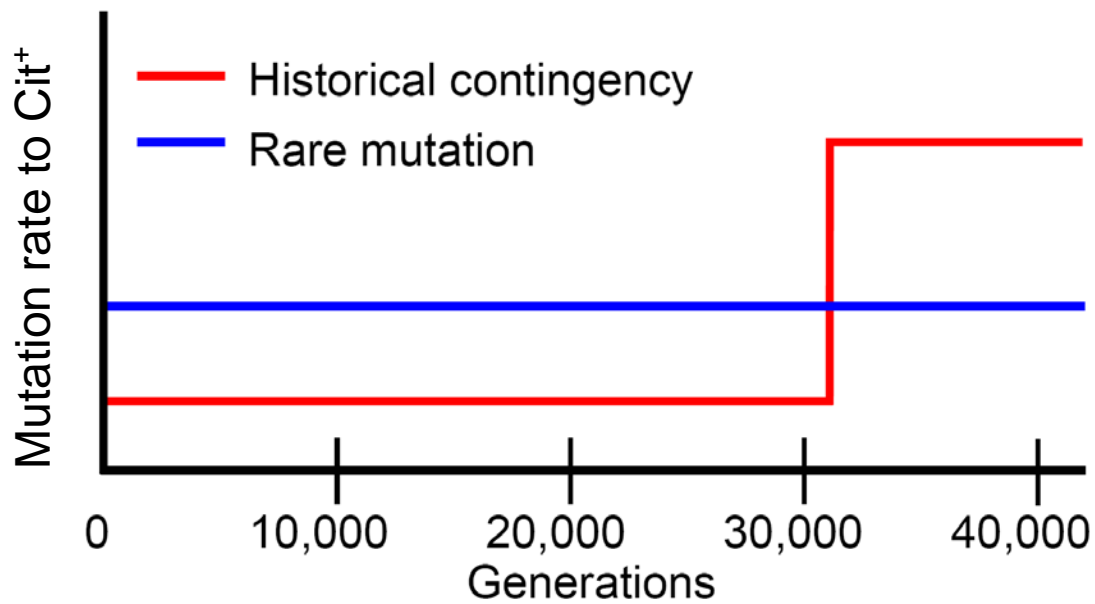


Figure 2.3. Alternative hypotheses for the origin of the Cit^+ function.

According to the rare-mutation hypothesis, the probability of mutation from Cit^- to Cit^+ was low but constant over time. Under the historical-contingency hypothesis, the probability of this transition increased when an earlier mutation arose that produced a genetic background with a higher mutation rate to Cit^+ .

Table 2.1. Summary of replay experiments						
	First experiment		Second experiment		Third experiment	
Generation	Reps	Independent Cit ⁺ mutants	Reps	Independent Cit ⁺ mutants	Reps	Independent Cit ⁺ mutants
Ancestor	6	0	10	0	200	0
5,000	–	–	–	–	200	0
10,000	6	0	30	0	200	0
15,000	–	–	–	–	200	0
20,000	6	0	30	0	200	2
25,000	6	0	30	0	200	0
27,000	–	–	–	–	200	2
27,500	6	0	30	0	–	–
28,000	–	–	–	–	200	0
29,000	6	0	30	0	200	0
30,000	6	0	30	0	200	0
30,500	6	1	30	0	–	–
31,000	6	0	30	0	200	1
31,500	6	1	30	0	200	1
32,000	6	0	30	4	200	2
32,500	6	2	30	1	200	0
Totals	72	4	340	5	2,800	8

were 72 replay populations, six from each generation, each founded by a single clone. These populations evolved for ~3,700 generations, and they were checked visually each day for the elevated turbidity indicative of the Cit⁺ phenotype. We also tested samples on MC and Christensen's citrate agar plates every 250 generations, with incubation for up to a week. New Cit⁺ variants evolved in four replay populations, all founded by clones from later generations of the original population (Table 2.1). These Cit⁺ variants emerged between 750 and 3,700 generations of the replay experiment.

Our second experiment also looked for Cit⁺ mutants derived from clones sampled at various times in the history of population Ara-3. This time, however, we incubated large populations of cells on MC plates, enabling us to test more clones and more cells of each clone. We also allowed a long incubation time to facilitate the growth and detection of very weak Cit⁺ mutants, as well as mutations that might occur as cells sat starving on the plates. In this experiment, $\sim 3.9 \times 10^8$ cells of each of the same 68 clones used in the first replay experiment were spread on each of five MC plates, and these 340 plates were then incubated for 59 days. Five plates produced Cit⁺ mutants, and all used clones from generations 32,000 or 32,500 of the original population (Table 2.1). None of the particular clones that evolved Cit⁺ in this experiment did so in the first one, although there was overlap in the generations from which those clones were sampled.

The third replay experiment was similar in design to the second, but on a larger scale. We isolated 20 clones from each of 13 time points in the history of population Ara-3, again through 32,500 generations. We generated and tested 10 replicate cultures of each evolved clone

and 200 replicates of the ancestor. Each culture grew to $\sim 1\text{--}2 \times 10^{10}$ cells, which were pelleted by centrifugation, spread on an MC plate, and incubated for 45 days. In total, $\sim 4.0 \times 10^{13}$ cells were tested for their ability to use citrate in this experiment. Eight additional Cit^+ mutants evolved from seven different clones (Table 2.1), with one clone yielding mutants in two replicate cultures. Four clones that produced Cit^+ mutants came from generations 31,000 and later, two were from generation 27,000, and one (the one that produced two mutants) was from generation 20,000. We found no Cit^+ mutants among any of the 200 ancestral cultures, nor among any of the other 600 cultures that used clones isolated before generation 20,000.

Interestingly, 7 of the 8 plates that yielded a mutant Cit^+ colony produced multiple colonies, including one with 137 colonies. This pattern illustrates the “jackpot” effect discovered by Luria and Delbrück (1943), and it implies that mutations arose during the population growth prior to plating on MC agar. On the other hand, the Cit^+ colonies were not observed until at least 8 days of incubation and, in one case, they were first seen after 28 days. Such late appearances suggest that the mutations to Cit^+ occurred after plating. One possible explanation for this apparent discrepancy is that the mutants grow very slowly but, in fact, they typically produce visible colonies in only two days when re-tested on MC plates. Another potential explanation is that the high density of Cit^- cells on the plates interfered with the growth and detection of emerging Cit^+ colonies. To test this possibility, we seeded dense Cit^- cultures with a few Cit^+ cells prior to plating on MC agar, but the Cit^+ colonies were seen after only two to three days. The rapid growth of Cit^+ colonies occurred even when the Cit^+ cells had grown

on glucose, and not on citrate, prior to plating. These results imply that mutations to Cit⁺ occurred after cultures were plated on MC agar. This conclusion, taken together with the jackpot distribution of Cit⁺ mutants, indicates that the phenotypic change required two mutations, one of which occurred during the culture growth prior to plating and the other after plating.

Statistical Analysis of the Replay Experiments.

All three experiments show the same tendency for Cit⁺ variants to evolve more often from clones sampled in later than earlier generations of population Ara-3 (Table 2.1). To calculate the significance of these data, we performed Monte Carlo re-sampling tests (shuffling without replacement) using the Statistics101 Re-sampling Simulator version 1.0.6 (www.statistics101.net). For each experiment, we compared the observed mean generation of those clones that yielded Cit⁺ variants to the mean expected under the null hypothesis that clones from all generations have equal likelihood. The null thus corresponds to the rare-mutation hypothesis laid out in the introduction. We ran one million re-sampling iterations for each experiment. The deviations from the null expectations range from marginally to highly significant in the three experiments, and in all cases they support the historical-contingency hypothesis, according to which clones from later generations have greater propensity to evolve the Cit⁺ phenotype (Table 2.2). Although the third experiment was the largest, it was the least significant owing primarily to the production of two Cit⁺ mutants by a 20,000-generation clone.

Table 2.2. Statistical analyses of three replay experiments			
Mean generation of clones yielding Cit ⁺	First experiment	Second experiment	Third experiment
Expected	24,917	28,382	22,571
Observed	31,750	32,100	27,563
Monte Carlo p- value	0.0085	0.0007	0.0823

We also used the Z-transformation method (Whitlock 2005) to combine the probabilities from our three experiments, and the result is extremely significant ($p < 0.0001$) whether or not the experiments are weighted by the number of independent Cit^+ mutants observed in each one. Furthermore, the potentiation effect in later generations is underestimated by these tests, because the number of cells trended lower in later-generation cultures owing to the evolution of larger cells that reach lower population density (Vasi et al 1994, Lenski and Mongold 2000, Hall 1982, Elena et al 1996).

These analyses compel us to reject the hypothesis that a rare mutation could have produced a Cit^+ variant with equal probability at any point in the LTEE. Some unusually rare mutation might be involved, but its rarity does not provide a sufficient explanation for the unique and exceptionally slow evolution of this new function during the LTEE. Our results instead support the hypothesis of historical contingency, in which a genetic background arose that had an increased potential to evolve the Cit^+ phenotype.

Rates of Mutation to Cit^+ and Quantification of the Potentiation Effect.

Given the fraction of cultures that produce no mutants (P_0) and the number of cells per culture (N), one can estimate the mutation rate as $\mu = -(\ln P_0) / N$. The corresponding 95% CI (confidence interval) can also be calculated from the Poisson distribution based on the uncertainty in P_0 , which is much greater than the uncertainty in N . By comparing the mutation rate to Cit^+ between clones, one can in principle quantify the potentiation effect in the evolved genetic background. However, we cannot unambiguously identify which evolved clones remain

non-potentiated owing to the very low mutation rate to Cit^+ even in the potentiated clones. We also cannot separate potentiated and non-potentiated clones by generation because some generational samples may be polymorphic.

We therefore performed an independent series of fluctuation tests using seven clones that yielded Cit^+ mutants in at least one replay experiment. These additional tests permit an unbiased estimate of the mutation rate to Cit^+ in the potentiated background. We prepared 40 replicate cultures for each potentiated clone, and 280 for the ancestor. Potentiated and ancestral cultures had, on average, 1.1 and 1.5×10^{10} cells, respectively, which were harvested and incubated on MC agar plates for 45 days.

None of the ancestral cultures yielded any Cit^+ mutants. We can nonetheless calculate an upper limit to the mutation rate by noting that the Poisson distribution has a 5% probability of yielding zero events when the expectation is three. With no more than 3 mutations among the 8.4×10^{12} cells tested here and in the third replay experiment, the upper bound on the ancestral mutation rate to Cit^+ is 3.6×10^{-13} per cell per generation (Figure 2.4). To the best of our knowledge, this value is the lowest upper bound ever reported for a mutation rate that has been experimentally measured. It is also probably too high because no mutations were actually observed for the ancestor, nor were any found among another 9.0×10^{12} cells of 60 clones sampled through 15,000 generations; and because some cell turnover and other DNA activity probably occurred during the many days that plates were incubated.

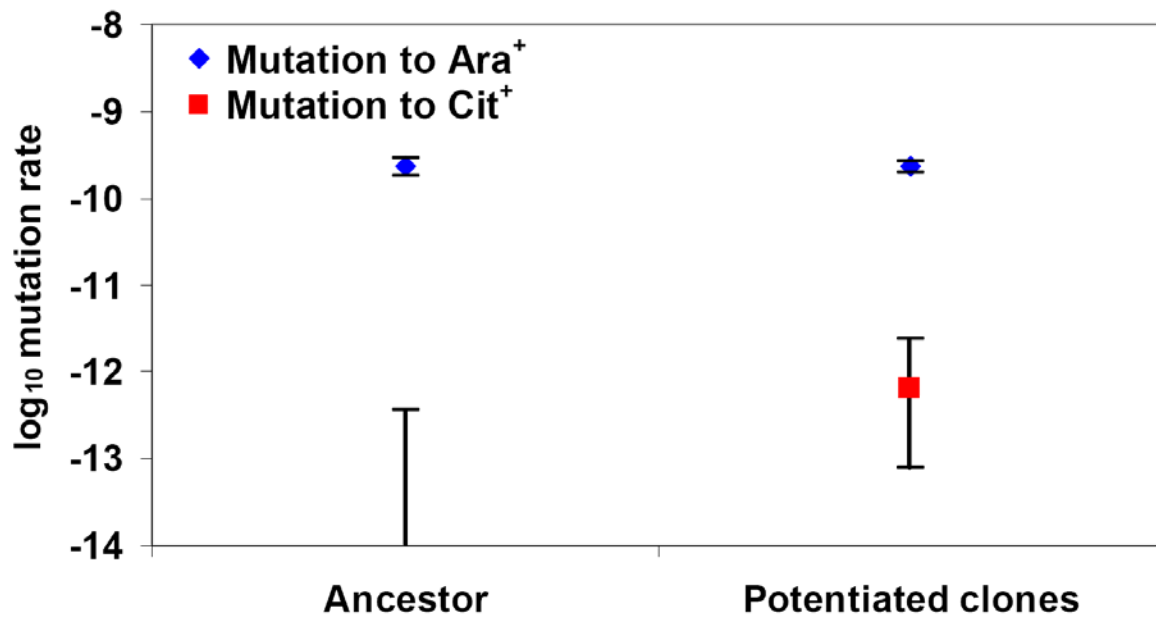


Figure 2.4. Mutation rates from Ara⁻ to Ara⁺ (blue diamonds) and Cit⁻ to Cit⁺ (red squares) of the ancestor and a set of potentiated clones.

Error bars are 95% confidence intervals. See text for details.

Even among the potentiated clones, the rate of mutation to Cit⁺ is extremely low. Cit⁺ mutants arose in two of the 280 new cultures, giving an estimate of 6.6×10^{-13} for the mutation rate, with the 95% CI extending from 7.9×10^{-14} to 2.4×10^{-12} (Figure 2.4). (Although the upper bound of the CI for the ancestor overlaps the lower bound of the CI for the potentiated clones, that upper bound does not overlap the point estimate for the potentiated clones, indicating a significant difference and adding further support to the replay experiments.) The potentiated genetic background thus increases the mutation rate to Cit⁺ at least two fold, and probably much more. However, even this potentiated value represents an unusually low mutation rate. A typical mutation rate in *E. coli* is $\sim 5 \times 10^{-10}$ per bp per generation (Drake 1991). Such a low rate suggests that the final mutation to Cit⁺ is not a point mutation but instead involves some rarer class of mutation or perhaps multiple mutations. The possibility of multiple mutations is especially relevant given our evidence that the emergence of Cit⁺ colonies on MC plates involved events both during the growth of cultures prior to plating and during prolonged incubation on the plates.

Another issue is whether the potentiating effect might indicate the evolution of a generalized hypermutability in population Ara-3. Previous surveys of mutation rates in the 12 LTEE lines found that four had become mutators by generation 20,000, although population Ara-3 retained the low ancestral mutability (Cooper and Lenski 2000, Sniegowski et al 1997). To investigate this issue further, we performed another series of fluctuation tests using the same seven potentiated clones to determine their rate of mutation from Ara⁻ to Ara⁺, which serves as a proxy for the background mutation rate (Sniegowski et al 1997). This mutation reverts a

phenotype that was knocked out during the derivation of a predecessor to the ancestor of the LTEE (Lenski et al 1991, Lederberg 1966). Sequence analysis indicates that most Ara^+ mutants have a mutation from GAC to GGC at codon 92 of the *araA* gene, restoring that codon to its distant ancestral state. Forty cultures of each clone were grown in DM1000, and the cells were spread on minimal-arabinose plates. At the same time, 280 cultures of the ancestor were tested in the same way. The ancestral mutation rate to Ara^+ is 2.3×10^{-10} per cell per generation using the P_0 method (95% CI from 1.8×10^{-10} to 2.9×10^{-10}), a typical rate for point mutations. The rate for potentiated clones is also 2.3×10^{-10} (95% CI from 2.0×10^{-10} to 2.7×10^{-10}); in fact, these estimates are within 1% of each other (Figure 2.4). We conclude that general hypermutability is not responsible for the elevated mutation rate to Cit^+ in the potentiated clones.

Frequency-dependent Selection Maintains Ecological Diversity.

To this point, we have examined the evolutionary origin of the Cit^+ function in population Ara-3. We now turn to its ecological consequence. The Cit^+ phenotype did not achieve fixation during the population expansion but, instead, Cit^- cells persisted as a minority. When we mixed Cit^- and Cit^+ clones at different initial frequencies, they stably coexisted over many serial transfers (Figure 2.5). In these mixtures, the Cit^- cells gradually approached an equilibrium frequency of roughly 1%, regardless of their initial frequency. We saw a transient jump in the frequency of the Cit^- subpopulation on days 7 and 8, which was probably caused by accidentally using a glucose-only medium on those days. After that perturbation, the populations

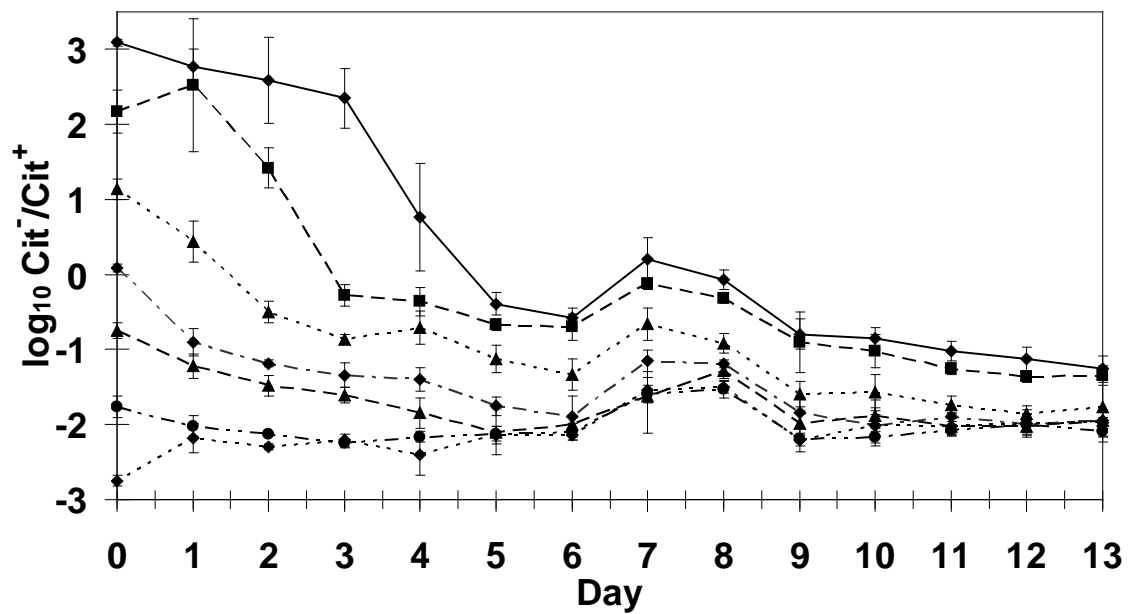


Figure 2.5. Frequency-dependent selection allows stable coexistence of Cit^- and Cit^+ clones.

Each trajectory shows the mean of five replicate cultures for seven different initial ratios of Cit^- and Cit^+ clones from generation 33,000. Error bars are 95% confidence intervals.

resumed their previous trajectories. Negative frequency-dependent selection thus maintains the polymorphism.

This stable coexistence suggests that the Cit^- cells are superior to the Cit^+ cells in competition for glucose, allowing the former to persist as glucose specialists. Indeed, the Cit^- cells have a shorter lag phase and higher growth rate on glucose than do the Cit^+ cells (Figure 2.2). These differences were also evident when we monitored the intra-day dynamics of mixtures of Cit^- and Cit^+ cells (data not shown).

DISCUSSION AND FUTURE DIRECTIONS

E. coli cells cannot grow on citrate under oxic conditions, and that inability has long been viewed as a defining characteristic of this important, diverse, and widespread species. In a long-term experiment, we propagated 12 populations of *E. coli*, all founded from the same ancestral strain, in a medium containing glucose, which is the limiting resource, and abundant citrate. For more than 30,000 generations, none of them evolved the capacity to use the citrate, although billions of mutations occurred in each population, such that any typical base-pair mutation would have been tested many times in each one. It is clearly very difficult for *E. coli* to evolve this function. In fact, the mutation rate of the ancestral strain from Cit^- to Cit^+ is immeasurably low; even the upper bound is 3.6×10^{-13} per cell generation, which is three orders of magnitude below the typical base-pair mutation rate. Nevertheless, one population eventually evolved the Cit^+ function, while all the others remain Cit^- after more than 40,000 generations.

We demonstrated that the evolution of this new function was contingent upon the history of the population in which it arose. In particular, we showed that one or more earlier mutations

potentiated the evolution of this function by increasing the mutation rate to Cit^+ , although even the elevated rate is much lower than a typical mutation rate. The potentiated cells are not generally hypermutable. Rather, their potentiation is specific to the Cit^+ function, which suggests two possible mechanisms. One mechanism is epistasis, whereby the functional expression of the mutation that finally yielded the Cit^+ phenotype requires interaction with one or more mutations that evolved earlier. A second possibility is that the physical production of the mutation that produced the Cit^+ phenotype requires some previous mutation that allows the final sequence to be generated. For example, the insertion of a mobile genetic element creates new sequences at its junctures, and one of these new sequences might then undergo a mutation that generates a final sequence that could not have occurred without the insertion. The *E. coli* genome has many insertion-sequence elements (Schneider et al 2002), some of which have been active in the LTEE (Papadopoulos et al 1999, Schneider et al 2000, Schneider and Lenski 2004). Whatever the mechanism, this potentiation made the Cit^+ function mutationally accessible, and a weak Cit^+ variant emerged by 31,500 generations.

The origin of the Cit^+ function also had profound consequences for the ecology and subsequent evolution of that population. This new capacity was refined over the next 2,000 generations, leading to a massive population expansion as the Cit^+ cells evolved to exploit more efficiently the abundant citrate in their environment. Although the Cit^+ cells continued to use glucose, they did not drive the Cit^- subpopulation extinct because the Cit^- cells were superior competitors for glucose. Thus, the overall diversity increased as one population gave rise

evolutionarily to an ecological community with two members, one a resource specialist and the other a generalist.

The evolution of the new Cit⁺ function represents a key innovation that involves multiple steps, and it provides an explicit demonstration of the importance of historical contingency in evolution. It also transcends the phenotypic boundaries of a diverse and well-studied species, and led to an ecological transition from a single population to a two-member community. Our future research on this fascinating case of evolution in action will revolve around four themes: genetics, physiology, ecology, and speciation.

What is the genetic basis of this evolutionary innovation? The emergence of the Cit⁺ phenotype in population Ara-3 indicates at least two important genetic events: the origin of the function in its weak form, and its subsequent refinement for efficient use of citrate. The replay experiments indicate an even more complex picture that must involve, at a minimum, three important genetic events. At least one mutation in the LTEE was necessary to produce a genetic background with the potential to generate Cit⁺ variants, while the distribution and dynamics of Cit⁺ mutants in fluctuation tests indicate at least two additional mutations are involved. To find the relevant mutations, we will perform whole-genome re-sequencing, which has become a powerful approach that is well suited to experimental evolution (Velicer et al 2006, Herring et al 2006, Hegreness and Kishony 2007). We expect to find dozens of mutations relative to the ancestor (Lenski 2004), which will complicate identification of those changes that were important specifically for the origin of the Cit⁺ function. However, some of the key changes should become apparent if we also re-sequence a Cit⁻ clone from the same population around the

time that the Cit⁺ variants first emerged. Once candidate genes and mutations have been identified, we can examine the other 19 Cit⁺ variants from our replay and mutation-rate experiments for parallel changes.

We are especially eager to find the potentiating mutation or mutations. We want to know whether the potentiating mutation interacts epistatically with a later mutation to allow expression of the Cit⁺ function or, alternatively, whether it was physically required for the later mutation to occur. We also want to test whether the potentiating mutation was itself beneficial or, alternatively, a neutral or deleterious change that fortuitously hitchhiked to high frequency. We anticipate that identifying the potentiating mutation will be especially challenging, however, because its only known phenotype is to increase the rate of production of certain mutants that are themselves extremely rare.

Once we have identified all the relevant mutations, it might be interesting to model the population dynamics that govern the emergence of this new function. Such a model would require not only all the relevant mutation rates but also the ecological phenotypes of the mutants, including their growth rates on glucose and citrate as well as their abilities to transition between the two resources. A satisfactory model should also reflect the stochastic origin of mutations, the role of random drift, and the possibility of alternative mutational paths to the phenotype of interest (Weinreich et al 2006, Lenski et al 2003).

What physiological mechanism has evolved that allows aerobic growth on citrate? *E. coli* should be able to use citrate as an energy source after it enters the cell, but it lacks a citrate transporter that functions in an oxygen-rich environment. One possibility is that the Cit⁺ lineage activated a “cryptic” transporter (Hall 1982), that is, some once-functional gene that has been

silenced by mutation accumulation. This explanation seems unlikely to us because the Cit^- phenotype is characteristic of the entire species, one that is very diverse and therefore very old. We would expect a cryptic gene to be degraded beyond recovery after millions of years of disuse. A more likely possibility, in our view, is that an existing transporter has been coopted for citrate transport under oxic conditions. This transporter may previously have transported citrate under anoxic conditions (Pos et al 1998) or, alternatively, it may have transported another substrate in the presence of oxygen. The evolved changes might involve gene regulation, protein structure, or both (Mortlock 1984).

What will be the long-term fates of the coexisting Cit^- and Cit^+ subpopulations? We showed they stably coexist owing to the inferiority of the Cit^+ cells in competition for glucose. However, the Cit^+ lineage might eventually acquire mutations that compensate for its inferior performance on glucose, thus undermining the coexistence. Compensation for maladaptive side-effects of adaptations, including resistance to phages and antibiotics, has been observed in many other experiments with bacteria (Lenski 1988, Bouma and Lenski 1988, Schrag et al 1997, Lenski 1997, Bjorkman et al 1998). Moreover, the Cit^+ subpopulation is much larger than the Cit^- subpopulation, so it should experience more beneficial mutations even without compensation. On the other hand, coexistence would be strengthened if selection in the Cit^+ subpopulation favors specialization on citrate. In the same way that we established multiple populations for retrospective replays in this study, we can establish multiple communities to examine their prospective evolution. We can also vary environmental factors, such as the

presence or absence of glucose, and the presence or absence of a Cit^- subpopulation, to investigate how they influence the future evolution of the Cit^+ lineage.

Will the Cit^+ and Cit^- lineages eventually become distinct species? According to the biological species concept widely used for animals and plants, species are recognized by reproductive continuity within species and reproductive barriers leading to genetic isolation between species (Mayr 1942). Although the bacteria in the LTEE are strictly asexual, we can nonetheless imagine testing this criterion by producing recombinant genotypes. In particular, we could move mutations that are substituted in the evolving Cit^+ lineage into a Cit^- background to test whether they reduce fitness in their ancestral context. One could also perform the reciprocal experiment, though we anticipate more rapid evolution in the Cit^+ lineage because it has acquired a key innovation that substantially changed its ecological niche. Such experiments would require, of course, controls to examine the fitness effects of the same mutations in the lineage where they arose. If the Cit^+ lineage is indeed evolving into a new species, then we expect, with time, that more and more of the beneficial mutations substituted in that lineage would be detrimental in the ecological and genetic context of its Cit^- progenitor.

In any case, our study shows that historical contingency can have a profound and lasting impact under the simplest, and thus most stringent, conditions in which initially identical populations evolve in identical environments. Even from so simple a beginning, small happenstances of history may lead populations along different evolutionary paths. A potentiated cell took the one less traveled by, and that has made all the difference.

MATERIALS AND METHODS

The Long-term Evolution Experiment

The LTEE is described in detail elsewhere (Lenski et al 1991, Lenski 2004). Briefly, two ancestral clones of *E. coli* B were each used to found six populations. The ancestors differ by a single mutation that allows one of them to use arabinose (Ara^+). Ara^- and Ara^+ cells make red and white colonies, respectively, on tetrazolium-arabinose (TA) plates, but the mutation is neutral in the environment of the LTEE (Lenski et al 1991). The twelve populations have been propagated for almost 20 years by daily serial dilution in DM25, a minimal salts medium that has 139 μM glucose and 1,700 μM citrate (Lenski et al 1991). Given 1:100 dilution and re-growth, the populations achieve ~ 6.64 generations per day, and they have evolved for over 40,000 generations in this experiment to date. Every 500 generations, population samples are frozen at -80°C with glycerol added as a cryoprotectant. These samples contain all of the diversity present in a population at that generation. Except as otherwise noted, all bacteria used in this study come from samples of population Ara-3.

Media and Culture Conditions

DM0 is identical to DM25 except that it contains no glucose. MC plates have the same formulation as liquid DM0, except that agar is added to solidify the medium and the concentration of citrate is higher to allow large colonies. Prior to starting an experiment, bacteria were preconditioned in the relevant experimental medium unless otherwise stated. For additional details, see Supporting Information (SI) Materials and Methods.

Analysis of the Population Expansion

When the increased turbidity was observed in population Ara-3, we froze samples of the two previous days' cultures, which had been temporarily stored at 4°C. We also froze samples after every transfer for several weeks. We later revived these and other samples taken between generations 30,000 and 35,000 to quantify the population expansion. For details, see SI Materials and Methods.

Searches for Cit⁺ Variants

In order to determine when the earliest Cit⁺ variant arose in population Ara-3, we randomly chose 1280 clones from samples frozen at generations 30,000, 30,500, 31,000, 31,500, 32,000, 32,500 and 33,000. We transferred colonies to Christensen's citrate agar plates, which were incubated for 10 days and examined daily for evidence of positive reactions. All putative Cit⁺ clones were retested on Christensen's agar for confirmation, and a subset was tested on MC agar. All clones that had positive reactions on Christensen's plates also grew on MC plates. In order to find Cit⁺ cells that might be present at low frequency in the other 11 LTEE populations, we tested their genetically heterogeneous full-population samples for growth in DM0. We allowed the top of each sample to thaw, then inoculated 50 µl into LB broth. After overnight growth, five 10-mL DM0 cultures were inoculated with 100 µL of 100-fold dilutions of each LB culture. The DM0 cultures were incubated for 28 days with periodic checks for turbidity.

Growth Curves

Three representative Cit⁻ clones and three representative Cit⁺ clones from generation 33,000 were separately preconditioned in DM25. The Cit⁻ clones were then combined, as were the Cit⁺ clones, and the mixtures were diluted in DM25 to $\sim 3.4 \times 10^4$ cells per mL. Eight 200- μ L aliquots of each mixture were dispensed into randomly assigned wells in a 96-well plate, and optical density (OD) at 420 nm wavelength was measured periodically using a VERSAmax automated plate reader.

Replay Experiments

In the first replay experiment, populations evolved under the same conditions as the LTEE. In the second and third replay experiments, cells were incubated on MC plates. We started the first replay experiment on the 3rd anniversary of Stephen Jay Gould's death; we ended it on the 66th anniversary of his birth. For further details, see SI Materials and Methods.

Confirmation of Cit⁺ Variants

Each putative Cit⁺ variant was streaked on MC agar and Christensen's citrate agar to confirm its phenotype. Additional tests are described in the SI Materials and Methods.

Test for Frequency-dependent Interaction

We constructed mixtures of clones from generation 33,000 to test whether Cit^- and Cit^+ clones coexisted and, if so, whether they interacted in a frequency-dependent manner. For details, see SI Materials and Methods.

ACKNOWLEDGEMENTS

We thank N. Hajela for years of outstanding technical assistance, and M. Adawe for help with media preparation; together they poured some 20,000 agar plates used in our experiments. We thank Francisco Ayala, Al Bennett, Rita Colwell, Bruce Levin, and Simon Levin for reviewing our paper prior to publication. This research has been supported in part by grants from the National Science Foundation (currently DEB-0515729) and the Defense Advanced Research Projects Agency “Fun Bio” Program (HR0011-05-1-0057).

SUPPORTING INFORMATION

SI Materials and Methods

Media and Culture Conditions

The various DM media used to grow cells in different experiments are all based on the same formulation as the DM25 used in the LTEE (Lenski et al 1991), with only glucose concentration varied. DM25 contains glucose at 25 mg/L. DM500, DM1000, and DM2000 have 20, 40, and 80 times, respectively, more glucose than DM25, while DM0 contains no glucose. MC plates have the same formulation as liquid DM0, except that agar is added at 16 g/L and

trisodium citrate dihydrate is increased to 4.5 g/L. Simmon's and Christensen's citrate agar media are described elsewhere (Atlas 2006).

Prior to starting an experiment, unless stated otherwise, clones or population samples were inoculated from frozen stocks into LB broth and incubated overnight. Next, the bacteria were preconditioned by diluting them 10,000-fold into 10 mL of the relevant experimental medium, and they were incubated for 24 h. Cultures were then diluted 100-fold into a second volume of experimental medium, and again incubated for 24 h.

Analysis of the Population Expansion

When the increased turbidity was observed in population Ara-3, we froze samples of the two previous days' cultures, which had been temporarily stored at 4°C. We also froze samples after every transfer for several weeks. We later revived the samples from generations 30,000, 31,000, 32,000, 32,500, 33,000, 33,113, 33,120, 33,127, 33,133, 33,140, 33,180, 33,200, 33,293, 33,413, 33,500, 33,760, 34,000, 34,260, 34,500, 34,760, and 35,000 in LB, and preconditioned them for 24 h in DM25. We then transferred each one into three replicate DM25 cultures. After 24 h, we moved 200-μL samples of each culture to a 96-well plate, and measured optical density at 420 nm wavelength using a VERSAmax automated plate reader.

First Replay Experiment

Three clones were chosen at random from Ara-3 population samples from 10,000, 20,000, 25,000, 27,500, 29,000, 30,000, 30,500, 31,000, 31,500, 32,000, and 32,500 generations. All these clones were Cit⁻, including those from samples taken after the weak Cit⁺ variant emerged. Spontaneous Ara⁺ mutants were derived from each clone by plating on minimal-arabinose agar (Lenski et al 1991). Two populations were founded by each evolved clone, one

using the Ara⁻ clone and one its Ara⁺ derivative. Six populations were founded by the ancestral strain, including three each using the Ara⁻ and Ara⁺ clones. These 72 populations were propagated by daily 1:100 dilutions in DM25, as in the LTEE except that these replays ran in unshaken test-tubes. Each culture was inspected at every transfer for increased turbidity that would indicate the re-evolution of the Cit⁺ phenotype. Undiluted samples from each culture were spread on MC and Christensen's citrate agar plates at ~250-generation intervals, at which time population samples were frozen for long-term storage. Excluding interruptions, this experiment ran for 557 days, or ~3,700 generations with 6.64 generations per day.

Second and Third Replay Experiments

The second replay experiment used the same 68 clones as the first one. For each clone, we inoculated five 10-mL cultures of DM500 with <60 cells. After 48 h, the cells in each culture were pelleted by centrifugation (~5000 g for 10 min), and then re-suspended, with half of each culture (~4 x 10⁹ cells) spread on an MC plate and the other half on an unamended agar plate as a control. The plates were incubated for 59 days with periodic checks for colonies.

For the third replay experiment, we isolated 20 Cit⁻ clones each from the Ara-3 population samples from 5,000, 10,000, 15,000, 20,000, 25,000, 27,000, 28,000, 29,000, 30,000, 31,000, 31,500, 32,000 and 32,500 generations. Clones were chosen in order to maximize the diversity of colony morphologies from each sample, in case only certain types could produce Cit⁺ mutants. For those generations included in the first and second experiments, the three clones used there were included among the 20 clones used here.

To facilitate handling and minimize possible confounding variables, we divided this third experiment into 20 blocks of 14 clones each. All of the clones within a block came from different generations, and the single ancestral clone was included in all 20 blocks. For every clone in each block, we inoculated ten replicate 10-mL cultures with <60 cells. Seven blocks used DM1000 medium, in which the bacteria achieved densities of $\sim 8 \times 10^8$ to 1.5×10^9 cells per mL. The other 13 blocks used DM2000, in which the cells reached densities about twice as high. Cultures were incubated for 48 h with shaking for aeration. We then diluted a small volume of every second culture on TA plates to estimate final cell numbers. Each culture was pelleted, re-suspended, and spread in its entirety on an MC plate. These plates were incubated for 45 days, with periodic checks for colonies.

The number of Cit^+ mutants found in the second experiment led us to expect many more mutants in our third replay experiment, but that expectation was not fulfilled. Subtle differences in procedures or conditions evidently may affect the mutation rate or selective enrichment of these mutants. We emphasize, however, that the same procedures and conditions were applied simultaneously to clones sampled from all generations in each experiment. Also, in the second and third replay experiments, some plates became contaminated during the repeated checks for mutants, and they were discarded before the experiment ended. Such contamination was infrequent and haphazard, and so it did not affect our analysis.

With hindsight, we should probably have grown the cells to be spread on the MC plates in the second and third experiments in medium without citrate, in order to avoid possible selection for Cit^+ mutants prior to plating. However, the presence of citrate in the growth medium made no practical difference for several reasons: (i) the final mutations that gave the

Cit⁺ phenotype arose on the plates; (ii) the same protocol was used for clones from all generations; and (iii) the P_0 method for estimating mutation rates is insensitive to the timing and growth rate of mutants relative to non-mutants.

Confirmation of Cit⁺ Variants

Each putative Cit⁺ variant was streaked on MC agar and Christensen's citrate agar to confirm its phenotype. One colony was then selected from the MC plate and its Ara marker status, sensitivity to phage T5, and resistance to phage T6 were checked to confirm that it was derived from the ancestral *E. coli* B strain (Lenski et al 1991). We also sequenced the *pykF* and *nadR* loci of Cit⁺ variants and their parental Cit⁻ clones to confirm single base-pair substitutions that uniquely identify the Ara-3 population (Woods et al 2006).

Test for Frequency-dependent Interaction

We constructed mixtures of six clones from generation 33,000 to test whether Cit⁻ and Cit⁺ clones coexisted and, if so, whether they interacted in a frequency-dependent manner. Three clones were Cit⁺ and three were Cit⁻, and the latter set carried the neutral Ara⁺ marker to help us distinguish the two types using TA agar. The six clones were revived, preconditioned, and combined in DM25, with $\sim 3.4 \times 10^6$ cells in total used to inoculate each mixed culture. To test for frequency-dependent effects, we made mixtures with seven different initial frequencies of Cit⁻ cells: 0.1%, 1%, 10%, 50%, 90%, 99%, and 99.9%. Each initial frequency was replicated five-fold, and all 35 mixed cultures were serially propagated with 100-fold dilution in DM25 for 13 days. As noted in the Results, we may have accidentally transferred these cultures into a

glucose-only medium for two days in the middle of this experiment. To estimate the abundances of the Cit^- and Cit^+ types, we plated aliquots of each mixed culture on TA agar at the start and at every daily transfer.

REFERENCES

REFERENCES

1. Atlas, R.M. *Handbook of Microbiological Media for the Examination of Food* (CRC Press, Boca Raton, FL, 2006).
2. Beatty, J. Replaying life's tape. *Journal of Philosophy* **7**, 336–362 (2006).
3. Bjorkman, J., Hughes, D., & Andersson, D.I. Virulence of antibiotic-resistant *Salmonella typhimurium*. *Proceedings of the National Academies of Science (USA)* **95**, 3949–3953 (1998).
4. Bouma, J.E. & Lenski, R.E. Evolution of a bacteria/plasmid association. *Nature* **335**, 351–352 (1988).
5. Conway Morris, S. *Life's Solution* (Cambridge Univ. Press, Cambridge, MA, 2003).
6. Cooper, V.S. & Lenski, R.E. The population genetics of ecological specialization in evolving *E. coli* populations. *Nature* **407**, 736–739 (2000).
7. Cooper, V.S., Schneider, S., Blot, M., & Lenski, R.E. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *E. coli* B. *Journal of Bacteriology* **183**, 2834–2841 (2001).
8. Cooper, T.F., Rozen, D.E., & Lenski, R.E. Parallel changes in gene expression after 20,000 generations of evolution in *E. coli*. *Proceedings of the National Academy of Sciences (USA)* **100**, 1072–1077 (2003).
9. Cooper, T.F., Remold, S.K., Lenski, R.E., & Schneider, D. Expression profiles reveal parallel evolution of epistatic interactions involving the CRP regulon in *Escherichia coli*. *PLoS Genetics* **4**, e35 (2008).
10. Crozat, E., Philippe, N., Lenski, R.E., Geiselmann, J., & Schneider, D. Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. *Genetics* **169**, 523–532 (2005).
11. Dawkins, R. *The Blind Watchmaker* (Norton, New York, NY, 1996).
12. Drake, J.W. A constant rate of spontaneous mutation in DNA-based microbes. *Proceedings of the National Academy of Sciences (USA)* **88**, 7160–7164 (1991).
13. Elena, S.F., Cooper, V.S., & Lenski, R.E. Punctuated evolution caused by selection of rare beneficial mutations. *Science* **272**, 1802–1804 (1996).
14. Foote, M.J. Contingency and convergence. *Science* **280**, 2068–2069 (1998).
15. Frost, G.E. & Rosenberg, H. The inducible citrate-dependent iron transport system in *Escherichia coli* K-12. *Biochimica et Biophysica ACTA* **330**, 90–101 (1973).

16. Gerrish, P.J. & Lenski, R.E. The fate of competing beneficial mutations in an asexual population. *Genetica* **102/103**, 127–144 (1998).
17. Gould, S.J. The paradox of the first tier: an agenda for paleobiology. *Paleobiology* **11**, 2–12 (1985).
18. Gould, S.J. *Wonderful Life* (Norton, New York, NY, 1989).
19. Gould, S.J. *The Structure of Evolutionary Theory* (Belknap, Cambridge, 2002).
20. Hall, B.G. Chromosomal mutation for citrate utilization by *Escherichia coli* K-12. *Journal of Bacteriology* **151**, 269–273 (1982).
21. Hegreness, M. & Kishony, R. Analysis of genetic systems using experimental evolution and whole-genome sequencing. *Genome Biology* **8**, 201 (2007).
22. Herring, C.D. *et al* Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nature Genetics* **38**, 1406–1412 (2006).
23. Hussein, S., Hantke, K., & Braun, V. Citrate-dependent iron transport system in *Escherichia coli* K-12. *European Journal of Biochemistry* **117**, 431–437 (1981).
24. Ishiguro, N., Oka, C., & Sato, G. Isolation of citrate positive variants of *Escherichia coli* from domestic pigeons, pigs, cattle, and horse. *Applied and Environmental Microbiology* **36**, 217–222 (1978).
25. Ishiguro, N., Oka, C., Hanazawa, Y., & Sato, G. Plasmids in *Escherichia coli* controlling citrate-utilizing ability. *Applied and Environmental Microbiology* **38**, 956–964 (1979).
26. Jablonski, D. Background and mass extinctions: the alternation of macroevolutionary regimes. *Science* **231**, 139–133 (1986).
27. Jacob F Evolution and tinkering. *Science* **196**, 1161–1166 (1977).
28. Koser, S.A. Correlation of citrate-utilization by members of the colon-aerogenes group with other differential characteristics and with habitat. *Journal of Bacteriology* **9**, 59–77 (1924).
29. Lara, F.J.S. & Stokes, J.L. Oxidation of citrate by *Escherichia coli*. *Journal of Bacteriology* **9**, 59–77 (1952).
30. Lederberg, S. Genetics of host-controlled restriction and modification of deoxyribonucleic acid in *Escherichia coli*. *Journal of Bacteriology* **91**, 1029–1036 (1966).

31. Lenski, R.E. Experimental studies of pleiotropy and epistasis in *Escherichia coli*. II. Compensation for maladaptive pleiotropic effects associated with resistance to virus T4. *Evolution* **42**, 433–440 (1988).
32. Lenski, R.E. The cost of antibiotic resistance – from the perspective of a bacterium. *CIBA Foundation Symposium* **207**, 131–140 (1997).
33. Lenski, R.E. Phenotypic and genomic evolution during a 20,000-generation experiment with the bacterium *Escherichia coli*. *Plant Breeding Reviews* **24**, 225–265 (2004).
34. Lenski, R.E., Rose, M.R., Simpson, S.C., & Tadler, S.C. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *American Naturalist* **138**, 1315–1341 (1991).
35. Lenski, R.E. & Travisano, M. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proceedings of the National Academy of Sciences (USA)* **91**, 6808–6814 (1994).
36. Lenski, R.E. & Mongold, J.A. in *Scaling in Biology*, eds Brown J, West G (Oxford Univ. Press, Oxford, UK), pp 221–235 (2000).
37. Lenski, R.E., Ofria, C., Pennock, R.T., & Adami, C. The evolutionary origin of complex features. *Nature* **423**, 139–144 (2003).
38. Lenski, R.E., Winkworth, C.L., & Riley, M.A. Rates of DNA sequence evolution in experimental populations of *Escherichia coli* during 20,000 generations. *Journal of Molecular Evolution* **56**, 498–508 (2003).
39. Lewontin, R.C. Is nature probable or capricious? *BioScience* **16**, 25–27 (1966).
40. Luria, S.E. & Delbrück, M. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**, 491–511 (1943).
41. Lutgens, M. & Gottschalk, G. Why a co-substrate is required for anaerobic growth of *Escherichia coli* on citrate. *J Gen Microbiol* **199**, 63–70 (1980).
42. Mani, G.S. & Clarke, B.C. Mutational order: a major stochastic process in evolution. *Proceedings of the Royal Society of London Series B: Biological Sciences* **240**, 29–37 (1990).
43. Mayr, E. *Systematics and the Origin of Species from the Viewpoint of a Zoologist* (Columbia Univ. Press, New York, 1942).
44. Mongold, J.A., Bennett, A.F., & Lenski, R.E. Evolutionary adaptation to temperature. VII. Extension of the upper thermal limit of *Escherichia coli*. *Evolution* **53**, 386–394 (1999).

45. Mortlock, R.P., ed. *Microorganisms as Model Systems for Studying Evolution* (Plenum, New York, 1984).
46. Novak, M., Pfeiffer, T., Lenski, R.E., Sauer, U., & Bonhoeffer, S. Experimental tests for an evolutionary trade-off between growth rate and yield in *E. coli*. *American Naturalist* **168**, 242–251 (2006).
47. Pelosi, L. *et al* Parallel changes in global protein profiles during long-term experimental evolution in *Escherichia coli*. *Genetics* **173**, 1851–1869 (2006).
48. Papadopoulos, D. *et al*. Genomic evolution during a 10,000-generation experiment with bacteria. *Proceedings of the National Academy of Sciences (USA)* **96**, 3807–3812 (1999).
49. Pos, K.M., Dimroth, P., & Bott, M. 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 (1998).
50. Reynolds, C.H., & Silver, S. Citrate utilization by *Escherichia coli*: plasmid- and chromosome-encoded systems. *Journal of Bacteriology* **156**, 1019–1024 (1983).
51. Scheutz, F. & Strockbine, N.A. in *Bergey's Manual of Systematic Bacteriology, Volume Two: The Proteobacteria*, eds Garrity GM, Brenner DJ, Kreig NR, Staley JR (Springer, New York, NY), pp 607–624 (2005).
52. Schneider, D. & Lenski, R.E. Dynamics of insertion sequence elements during experimental evolution of bacteria. *Research in Microbiology* **155**, 319–327 (2004).
53. Schneider, D., Duperchy, E., Coursange, E., Lenski, R.E., & Blot, M. Long-term experimental evolution in *Escherichia coli*. IX. Characterization of IS-mediated mutations and rearrangements. *Genetics* **156**, 477–488 (2000).
54. Schneider, D. *et al*. Genomic comparisons among *Escherichia coli* strains B, K-12, and O157:H7 using IS elements as molecular markers. *BMC Microbiology* **2**, 18 (2002).
55. Schrag, S.J., Perrot, V., & Levin, B.R. Adaptation to the fitness costs of antibiotic resistance in *Escherichia coli*. *Proceedings of the Royal Society of London Series B: Biological Sciences* **264**, 1287–1291 (1997).
56. Sniegowski, P.D., Gerrish, P.J., & Lenski, R.E. Evolution of high mutation rates: separating causes from consequences. *Nature* **387**, 703–705 (1997).
57. Travisano, M., Vasi, F., & Lenski, R.E. Long-term experimental evolution in *Escherichia coli*. III. Variation among replicate populations in correlated responses to novel environments. *Evolution* **49**, 189–200 (1995).
58. Travisano, M., Mongold, J.A., Bennett, A.F., & Lenski, R.E. Experimental tests of the roles of adaptation, chance, and history in evolution. *Science* **267**, 87–90 (1995).

59. Van Valen, L.M. How far does contingency rule? *Evolutionary Theory* **10**, 47–52 (1991).
60. Vasi, F., Travisano, M., & Lenski, R.E. Long-term experimental evolution in *Escherichia coli*. II. Changes in life-history traits during adaptation to a seasonal environment. *American Naturalist* **144**, 432–456 (1994).
61. Velicer, G.J. *et al.* Comprehensive mutation identification in an evolved bacterial cooperator and its cheating ancestor. *Proceedings of the National Academy of Sciences (USA)* **103**, 8107–8112 (2006).
62. Vermeij, G.J. Historical contingency and the purported uniqueness of evolutionary innovations. *Proceedings of the National Academy of Sciences (USA)* **103**, 1804–1809 (2006).
63. Weinreich, D.M., Watson, R.A., & Chao, L. Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59**, 1165–1174 (2005).
64. Weinreich, D.M., Delaney, N.F., DePristo, M.A., & Hartl, D.L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).
65. Whitlock, M.C. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *Journal of Evolutionary Biology* **18**, 1368–1373 (2005).
66. Woods, R., Schneider, D., Winkworth, C.L., Riley, M.A., & Lenski, R.E. Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proceedings of the National Academy of Sciences (USA)* **103**, 9107–9122 (2006).
67. Wright, S. Surfaces of selective value revisited. *American Naturalist* **131**, 115–123 (1988).
68. Yedid, G. & Bell, G. Macroevolution simulated with autonomously replicating computer programs. *Nature* **420**, 810–812 (2002).

CHAPTER 3

Blount, Z.D., Barrick, J.E., Davidson, C.J., & Lenski, R.E. Genomic analysis of a key innovation in an experimental *E. coli* population. *Submitted for review*.

GENOMIC ANALYSIS OF A KEY INNOVATION IN AN EXPERIMENTAL *E. COLI* POPULATION

ABSTRACT

Evolutionary novelties have been important in the history of life, but their origins are usually difficult to examine in detail. We previously described the evolution of a novel trait, aerobic citrate utilization (Cit^+), in an experimental population of *Escherichia coli*. Here we analyze 29 genomes to investigate the history and genetic basis of this trait. At least three distinct clades coexisted for more than 10,000 generations prior to its emergence. The Cit^+ trait originated in one clade by a tandem duplication that captured an aerobically-expressed promoter for the expression of a previously silent citrate transporter. The clades varied in their propensity to evolve this novel trait, although genotypes able to do so existed in all three clades, implying that multiple potentiating mutations arose during the population's history. Our findings illustrate the importance of promoter capture and altered gene regulation in mediating the exaptation events that often underlie evolutionary innovations.

BACKGROUND AND INTRODUCTION

Evolutionary novelties are qualitatively new traits that open up ecological opportunities and thereby promote diversification (Mayr 1960, Piliucci 2008). These traits are thought to arise largely by the exaptation of genes that previously encoded other functions (Pigliucci 2008, Jacob 1977, 1982, Gould and Vrba 1982) via such processes as domain shuffling (Patthy 1999), altered regulation (True and Carroll 2002), and duplication followed by neo-functionalization (Zhang 2003, Bergthorsson et al 2007). Multiple mutations may be necessary to produce the new function (Zhang 2003, Lenski et al 2003), and thus the potential to evolve that function may be contingent on subtle genetic differences between species, populations, or even genotypes. A complete understanding of the evolution of a novel trait requires explanation of its ecological function, its mechanistic basis, the genetic changes that underlie its evolution, and the history of the accumulated changes (Pigliucci 2008).

Evolution experiments with bacteria present unparalleled opportunities to assess these issues (Elena and Lenski 2003, Lenski 2004). Bacteria have rapid generations and large populations, making it possible to observe the emergence of novel traits, while new technologies allow discovery of mutations throughout their genomes (Bentley 2006, Hegreness and Kishony 2007, Barrick et al 2009, Barrick and Lenski 2009). Samples of evolving populations can be frozen and later revived, allowing phylogenies and mutational histories to be constructed from ancestral, intermediate, and evolved genome sequences (Barrick et al 2009, Barrick and Lenski 2009).

Twelve populations of *Escherichia coli* have been propagated in the long-term evolution experiment (LTEE) for over 40,000 generations in a glucose-limited minimal medium (Lenski 2004). The medium also contains abundant citrate, which is present as a chelating agent (Blount

et al 2008), but *E. coli* cannot exploit citrate as a carbon and energy source in the well-aerated conditions of the experiment (Koser 1924, Lutgens and Gottschalk 1980). Indeed, the inability to grow aerobically on citrate is a long-recognized trait that, in part, defines *E. coli* as a species²¹. Spontaneous citrate-using (Cit^+) mutants are extraordinarily rare (Hall 2982), but a Cit^+ variant evolved in one LTEE population (designated Ara-3) after 31,000 generations (Blount et al 2008). Cit^+ cells became numerically dominant after 33,000 generations, although Cit^- cells persisted. This shift was accompanied by a substantial increase in total population size owing to the high concentration of citrate relative to glucose in the medium.

The eventual origin of the Cit^+ trait was contingent upon one or more earlier mutations in the population's history. When evolution was replayed from clones isolated at various time-points, clones from later generations had a significantly greater propensity to become Cit^+ than did the ancestor and other early clones (Blount et al 2008). This finding implied that a genetic background had evolved that “potentiated” the evolution of this trait. In principle, this effect could involve two distinct mechanisms. One possibility is that the rate of mutation, or certain types of mutation, increased such that the required event was more likely to occur in later generations. The other possibility is an epistatic interaction, such that the phenotypic expression of a mutation that produced the Cit^+ trait required one or more preceding mutations. Here we report the results of extensive whole-genome re-sequencing that allowed us to reconstruct the population's history, identify and manipulate mutations underlying the Cit^+ phenotype, and elucidate the mechanistic basis for this novel trait.

GENOME SEQUENCING AND PHYLOGENETIC HISTORY

We sequenced 29 clones sampled at various generations from population Ara-3, including 9 Cit⁺ and 3 Cit⁻ clones known to be potentiated (Blount et al 2008) (Table 3.1). Mutations including single-nucleotide polymorphisms (SNPs) as well as deletions, insertions, and other rearrangements were identified by comparing DNA sequence reads to the complete genome of the ancestral strain, REL606 (Jeong et al 2009).

We constructed the population's phylogenetic history from the presence-absence matrices of all mutations in the sequenced genomes. Figure 3.1 shows that the population was polymorphic for most of its history. At least four major clades, each represented by multiple clones, emerged before 20,000 generations. One of them, which we call UC (Unsuccessful Clade), was not seen after 15,000 generations. The remaining clades C1, C2, and C3 coexisted through the evolution of Cit⁺ and beyond. C1 includes four sequenced clones, the earliest from generation 25,000, although a molecular clock (Figure 3.2) implies that C1 diverged from the ancestor of C2 and C3 before 15,000 generations. C2 and C3 had diverged by generation 20,000. C3 includes both Cit⁻ clones and all the Cit⁺ clones, although the first Cit⁺ cells did not arise until ~31,000 generations. Two mechanisms might explain the long-term coexistence of the Cit⁻ lineages. First, they may all have acquired beneficial mutations without one gaining enough advantage to fix (Fogle et al 2008). Alternatively, they might have filled subtly different ecological niches, such as through the differential production of and growth on secreted metabolites (Treves et al 1998, Rozen et al 2005).

Supplementary Table 3.1 Historical Ara-3 Clones Subjected to Whole Genome Sequencing		
Generation	Clones	Clade
0	REL606	N/A
2,000	REL1166A	?
5,000	ZDB409	?
10,000	ZDB429	UC
15,000	ZDB446	UC
20,000	ZDB458	(C1,C2)
	ZDB464*	(C1,C2)
	ZDB467	(C1,C2)
25,000	ZDB477	C1
	ZDB483	C3
30,000	ZDB16	C1
	ZDB357	C2
31,500	ZDB199*	C1
	ZDB200	C2
	ZDB564	Cit ⁺
32,000	ZDB30*	C3
	ZDB172	Cit ⁺
32,500	ZDB158	C2
	ZDB143	Cit ⁺
33,000	CZB199	C1
	CZB152	Cit ⁺
	CZB154	Cit ⁺
34,000	ZDB83	Cit ⁺
	ZDB87	C2
36,000	ZDB96	Cit ⁺
	ZDB99	C1
38,000	ZDB107	Cit ⁺
	ZDB111	C2
40,000	REL10979	Cit ⁺
	REL10988	C2

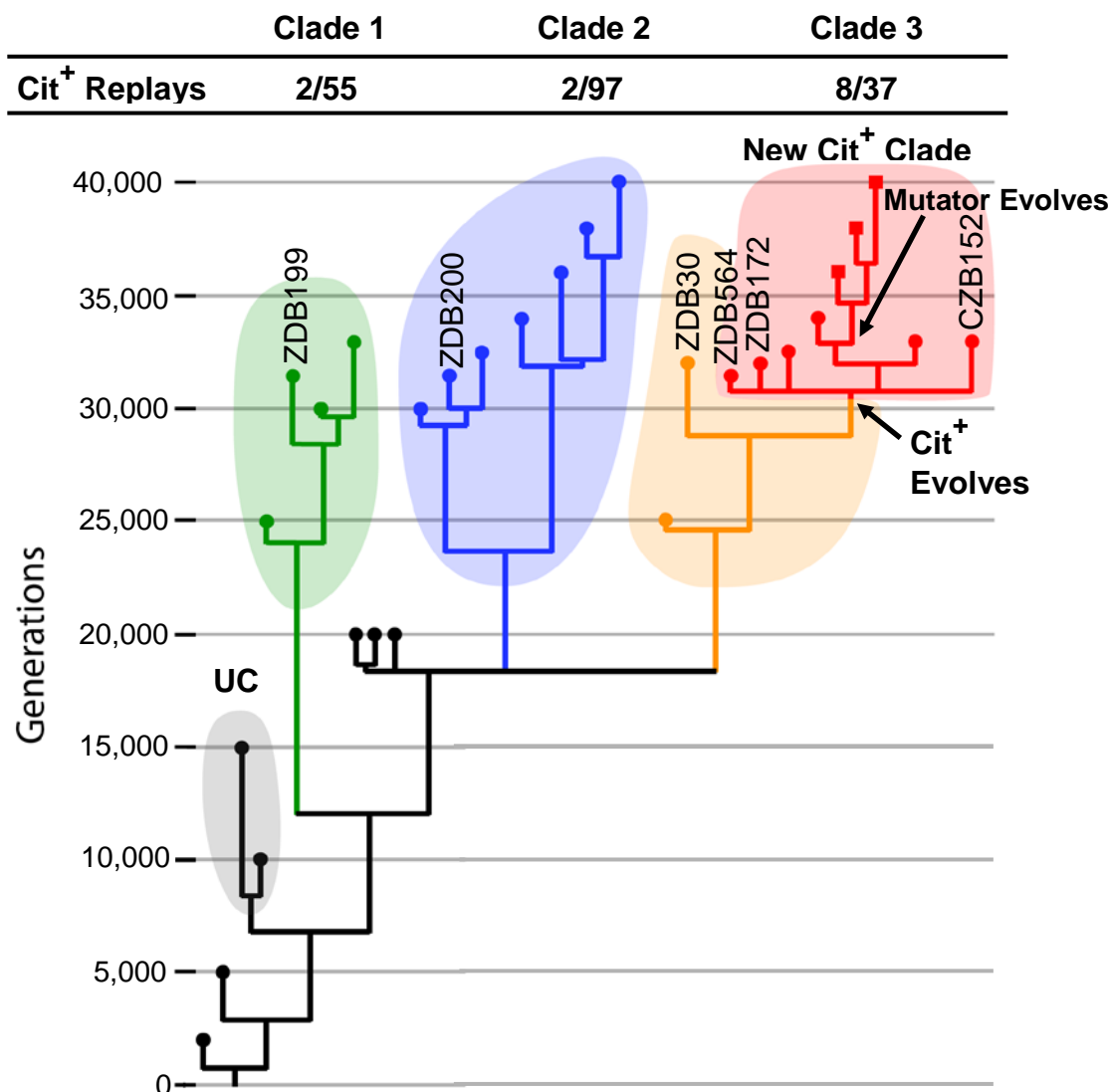


Figure 3.1 | Phylogeny of Ara-3 population. Symbols at branch tips indicate placement of 29 sequenced clones; identifying labels are shown for clones mentioned in main text and figures. Shaded areas and colored symbols identify major clades. Fractions above the tree show the number of clones identified as belonging to the clade that yielded Cit⁺ mutants during replay experiments (numerator) and the corresponding total number of clones used in those experiments (denominator).

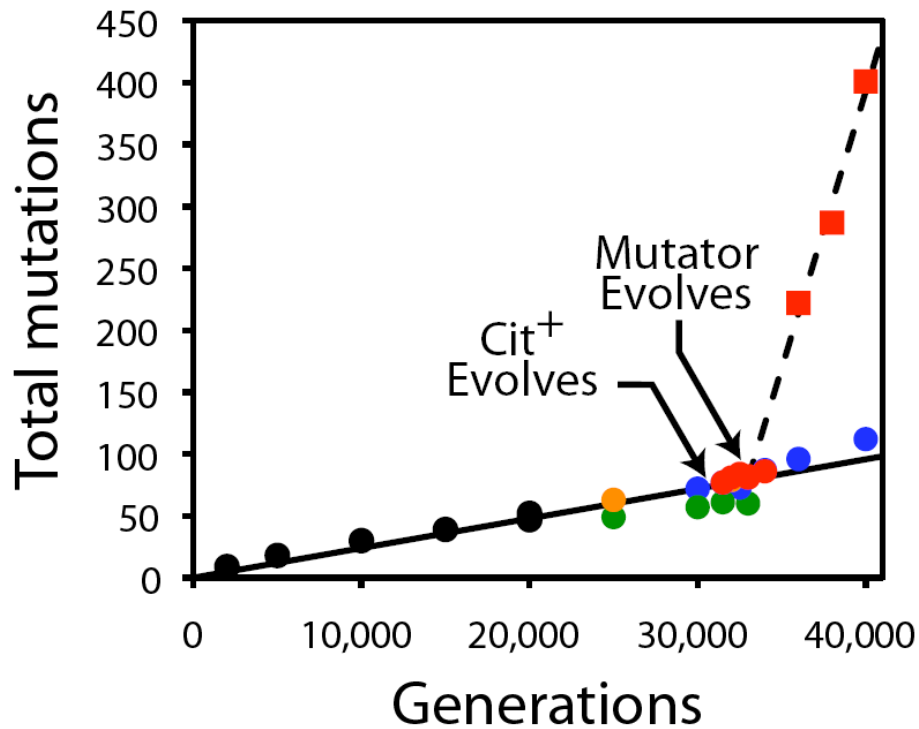


Figure 3.2 | Total number of mutations relative to the ancestor. Number of mutations identified in each sequenced genome relative to the ancestor. Colors correspond to clade of origin (green = Clade 1, blue = Clade 2, orange = Clade 3, red = Cit⁺). Circles are non-mutator genomes, while squares correspond to genomes with mutator genotypes. Lines are maximum likelihood estimates of Poisson rates of mutation accumulation in non-mutator (solid) and mutator (dashed) genomes from a model used to estimate branching times.

The sequenced Cit⁺ clones from generation 36,000 and later have a SNP in the *mutS* gene that produces a premature stop codon and truncates the MutS protein, thereby causing a defect in methyl-directed mismatch DNA repair (Glickman and Radman 1980). We sequenced *mutS* from three Cit⁺ clones from each of generations 33,000, 34,000, 35,000, 36,000, 37,000 and 38,000, and found this SNP in 0, 0, 2, 3 and 3 clones, respectively, indicating that the mutation arose after the origin of the Cit⁺ lineage. None of the 20 Cit⁻ genomes have this mutation. As a consequence, Cit⁺ genomes from later generations accumulated SNPs much faster than the Cit⁻ and early Cit⁺ genomes (Figure 3.2). Mutator phenotypes evolved in several other populations in the LTEE (Sniegowski et al 1997, Barrick et al 2009), so this change was not unique to the Cit⁺ lineage.

EVOLUTION OF THE CIT⁺ FUNCTION

The evolution of the Cit⁺ trait involved three successive processes: potentiation, actualization and refinement. Potentiation refers to the evolution of a genetic background in which the Cit⁺ function became accessible by mutation; the ancestor's rate of mutation to Cit⁺ was inaccessibly low, on the order of 10^{-13} per cell-generation or lower (Blount et al 2008). An extremely weak Cit⁺ variant had emerged by 31,500 generations, which represents the actualization step. Finally, the new function was refined, allowing the efficient exploitation of citrate, the rise of the Cit⁺ subpopulation to numerical dominance, and the substantial expansion

of the population size. In the following sections, we identify and examine the effects of mutations involved in these processes. We begin with the second and third processes, before turning to potentiation.

Actualization of the Cit⁺ function

One reason that *E. coli* cannot grow aerobically on citrate is its inability to transport citrate (Hall 1982, Lara and Stokes 1952, Pos et al 1998). The origin of the Cit⁺ phenotype therefore required expression of a citrate transporter. All 9 sequenced Cit⁺ genomes have multiple (2 to 9) tandem copies of a 2933-bp segment that includes part of the citrate fermentation (*cit*) operon (Figure 3.3a). The amplified segment contains two genes: *rna* encodes rnaase I (Zhu and Deutscher 1992), and *citT* encodes a broad-spectrum C4-di- and tri-carboxylic acid transporter that functions in fermentation as a citrate-succinate antiporter (Pos et al 1998). The boundary upstream of *citT* is in the 3' end of the *citG* gene, which encodes triphosphoribosyl-dephospho-CoA synthase, while the boundary downstream of *rna* is in the 5' end of *rnk*, which encodes a regulator of nucleoside diphosphate kinase (Shankar et al 1995). This tandem amplification is not present in the ancestor or any of the 20 sequenced Cit⁻ genomes, and it is only found in population samples after the evolution of the Cit⁺ lineage (Table 3.2). PCR screens also failed to detect this segment in any of 27 Cit⁻ clones from generations 33,000 through 40,000, whereas it was found in all 33 Cit⁺ clones from the same generations (Table 3.3).

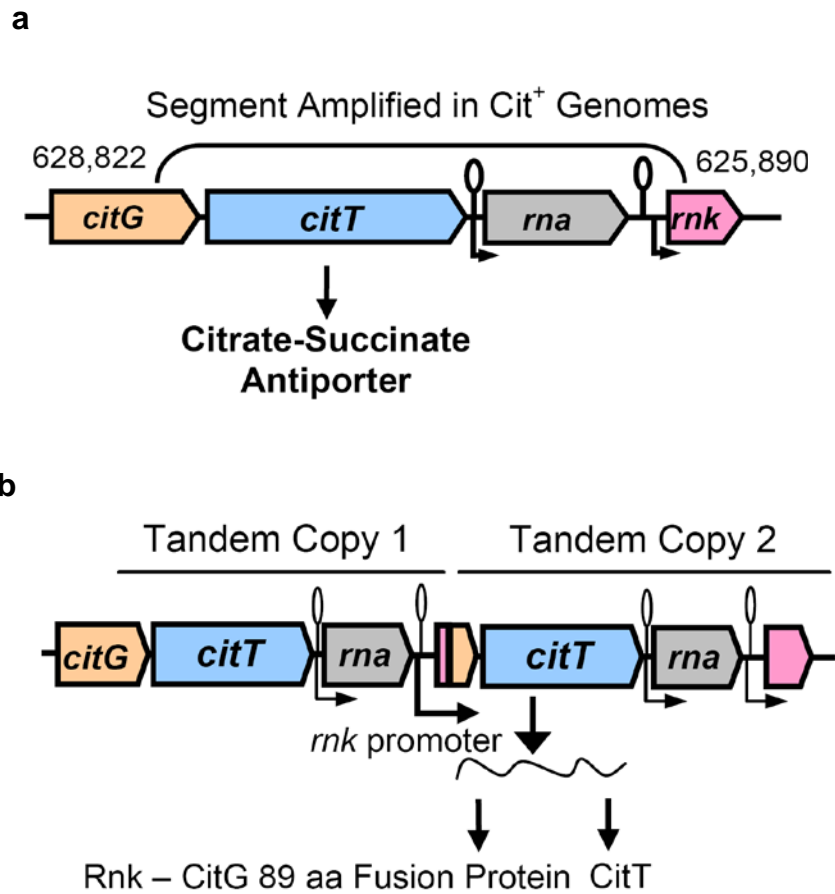


Figure 3.3 | Tandem amplification in Cit⁺ genomes. **a.** Ancestral arrangement of *citG*, *citT*, *rna*, and *rnk* genes. **b.** Altered spatial and regulatory relationships generated by the tandem amplification.

Supplementary Table 3.2 Results of PCR screens on whole-population samples for <i>cit</i> amplification		
Generation	Replicate	<i>cit</i> Amplification Detected?
25,000	A	–
	B	–
	C	–
30,000	A	–
	B	–
	C	–
31,500	A	–
	B	–
	C	–
32,000	A	–
	B	–
	C	+
32,500	A	–
	B	–
	C	–
33,000	A	+
	B	+
	C	+
33,500	A	+
	B	+
	C	+
34,000	A	+
	B	+
	C	+
36,000	A	+
	B	+
	C	+

Table 3.3 Detection of <i>cit</i> amplification in Ara-3 clones			
Generation	Clone	Phenotype	<i>cit</i> amplification Detected?
31,500	ZDB25	-	No
	ZDB26	-	No
	ZDB27	-	No
	ZDB564	+	Yes
	ZDB565	+	Yes
	ZDB566	+	Yes
32,000	ZDB28	-	No
	ZDB29	-	No
	ZDB30	-	No
	ZDB172	+	Yes
	ZDB173	+	Yes
	ZDB179	+	Yes
32,500	ZDB31	-	No
	ZDB32	-	No
	ZDB33	-	No
	ZDB143	+	Yes
	ZDB144	+	Yes
	ZDB145	+	Yes
33,000	CZB199	-	No
	CZB204	-	No
	CZB205	-	No
	CZB151	+	Yes
	CZB152	+	Yes
	CZB154	+	Yes
34,000	ZDB83	-	No
	ZDB84	-	No
	ZDB85	-	No
	ZDB86	+	Yes
	ZDB87	+	Yes
	ZDB88	+	Yes
35,000	ZDB89	+	Yes
	ZDB90	+	Yes
	ZDB91	+	Yes

Table 3.3 Detection of <i>cit</i> amplification in Ara-3 clones			
Generation	Clone	Phenotype	<i>cit</i> amplification Detected?
	ZDB92	-	No
	ZDB93	+	Yes
	ZDB94	+	Yes
36,000	ZDB95	+	Yes
	ZDB96	+	Yes
	ZDB97	+	Yes
	ZDB98	+	Yes
	ZDB99	-	No
	ZDB100	-	No
37,000	ZDB101	+	Yes
	ZDB102	+	Yes
	ZDB103	-	No
	ZDB104	-	No
	ZDB105	-	No
	ZDB106	-	No
38,000	ZDB107	+	Yes
	ZDB108	+	Yes
	ZDB109	+	Yes
	ZDB110	-	No
	ZDB111	-	No
	ZDB112	+	Yes
40,000	REL10979	+	Yes
	REL10980	+	Yes
	REL10981	+	Yes
	REL10988	-	No
	REL10989	-	No
	REL10990	-	No

Amplification mutations can alter the spatial relationship between structural genes and regulatory elements, potentially causing altered regulation and novel traits (Usakin et al 2005, Adam et al 1993, Bock and Timmis 2008, Whoriskey et al 1987). The structure of the *cit* amplification suggests that the Cit^+ trait arose from an amplification-mediated promoter capture (Figure 3.3b, Figure 3.4). The amplification joined upstream *rnk* and downstream *citG* fragments, producing an *rnk-citG* hybrid gene expressed from the upstream *rnk* promoter. Because the *citT* and *citG* genes are normally monocistronic, the downstream copy of *citT* should therefore be co-transcribed with the hybrid gene. If the *rnk* promoter directs transcription under oxic conditions, then the new *rnk-citT* regulatory module might allow Cit^+ expression during aerobic metabolism and thereby confer a Cit^+ phenotype (Pos et al 1998).

To test this hypothesis, we first examined the capacity of the *rnk-citT* module to support *citT* expression under oxic conditions. We engineered a low-copy (1-2 per cell), *lux*-based reporter plasmid, $P_{lux}::rnk-citT$, that has an *rnk-citT* module in which *citT* was replaced by the *luxCDABE* reporter operon. We constructed two additional plasmids, $P_{lux}::rnk$ and $P_{lux}::citT$, in which the reporter was under the control of the native upstream regulatory regions of *rnk* and *citT*, respectively. We transformed each plasmid into REL606, the ancestral strain; ZDB30, a potentiated C3 clone from generation 32,000; and ZDB172, a weakly Cit^+ clone from generation 32,000. We measured Lux expression during growth and stationary phase under oxic conditions (Figure 3.5). The native *citT* regulatory region showed no Lux expression (above background) in any strain, indicating that *citT* is normally silent under oxic conditions. The native *rnk* regulatory region expressed Lux, with a peak around the transition into stationary phase. Lux expression

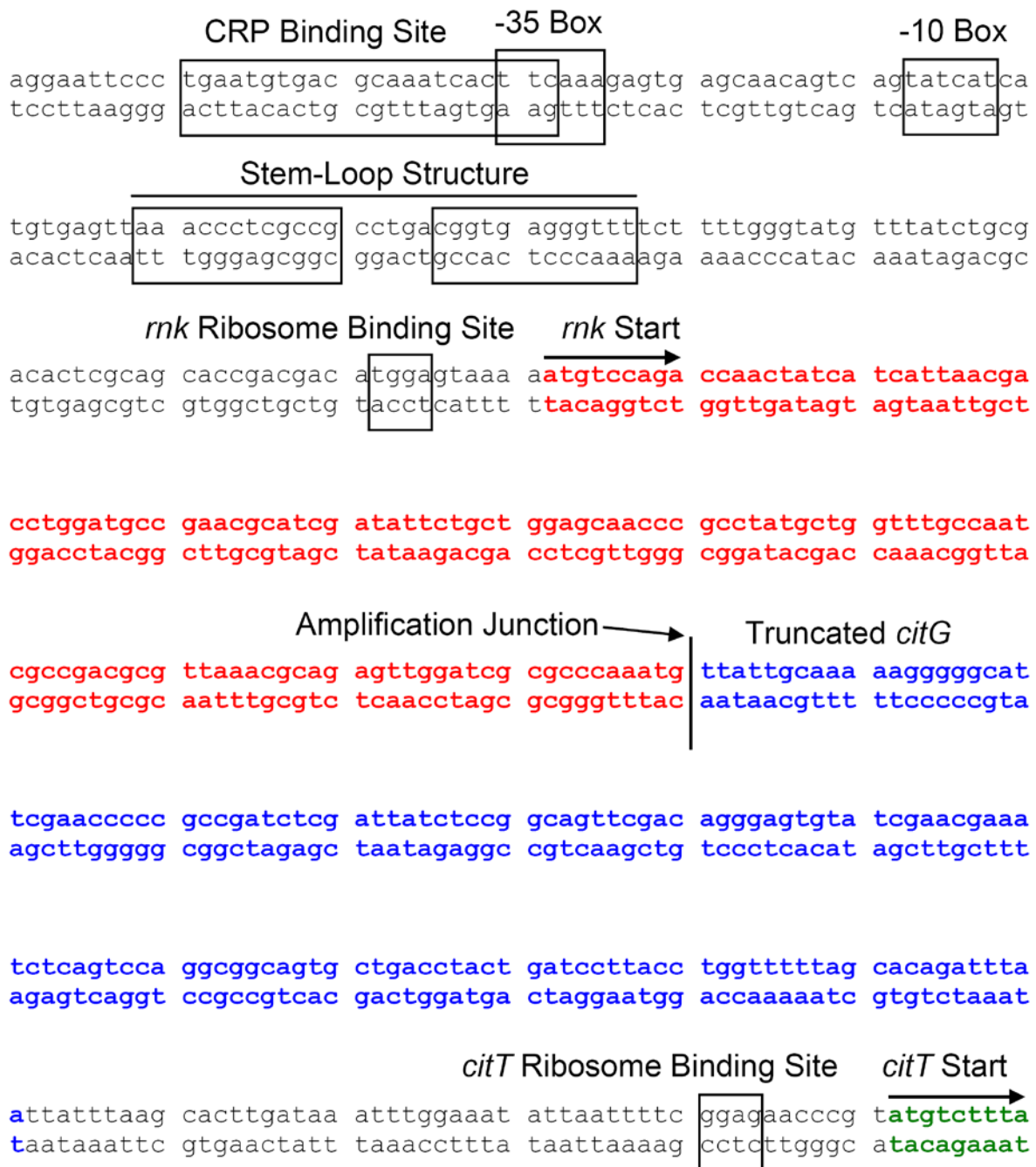


Figure 3.4 | Annotation of sequence adjacent to the boundary of the *cit* amplification found in *Cit*⁺ genomes.

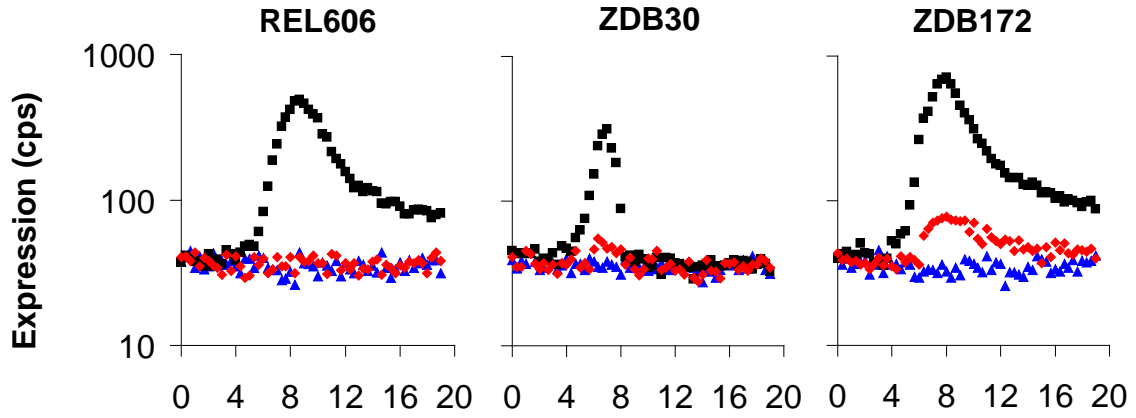


Figure 3.5 | Expression levels from native *citT*, native *rnk*, and evolved *rnk-citT* regulatory regions during aerobic metabolism. Average time course of luciferase expression for the ancestral strain REL606 (left) and evolved clones ZDB30 (center) and ZDB172 (right), each transformed with reporter plasmids $P_{lux}::citT$ (blue), $P_{lux}::rnk$ (black) and $P_{lux}::rnk-citT$ (red). Each curve shows the average of four replicates.

from the evolved *rnk-citT* module was much weaker, but there were small spikes in expression in both ZDB30 and ZDB172 coincident with peak expression from the native *rnk* regulatory region. These results confirm that the *rnk-citT* module can support *citT* expression during aerobic metabolism.

We also tested whether the *rnk-citT* module could confer a Cit⁺ phenotype by inserting a single copy of an amplification fragment containing the *rnk* promoter immediately upstream of the chromosomal copy of *citT* (Figure 3.6a) in potentiated clone ZDB30. This construct, ZDB595, has a Cit⁺ phenotype, but it is extremely weak, similar to the earliest evolved Cit⁺ variants (Figure 3.6b). In the same medium used in the evolution experiment, ZDB595 experienced a lag of >60 hours after glucose depletion, followed by a short period of abortive and inconsistent growth on citrate (Figure 3.6b,c). These data imply that the initial Cit⁺ type was too weak to permit exploitation of citrate under the daily-transfer regime of the long-term experiment. Nonetheless, ZDB595 had a small (1.0%) but significant fitness advantage over ZDB30 in competition assays performed in the same environment ($n = 10$, $t = 3.09$, two-tailed $P = 0.0128$).

Refinement of the Cit⁺ function

Given the extremely weak initial Cit⁺ phenotype, the rise of Cit⁺ cells to high frequency required further mutations to refine the new function. These mutations are evidenced by the substantial improvement on citrate of variants between 31,500 and 33,000 generations (Figure 3.7), shortly before the population expansion. To understand this refinement, we examined the genomes of

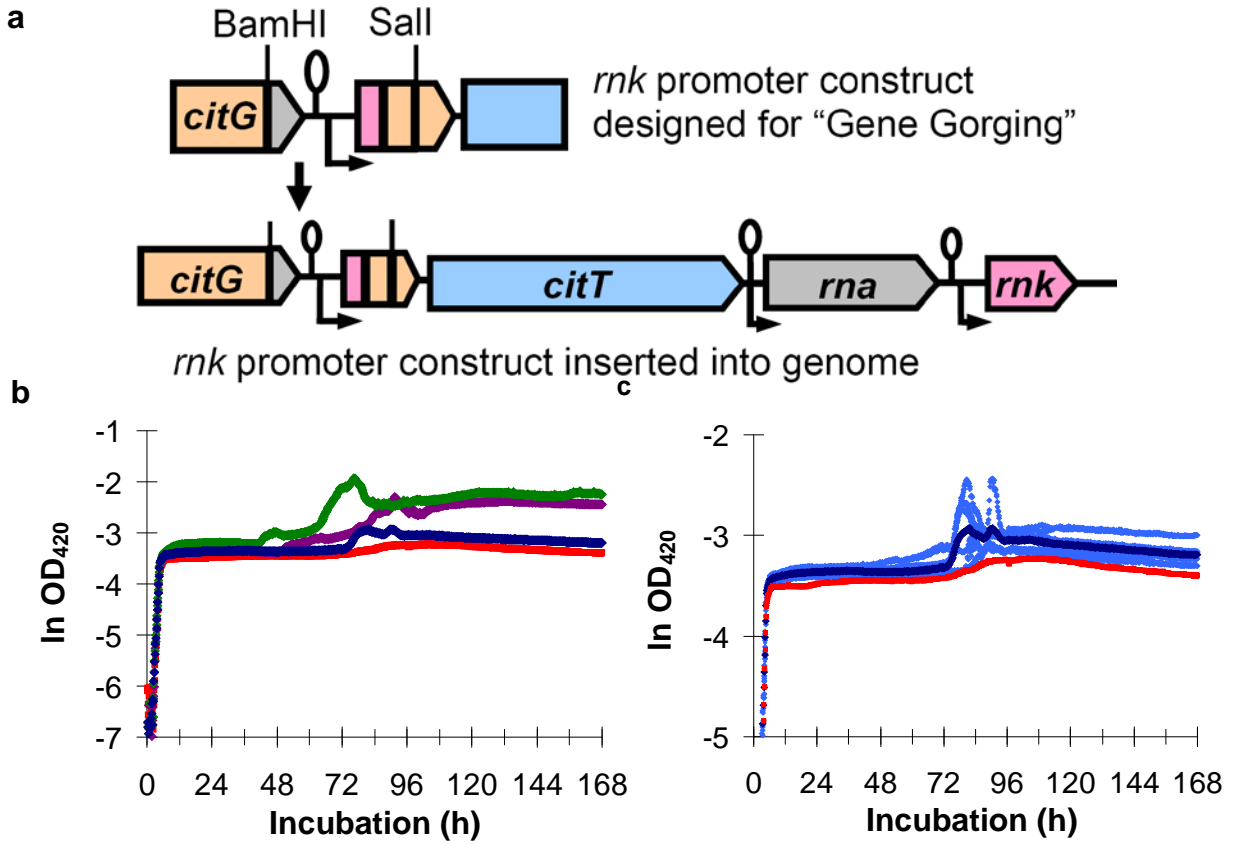


Figure 3.6 | New *rnk-citT* module confers Cit⁺ phenotype in potentiated background. a. Engineered construct containing the *cit* amplification junction with the *rnk* promoter, and the gene arrangement following insertion of the construct into the chromosome. **b.** Average growth trajectories in DM25 of potentiated clone ZDB30 (red), its isogenic construct with chromosomal integration of the *rnk-cit* module ZDB595 (blue), 31,500-generation Cit⁺ clone ZDB564 (purple), and 32,000-generation Cit⁺ clone ZDB172 (green). **c.** Growth trajectories of ZDB595 compared to its parent. Light blue trajectories show heterogeneity among replicate cultures of ZDB595; dark blue and red are averages of trajectories for ZDB595 and its parent, ZDB30, as in panel b (except on different scale).

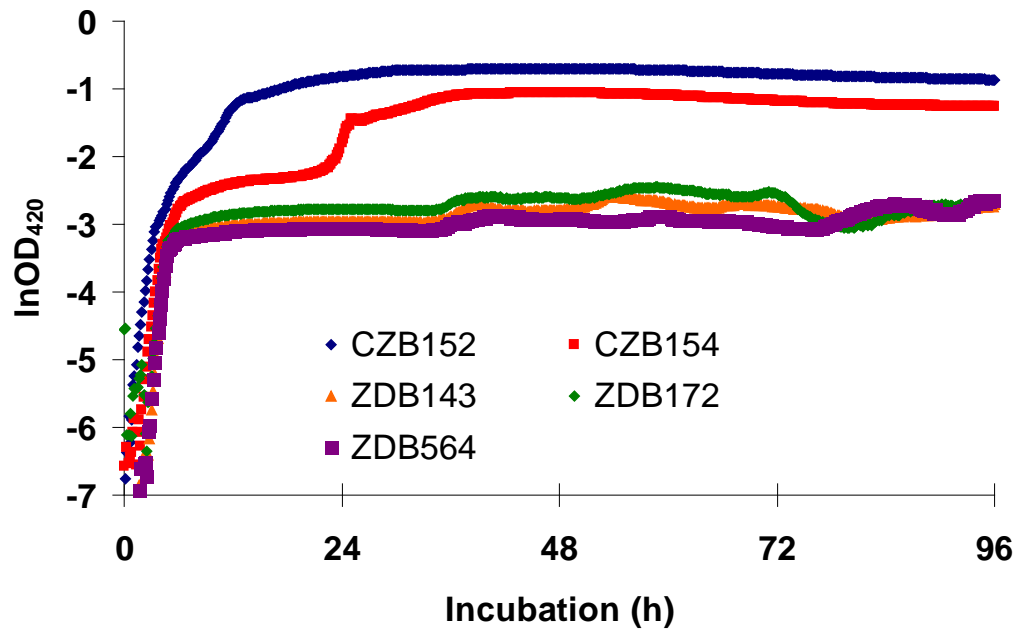


Figure 3.7 | Growth of early Cit⁺ clones in DM25.

Early Cit⁺ clones from population Ara-3 improved greatly in their capacity to grow on the citrate in DM25 over time. Cit⁺ clones from generations 31,500 (ZDB564), 32,000 (ZDB172), and 32,500 (ZDB143) showed little growth on citrate even after 96 h, while two clones from generation 33,000 (CZB152 and CZB154) achieved substantial growth on citrate within 24 h. This improvement in citrate utilization allowed Cit⁺ clones to rise from a minority of the population to numerical dominance.

five Cit⁺ clones from before the population expansion, including one each from generations 31,500, 32,000 and 32,500 and two from generation 33,000. We focused on those mutations on the line of descent for the Cit⁺ subpopulation, where that line is defined by the presence of the same mutations in Cit⁺ clones from generations 34,000, 36,000 and 38,000 (Tables 3.4 – 7). Among those mutations, two SNPs and one IS-element insertion were present in only one of the 33,000-generation genomes. The IS insertion and one SNP appear unrelated to growth on citrate. The remaining SNP is in the regulatory region of *dctA*, which encodes a transporter of succinate and other C4-di-carboxylic acids (Janausch et al 2002). This mutation may improve recovery of succinate exported in exchange for citrate. However, the other 33,000-generation clone showed stronger growth on citrate (Figure 3.7). Thus, the *dctA* mutation does not appear to be responsible for the population expansion.

These early Cit⁺ genomes also underwent a progressive increase in *cit* copy number. The earliest one had a tandem duplication, whereas later genomes had a tandem duplication within a larger tandem duplication, a four-copy tandem array, a three-copy tandem array within a larger duplication, and a nine-copy tandem array, respectively. Changes in amplification copy number readily occur by recombination, and have been implicated in the refinement of other weak functions (Andersson et al 1998, Reams et al 2010). These early changes all increased the number of *rnk-citT* modules (Figure 3.8a, Table 3.8) and should therefore increase expression of CitT. To test whether this increase caused the population expansion, we cloned the *rnk-citT* module (Figure 3.8b) into the high-copy plasmid pUC19 (Yanisch-Perron et al 1985) and moved it into the 32,000-generation potentiated clone ZDB30. The resulting strain, ZDB612, is strongly Cit⁺, rapidly transitions from glucose to citrate, and grows similarly to the 33,000-generation

Table 3.4 Point mutations in early Cit ⁺ genomes										
Genome Position	Nucleotide Change	Type	Amino-acid Change	Gene	Product	Clone (generation)	ZDB564 (31,500)	ZDB172 (32,000)	ZDB143 (32,500)	CZB152 (33,000)
447290	T → C	NC	N/A	<i>tesB/ybaY</i>	Acyl-CoA thioesterase II / predicted outer membrane lipoprotein					
1399744	T → C	NS	I228V	<i>abgB</i>	Predicted peptidase / aminobenoyl-glutamate utilization protein					
2241625	C → A	NS	A259S	<i>ccmH</i>	Heme lyase subunit					
2443160	C → A	S	N/A	<i>glk</i>	Glucokinase					
3612959	C → T	NC	N/A	<i>dctA/yhjK</i>	C4-dicarboxylate transporter / Predicted diguanylate cyclase					

Mutations shown are those that were not uniformly found in all five Cit⁺ clones under study. Red fill indicates presence of mutation. Crosshatching further indicates mutation was also present in Cit⁺ clones from generations 34,000, 36,000, and 38,000. Gray fill indicates absence of mutation.

Table 3.5 IS-element insertions in early Cit ⁺ genomes									
Genome Position	Element	Gene	Product	Clone (generation)	ZDB564 (31,500)	ZDB172 (32,000)	ZDB143 (32,500)	CZB152 (33,000)	CZB154 (33,000)
595335	IS 150	<i>fepA/fes</i>	Outer membrane transporter/Ferric enterobactin esterase						
620126	IS 150	<i>dsbG</i>	Periplasmic disulfide isomerase/Thiol-disulphide oxidase						
1028311	IS 150	<i>uup</i>	Fused predicted transporter subunits of ABC superfamily: ATP-binding components						
2322345	IS 186	<i>menC</i>	O-succinylbenzoate						
2877315	IS 150	<i>kduD</i>	2-deoxy-D-fluconate/3-dehydrogenase						
4252526	IS 150	<i>uvrA</i>	Excinuclease ABC subunit A						

Mutations shown are those that were not uniformly found in all five Cit⁺ clones under study. Red fill indicates presence of mutation. Crosshatching further indicates mutation was also present in Cit⁺ clones from generations 34,000, 36,000, and 38,000. Gray fill indicates absence of mutation.

Table 3.6 Deletions in early Cit ⁺ genomes									
Genome Start	Genome End	Size (bp)	Genes Deleted	Clone (generation)	ZDB564 (31,500)	ZDB172 (32,000)	ZDB143 (32,500)	CZB152 (33,000)	CZB154 (33,000)
590472	599560	9088	<i>hokE, insL-3, entD, fepA, fes, ybdZ</i>						
590472	595335	4863	<i>hokE, insL-3, entD, fepA</i>						
1345210	1345211	1	<i>yciR/rnb</i>						
3786737	3786738	1	Noncoding						

Mutations shown are those that were not uniformly found in all five Cit⁺ clones under study. Red fill indicates presence of mutation. Crosshatching further indicates mutation was also present in Cit⁺ clones from generations 34,000, 36,000, and 38,000. Gray fill indicates absence of mutation.

Table 3.7 Amplifications in early Cit ⁺ genomes									
Genome Start	Genome End	Size (bp)	Genes Duplicated	Clone (Generation)	ZDB564 (31,500)	ZDB172 (32,000)	ZDB143 (32,500)	CZB152 (33,000)	CZB154 (33,000)
12452	12455	3 (x2)	<i>dnaK</i> (internal fragment)						
599561	666130	66569 (x2)	<i>[entF], fepE, fepC, fepG, fepD, ybdA, fepB, entC, entE, entB, entA, ybdB, cstA, cstA, ybdD, ybdH, ybdL, ybdM, ybdN, insB-8, insA-8, dsbG, ahpC, ahpF, ybdQ, ybdR, rnk, rna, citT, citG, citX, citF, citE, citD, citC, insB-9, insA-9, citA, citB, dcuC, crcA, cspE, ccrB, ybeM, tatE, lipA, ybeF, lipB, ybeD, dacA, rlpA, mrdB, mrdA, ybeA, ybeB, phpB, nadD, holA, rlpB, leuS, ybeL, ybeQ, ybeQ, ybeR, ybeV, hscC, rihA, gltL, insJ-2, insK-2, [gltL]</i>						
619885	634746	14861 (x3)	<i>[dsbG], ahpC, ahpF, ybdQ, ybdR, rnk, rna, citT, citG, citX, citF, citE, citD, citC, insB-9, insA-9</i>						
1729052	1995783	266731 (x3)	<i>[ydhV], ydhY, ydhZ, pykF, lpp, ynhG, sufE, sufS, sufD, sufC, sufB, sufA, ydiH, ydil, ydiJ, ydiK, ydiL, ydi</i>						

Table 3.7 | Amplifications in early Cit⁺ genomes

Genome Start	Genome End	Size (bp)	Genes Duplicated	Clone (Generation)	ZDB564 (31,500)	ZDB172 (32,000)	ZDB143 (32,500)	CZB152 (33,000)	CZB154 (33,000)
			M, ydiM, ydiN, ydiB, aroD, ydiF, ydiO, ydiP, ydiQ, ydiR, ydiS, ydiT, ydiD, ppsA, ydiA, aroH, ydiE, ydiU, ydiV, nlpC, btuD, btuE, btuC, himA, pheT, pheS, pheM, rplT, rpml, infC, thrS, arpB, a rpB, ECB_01690, ydiY, pfkB, ydiZ, yniA, yniB, yniC, ydjM, ydjN, ydjO, ce dA, katE, ydjC, celF, cel D, celC, celB, celA, osmE, nadE, ydjQ, ydjR , spy, astE, astB, astD, a stA, astC, xthA, ydjX, yd jY, ydjY, ydjZ, ynjA, ynjB, ynjC, y njD, ynjE, ynjF, nudG, ynjH, gdhA, ynjI, topB, selD, ydjA, sppA, a nsA, pncA, ydjE, ydjF, y djG, ydjH, ydjI, ydjJ, ydjK , ydjL, yeaC, yeaA, gapA, yeaD, yeaE, mipA, yeaG, yeaH, yeal, yeaJ, yeaK, ECB_01757, yeaL, yeaM, yeaN, yeaO, yoaF, yeaP, yeaQ, yoa G, yeaR, insL-4, yeaS, yeaT, yeaU, yeaV, yeaW, yeaX, rnd, fadD, yeaY, yeaZ, yoaA, yoaB, yoaC, yoaH, pabB,						

Table 3.7 Amplifications in early Cit ⁺ genomes									
Genome Start	Genome End	Size (bp)	Genes Duplicated	Clone (Generation)	ZDB564 (31,500)	ZDB172 (32,000)	ZDB143 (32,500)	CZB152 (33,000)	CZB154 (33,000)
			yeaB,sdaA, yoaD, yoaE, manX, manY, manZ,yobD, yebN, rrmA, cspC, yobF, yebO,yobG, ECB_0179 7, kdgR, yebQ, htpX,prc, proQ, yebR, yebS, yebT, yebU,yebV, yebW, pphA, yebY, yebZ, yobA, holE, yobB, exoX, ptrB, yebE, yebF, yebG, purT, eda, edd, zwf, yebK, pykA, msbB, yebA, znuA, znuC, znuB, ruvB ,ruvA, yebB, ruvC, yebC, ntpA, aspS, yecD, yecE, yecN, yecO, yecP, torZ,torY, cutC, yecM, argS, yecT, flhE, flhA,flhB, cheZ ,cheY, cheB, cheR, tap, tar, cheW, cheA, motB, motA, flhC, flhD,yecG, otsA, otsB, araH, araG, araF, yecl, yecJ, yecR, ftn, yecH, tyrP, yecA, leuZ, cystT, glyW, pgsA, uvrC, yvrY, insA-13, insB-13, yedU, yedV, yedW, yedX, yedY, yedZ, yodA, yodB, serU, yeel, asnT, yeeJ, yeeL, yeeL, shiA, amn, yeeN						

Table 3.7 | Amplifications in early Cit⁺ genomes

Genome Start	Genome End	Size (bp)	Genes Duplicated	Clone (Generation)	ZDB564 (31,500)	ZDB172 (32,000)	ZDB143 (32,500)	CZB152 (33,000)	CZB154 (33,000)
3268052	2086894	3268052 (x2)	<i>rpmA, rplU, ispB, nlp, murA, yrbA, yrbB, yrbC, yrbD, yrbE, yrbF, yrbG, yrbH, yrbl,yrbK, yhbN, yhbG, rpoN, yhbH, ptsN, yhbJ, ptsO, yrbL, mtgA, yhbL, arcB,yhcC gltB, gltD, yhcG, ECB_03080,yhcH, nanK, nanE, nanT, nanA, nanR,dcuD, sspB, sspA, rpsl, rplM, yhcM,yhcB, degQ degS, mdh, argR, yhcN ,yhcO, yhcP, yhcQ, yhcR, yhcS, tldD, yhdP, rng, maf, mreD, mreC, mreB, yhdA, yhdH, accB, accC, yhdT, panF, prmA, yhdG, fis, yhdJ, yhdU, envR, acrE,acrF, yhdV, yhdW, yhdX, yhdY, yhdZ,rrfF, thrV, rrfD, rrlD, alaU, ileU, rrsD,yrdA, yrdB, aroE, yrdC, yrdD, smg, smf, def, fmt, rrmB, trkA, mscL, yhdL, zntR,yhdN, rplQ, rpoA, rpsD, rpsK, rpsM,rpmJ, prlA, rplO, rpmD, rpsE, rplR, rplF,rpsH, rpsN, rplE, rplX, rplN, rpsQ, rpmC,rplP, rpsC, rplV, rpsS, rplB, rplW, rplD, rplC, rpsJ, pioO, gspA,</i>						

Table 3.7 Amplifications in early Cit ⁺ genomes								
Genome Start	Genome End	Size (bp)	Genes Duplicated	Clone (Generation)	ZDB564 (31,500)	ZDB172 (32,000)	ZDB143 (32,500)	CZB152 (33,000)
			<i>gspC, gspD, gspE, gspF, gspG, gspH, gspI, gspJ, gspK, gspL, gspM, gspO, bfr, bfd, chiA, tuf, fusa, rpsG, rpsL, yheL, yheM, yheN, yheO, fkpA, slyX, slyD, kefB, yheR, yheS, yheT, yheU, prkB, yhfA, crp, yhfK, argD, pabA, fic, yhfG, ppiA, yhfC, nirB, nirD, nirC, cysG, yhfL, yhfM, yhfN, frlC, yhfQ, yhfR, yhfS, insB-23, insA-23, yhfT, yhfU, yhfV, yhfW, yhfX, yhfY, yhfZ, trpS, gph, rpe, dam, damX, aroB, aroK, hofQ, yrfA, yrfB, yrfC, yrfD, mrcA, yrfE, yrfF, yrfG, hslR, hslO, yhgE, pckA, envZ, ompR, greB, yhgF, feoA, feoB, yhgG, yhgA, bioH, yhgH, yhgI, gntT, malQ, malP, malT, rtcA, rtcB, rtcR, glpR, glpG, glpE, glpD, yzgL, ECB_03279, yzgL, glgP, glgA, glgC, glgX, glgB, asd, yhgN, gntU, gntK, gntR, yhhW, yhhX, yhhY, yhhZ, yrhA, yrhB, ggt, yhhA, ugpQ, ugpC, ugpE, ugpA, ugp, livF, livG, livM, livH, livK, yhhK, livJ, rpoH, ftsX,</i>					

Table 3.7 Amplifications in early Cit ⁺ genomes								
Genome Start	Genome End	Size (bp)	Genes Duplicated	Clone (Generation)	ZDB564 (31,500)	ZDB172 (32,000)	ZDB143 (32,500)	CZB152 (33,000)
			<i>ftsE, ftsY,yhhF, yhhL, yhhM, yhhN, zntA, sirA,yhhQ, dcrB, yhhS, yhhT, acpT, nika,nikB, nikC, nikD, nikE, nikR, rhsB, yhhH, yrhC, yhhI, yhhJ, yhiH, yhiI, yhiJ, yhiKL,yhiM, yhiN, pitA, yhiO, uspA, yhiP, yhiQ,prlC, yhiR, gor, arsR, arsB, arsC, yhiS,slp, yhiF, yhiD, hdeB, hdeA, hdeD, yhiE,yhiU, yhiV, yhiW, gadX, gadA, yhjA, treF, yhjB, yhjC, yhjD, yhjE, yhjG, yhjH, kdgK,yhjJ, dctA, yhjK, bcsC, bcsZ, yhjN, yhjO,yhjQ, yhjR, yhjS, yhjT, yhjU, ldrD, ldrD,ldrD, yhjV, dppF, dppD, dppC, dppB, dppA, proK, yhjW, yhjX, yhjY, tag, yiaC, bisC, yiaD, tkrA, yiaF, yiaG, cspA, hokA, insJ-4, insK-4</i>					

Mutations shown are those that were not uniformly found in all five Cit⁺ clones under study. Red fill indicates presence of mutation. Crosshatching further indicates mutation was also present in Cit⁺ clones from generations 34,000, 36,000, and 38,000. Gray fill indicates absence of mutation.

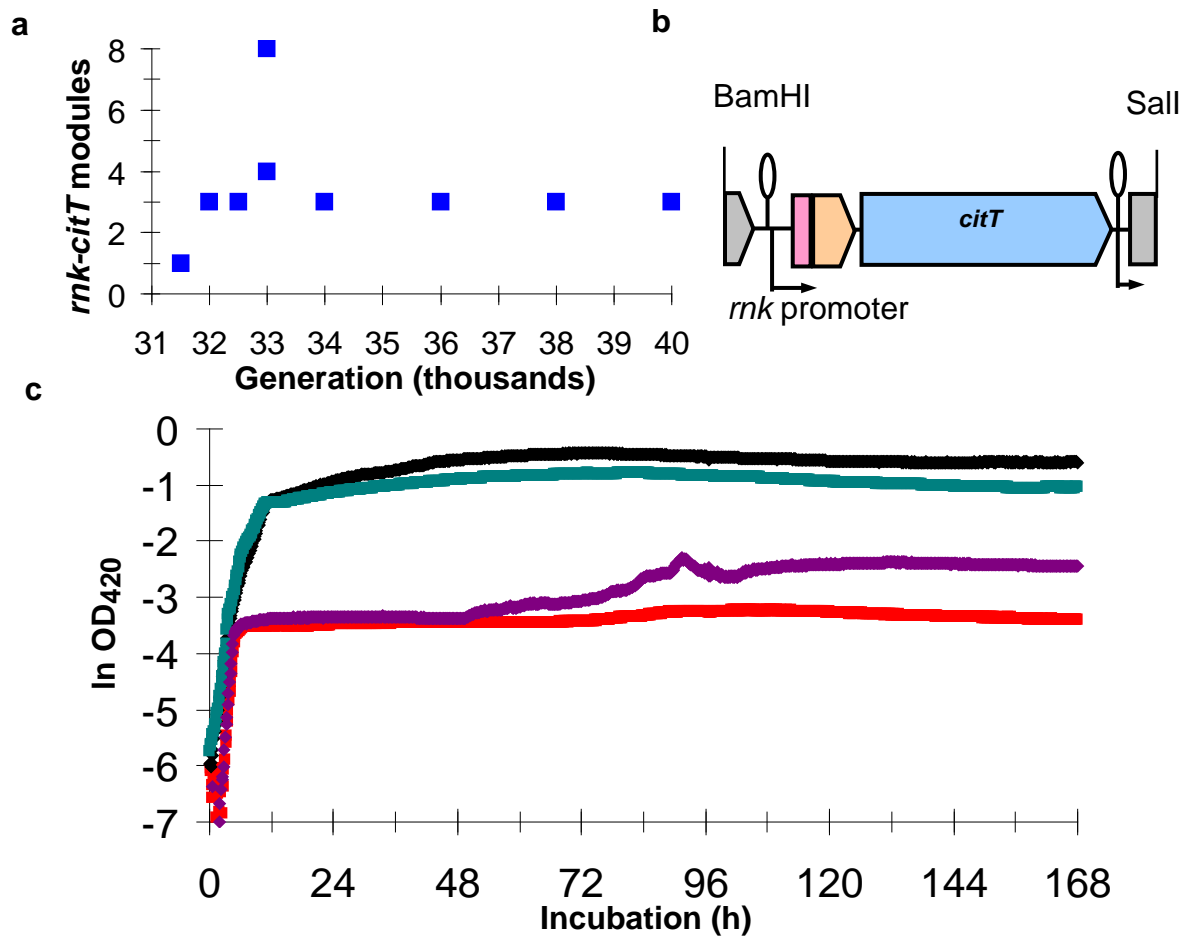


Figure 3.8 | Refinement of Cit⁺ phenotype by increased copy number of *rnk-citT* module.

a. Change in *rnk-citT* module copy number in sequenced Cit⁺ clones over time. **b.** Structure of *rnk-citT* module cloned into the high-copy plasmid pUC19. **c.** Growth trajectories in DM25 of potentiated Cit⁻ clone ZDB30 (red), 31,500-generation Cit⁺ clone ZDB564 (purple), 33,000-generation Cit⁺ clone CZB152 (black), and ZDB612, a pUC19::*rnk-citT* transformant of ZDB30 (green). Trajectories are averages of six replicates for each strain.

Table 3.8 Estimated <i>citT</i> module copy number									
Gen (k)	Clone	Junction Number				Coverage			Predicted Configuration
		Anc <i>rnk</i>	Anc <i>citT</i>	<i>rnk-citT</i>	Relative <i>rnk-citT</i>	<i>citT</i> module	Reference Regions	Relative <i>citT</i>	
31.5	ZDB564	52	49	42	0.83	120.8	64.9	1.86	2×
32	ZDB172	45	39	82	1.95	139.3	23.8	5.86	2× (3×)
32.5	ZDB143	116	76	301	3.14	457.2	112.8	4.05	4×
33	CZB152	22	23	192	8.53	267.4	32.7	8.18	9×
	CZB154	398	252	345	1.06	797.6	130.0	6.14	3× (2×)
34	ZDB83	57	43	125	2.50	214.7	62.9	3.42	4×
36	ZDB96	26	22	87	3.63	97.4	28.8	3.39	4×
38	ZDB107	51	32	133	3.20	204.7	60.0	3.41	4×
40	REL10979	36	32	73	2.15	115.6	32.5	3.56	4×

For each Cit^+ genome, the number of new junctions per genome was estimated from the relative number of reads supporting the new *rnk-citT* junction produced by the amplification versus the number of reads supporting the ancestral *rnk* and *citT* junctions. The total number of *citT* modules per genome was estimated by comparing read-depth coverage of the amplified *citT* region to coverage of regions that appear to be single copy in all genomes (comprising ~20 kb total including the *ara* operon and *tufB* gene). Together these data can be used to predict the likely configuration of *citT* amplification copies in each genome. Examination of read-depth coverage over a larger area supports the observation that there are nested amplifications in CZB154 and ZDB172. For example, the CZB154 genome contains three copies of a larger region, and each copy of that region contains two tandem copies with the usual *rnk-citT* junction.

clone CZB152 (Figure 3.8c). The increased number of *rnk-citT* modules thus appears to explain the initial refinement of the Cit⁺ phenotype that led to the population expansion. In contrast to the early variation in *cit* amplification, later Cit⁺ genomes have four-copy tandem arrays (Figure 3.8). Amplification mutations tend to be unstable (Andersson et al 1998, Reams et al 2010), and further refinement of the Cit⁺ phenotype may have favored more stable mutations. The evolution of the mutator phenotype in the Cit⁺ lineage complicates efforts to identify these later mutations, but interesting candidates include SNPs in *citT* itself; *gltA*, which encodes citrate synthase; and *aceA*, which encodes isocitrate lyase.

Potential of Cit⁺ evolution

Before the Cit⁺ phenotype appeared, the Ara-3 population had to evolve a background in which the requisite new function was accessible by mutation. The ancestor's mutation rate to Cit⁺ was immeasurably low, with an upper bound of 3.6×10^{-13} per cell-generation. Potentiation was demonstrated by 'replay' experiments using 270 clones sampled throughout this population's history (Blount et al 2008). The replays produced 17 additional Cit⁺ mutants that derived from 13 clones, all from generation 20,000 or later. Fluctuation tests confirmed that potentiated clones had increased mutation rates to Cit⁺, although such mutations were still extremely rare (Blount et al 2008).

The potentiating mutations are not known to confer any other phenotype, hence there is no simple way to distinguish between potentiated and non-potentiated clones. Therefore, we first

Table 3.9 Phylogenetically informative mutations						
Gene	Gene Product	Genome Position	Gene Position	Ancestral Nucleotide	Evolved Nucleotide	Associated Clades
<i>ybaL</i>	Predicted transporter with NAD(P)-binding Rossmann-fold domain	475173	133	G	C	UC
<i>nadR</i>	Nicotinamide-nucleotide adenyltransferase	4616538	1010	A	C	C1,C2, C3
<i>hemE</i>	Uroporphyrinogen decarboxylase	4177963	636	T	G	C1
<i>cspC</i>	Stress protein, member of the CspA-family	1886011	4	C	A	C1
<i>yaaH</i>	Conserved inner membrane protein associated with acetate transport	9972	521	T	G	C2,C3
<i>leuA</i>	2-isopropylmalate synthase	85556	778	C	T	C2
<i>tolR</i>	Membrane spanning protein in TolA-TolQ-TolR complex	756799	69	C	T	C2
<i>arcB</i>	Hybrid sensory histidine kinase in two-component regulatory system with ArcA. Aerobic respiration control sensor	3288053	208	G	C	C2
		3288026	236	T	A	C3
<i>gltA</i>	Citrate synthase	734488	772	G	A	C3

examined the distribution of the 13 potentiated clones identified by the replay experiments using ten mutations (Tables 3.9, 10) that differentiated clades UC, C1, C2 and C3 (Figure 3.1). We also determined the distribution of the other 256 evolved clones used in the replays to assess the coverage of the clades in those experiments (Figure 3.9, 10). Overall, 205 clones were assigned to clades, including 12 potentiated clones (Appendix Table A.1). Sixteen clones from generations 15,000 and earlier were in clade UC. The others came from generations 20,000 and later including 55 in C1, 97 in C2, and 37 in C3. Potentiated clones occurred in all three clades with 8 in C3 and 2 each in C1 and C2. Nonetheless, this distribution is highly non-random (two-tailed Fisher's Exact Test comparing C3 with C1 and C2 combined, $P < 0.0003$; that test finds no difference between C1 and C2, $P = 0.6206$), with most Cit^+ mutants from the C3 clade that produced the original Cit^+ mutant. These data suggest that potentiation involved at least two mutations, with one before these clades diverged and another in C3 (Figure 3.1).

Two distinct mechanisms might explain the potentiation effect. One is epistasis, whereby an interaction between the potentiating background and the actualizing mutation was necessary to produce the Cit^+ phenotype. The second is that the background physically promoted the final mutation; for example, a later rearrangement might require some prior genomic rearrangement. If expression of Cit^+ required earlier mutations, then the *rnk-citT* module should confer a weaker Cit^+ phenotype when moved to a non-potentiated background than to a potentiated background. Alternatively, if potentiation facilitated the amplification event itself, then that module should produce an equally strong Cit^+ phenotype in both potentiated and non-potentiated backgrounds.

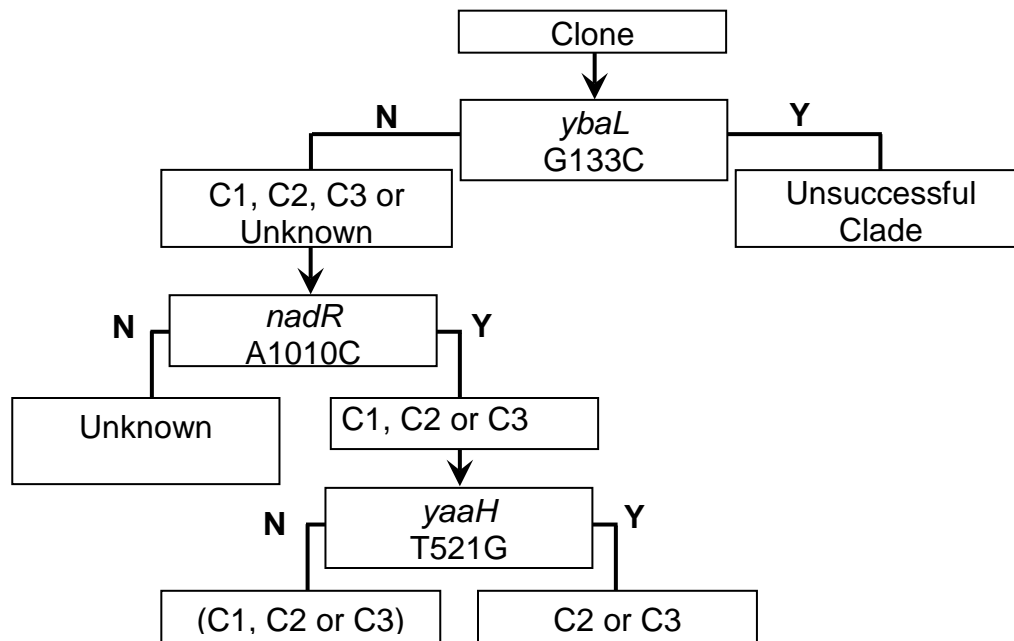


Figure 3.9 | Key for placement of clones used in replay experiments that were sampled before generation 20,000.

The phylogenetic placement of each clone isolated from the population prior to generation 20,000 for use in the replay experiments was determined by scoring the presence or absence of informative mutations according to the key above.

b.

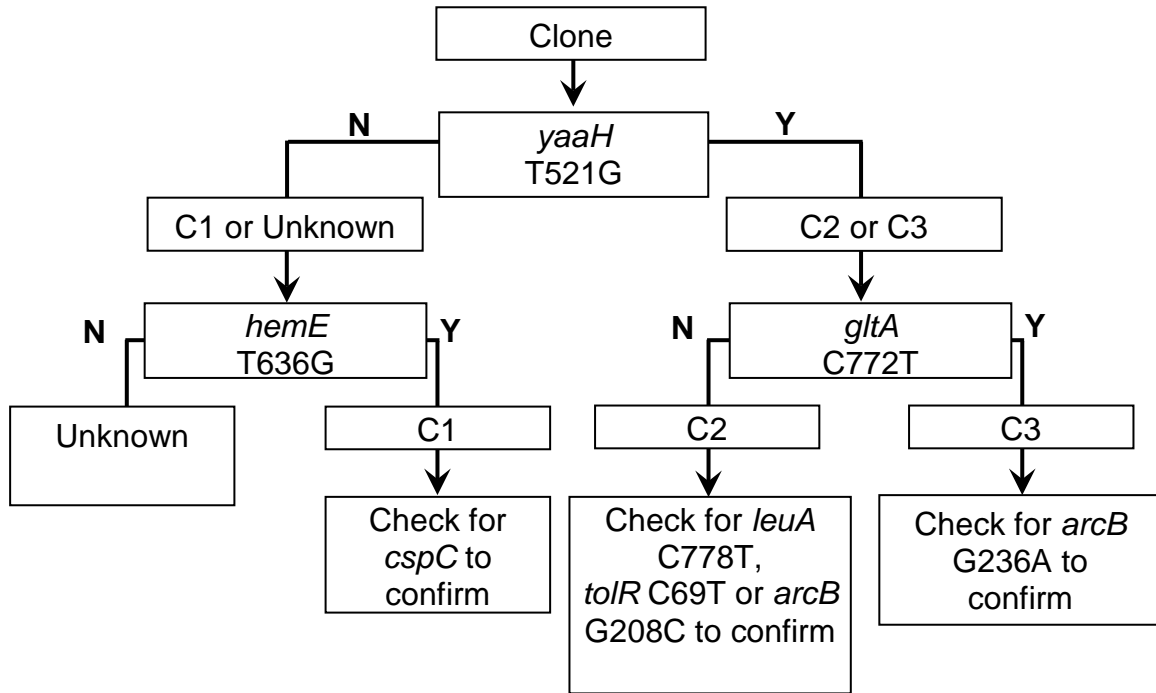


Figure 3.10 | Key for placement of clones used in replay experiments that were sampled from generations 20,000 or later.

The phylogenetic placement of each clone isolated from the population at generations 20,000 or higher for use in the replay experiments was determined by scoring the presence or absence of informative mutations according to the key above. Note the increased resolution that results from greater divergence.

To test these predictions, we moved pUC19::*rnk-citT* into the ancestor and clones from clades C1, C2 and C3, and examined their growth trajectories (Figure 3.11). All four transformants grow on citrate after depleting the available glucose. However, the transformants of the ancestor and clones from C1 and C2 grow poorly on citrate, even with this high-copy-number plasmid, as evidenced by long lags while transitioning from glucose to citrate, low yields, and inconsistent trajectories across replicates (Figure 3.10a-c). By contrast, ZDB612, a transformant of the potentiated C3 clone ZDB30, grows much faster, more extensively, and consistently across replicates (Figure 3.10d). These differences indicate epistatic interactions between the *rnk-citT* module and mutations that distinguish the backgrounds. These data also support the phylogenetic association between clade C3 and the strength of potentiation observed in the replay experiments. We examined C3 for candidate mutations that may contribute to its potentiation. A mutation in *arcB*, which encodes a histidine kinase (Gunsalus and Park 1994), is noteworthy because disabling that gene up-regulates the TCA cycle (Niam et al 2009). That mutation might interact with the *rnk-citT* module by allowing efficient use of citrate that enters the cell via the CitT transporter.

Given the extremely weak Cit⁺ phenotype when the *rnk-citT* module was moved onto the chromosome of the same potentiated C3 clone (Figure 3.6), any mutation that produced a single copy of that module in the ancestor would presumably give either no growth on citrate or such weak growth that no mutants would have been detected in the replay experiments. Indeed, we were unsuccessful in moving the *rnk-citT* module to the ancestral chromosome. One step of that failed process was to screen potential recombinants for a Cit⁺ phenotype; that approach was successful, albeit difficult, in the potentiated C3 clone. Thus, the relative strength of the Cit⁺

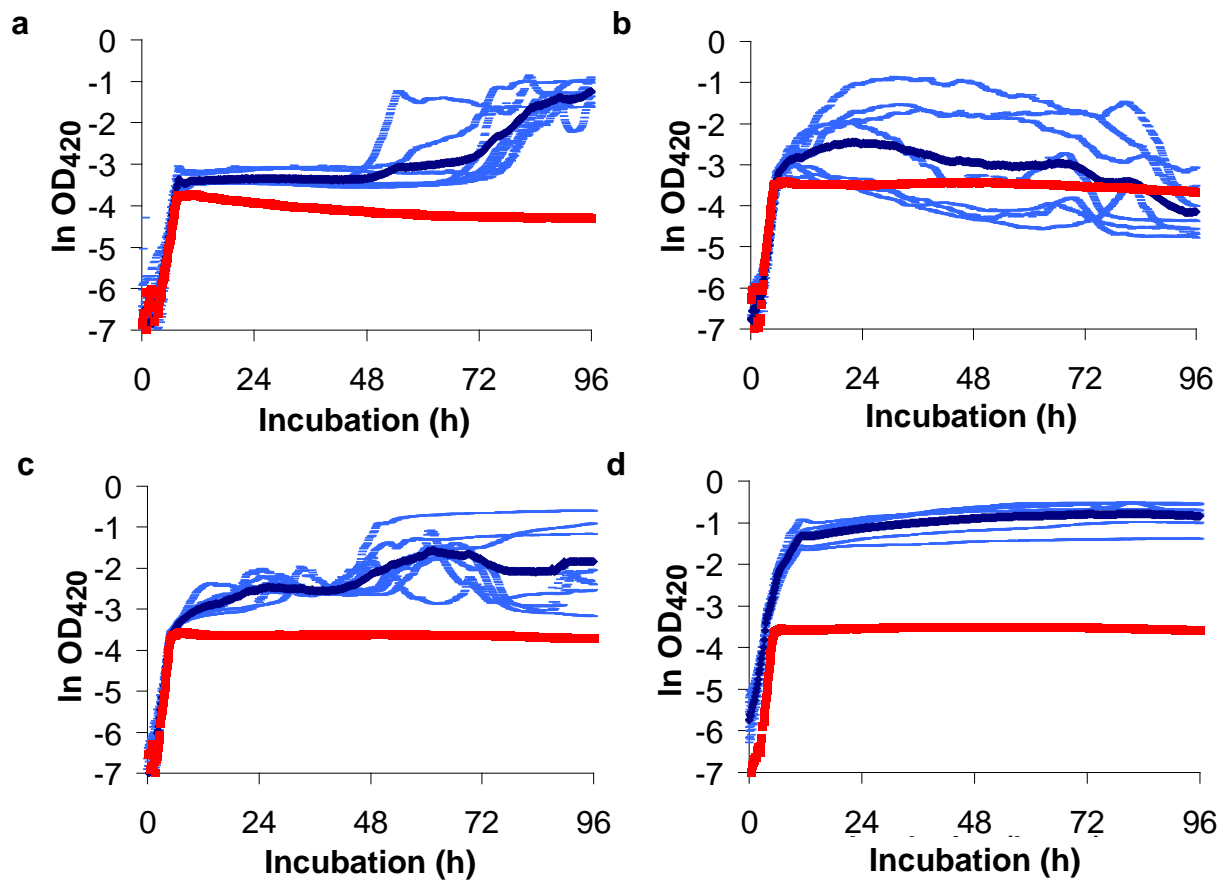


Figure 3.11 | Evidence of epistatic interactions in potentiation of Cit⁺ phenotype. Growth trajectories in DM25 of phylogenetically diverse clones transformed with the pUC19::*rnk-citT* plasmid. **a.** Ancestral strain REL606 and its transformant ZDB611. **b.** C1 clone ZDB199 and its transformant ZDB614. **c.** C2 clone ZDB200 and its transformant ZDB615. **d.** C3 clone ZDB30 and its transformant ZDB612. In each panel, red and dark blue trajectories are averages for the parent clone and its Cit⁺ transformant; light blue trajectories show the replicates for the transformant.

phenotype produced by the high-copy plasmid in the various backgrounds (Figure 3.11) and the relative success in generating any Cit^+ phenotype with a chromosomal copy of the *rnk-citT* module both support the hypothesis that potentiation depends, at least partly, on epistatic interactions between the genetic background and the actualizing mutation that first generated that module.

Further evidence regarding potentiation comes from the Cit^+ mutants observed during the replay experiments (Blount et al 2008). The physical-promotion hypothesis predicts that these mutants would all have *cit* amplifications similar or identical to the original one. If epistatic potentiation enhanced *citT* expression only from the *rnk* promoter, then the prediction would be the same. However, if epistatic potentiation operated at some broader physiological level, then the Cit^+ replay mutants should have a variety of mutations that share only the property that they enable expression of the citrate transporter in the oxic environment of the long-term experiment. We examined 19 re-evolved Cit^+ mutants to identify the relevant mutations. The region containing *citT* was examined in all of them, and the genomes of six were sequenced. All 19 have mutations affecting *citT*, and most clearly put that gene downstream of a new promoter (Table 3.10). Four fully sequenced Cit^+ mutants were derived from Cit^- clones that were also sequenced. Besides *citT*-related mutations, these sequenced mutants had 1-3 other mutations; no gene was mutated in multiple cases, and none appear related to citrate utilization (Table 3.11), supporting the inference that the *citT* mutations were responsible for all re-evolved Cit^+ phenotypes.

Table 3.10 Mutations affecting <i>cit</i> region in Cit ⁺ replay mutants				
Gen	Cit ⁻ Parent	Clade	Cit ⁺ Mutant	Mutation affecting <i>cit</i> Region
20,000	ZDB464*	C3	ZDB285*	2978-bp tandem <i>cit</i> duplication that creates <i>rnk-citT</i> regulatory module.
			ZDB286	2656-bp tandem <i>cit</i> duplication that creates <i>rnk-citT</i> regulatory module.
27,000	ZDB309	C3	ZDB288	2070-bp tandem <i>cit</i> duplication that creates <i>rna-citT</i> regulatory module.
	ZDB310	C3	ZDB290*	~5,000-bp tandem duplication. Basis of <i>citT</i> activation unknown.
30,500	ZDB20	C3	ZDB547	IS3 insertion in <i>citG</i> .
31,000	ZDB390	C1	ZDB292	2706-bp tandem <i>cit</i> duplication that creates <i>rnk-citT</i> regulatory module.
31,500	ZDB25	C3	ZDBr218	IS3 insertion in <i>citG</i>
	ZDB199	C1	ZDB279	2745-bp tandem <i>cit</i> duplication that creates <i>rnk-citT</i> regulatory module.
32,000	ZDB28	C3	ZDB161	IS3 insertion in <i>citG</i> .
	ZDB29	C2	ZDB166	IS3 insertion in <i>citG</i> .
	ZDB30*	C3	ZDB164	IS3 insertion in <i>citG</i> .
			ZDB165*	IS3 insertion in <i>citG</i> .
			ZDB283*	~568-kbp inversion that places much of the <i>cit</i> operon under control of the <i>fimB</i> promoter.
			ZDB294*	2663-bp tandem <i>cit</i> duplication that creates <i>rnk-citT</i> regulatory module.
	ZDB183	C3	ZDB281*	~14.3-kbp duplication. Basis of <i>citT</i> activation unresolved.
32,500	ZDB31	C3	ZDB163	2990-bp tandem <i>cit</i> duplication that creates <i>rnk-citT</i> regulatory module.
			ZDB546	Unknown rearrangement or duplication affecting <i>citT</i> .
			ZDB549	422-bp deletion in <i>citG</i> . Basis of <i>citT</i> activation unresolved.
	ZDB32	C2	ZDB548	3144-bp tandem <i>cit</i> duplication that creates <i>rnk-citT</i> regulatory module.

*Genome has been sequenced.

Table 3.11 | Annotated differences between genomes of four Cit⁺ mutants and their Cit⁻ parent clones

Supplementary Table 3.11a Pair 1: ZDB464 (Generation 20,000) and Cit ⁺ Mutant ZDB285					Cit ⁻ Parent ZDB464	Cit ⁺ Mutant ZDB285
Position	Mutation	Description	Gene or Genes Involved	Product(s)		
626102	2794 bp duplication	Tandem duplication with one junction in <i>citG</i> and the other between <i>rna</i> and <i>rnk</i> ; presumed Cit ⁺ actualizing mutation	<i>citG</i>	Triphospho-ribosul-dephospho-CoA transferase		
			<i>citT</i>	Citrate transporter		
			<i>rna</i>	Ribonuclease I		
848202	IS150 Insertion	Insertion in a coding region	<i>ybiS</i>	Hypothetical protein		
1360374	IS150 Insertion	Insertion in a coding region	<i>ycjC</i>	DNA-binding transcriptional repressor		
3270443	IS150 Insertion	Insertion in a non-coding, intergenic region; promoters not disrupted	Site between <i>nlp/murA</i>	DNA-binding transcriptional activator of maltose metabolism/ UDP-N-acetylglucosamine 1-carboxyvinyl-transferase		

Table 3.11b Pair 2: ZDB30 (Generation 32,000) and Cit ⁺ Mutant ZDB165					Cit ⁻ Parent	Cit ⁺ Mutant
Position	Mutation	Description	Gene or Genes Involved	Product(s)	ZDB30	ZDB165
628716	IS3 Insertion	Insertion in a coding region; presumed Cit ⁺ actualizing mutation	<i>citG</i>	Triphospho-ribosul-dephospho-CoA transferase		
1651966	Δ795 bp	Deletion of multiple coding regions	<i>ydgG</i>	Transporter of quorum signal AI-2		
			<i>pntB</i>	Pyridine nucleotide trans-hydrogenase, β subunit		

Table 3.11c Pair 3: ZDB30 (Generation 32,000) and Cit ⁺ ZDB293					Cit ⁻ Parent	Cit ⁺ Mutant
Position	Mutation	Description	Gene or Genes Involved	Gene Product(s)	ZDB30	ZDB293
632864	568476 bp inversion	Most of <i>cit</i> operon structural genes placed downstream of <i>fimB</i> promoter; new junctions in <i>citC</i> and between <i>yjhA</i> and <i>fimB</i> ; presumed Cit ⁺ actualizing mutation	<i>citC</i>	Citrate lyase synthetase		
			<i>fimB</i>	Regulator of <i>fimA</i> pili subunit		
687047	IS150 Insertion	Insertion in a coding region	<i>nagE</i>	Fused N-acetyl glucosamine specific PTS enzyme: IIC, IIB, and IIA components		

Table 3.11d Pair 4: ZDB30 (Generation 32,000) and Cit ⁺ Mutant ZDB294					Cit ⁻ Parent	Cit ⁺ Mutant
Position	Mutation	Description	Gene or Genes Involved	Gene Product(s)	ZDB30	ZDB294
626107	2661 bp duplication	Tandem duplication with junctions in <i>citG</i> and between <i>rna</i> and <i>rnk</i> ; presumed Cit ⁺ actualizing mutation	<i>citG</i>	Triphospho-ribosul-dephospho-CoA transferase		
			<i>citT</i>	Citrate transporter		
			<i>rna</i>	Ribonuclease I		
1607917	Δ1 bp	Single nucleotide deletion in a coding region	<i>ECB_1510</i>	Putative tail component of prophage		
1729055	Δ2462 bp	IS150-mediated deletion of multiple coding regions	<i>ydhV</i>	Predicted oxido-reductase		
			<i>ydhY</i>	Predicted 4Fe-4S ferredoxin-type protein		
2157881	Δ12781 bp	IS150-mediated deletion of multiple coding regions.	<i>yehM</i>	Conserved protein		
			<i>yehP</i>	Conserved protein		
			<i>yehQ</i>	Possible pseudogene		
			<i>yehR</i>	Conserved protein		
			<i>yehS</i>	Conserved protein		
			<i>yehT</i>	Predicted sensory kinase in two-component system with YehU		

Table 3.11d Pair 4: ZDB30 (Generation 32,000) and Cit ⁺ Mutant ZDB294					Cit ⁻ Parent	Cit ⁺ Mutant
Position	Mutation	Description	Gene or Genes Involved	Gene Product(s)	ZDB30	ZDB294
			<i>yehU</i>	Predicted sensory kinase in two-component system with YehT		
			<i>yehV</i>	MerR-like regulator		
			<i>ECB 02057</i>	Unknown function		
			<i>yehW</i>	Membrane component of an ABC transporter involved in osmoprotection		
			<i>yehX</i>	Membrane component of an ABC transporter involved in osmoprotection		
			<i>yehY</i>	Membrane component of an ABC transporter involved in osmoprotection		
			<i>yehZ</i>	Periplasmic component of an ABC transporter involved in osmoprotection		

Red fill indicates presence of mutation. Gray fill indicates absence of corresponding mutation.

The re-evolved Cit⁺ mutants arose by diverse mutational processes, though all affected *citT* (Table 3.10). Eight have *citT* duplications similar to the original case, although no two share the same boundaries (Figure 3.12). In seven of these, the duplication generated a version of the *rnk-citT* module; in the other, the second *citT* occurs downstream of the *rna* promoter. Six mutants have an IS3 element inserted in the 3' end of *citG* (Figure 3.11). IS3 carries outward-directed promoter elements that can activate adjacent genes (Treves et al 1998, Charlier et al 1982). Two mutants have large duplications that encompass all or part of the *cit* operon. One mutant has a large inversion that places most of that operon downstream of the promoter for the fimbrial regulatory gene *fimB*, while another has a deletion in *citG* that presumably formed a new promoter. Thus, the Cit⁺ phenotype can evolve in potentiated backgrounds by a variety of mutational processes that recruit several different promoters that enable CitT expression during aerobic metabolism. It is also interesting that most of these replay mutants exhibit stronger phenotypes (Figure 3.13) than the earliest Cit⁺ clones in the main experiment (Figure 3.6 – 7).

PERSPECTIVE

The evolution of citrate-utilization in an experimental *E. coli* population provided an unusual opportunity to examine in detail the multi-step origin of a key innovation. From comparative studies it has long been recognized that gene duplications play an important creative role in evolution, particularly by generating redundancies that allow neo-functionalization (Gould and Vrba 1982, Taylor and Raes 2004, True and Carroll 2002, Zhang 2003, Bergthorsson et al 2007). Our findings highlight the less-appreciated capacity of gene

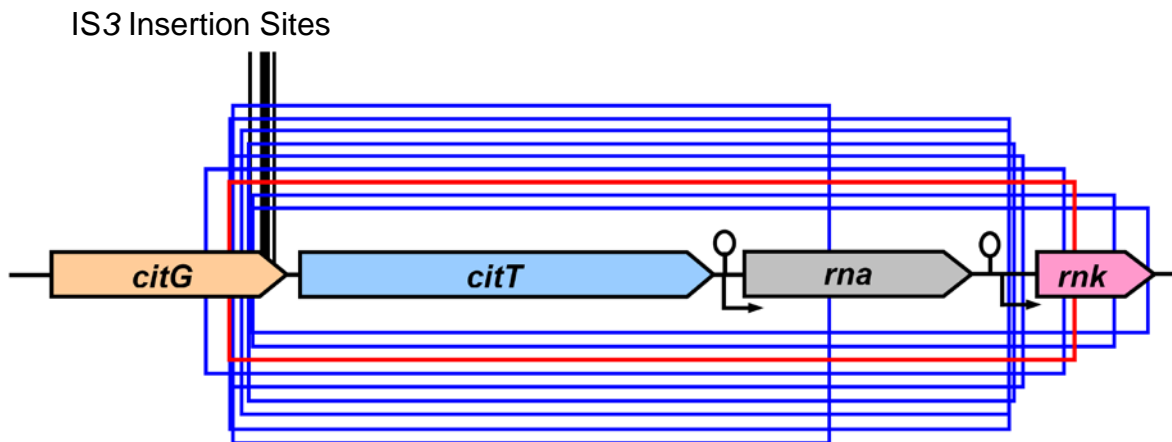


Figure 3.13 | Mutations that produced Cit^+ phenotype in 14 replay experiments. The red box shows the boundaries of the 2,933-bp amplified segment that actualized the Cit^+ function in the original long-term population. Blue boxes show *citT*-containing regions amplified in 8 replays that produced Cit^+ mutants. Vertical black lines mark 5 locations in *citG* where IS3 insertions produced Cit^+ mutants in 6 other replays.

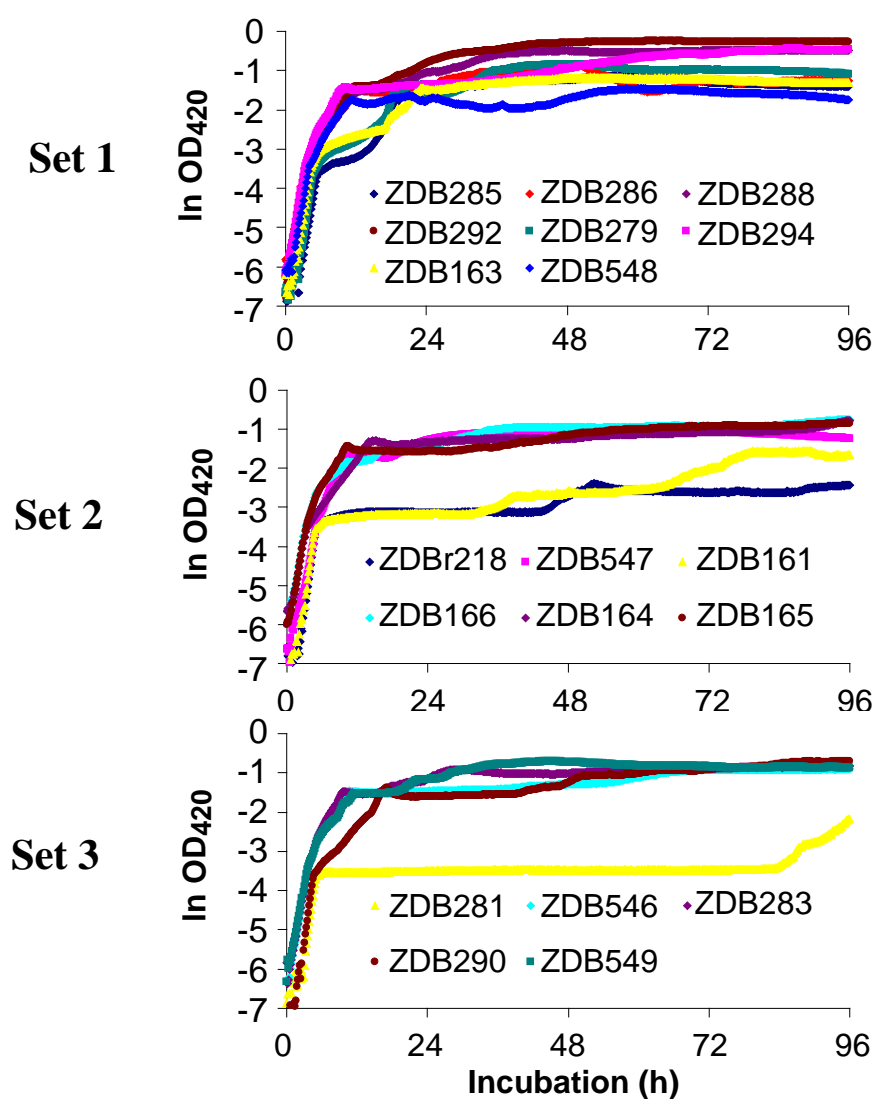


Figure 3.12 | Growth of Cit^+ mutants derived in replay experiments. The 19 spontaneous Cit^+ mutants isolated during the course of replay experiments vary in how quickly they transition from the glucose to the citrate and how well they grow on the citrate in DM25. Here mutants are grouped according to the mutations responsible for the Cit^+ phenotype. The eight mutants in Set 1 have tandem *citT* duplications similar to that which evolved in the original population. The six mutants in Set 2 mutants have IS3 insertions in *citG*. The five mutants in Set 3 have a variety of mutations, as described in the main text and Supplementary Table 12.

duplications to mediate exaptation events by altering gene-regulatory networks, especially by promoter capture (Whoriskey et al 1987). The evolution of citrate-utilization also highlights that such actualizing mutations are only part of the process by which novelties arise. Before some new function can arise, it may be essential for a lineage to evolve a potentiating genetic background that allows the actualizing mutation to occur or the new function to be expressed. Finally, novel functions often emerge in rudimentary forms that must be refined by further mutations and selection to exploit fully the ecological opportunities. This three-step process – in which potentiation makes a trait possible, actualization makes the trait manifest, and refinement makes it effective – is likely typical of many new functions.

METHODS

Evolution Experiment

The long-term experiment has been described in detail elsewhere (Lenski 2004, Lenski et al 1991). In brief, twelve populations of *E. coli* B were started in 1988, and they have evolved for more than 20 years under conditions of daily 100-fold dilutions in a minimal medium, DM25, containing 139 μM glucose and 1700 μM citrate (Lenski et al 1991). The populations experience ~6.64 generations per day, and they had evolved for over 40,000 generations when this study began. Every 500 generations, viable samples of each population were frozen at –80°C with glycerol added as a cryoprotectant. The focus of this study was population Ara–3, in which the ability to grow aerobically on citrate evolved after more than 30,000 generations (Blount et al 2008).

Genomic DNA Isolation

Clones from the Ara-3 population were revived from frozen stocks by overnight growth in LB medium at 37°C with aeration. DNA was extracted and purified from several mL of each culture using the Qiagen Genomic-tip 100/G kit (Qiagen, Hilden, DE).

Whole-genome Shotgun Re-sequencing and Mutation Detection

Clones were sequenced on Illumina GA, GA II, and GA IIx instruments. The resulting read data have been deposited in the NCBI SRA database (SRP004752). Most genomic datasets consist of single-end reads only, but additional mate-paired libraries were sequenced for ZDB30 and ZDB172. Reads were mapped to the reference genome of the ancestral strain (REL606) (Jeong et al 2009), and mutations were predicted using the *breseq* computational pipeline (Barrick and Knoester 2010). This pipeline detects point mutations, deletions, and new sequence junctions that may indicate IS-element insertions or other rearrangements, as described in its online documentation. Large duplications and amplifications were predicted manually by examining the depth of read coverage across each genome.

Mutation lists were further refined by manually reconstructing the most plausible series of mutational events generating the genomic differences observed in each clone. This procedure involved: (1) splitting predicted changes into multiple mutations based on phylogenetic relationships, (2) assigning mutations to genomes when a subsequent mutational event prevented their detection (e.g., a SNP within a later deletion), and (3) correcting false-positives for a handful of difficult-to-predict mutations by examining read alignments and coverage in all clones. After these procedures, 17 homoplasies remained in the phylogenetic tree, most of which

appear to indicate mutational hot-spots: 9 are IS-element insertions at specific sites, 7 are insertions or deletions at the boundaries of IS-elements, and 1 is a single-base substitution not associated with IS-elements. Given these signatures, it is likely that most or all of these mutations occurred independently in multiple lineages.

The copy number and configuration of the *citT* module were estimated from the number of reads overlapping the new *rnk-citT* sequence junction relative to the two original flanking-sequence junctions, and from the average read-depth in this region relative to a single-copy region (Table 3.8).

Phylogenetic Reconstruction

An initial parsimony-based tree was calculated using presence-absence data for all mutational events in each genome using the *dnapars* program from PHYLIP. Branch lengths were recalculated for the tree using an irreversible Camin-Sokal model (Camin and Sokal 1965), because the ancestral states are known and reversions are extremely unlikely given the genome size and number of mutations observed. A maximum-likelihood model was then used to estimate all branching times and two mutation rates, one for non-mutator branches and one for mutator branches. This model fixed the generation when each clone was sampled and assumed that mutations accumulated on each branch according to a Poisson process.

Ten mutations in nine genes (*ybaL*, *nadR*, *hemE*, *cspC*, *yaaH*, *leuA*, *tolR*, *arcB*, and *gltA*) were identified as phylogenetically informative based on their association with particular clades in the Ara-3 population (Table 3.9). Sanger sequencing of PCR-amplified gene fragments was used to detect the presence or absence of these mutations in replay clones, which were then mapped onto the population phylogeny according to the keys shown in Figure 3.9. Owing to the

large number of clones and genes under consideration, not all genes were sequenced for all clones. Appendix Table 1 shows sequencing results and phylogenetic assignments. Table 3.16 provides sequences of the primer pairs used to amplify each locus.

PCR Screens for *cit* Amplifications

The *cit* amplification was detected in both population samples and clones by PCR amplification across the *rnk-citG* junction using outward-directed primers specific to *citT* (Table 3.12). When screening population samples, three separate PCR reactions were run for each time point, and the template used was a 1:10 dilution of the population sample frozen for that generation.

Expression Experiments

The regulation of expression of the Cit⁺ phenotype was measured using luciferase-based reporter constructs. The complete upstream regions for the native *citT* and *rnk* genes were PCR-amplified from the cognate reporters from the *E. coli* transcriptional library (Zaslaver et al 2006) using the primers pZE05 and pZE07. The intergenic region of the *rnk-citT* module was amplified from Cit⁺ clone ZDB172 using primers nctForward and nctReverse (Table 3.12). The PCR products were cloned into the low-copy (1-2) plasmid pCS26-pac, which contains a kanamycin-resistance gene and the luciferase operon (*luxCDABE*) (Bjarnason et al 2003). Each reporter plasmid was transformed into clones REL606, ZDB30 and ZDB172.

Prior to the expression assays, strains were grown in a 96-well plate (BD Biosciences, Bedford, MA, USA) in 200 µL per well of DM25 supplemented with 50 µg/mL kanamycin, with

Table 3.12 Primer pairs used in this study			
Gene or Region Amplified	Primer Name	Primer Sequence	PCR Product Size
<i>ybaL</i> mutation	<i>ybaL</i> mut F	5' CATCGCCCTGTTCCATCATTCT 3'	503 bp
	<i>ybaL</i> mut R	5' ACCCCGCTTATCACCACCATTGTT 3'	
<i>nadR</i> mutation	<i>nadR</i> mut F	5' ATGGTCGCGATTATGTCTTTTCAC 3'	459 bp
	<i>nadR</i> mut R	5' CGTTTCATCGCGGTTATCTCTG 3'	
<i>hemE</i> mutation	<i>hemE</i> mut F	5' GTGCCGGACGCGATGGGGTTAG 3'	524 bp
	<i>hemE</i> mut R	5' CACTGTCCGCCGCCCTTTGGTA 3'	
<i>cspC</i> mutation	<i>cspC</i> mut F	5' GGGCAAATATCCGAACG 3'	412 bp
	<i>cspC</i> mut R	5' AGCCTTATATTGGTGCCTCAT 3'	
<i>yaaH</i> mutation	<i>yaaH</i> mut F	5' CTTTCGCGTCAGGTTGGTGTG 3'	1030 bp
	<i>yaaH</i> mut R	5' CCTGCCTGCGCCGGATGGTTAG 3'	
<i>leuA</i> mutation	<i>leuA</i> mut F	5' GAATGCGCCGCTGCCAACA 3'	497 bp
	<i>leuA</i> mut R	5' GCCTCAACCAGCGCGTAAACAAA 3'	
<i>tolR</i> mutation	<i>tolR</i> mut F	5' GCCTCAACCAGCGCGTAAACAAA 3'	400 bp
	<i>tolR</i> mut R	5' ACTTCCGCCACCACCTGCTCTG 3'	
<i>arcB</i> mutation	<i>arcB</i> mut F	5' TGTCGCGACCAAAGCCCATCA 3'	709 bp
	<i>arcB</i> mut R	5' GCCCTCGTCGTTCTTGCCATTGT 3'	
<i>gltA</i> mutation	<i>gltA</i> mut F	5' TGTGTTTAACGGAGCTGATTTCTT 3'	626 bp
	<i>gltA</i> mut R	5' GCTGGCGACCGATTCTAACTACCT 3'	
<i>cit</i> amplification	<i>cit</i> Tout F	5' GTCCTGGGGTGATTATTTACGGCT 3'	1807 bp
	<i>cit</i> Tout R	5' CAATAACGCAAATAGTAACCGCAA 3'	
<i>citAmpJ</i> fragment	<i>citAmpJ</i> F	5' TTTTTTGGATCCGGTTCGAATGCCCC-CTTTTT 3'	529 bp
	<i>citAmpJ</i> R	5' TTTTTTGTGCGACGGTAACCCTGCGTATTGACTGAA 3'	
<i>rnk</i> promoter region of <i>rnk-citT</i> module for expression studies	nctForward	5' AAAAAAGGATCCGACACCCATCACC-ACCAGT 3'	707 bp
	nctReverse	5' AAAAAACTCGAGACGCCATCAACGC-TCCGCTTTCT 3'	
<i>citT-citG</i> fragment for gene-gorging	<i>citT-citG</i> F	5' AACCAGCCAGGCCCCATTTTCAGC 3'	648 bp
	<i>citT-citG</i> R	5' AAAAAAGGATCCCACGCCTTGCCGC-ATTACCTCACT 3'	
<i>citG</i> frag fragment for gene-gorging	<i>citG</i> frag F	5' TTTTTTGGATCCGGGGGTTCGAATG-CCCCCTTTTT 3'	694 bp
	<i>citG</i> frag R	5' GCACAAAGATATGGCGCTGGAAGA 3'	

Table 3.12 Primer pairs used in this study			
Gene or Region Amplified	Primer Name	Primer Sequence	PCR Product Size
<i>rnk</i> promoter and <i>cit</i> amplification junction construct for gene-gorging	<i>citT</i> – <i>citG</i> Gorge F	5' TAGGGATAACAGGGTAATAACCAGC-CAGGCCCCATTTTCAGC 3'	1889 bp*
	<i>citG</i> frag R	5' GCACAAAGATATGGCGCTGGAAGA 3'	
<i>rnk</i> - <i>citT</i> module for cloning into pUC19	<i>citT</i> AmpX F	5' AAAAAAGGATCCGGGCAGCAACCGAT-TTAGG 3'	2490 bp
	<i>citT</i> AmpX R	5' AAAAAAGTTCGACAACGCTCCGCTTTC-TGC 3'	
<i>citT</i> internal fragment for Southern hybridizations	<i>citT</i> probe F	5' AGCCGTAAATAATCACCCCAGGAC 3'	1173 bp
	<i>citT</i> probe R	5' TTGCGGTTACTATTTGCGTTATTG 3'	
Amplification of genomic region immediately upstream of <i>citT</i>	<i>citT</i> upstrm R	5' CTCTCCCGCCGCGACTATTCA 3'	1264 bp [†]
	<i>citT</i> upstrm F	5' CAATAACGCAAATAGTAACCGCAA 3'	

*Length for fully assembled construct.

[†]Length without deletions or insertions.

constant shaking at 37°C for two 24-h cycles to ensure proper preconditioning. Fresh overnight cultures were then diluted 100-fold into a black, clear-bottomed 96-well plate (9520 Costar; Corning, Lowell, MA, USA), with 150 µL per well of DM25, and covered with breathable sealing membrane (Nunc, Rochester, NY, USA) to prevent evaporation. Luciferase activity was read in a Wallac Victor² plate reader (Perkin Elmer Life Sciences, Boston, MA, USA) every 20 minutes for 19 h with 90 s of 2-mm orbital shaking prior to each reading. Assays were performed in quadruplicate in the same plate.

Isogenic Strain Construction

A single-copy chromosomal *rnk-citT* module was made in a Cit⁻ background using the “gene gorging” allele transfer technique (Herring et al 2003). Owing to difficulties inherent to the manipulation of amplified genes, we did not attempt to move an entire *citT* amplification segment. Instead, we engineered a construct of the *cit* amplification junction containing the *rnk* promoter from three smaller fragments that were PCR-amplified from the 32,000-generation Cit⁺ clone ZDB172. The first fragment contained the *cit* amplification junction (*citAmpJ*), including the *rnk* promoter region, and was PCR-amplified using the primers *citTAmpJ* F and *citTAmpJ* R (Table 3.12). The second fragment contained the *citT-citG* junction (*citT-citG*) and was PCR-amplified using the primers *citT-citG* F and *citT-citG* R. The third fragment contained sequence internal to the *citG* gene (*citGfrag*) and was PCR-amplified using the primers *citGfrag* F and *citGfrag* R. Each primer pair was designed with restriction sites that allowed ligation of the fragments to a hybrid construct in which the *cit* amplification junction, including the *rnk* promoter, was embedded within >500 bp of *citG* flanking sequence. The assembled fragment

was then PCR-amplified using the primers *citT-citG* Gorge F and *citGfrag* R (Table 3.12). The primer *citT-citG* Gorge F incorporated an *I-Sce-I* restriction site necessary for the gene-gorging procedure, which was then performed as described elsewhere (Herringg et al 2003). We screened for putative transformants by performing PCR and by testing for expression of a Cit⁺ phenotype as indicated by a positive reaction on Cristensen's Citrate Agar (Christensen 1949). Successful constructs were then confirmed by Sanger sequencing.

Growth Trajectories

Strains of interest (Table 3.13) were revived from frozen stocks by growing them in LB, and they were then preconditioned by two 24-h culture cycles in DM25. Following preconditioning, 100 μ L of each culture was diluted into 9.9 mL of DM25, and then dispensed as 200 μ L aliquots into randomly assigned wells in a 96-well plate. For all pUC19::*rnk-citT* transformants, the medium was supplemented with 100 μ g/mL ampicillin to ensure plasmid retention. Growth trajectories were replicated 6- to 8-fold for every strain. To reduce evaporation, strains were grown in the innermost 60 wells, while the outermost 36 wells were filled with 300 μ L of a saline buffer. In those cases when the assays lasted longer than 96 hours, the buffer was replenished after 96 hours. OD₄₂₀ was measured every 10 minutes by a VersaMax automated plate reader (Molecular Devices, Sunnyvale, CA, USA). Plates were shaken orbitally for 5 seconds prior to each measurement, but were otherwise stationary.

Fitness Assays

Relative fitness was measured in competition experiments described in detail elsewhere (Lenski et al 1991). In brief, we inoculated 0.05 mL each of preconditioned cultures of ZDB595

and ZDB63, an Ara⁺ revertant of ZDB30, into 9.9 mL of fresh DM25 with 10-fold replication. Starting densities of each strain were determined by dilution plating on tetrazolium arabinose (TA) plates, on which the Ara⁻ strain ZDB595 and the Ara⁺ strain ZDB63 made red and white colonies, respectively. The competition cultures were then propagated for three daily transfer cycles, after which the final densities of each strain were again determined by dilution plating on TA plates. Relative fitness was calculated as the ratio of the realized net growth rates of the two strains over the course of the competition experiment (Lenski et al 1991).

Plasmid Construction

A fragment containing the complete *rnk-citT* module was PCR-amplified from the genome of Cit⁺ clone ZDB172 using the primers *citTAmpX F* and *citTAmpX R* (Table 3.12). The fragment was then inserted into the cloning site of pUC19 (NEB, Ipswich, MA, USA). The corresponding region of the plasmid, pUC19::*rnk-citT*, was confirmed by Sanger sequencing. The plasmid was transformed into strains REL606, ZDB30, ZDB199 and ZDB200. Cit⁺ transformants were identified by positive reactions on Cristensen's Citrate Agar (Christensen 1949).

Identification of Cit⁺-conferring Mutations in Replays

Nineteen Cit⁺ mutants isolated during replay and related experiments (Blount et al 2008) were analyzed to identify the responsible events. These mutants were first checked for large changes in the region containing *citT* by Southern hybridization with *citT*-specific probes.

Table 3.13 Clones used in growth-trajectory experiments			
Clone	Gen	Description	Growth Curve Locations
REL606	0	Ancestor	Figure 3.10
CZB152	33,000	Cit ⁺ clone from main population	Figure 3.7, 8
CZB154	33,000	Cit ⁺ clone from main population	Figure 3.7
ZDB30	32,000	Potentiated Cit ⁻ clone from Clade 3	Figure 3.6, 7, 10
ZDB143	32,500	Cit ⁺ clone from main population	Figure 3.7
ZDB161	32,000	Cit ⁺ mutant of ZDB28	Figure 3.12
ZDB163	32,500	Cit ⁺ mutant of ZDB31	Figure 3.12
ZDB164	32,000	Cit ⁺ mutant of ZDB30	Figure 3.12
ZDB165	32,000	Cit ⁺ mutant of ZDB30	Figure 3.12
ZDB166	32,000	Cit ⁺ mutant of ZDB29	Figure 3.12
ZDB172	32,000	Cit ⁺ clone from main population	Figure 3.6, 7
ZDB199	31,500	Potentiated Cit ⁻ clone from Clade 1	Figure 3.10
ZDB200	31,500	Cit ⁻ clone from Clade 2	Figure 3.10
ZDB279	31,500	Cit ⁺ mutant of ZDB199	Figure 3.12
ZDB281	32,000	Cit ⁺ mutant of ZDB183	Figure 3.12
ZDB283	32,000	Cit ⁺ mutant of ZDB30	Figure 3.12
ZDB285	20,000	Cit ⁺ mutant of ZDB464	Figure 3.12
ZDB286	20,000	Cit ⁺ mutant of ZDB464	Figure 3.12
ZDB288	27,000	Cit ⁺ mutant of ZDB309	Figure 3.12
ZDB290	27,000	Cit ⁺ mutant of ZDB310	Figure 3.12
ZDB292	31,000	Cit ⁺ mutant of ZDB390	Figure 3.12
ZDB294	32,000	Cit ⁺ mutant of ZDB30	Figure 3.12
ZDB546	32,500	Cit ⁺ mutant of ZDB31	Figure 3.12
ZDB547	30,500	Cit ⁺ mutant of ZDB20	Figure 3.12
ZDB548	32,500	Cit ⁺ mutant of ZDB32	Figure 3.12
ZDB549	32,500	Cit ⁺ mutant of ZDB31	Figure 3.12
ZDB564	31,500	Cit ⁺ clone from main population	Figure 3.6, 7, 8
ZDB595	32,000	Cit ⁺ isogenic construct of ZDB30 in which <i>rnk-citT</i> module promoter was inserted into the chromosome	Figure 3.6

Table 3.13 Clones used in growth-trajectory experiments			
Clone	Gen	Description	Growth Curve Locations
ZDB611	0	pUC19:: <i>rnk-citT</i> transformant of REL606	Figure 3.10
ZDB612	32,000	pUC19:: <i>rnk-citT</i> transformant of ZDB30	Figure 3.8, 10
ZDB614	31,500	pUC19:: <i>rnk-citT</i> transformant of ZDB199	Figure 3.10
ZDB615	31,500	pUC19:: <i>rnk-citT</i> transformant of ZDB200	Figure 3.10
ZDBr218	31,500	Cit ⁺ mutant of ZDB25	Figure 3.12

Genomic DNA was digested with EcoRV (NEB, Ipswich, MA, USA), and fragments were separated on 0.8% agarose gels with a 1-kb ladder (NEB, Ipswich, MA, USA), and then transferred to nylon membranes. Hybridizations were performed at 68°C. The *citT*-specific probe was an internal fragment amplified by PCR using the primers described in Table 3.16, purified using an Illustra GFX PCR DNA purification kit (GE Healthcare, Little Chalfont, Buckinghamshire, UK), and labeled using the DIG DNA labeling and detection kit (Roche, Basel, Switzerland). Most, but not all, mutants showed enlarged *citT* bands. We then tried to PCR-amplify across possible amplification boundaries of each mutant with the same outward-directed *citT* primers used to screen for the amplification in the original Ara-3 population (Table 3.12). PCR products were thus obtained for 8 of the 19 clones. Sanger sequencing of these products permitted resolution of novel junctions that were consistent with amplifications similar, but not identical, to that in the original population. For 7 of the remaining 11 clones, PCR products of altered size were obtained when the region immediately upstream of *citT* was amplified, and Sanger sequencing determined that each altered product was caused by either a small deletion or an IS3 insertion in *citG*. To identify mutations at other loci, as well as mutations affecting the *cit* region, we re-sequenced the genomes of six mutants, as described above. These genomes included three of the four mutants for which the mutations conferring the Cit⁺ phenotype were not resolved using the approaches described above.

ACKNOWLEDGEMENTS

We thank N. Hajela, M. Kauth and S. Sleight for laboratory assistance, J. Meyer for discussion, and C. Turner for comments on the manuscript. We thank the MSU Research Technology Support Facility for sequencing services. We acknowledge support from the

DARPA “Fun Bio” Program (HR0011-09-1-0055 to R.E.L), the US National Institutes of Health (K99GM087550 to J.E.B.); the US National Science Foundation (DEB-1019989 to R.E.L) including the BEACON Center for the Study of Evolution in Action (DBI-0939454), a Rudolf Hugh Fellowship (to Z.D.B.), a DuVall Family Award (to Z.D.B.), and a Barnett Rosenberg Fellowship (to Z.D.B).

APPENDIX

APPENDIX: Supplemental Data Table for Chapter 3

Appendix Table A.1 Phylogenetic placement of Cit⁻ replay clones													
Gen (k)	Clone	Phylogenetically Informative Mutations											Clade
		Gene	<i>ybaL</i>	<i>nadR</i>	<i>hemE</i>	<i>cspC</i>	<i>yaaH</i>	<i>leuA</i>	<i>tolR</i>	<i>arcB</i>	<i>arcB</i>	<i>gltA</i>	
		Gene Position	133	1010	636	4	521	778	69	208	236	772	
		Ancestral Nucleotide	G	A	T	C	T	C	C	G	T	G	
		Evolved Nucleotide	C	C	G	A	G	T	T	C	A	A	
0	REL606*												Ancestor
5	ZDB400												n.d.
5	ZDB401												n.d.
5	ZDB402												n.d.
5	ZDB403												n.d.
5	ZDB404												n.d.
5	ZDB405												n.d.
5	ZDB406												n.d.
5	ZDB407												n.d.
5	ZDB408												n.d.
5	ZDB409*												n.d.
5	ZDB410												n.d.
5	ZDB411												n.d.
5	ZDB412												n.d.
5	ZDB413												n.d.
5	ZDB414												n.d.
5	ZDB415												n.d.
5	ZDB416												n.d.
5	ZDB417												n.d.
5	ZDB418												n.d.
5	ZDB419												n.d.
10	ZDB1												(C1,C2, C3)
10	ZDB2												(C1,C2, C3)
10	ZDB3												(C1,C2, C3)
10	ZDB420												n.d.
10	ZDB421												UC

Appendix Table A.1 Phylogenetic placement of Cit ⁻ replay clones													
Gen (k)	Clone	Phylogenetically Informative Mutations											Clade
		Gene	<i>ybaL</i>	<i>nadR</i>	<i>hemE</i>	<i>cspC</i>	<i>yaaH</i>	<i>leuA</i>	<i>tolR</i>	<i>arcB</i>	<i>arcB</i>	<i>gltA</i>	
		Gene Position	133	1010	636	4	521	778	69	208	236	772	
		Ancestral Nucleotide	G	A	T	C	T	C	C	G	T	G	
		Evolved Nucleotide	C	C	G	A	G	T	T	C	A	A	
10	ZDB422												UC
10	ZDB423												n.d.
10	ZDB424												(C1,C2, C3)
10	ZDB425												(C1,C2, C3)
10	ZDB426												UC
10	ZDB427												n.d.
10	ZDB428												UC
10	ZDB429*												UC
10	ZDB430												UC
10	ZDB431												UC
10	ZDB432												n.d.
10	ZDB433												n.d.
10	ZDB434												UC
10	ZDB435												n.d.
10	ZDB436												UC
15	ZDB437												n.d.
15	ZDB438												(C1,C2, C3)
15	ZDB439												n.d.
15	ZDB440												UC
15	ZDB441												UC
15	ZDB442												n.d.
15	ZDB443												UC
15	ZDB444												n.d.
15	ZDB445												(C2,C3)
15	ZDB446*												UC
15	ZDB447												UC
15	ZDB448												(C2,C3)
15	ZDB449												n.d.

Appendix Table A.1 Phylogenetic placement of Cit ⁻ replay clones													
Gen (k)	Clone	Phylogenetically Informative Mutations											Clade
		Gene	<i>ybaL</i>	<i>nadR</i>	<i>hemE</i>	<i>cspC</i>	<i>yaaH</i>	<i>leuA</i>	<i>tolR</i>	<i>arcB</i>	<i>arcB</i>	<i>gltA</i>	
		Gene Position	133	1010	636	4	521	778	69	208	236	772	
		Ancestral Nucleotide	G	A	T	C	T	C	C	G	T	G	
		Evolved Nucleotide	C	C	G	A	G	T	T	C	A	A	
15	ZDB450												UC
15	ZDB451												n.d.
15	ZDB452												UC
15	ZDB453												n.d.
15	ZDB454												n.d.
15	ZDB455												(C2,C3)
15	ZDB456												n.d.
20	ZDB4												(C2,C3)
20	ZDB5												(C2,C3)
20	ZDB6												C1
20	ZDB457												(C2,C3)
20	ZDB458*												(C2,C3)
20	ZDB459												C1
20	ZDB460												(C2,C3)
20	ZDB461												(C2,C3)
20	ZDB462												(C2,C3)
20	ZDB463												(C2,C3)
20	ZDB464*												(C2,C3)
20	ZDB465												(C2,C3)
20	ZDB466												C1
20	ZDB467*												(C2,C3)
20	ZDB468												C1
20	ZDB469												(C2,C3)
20	ZDB470												(C2,C3)
20	ZDB471												(C2,C3)
20	ZDB472												(C2,C3)
20	ZDB473												(C2,C3)
25	ZDB7												(C2,C3)
25	ZDB8												C2
25	ZDB9												C1

Appendix Table A.1 Phylogenetic placement of Cit ⁻ replay clones													
Gen (k)	Clone	Phylogenetically Informative Mutations										Clade	
		Gene	<i>ybaL</i>	<i>nadR</i>	<i>hemE</i>	<i>cspC</i>	<i>yaaH</i>	<i>leuA</i>	<i>tolR</i>	<i>arcB</i>	<i>arcB</i>		<i>gltA</i>
		Gene Position	133	1010	636	4	521	778	69	208	236		772
		Ancestral Nucleotide	G	A	T	C	T	C	C	G	T		G
		Evolved Nucleotide	C	C	G	A	G	T	T	C	A		A
25	ZDB474												C2
25	ZDB475												C1
25	ZDB476												C3
25	ZDB477												C1
25	ZDB478												C3
25	ZDB479												C3
25	ZDB480												(C2,C3)
25	ZDB481												C3
25	ZDB482												(C2,C3)
25	ZDB483												C3
25	ZDB484												(C2,C3)
25	ZDB485												C3
25	ZDB486												C3
25	ZDB487												C3
25	ZDB488												C3
25	ZDB489												C3
25	ZDB490												C3
27	ZDB300												C3
27	ZDB301												C2
27	ZDB302												C1
27	ZDB303												C1
27	ZDB304												C2
27	ZDB305												C2
27	ZDB306												C2
27	ZDB307												C2
27	ZDB308												C2
27	ZDB309												C3
27	ZDB310												C3
27	ZDB311												C3
27	ZDB312												C3

Appendix Table A.1 Phylogenetic placement of Cit ⁻ replay clones													
Gen (k)	Clone	Phylogenetically Informative Mutations										Clade	
		Gene	<i>ybaL</i>	<i>nadR</i>	<i>hemE</i>	<i>cspC</i>	<i>yaaH</i>	<i>leuA</i>	<i>tolR</i>	<i>arcB</i>	<i>arcB</i>		<i>gltA</i>
		Gene Position	133	1010	636	4	521	778	69	208	236		772
		Ancestral Nucleotide	G	A	T	C	T	C	C	G	T		G
		Evolved Nucleotide	C	C	G	A	G	T	T	C	A		A
27	ZDB313												C1
27	ZDB314												C3
27	ZDB315												C2
27	ZDB316												C2
27	ZDB317												C3
27	ZDB318												C3
27	ZDB319												C2
27	ZDB10												C2
27	ZDB11												C1
27	ZDB12												C2
28	ZDB320												C2
28	ZDB321												C2
28	ZDB322												C1
28	ZDB323												C2
28	ZDB324												C2
28	ZDB325												C2
28	ZDB326												C2
28	ZDB327												C2
28	ZDB328												C2
28	ZDB329												C1
28	ZDB330												C2
28	ZDB331												C2
28	ZDB332												C2
28	ZDB333												C2
28	ZDB334												C3
28	ZDB335												C2
28	ZDB336												C2
28	ZDB337												C2
28	ZDB338												C2
28	ZDB339												C3

Appendix Table A.1 Phylogenetic placement of Cit ⁻ replay clones													
Gen (k)	Clone	Phylogenetically Informative Mutations										Clade	
		Gene	<i>ybaL</i>	<i>nadR</i>	<i>hemE</i>	<i>cspC</i>	<i>yaaH</i>	<i>leuA</i>	<i>tolR</i>	<i>arcB</i>	<i>arcB</i>		<i>gltA</i>
		Gene Position	133	1010	636	4	521	778	69	208	236		772
		Ancestral Nucleotide	G	A	T	C	T	C	C	G	T		G
		Evolved Nucleotide	C	C	G	A	G	T	T	C	A		A
29	ZDB13												C3
29	ZDB14												C3
29	ZDB15												C2
29	ZDB340												C2
29	ZDB341												C2
29	ZDB342												C1
29	ZDB343												C2
29	ZDB344												C2
29	ZDB345												C2
29	ZDB346												C1
29	ZDB347												C1
29	ZDB348												C2
29	ZDB349												C1
29	ZDB350												C1
29	ZDB351												C2
29	ZDB352												C1
29	ZDB353												C2
29	ZDB354												C2
29	ZDB355												C1
29	ZDB356												C2
30	ZDB16*												C1
30	ZDB17												C3
30	ZDB18												C3
30	ZDB357*												C2
30	ZDB358												C2
30	ZDB359												C2
30	ZDB360												C1
30	ZDB361												C2
30	ZDB362												C2
30	ZDB363												C2

Appendix Table A.1 | Phylogenetic placement of Cit⁻ replay clones

Gen (k)	Clone	Phylogenetically Informative Mutations											Clade
		Gene	<i>ybaL</i>	<i>nadR</i>	<i>hemE</i>	<i>cspC</i>	<i>yaaH</i>	<i>leuA</i>	<i>tolR</i>	<i>arcB</i>	<i>arcB</i>	<i>gltA</i>	
		Gene Position	133	1010	636	4	521	778	69	208	236	772	
		Ancestral Nucleotide	G	A	T	C	T	C	C	G	T	G	
		Evolved Nucleotide	C	C	G	A	G	T	T	C	A	A	
30	ZDB364												C2
30	ZDB365												C2
30	ZDB366												C2
30	ZDB367												C2
30	ZDB368												C2
30	ZDB369												C2
30	ZDB370												C1
30	ZDB371												C2
30	ZDB372												C2
30	ZDB373												C2
30.5	ZDB19												C3
30.5	ZDB20												C3
30.5	ZDB21												C2
31	ZDB22												C2
31	ZDB23												C3
31	ZDB24												C1
31	ZDB374												C2
31	ZDB375												C2
31	ZDB376												C2
31	ZDB377												C1
31	ZDB378												C2
31	ZDB379												C2
31	ZDB380												C2
31	ZDB381												C1
31	ZDB382												C2
31	ZDB383												C2
31	ZDB384												C2
31	ZDB385												C1
31	ZDB386												C2
31	ZDB387												C2

Appendix Table A.1 Phylogenetic placement of Cit ⁻ replay clones													
Gen (k)	Clone	Phylogenetically Informative Mutations										Clade	
		Gene	<i>ybaL</i>	<i>nadR</i>	<i>hemE</i>	<i>cspC</i>	<i>yaaH</i>	<i>leuA</i>	<i>tolR</i>	<i>arcB</i>	<i>arcB</i>		<i>gltA</i>
		Gene Position	133	1010	636	4	521	778	69	208	236		772
		Ancestral Nucleotide	G	A	T	C	T	C	C	G	T		G
		Evolved Nucleotide	C	C	G	A	G	T	T	C	A		A
31	ZDB388												C2
31	ZDB389												C2
31	ZDB390												C1
31.5	ZDB25												C3
31.5	ZDB26												C1
31.5	ZDB27												C3
31.5	ZDB197												C2
31.5	ZDB198												C1
31.5	ZDB199*												C1
31.5	ZDB200*												C2
31.5	ZDB201												C1
31.5	ZDB202												C2
31.5	ZDB203												C2
31.5	ZDB204												C1
31.5	ZDB205												C1
31.5	ZDB206												C2
31.5	ZDB207												C2
31.5	ZDB208												C2
31.5	ZDB209												C2
31.5	ZDB210												C2
31.5	ZDB211												C2
31.5	ZDB212												C2
31.5	ZDB213												C2
31.5	ZDB214												C2
31.5	ZDB215												C1
31.5	ZDB216												C2
32	ZDB28												C3
32	ZDB29												C2
32	ZDB30*												C3
32	ZDB183												C3

Appendix Table A.1 Phylogenetic placement of Cit ⁻ replay clones													
Gen (k)	Clone	Phylogenetically Informative Mutations										Clade	
		Gene	<i>ybaL</i>	<i>nadR</i>	<i>hemE</i>	<i>cspC</i>	<i>yaaH</i>	<i>leuA</i>	<i>tolR</i>	<i>arcB</i>	<i>arcB</i>		<i>gltA</i>
		Gene Position	133	1010	636	4	521	778	69	208	236		772
		Ancestral Nucleotide	G	A	T	C	T	C	C	G	T		G
		Evolved Nucleotide	C	C	G	A	G	T	T	C	A		A
32	ZDB184												C1
32	ZDB185												C3
32	ZDB186												C1
32	ZDB187												C1
32	ZDB188												C2
32	ZDB189												C1
32	ZDB190												C2
32	ZDB191												C3
32	ZDB192												C2
32	ZDB391												C2
32	ZDB392												C2
32	ZDB393												C2
32	ZDB394												C1
32	ZDB395												C2
32	ZDB396												C2
32	ZDB397												C1
32.5	ZDB31												C3
32.5	ZDB32												C2
32.5	ZDB33												C1
32.5	ZDB146												C1
32.5	ZDB147												C1
32.5	ZDB148												C1
32.5	ZDB149												C1
32.5	ZDB150												C1
32.5	ZDB151												C1
32.5	ZDB152												C1
32.5	ZDB153												C1
32.5	ZDB154												C1
32.5	ZDB155												C1
32.5	ZDB156												C1

Appendix Table A.1 Phylogenetic placement of Cit ⁻ replay clones													
Gen (k)	Clone	Phylogenetically Informative Mutations										Clade	
		Gene	<i>ybaL</i>	<i>nadR</i>	<i>hemE</i>	<i>cspC</i>	<i>yaaH</i>	<i>leuA</i>	<i>tolR</i>	<i>arcB</i>	<i>arcB</i>		<i>gltA</i>
		Gene Position	133	1010	636	4	521	778	69	208	236		772
		Ancestral Nucleotide	G	A	T	C	T	C	C	G	T		G
		Evolved Nucleotide	C	C	G	A	G	T	T	C	A		A
32.5	ZDB157												C3
32.5	ZDB158*												C2
32.5	ZDB159												C2
32.5	ZDB160												C1
32.5	ZDB398												C1
32.5	ZDB399												C2

*Genome has been sequenced.

Red fill indicates presence of mutation has been established by sequencing. Gray fill indicates absence of mutation has been established by sequencing. No fill indicates that the presence or absence of mutation was not examined.

Clade refers to UC, C1, C2, or C3 as shown in Fig. 1 of the main text. When two or more clades are grouped by parentheses, either the clone belongs to the basal group or the clone's placement could not be resolved further based on the available data. n.d. indicates that the clone belongs to some other early clade or its placement could not be resolved based on the available data.

REFERENCES

REFERENCES

1. Adam, D., Dimitrijevic, N. & Scharl, M. Tumor suppression in *Xiphophorus* by an accidentally acquired promoter. *Science* **259**, 816–819 (1993).
2. Andersson, D.I., Slechta, E.S. & Roth, J.R. Evidence that gene amplification underlies adaptive mutability of the bacterial *lac* operon. *Science* **282**, 1133–1135 (1998).
3. Barrick, J.E. & Knoester, D.B. breseq. <http://barricklab.org/breseq> (2010).
4. Barrick, J.E. & Lenski, R.E. Genome-wide mutational diversity in an evolving population of *Escherichia coli*. *Cold Spring Harbor Symposia on Quantitative Biology* **74**, 1–11 (2009).
5. Barrick, J.E., Yu, D.S., Yoon, S.H., Jeong, H., Oh, T.K., Schneider, D., Lenski, R.E., & Kim, J.F. Genome evolution and adaptation in a long-term experiment with *E. coli*. *Nature* **461**, 1243–1247 (2009).
6. Bentley, D.R. Whole-genome resequencing. *Current Opinion in Genetics and Development* **16**, 545–552 (2006).
7. Bergthorsson, U., Andersson, D.I. & Roth, J.R. Ohno's dilemma: evolution of new genes under continuous selection. *Proceedings of the National Academy of Sciences (USA)* **104**, 17004–17009 (2007).
8. Bjarnason, J., Southward, C.M. & Surette, M.G. Genomic profiling of iron-responsive genes in *Salmonella enterica* Serovar Typhimurium by high-throughput screening of a random promoter library. *Journal of Bacteriology*, **185**, 4973–4982 (2003).
9. Blount, Z.D., Borland, C.Z. & Lenski, R.E. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proceedings of the National Academy of Sciences (USA)* **105**, 7899–7906 (2008).
10. Bock, R. & Timmis, J.N. Reconstructing evolution: gene transfer from plastics to the nucleus. *BioEssays* **30**, 556–566 (2008).
11. Camin, J.H. & Sokal, R.R. A method for deducing branching sequences in phylogeny. *Evolution* **19**, 311–326 (1965).
12. Charlier, D., Piette, J. & Glansdorff, N. IS3 can function as a mobile promoter. *Nucleic Acids Research* **10**, 5935–5948 (1982).
13. Christensen, W.B. Hydrogen sulfide production and citrate utilization in the differentiating of enteric pathogens and coliform bacteria. *Research Bulletin Weld County Health Department of Greeley Colorado* **1**, 3–16 (1949).
14. Elena, S.F. & Lenski, R.E. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Reviews Genetics* **4**: 457–469 (2003).

15. Fogle, C.A., Nagle, J.L. & Desai, M.M. Clonal interference, multiple mutations and adaptation in large asexual populations. *Genetics* **180**, 2163–2173 (2008).
16. Glickman, B.W. & Radman, M. *Escherichia coli* mutator mutants deficient in methyl-instructed DNA mismatch correction. *Proceedings of the National Academy of Sciences (USA)* **77**, 1063–7 (1980).
17. Gould, S.J. & Vrba, E.S. Exaptation – a missing term in the science of form. *Paleobiology* **8**, 4–15 (1982).
18. Gunsalus, R.P. & Park, S.J. Aerobic-anaerobic gene regulation in *Escherichia coli*: control by the ArcAB and Fnr regulons. *Research in Microbiology*. **145**, 437–450 (1994).
19. Hall, B.G. Chromosomal mutation for citrate utilization by *Escherichia coli* K-12. *Journal of Bacteriology* **151**, 269–273 (1982).
20. Hegreness, M. & Kishony, R. Analysis of genetic systems using experimental evolution and whole-genome sequencing. *Genome Biology* **8**, 201 (2007).
21. Herring, C.D., Glasner, J.D. & Blattner, F.R. Gene replacement without selection: regulated suppression of amber mutations in *Escherichia coli*. *Gene* **311**, 153–63 (2003).
22. Jacob, F. Evolution and tinkering. *Science* **196**, 1161–1166 (1977).
23. Jacob, F. *The Possible and the Actual*. (Univ. Washington Press, Seattle, 1982).
24. Janausch, I.G., Zientz, E., Tran, Q.H., Kroger, A. & Uden, G. C4-dicarboxylate carriers and sensors in bacteria. *Biochimica et Biophysica Acta* **1553**, 39–56.
25. Jeong, H., Barbe, V., Lee, C.H., Vallenet, D., Yu, D.S., Choi, S.H., Couloux, A., Lee, S.W., Yoon, S.H., Cattolico, L., Hur, C.G., Park, H.S., Seurens, B., Kim, S.C, Oh, T.K., Lenski, R.E., Studier, F.W., Daegelen, P., & Kim, J.F. 2009. Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *Journal of Molecular Biology* **394**, 644–652 (2009).
26. Koser, S.A. Correlation of citrate-utilization by members of the colon-aerogenes group with other differential characteristics and with habitat. *Journal of Bacteriology* **9**, 59–77 (1924).
27. Lara, F.J.S. & Stokes, J.L. Oxidation of citrate by *Escherichia coli*. *Journal of Bacteriology* **63**, 415–420 (1952).
28. Lenski, R.E. Phenotypic and genomic evolution during a 20,000-generation experiment with the bacterium *Escherichia coli*. *Plant Breeding Reviews* **24**, 225–265 (2004).
29. Lenski, R.E., Ofria, C., Pennock, R.T. & Adami, C. The evolutionary origin of complex features. *Nature* **423**, 139–144 (2003).

30. Lenski, R.E., Rose, M.R., Simpson, S.C. & Tadler, S.C. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *American Naturalist* **138**, 1315–1341 (1991).
31. Lutgens, M. & Gottschalk, G.. Why a co-substrate is required for anaerobic growth of *Escherichia coli* on citrate. *Journal of General Microbiology* **199**, 63–70 (1980).
32. Mayr, E. “The emergence of evolutionary novelties”. In *Evolution after Darwin* Tax, S. (Ed.) (University of Chicago Press, Chicago, 1960).
33. Nizam, S.A., Zhu, J, Ho, P.Y. & Shimizu, K. Effects of *arcA* and *arcB* genes knockout on the metabolism in *Escherichia coli* under aerobic condition. *Biochemical Engineering Journal* **44**, 240–250 (2009).
34. Patthy, L. Genome evolution and the evolution of exon-shuffling – a review. *Gene* **238**, 103–114 (1999).
35. Pigliucci, M. What, if anything, is an evolutionary novelty? *Philosophy of Science* **75**, 887–898 (2008).
36. Pos, K., Dimroth, P. & Bott, M. 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 (1998).
37. Reams, D., Kofoed, E., Savageau, M. & Roth, J. R. Duplication frequency in a population of *Salmonella enterica* rapidly approaches steady state with or without recombination. *Genetics* **184**, 1077–1094 (2010).
38. Rozen, D.E., Schneider, D. & Lenski, R.E. Long-term experimental evolution in *Escherichia coli*. XIII. Phylogenetic history of a balanced polymorphism. *Journal of Molecular Biology* **61**, 171–180 (2005).
39. Scheutz, F. & Strockbine, N.A. Genus I. *Escherichia*, Castellani and Chalmers 1919, *Bergey's Manual of Systematic Bacteriology, Volume 2: The Proteobacteria*, Garrity, G.M., Brenner, D.J., Kreig, N.R. & Staley, J.R. (Eds.). (Springer, New York, NY, 2005).
40. Shankar, S., Schlichtman, D., & Chakrabarty, A.M. Regulation of nucleoside diphosphate kinase and an alternative kinase in *Escherichia coli*: role of the *sspA* and *rnk* genes in nucleoside triphosphate formation. *Molecular Microbiology* **17**, 935–943 (1995).
41. Sniegowski, P.D., Gerrish, P.J. & Lenski, R.E. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* **387**, 703–705 (1997).
42. Taylor, J.S. & Raes, J. Duplication and divergence: the evolution of new genes and old ideas. *Annual Review of Genetics* **38**, 615–643 (2004).

43. Treves, D.S., Manning, S. & Adams, J. Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli*. *Molecular Biology and Evolution* **15**, 789–797 (1998).
44. True, J.R. & Carroll, S.B. Gene co-option in physiological and morphological evolution. *Annual Review of Cell and Developmental Biology* **18**, 53–80 (2002).
45. Usakin, L.A., Kogan, G.L., Kalmykova, A.I. & Gvozdev, V.A. An alien promoter capture as a primary step of the evolution of testes-expressed repeats in the *Drosophila melanogaster* genome. *Molecular Biology and Evolution* **22**, 1555–1560 (2005).
46. Whoriskey, S.K., Nghiem, V., Leong, P., Masson, J. & Miller, J.H. Genetic rearrangements and gene amplification in *Escherichia coli*: DNA sequences at the junctures of amplified gene fusions. *Genes Development* **1**, 227–237 (1987).
47. Yanisch-Perron, C., Vieira, J., & Messing, J. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors. *Gene* **33**, 103–119 (1985).
48. Zaslaver, A., Bren, A., Ronen, M., Itzhovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M.G., & Alon, U. A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nature Methods* **3**, 623–628 (2006).
49. Zhang, J. Evolution by gene duplication: an update. *Trends in Ecology and Evolution* **18**, 282–298 (2003).
50. Zhu, L. & Deutscher, M.P. The *Escherichia coli rna* gene encoding RNase I: sequence and unusual promoter structure. *Gene* **119**, 1–6 (1992).