TRAIT VARIABILITY IN BACTERIA: STUDIES IN DNA METHYLATION AND SELECTIVE AMPLIFICATION

A Dissertation

Presented in Partial Fulfillment of the Requirements for the

Degree of Doctorate of Philosophy

with a

Major in Microbiology, Molecular Biology, and Biochemistry

in the

College of Graduate Studies

University of Idaho

by

Dustin J. Van Hofwegen

Major Professor: Scott A. Minnich, Ph.D.

Committee Members: Carolyn J. Hovde, Ph.D., Patricia L. Hartzell, Ph.D., and

Patrick J. Hrdlicka, Ph.D.

Department Head: James J. Nagler, Ph.D.

August 2016

Authorization to Submit Dissertation

This dissertation of Dustin J. Van Hofwegen, submitted for the degree of Doctor of Philosophy with a Major in Microbiology, Molecular Biology, and Biochemistry and titled “Trait variability in bacteria: studies in DNA methylation and selective amplification,” has been reviewed in final form. Permission, as indicated by the signatures and dates below, is now granted to submit final copies to the College of Graduate Studies of approval.

Major Professor: _____ Date: _____
Scott A. Minnich, Ph.D.

Committee Members: _____ Date: _____
Carolyn J. Hovde, Ph.D.

_____ Date: _____
Patricia L. Hartzell, Ph.D.

_____ Date: _____
Patrick J. Hrdlicka, Ph.D.

Department Chair: _____ Date: _____
James J. Nagler, Ph.D.

Abstract

Eubacteria recognize environmental cues to vary gene expression. The mechanisms used are often determined by modifications to their DNA, including variations to the linear sequence and structure, and epigenetic modifications. Using two bacterial model systems, *Escherichia coli* and pathogenic *Yersinia enterocolitica*, we investigated two modification mechanisms— selective amplification and epigenetic DNA methylation.

Y. enterocolitica respond when encountering a shift to host temperature by immediately repressing motility and activating plasmid-encoded virulence operons. In contrast, shifting from host to ambient temperature requires 2.5 generations to restore motility suggesting a link to the cell cycle. We hypothesized that differential DNA methylation contributes to temperature-regulated gene expression. We tested this hypothesis by comparing single-molecule real-time (SMRT) sequencing of *Y. enterocolitica* DNA from cells growing exponentially at 22°C and 37°C. Among the 27,118 DNA adenine methylase (Dam) sites, we identified 42 had differential methylation patterns while 17 remained unmethylated regardless of temperature. A subset of the differentially methylated Dam sites localized to promoter regions of predicted regulatory genes including LysR-type and PadR-like transcriptional regulators, and a cyclic-di-GMP phosphodiesterase. Identification of these differences in methylation provide likely candidates for regulators responsible for temperature-dependent phenotypes.

In *E. coli*, the isolation of an aerobic citrate-utilizing (Cit^+) strain in long-term evolution experiments (LTEE) has been termed a rare, innovative, presumptive speciation event. We hypothesized that direct selection would rapidly produce the selective amplification events necessary to yield the same class of *E. coli* Cit^+ mutants. Using similar media, 46 independent citrate-utilizing mutants were isolated in as few as 12 to 100 generations. Genomic DNA sequencing revealed an amplification of the *citT* and *dctA* loci and DNA rearrangements to capture a promoter to express CitT, aerobically. These are members of the same class of mutations identified by the LTEE. We conclude that the rarity of the LTEE mutant was an artifact of the experimental conditions and not a unique evolutionary event. No new genetic information (novel gene function) evolved.

These two systems show how bacteria respond to adverse environmental conditions— the host immune system (*Yersinia*) and starvation (*E. coli*).

Acknowledgements

Countless people have been enormously helpful during my doctoral education and initiation into this area of scientific research. Paramount in this has been my advisor and mentor, Dr. Scott A. Minnich— from whom I have had the fantastic privilege of discovering microbiology and microbial genetics. His immense depth of knowledge, wisdom, and experience overflows to those who have the privilege of knowing him. Sam's passion for uncovering nature's secrets is infectious and has significantly expanded my fascination for this concealed, yet distinguishable world. I am deeply thankful for Sam's patience, encouragement, guidance, and confidence in my abilities.

To my committee members, Carolyn J. Hovde, Patricia L. Hartzell, and Patrick J. Hrdlicka, for their recommendations and advice. Committee meetings were often the most stressful part of my doctoral studies, yet the advice I received from my committee always helped to enhance my projects. They helped make my time more productive and my endeavors more successful. Thank you! Special thanks to Carolyn Hovde, who has also provided invaluable editorial assistance. Carolyn is a fantastic writer and editor, and this dissertation and published manuscripts are all the much better because of her contribution.

Several people have provided invaluable technical assistance during my research projects. I thank research assistants Claudia Deobald and Harold Rohde for their masterful, yet patient instruction in laboratory techniques— I am a successful scientist today because of their assistance. Sequencing services were

provided by Derek Pouchnik and Mark Wildung at Washington State University— I thank them for their many discussions about and enthusiasm for DNA sequencing technologies. I thank Robert Sebra and the staff at the Genomics Core facility at the Icahn School of Medicine at Mt. Sinai Hospital and Jonas Korlach and Khai Luong at Pacific Biosciences for additional sequencing services and expertise. I also thank Lee Deobald, University of Idaho's adept mass spectrometry expert, for providing superior services.

I would also like to thank the makers of several incomparable programs and services that were indispensable in writing this dissertation. It was written on a Mac using Word, backed up constantly onto 9 different drives, on 3 different servers, and on 3 separate Macs in different locations with the help of Dropbox, iCloud, Time Machine, and Backblaze. All citations were handled using Papers, sequencing data was analyzed using R, 4Peaks, and CLC Genomics Workbench, and graphs and spreadsheets were prepared using Excel and Keynote.

Much appreciation to the faculty of what was the MMBB department here at the University of Idaho, who have provided knowledgeable education. This project would not have been conceived were it not for the necessary depth of knowledge their instruction provided.

To the Idaho INBRE program for continued financial support. May this program continue to successfully foster scientists and cultivate cutting-edge research in this great state of Idaho!

Special thanks to my two sets of parents for unwavering support, even in the face of uncertainty and seeming absurdity. To my parents, Brad and Marion, who

always encouraged me to pursue my education, even if it took me away from home. My dad taught me many very important lessons in life, including one in particular that has helped shape my scientific skillset— that there is more than one side to any argument. To my wife's parents, Brian and Jackie, whose steadfast support has enriched our family. Thank you for all the many blessings, counsel, and encouragement.

Finally, I am most grateful to the One who made the universe, including the mind with which I seek to comprehend it. I have accomplished nothing apart from Him.

Dedication

To the three ladies in my life, who keep me sane yet avid, humbled yet fortunate.

To Dena, for her immense courage, faithful support, and selfless devotion. We could not have continued in this endeavor were it not for you. Thank you for your commitment, encouragement, and love. To Carina, your fervor and happiness are infectious. Thank you for making us laugh! To Anika, your sweetness and grace are a blessing to us all. Thank you for all the hugs, songs, and smiles! I love you all!!

Table of Contents

List of Figures

List of Tables

Preface

This dissertation is the culmination of my years of doctoral study under the direction of Scott A. Minnich on two topics. First, I analyzed DNA methylation as a mechanism of temperature-regulated gene regulation in *Yersinia*. The background for this investigation is presented in Chapter 1. Chapter 2 describes our investigation of methylation modifications to the *Y. enterocolitica* chromosome as it relates to temperature-regulated phenotypes. This work was published in *Microbial Genomics*. Second, I challenged the interpretation of Richard Lenski's long-term evolution experiment (LTEE) that describe *E. coli's* adaptation to use citrate as a carbon source, aerobically. The LTEE *E. coli* mutant(s) that use citrate aerobically have been heralded as a breakthrough result in understanding evolution as well as a rare speciation event. The background for our investigation is in Chapter 3. Chapter 4 describes our experiments that debunk the LTEE interpretation and show the rapid and repeated isolation of *E. coli* mutants that use citrate as a carbon source, aerobically. This work was published in the *Journal of Bacteriology*, and has been the journal's most-read article for three of the first four and in the top 20 for each of the first five months since its availability. Our publication was highlighted by the editor-in-chief, Thomas Silhavy, with a Commentary, a practice reserved for articles of broad interest and significant impact. This Commentary was written by the eminent microbial geneticist, and National Academy member, John Roth and his postdoctoral fellow, Sophie Maisnier-Patin. Also, Richard Lenski, and his postdoctoral fellow, Zachary Blount, responded to our article and the associated

Commentary on Lenski's blog. Because of the high visibility of our publication, the commentary, and the blog, Chapter 5 is an Epilogue outlining their comments and giving our response.

CHAPTER 1

LITERATURE REVIEW—Thermoregulation in *Yersinia*

Yersinia pathogenesis

The genus *Yersinia* is classified in the Enterobactieriacaea family of Gammaproteobacteria. Like all organisms within this family, *Yersinia* are Gram-negative, rod-shaped, facultative anaerobes. Within this genus there are 14 species, four of which are pathogenic. *Y. ruckeri* is the causative agent of 'red mouth' disease in salmonid fish. Until the advent of an effective vaccine, *Y. ruckeri* had an important economic impact on the aqua culture industry, particularly in Idaho (Sulakvelidze, 2000). Three of four pathogenic strains are agents of human disease: *Y. pestis*, *Y. enterocolitica*, and *Y. pseudotuberculosis*. Both *Y. enterocolitica*, and *Y. pseudotuberculosis* cause severe, but self-limiting, Yersiniosis which is marked by diarrhea, abdominal pain caused by enteritis or ileitis, and fever (Carniel and Mollaret, 1990; Bottone, 1999). Clinical manifestations of infection are usually enteric, with symptoms ranging from abdominal pain and mild diarrhea to severe ulcerative lesions within the gastrointestinal tract, fever and vomiting. In contrast, *Y. pestis* is the etiological agent of bubonic, pneumonic, and septicemic plague, a systemic disease with mortality reaching 100% in untreated cases (Perry and

Fetherston, 1997). *Y. enterocolitica* is the most prevalent pathogen of the genus, and the CDC lists it as one of the leading causes of foodborne illness in the United States. Infection is thought to be contracted via the consumption of undercooked meat products, unpasteurized milk, and contaminated water (Tauxe *et al.*, 1987; Huovinen *et al.*, 2010). A recent study has shown that 69% of supermarket packages of pork were contaminated with *Y. enterocolitica* (What's in that pork? We found antibiotic-resistant bacteria--and traces of a veterinary drug., 2013).

Y. enterocolitica has served as the primary model organism in evaluating Gram-negative bacterial pathogenesis (Cornelis *et al.*, 1987). This is because there is a good animal model of infection and it is easily manipulated genetically. Several significant findings have emerged from the *Yersinia* model: 1) affirming plasmids essential for bacterial virulence (Ben-Gurion and Shafferman, 1981), 2) defining islands of pathogenicity (Carniel, 2001; Koczura and Kaznowski, 2003a; Koczura and Kaznowski, 2003b), 3) identifying the first invasin protein (Isberg and Leong, 1990), 4) providing the archetype for type III secretion systems (Perry and Fetherston, 1997; Minnich and Rohde HN, 2007), 5) describing the role of iron sequestration during infection (Carniel, 2001), 6) revealing how DNA topology contributes to virulence gene activation (Cornelis *et al*., 1989; Rohde JR *et al.*, 1994), and 7) temperature regulation of gene expression (Rohde JR *et al.*, 1994; Kapatral and Minnich, 1995; Straley and Perry, 1995; Kapatral *et al.*, 1996; Rohde JR *et al.*, 1999; Han *et al.*, 2007). Additionally, pathogenic *Yersinia* has been a model of both macro and micro evolution among bacterial species, significantly in regards to its role in genome reduction in bacterial pathogenesis (Achtman *et al.*,

1999; Achtman *et al.*, 2004; Reuter *et al.*, 2014; McNally *et al.*, 2016). Furthermore, each of the pathogenic strains of invade lymphoid tissue and share a common ability to evade the innate host immune system (Cornelis *et al.*, 1987; Cornelis *et al.*, 1998). *Y. enterocolitica*, due to its genetic relatedness to *Y. pestis*, has served as an important model for studying *Yersinia* pathogenesis. The impact of understanding bacterial pathogenesis using the *Yersinia* paradigm has been profound.

Temperature Regulation

Early studies of temperature-responsive gene regulation in *Yersinia* were prompted by reciprocal regulation of plasmid-encoded virulence factors. Most virulence factors are thermally regulated, active at either 28°C (the optimal growth temperature) or 37°C (the host temperature). Motility is also thermally regulated, *Y. enterocolitica* and *Y. pseudotuberculosis* are motile at temperatures at or below 28°C and nonmotile at 37°C (Cornelis *et al.*, 1987; Kapatral *et al.*, 1996). Both strains, upon encountering a shift to host temperature of 37°C, immediately suppress motility. Conversely, upon encountering a shift from high to low temperature, motility is not immediately restored, rather, reactivation of the motility operon is achieved only following 2.5 generations of growth (Rohde JR *et al.*, 1994; Kapatral and Minnich, 1995), indicating a link to the cell cycle.

Using temperature as the stimulus confers a sensing mechanism whereby the organisms can react to changing and often-hazardous environmental conditions (Steinmann and Dersch, 2013). Like other bacterial species, *Yersinia* employ many

sophisticated strategies to monitor their local environment and respond with defensive or adaptive measures. Other environmental cues that the bacteria monitor include pH, nutrient availability, osmolarity, and ion concentration (Straley and Perry, 1995; Chen *et al.*, 2016). Pathogenic *Yersinia* are zoonotic pathogens, frequently circulating between vectors or environmental reservoirs and warm-blooded hosts, and therefore frequently encountering variations in temperature. It is not surprising, then, that these mammalian pathogens respond to elevated temperature to activate virulence functions.

Thermosensing via DNA

Temperature-induced changes in DNA are exhibited by alterations of DNA topology, supercoiling, or curvature of intrinsic DNA bending on both the chromosome and virulence plasmid (Dorman, 1991; Hurme and Rhen, 1998; Travers and Muskhelishvili, 2005). DNA physically changes the number of helical turns (twist) and the number times the helix winds around itself (writhe) in response to shifts in temperature. Intrinsic bends, whereby the DNA physically curves, occur predominantly in the 5'-regulatory regions of genes at AT-rich sequences (Ross *et al.*, 1993; Nickerson and Achberger, 1995; Prosseda *et al.*, 2010). In *Yersinia*, the virulence regulator *virF* contains intrinsic bends within its promoter, which 'melt' at increased temperature to activate its expression (Rohde JR *et al.*, 1999). This mechanism provides a rheostatic, or analog, level of control of gene expression

atop the digital nature of the DNA code (Travers and Muskhelishvili, 2005; Travers *et al.*, 2012).

Interactions with DNA-binding proteins have been demonstrated to modulate gene expression. DNA-binding proteins possess the ability to manipulate supercoiling or DNA bending in response to temperature. Most DNA-binding proteins are involved in chromosomal compaction, facilitating the genome's tight packaging. However, several nucleoid-associated proteins (NAPs) have been identified that interact with promoter regions, inducing and enhancing pre-existing bends in DNA. The temperature-induced changes in DNA topology mentioned above are further modified by interactions between DNA-binding proteins and their target promoters. This is exemplified in the regulation of the virulence regulator *virF* in *Shigella.* The promoter of *virF* contains binding sites for two NAPs— H-NS (two binding sites) and FIS (four binding sites). H-NS preferentially binds to bent DNA regions with AT-rich sequences. At ambient temperature, the DNA curvature is more pronounced, and H-NS binds to the binding sites present (Falconi *et al.*, 1998). H-NS exists as a dimer and forms a bridge within the bent region of the promoter, further enhancing the intrinsic curvature and repressing activation of the *virF* promoter (Dame *et al.*, 2001; Arold *et al.*, 2010). However at temperatures above 32°C, inhibition is relieved as a conformation change in H-NS occurs rendering it unable to bind to DNA (Maurelli and Sansonetti, 1988). This is coupled with a relaxation of the DNA curvature, allowing binding of FIS which counteracts H-NS, activating expression of *virF* (Falconi *et al.*, 2001; Prosseda *et al.*, 2004).

Activation of *virF* then triggers a signaling cascade inducing expression of several virulence associated genes (Tobe *et al.*, 1993).

In pathogenic *Yersinia*, temperature-induced changes in both DNA supercoiling and curvature acts together with NAPs to module expression of virulence genes. To activate virulence-associated genes encoded on the *Yersinia* virulence plasmid (pYV), *Yersinia* sense temperature-induced changes to DNA topology (Rohde JR *et al.*, 1994; Rohde JR *et al.*, 1999). This is mediated by using this mechanism to activate the virulence-inducing transcription activator VirF (LcrF in *Y. pestis*). VirF/LcrF is only expressed at 37°C. At ≤28°C, the *virF*/*lcrF* promoter contains regions of intrinsic curvature which repress expression. Furthermore, changes in NAP binding seem to act in concert with DNA topology at the *virF*/*lcrF* promoter, further facilitating temperature-controlled expression. *virF*/*lcrF* activation can occur at lower temperature in the absence of the NAP YmoA (Cornelis *et al.*, 1991; Böhme *et al.*, 2012). YmoA represses *virF*/*lcrF* by influencing DNA supercoiling and forming heterodimers with H-NS (Madrid *et al.*, 2002; Nieto *et al.*, 2002). Taken together, temperature-inducible changes to the *Yersinia* chromosome serves as an important ‘thermostat’, with changes modulating virulence gene expression upon encountering environments with elevated temperature.

Thermosensing via RNA

RNA thermosensors respond to temperature in a rapid manner. The immediacy of the response is a result of the presence of the RNA thermosensor,

having already been transcribed. Single-stranded mRNAs base pair intramolecularly to form secondary stem-loop structures immediately upon transcription. These RNA thermometers, as they are called, are a common post-transcriptional regulatory mechanism employed by bacteria (Kortmann and Narberhaus, 2012). Bacteria use these physical alterations to control differential virulence gene expression in response to temperature. Temperature-induced conformational changes in the RNA molecule often modulate exposure of binding elements. The vast majority of RNA thermosensors are located within the 5'-untranslated region (UTR) of mRNA and function by either revealing or concealing ribosome binding sites (RBS). At low temperature, thermolabile mRNA often exist in a closed conformation, concealing the RBS. At high temperature, there is a rapid and immediate disruption in the stem-loop causing conformational change revealing the RBS or other binding elements (Narberhaus *et al.*, 2006; Righetti *et al.*, 2016). Therefore, a temperature upshift encountered upon entering a warm-blooded host organism, immediately triggers translation of virulence-associated proteins.

Pathogenic *Yersinia* employ such a strategy in the regulation of the VirF/LcrF, mentioned above (Hoe and Goguen, 1993). The promoter for *virF*/*lcrF* transcribes a polycistronic transcript consisting of *yscW*, located upstream of *virF*/*lcrF*. The RBS for *virF*/*lcrF* is located within the intergenic region of the transcript. The expression of VirF/LcrF is contingent on exposure of the RBS, which is concealed by a hairpin RNA thermometer concealing the site. This RNA thermometer is composed of four uracil nucleotides which pair with four AGGA residues within the RBS, forming a stem-loop conformation blocking translation at low temperature. This stem-loop

opens after an upshift to 37°C revealing the RBS and allowing translation of the virulence activator (Böhme *et al.*, 2012).

Thermosensing via Proteins

Invasion into epithelial cells is also thermally regulated in *Yersinia*. Invasin is responsible for adherence to host epithelial cells. At low temperatures, invasin (*inv*) is maximally expressed, and at high temperature, expression is minimal (Pepe *et al.*, 1994). Invasin is activated by RovA, which binds to the promoter at only low temperature (Lawrenz and Miller VL, 2007). At high temperature, RovA undergoes a conformational change rendering it incapable of binding to the promoter of *inv*. This conformation change in RovA also makes it susceptible to degradation by Lon protease, leading to decreased concentration within the cell (Herbst *et al.*, 2009; Quade *et al.*, 2012). Augmenting this post-translational regulation, transcription of *inv* is repressed by the NAPs YmoA and H-NS acting in concert which out-compete RovA binding-mediated activation at high temperature (Ellison *et al.*, 2003; Ellison and Miller VL, 2006). Furthermore, RovA is positively autoregulated, i.e. activating its own expression (Heroven *et al.*, 2004). RovA binding to the *rovA* promoter is also repressed by H-NS and/or YmoA complexed with the LysR-type transcriptional regulator RovM (Heroven *et al.*, 2004).

Multiple Overlapping Systems

Taken together, regulation of *virF*/*lcrF* and *inv* uses multiple methods of temperature-inducible regulation. By using several overlapping methods of regulation, *Yersinia* can fine-tune expression of these important virulence factors in response to temperature fluctuations. Interestingly, additional regulatory factors have been described which further balance expression in response to other environmental cues. VirF/LcrF levels are furthered adjusted by additional regulatory factors: activated by the two-component regulators CpxR and RscB (Liu *et al.*, 2012; Li *et al.*, 2015) and the IscR iron-sulfur cluster regulator (Miller HK *et al.*, 2014), and repressed by the LysR-like transcriptional regulator YtxR (Axler-DiPerte *et al.*, 2009). By using multiple overlapping strategies, pathogenic organisms are able to strictly adjust virulence gene expression upon encountering a warm-blooded host.

Bacterial Epigenetic Regulation

Epigenetic modifications to the genome provide another method of response to environmental conditions. The term ‘epigenetic’ is derived from the Greek ‘epi’ meaning ‘above', ‘upon', or 'in addition’ and ‘genetic’ being 'of or relating to genes or heredity’. Biologists currently use the term to describe heritable changes in gene expression that do not involved modifications to the DNA sequence. In general, this would include any additional information outside the DNA digital sequence. In bacterial systems, epigenesis specifically refers to the role of DNA methylation in

gene activity. DNA methylation, whereby chemical methyl (-CH3) groups are covalently bound to nucleotides, provides a signal that regulates DNA-protein interactions, fine-tuning the binding affinity and creating alternative recognition states of protein binding sites.

Background

Historically, DNA methylation is known to protect bacteria against foreign DNA encountered by horizontal gene transfer such as conjugation, transformation, or viral infection. Host DNA is site-specifically methylated by a methyltransferase (Arber and Linn, 1969; Wilson, 1991). Foreign DNA (unmethylated) is cleaved by a cognate restriction enzyme endonuclease that recognizes the same site. This was first described by Arber and Dussoix following observations that *E. coli* would 'restrict' acceptance of phage P1 (Arber and Dussoix, 1962; Dussoix and Arber, 1962). Arber further elucidated that the 'modification' event required methionine, a known methyl group donor (Arber, 1965b). If phage grown in a host species was transduced into the same species, the viral genome would be insensitive to restriction (Arber, 1965a).

These systems come in three main groups (Roberts RJ *et al.*, 2003). Types I and III enzymes function as a single polypeptide with both restriction and modification functionality, while Type II systems have two separate polypeptides. Type II and III restriction enzymes cleave at or near the recognition sequence, and at variable positions away from their target with Type I restriction enzymes. Bacteria

express each system at levels appropriate to fully modify the genome so as to prevent potentially catastrophic lethal cleavage of its own genome. In addition to these R-M systems, there are two DNA adenine methylases which lack a cognate restriction enzyme. The orphan methylases include DNA adenine methylase (Dam) and DNA cytosine methylase (Dcm), and produce the modified bases N6-methyladenine (6mA), 5-methylcytosine (5mC), and N4-methylcytosine (4mC) respectively (Palmer and Marinus, 1994). Dam is present in gammaproteobacteria, and methylates the adenosine moiety of 5'-GATC-3' sites (Barras and Marinus, 1989). The enzyme is highly processive in that it slides along the DNA molecule following each modification so as to cover the entire genome (Urig *et al.*, 2002). The presence of 6mA at Dam sites alters DNA topology by lowering the thermodynamic stability, thereby increasing DNA curvature (Diekmann, 1987; Wion and Casadesús, 2006). This structural modification influences DNA-protein interactions and DNA-binding proteins can discriminate methylation states.

DNA Adenine Methylation-Dependent Gene Regulation

The role DNA methylation plays in governing transcription was first described in *E. coli*, where work in Nancy Kleckner's lab showed transposase Tn10 was regulated by DNA methylation (Roberts D *et al.*, 1985). Tn10 has two Dam sites, one at the promoter and one near its 3'-terminus. IS10 transposition is activated when the Dam site in the promoter is hemimethylated and target insertion at the termini proceeds only from a hemimethylated Dam site. Due to the activity and processivity

of Dam, most GATC sites are fully methylated. However during replication, as the replication fork proceeds, fully methylated sites become hemimethylated, i.e. methylated on one strand. This transient wake of hemimethylation behind a replication fork provides a short window of hemimethylation where Tn10 may be activated (Campbell and Kleckner, 1988).

Bacteria also use DNA methylation to regulate several biochemical processes, such as chromosomal replication, mismatch repair, and transcription initiation (Casadesús and Low, 2006; Marinus and Casadesús, 2009). The hemimethylated state of Dam sites following replication is leveraged in directing mismatch repair. In methyl-directed mismatch repair, MutH follows DNA polymerase and is responsible for differentiating the methylated (parent) strand from the unmethylated (daughter) strand. Gene regulation using these alternative binding states are best described in *Caulobacter crescentus* were they govern regulation of the cell cycle.

Caulobacter uses the Dam homolog CcrM to guide progress through the cell cycle as the chromosome is replicated (Mohapatra *et al.*, 2014). CcrM recognizes and methylates the adenosine moiety of 5'-GANTC-3' (Reisenauer *et al.*, 1999; Wion and Casadesús, 2006). Chromosomal replication is initiated at the origin (*Cori*) when DnaA is expressed and binds to *Cori* (Marczynski and Shapiro, 2002). DnaA is preferentially expressed when GANTC sites in its promoter are fully methylated, thus requiring expression of CcrM (Collier *et al.*, 2007). DnaA also induces expression of the next cell cycle regulator GcrA. As the replication fork proceeds, fully methylated sites become hemimethylated (Reisenauer *et al.*, 1999). DnaA is not

expressed from a hemimethylated promoter, and levels of this activator drop. GcrA induces the expression of the next cell cycle regulator, CtrA, along with several genes involved in DNA metabolism (Holtzendorff *et al.*, 2004). GcrA preferentially activates expression of *ctrA* from a hemimethylated promoter, occurring only after the replication fork passes (Reisenauer and Shapiro, 2002). CtrA activates expression of CcrM at the end of chromosomal replication (Laub *et al.*, 2002). CcrM then remethylates all hemimethylated sites back to fully methylated including Cori and the promoter of DnaA signally initiation of the next round of replication. This temporal modulation of binding affinity by DNA methylation serves to synchronize the transcription of genes with the cell cycle.

Bacterial Virulence

Pathogenic bacteria use DNA methylation to regulate motility, phase variation, and expression of virulence factors (Heusipp *et al.*, 2007). The best described example of this is in uropathogenic *E. coli*, where expression of the adhesion-mediating Pap pili is subject to DNA methylation-controlled phase variation when the Pap pili is on or off (Hernday *et al.*, 2002). The promoter of *pap* contains two Dam sites that are differentially methylated (Blyn *et al.*, 1990). In phase OFF, the GATC proximal to the gene is unmethylated and the distal GATC is methylated. The phase switches ON when the proximal GATC is methylated and the distal GATC is unmethylated. The methylation patterns of these two sites are influenced by binding of two regulatory proteins, Lrp and PapI, whose binding sites

within the *pap* promoter overlap the GATC sites and block Dam methylation. Furthermore, the phase variation of Pap pili is thermally regulated, OFF at 26°C and ON at 37°C. This is controlled by the nucleoid-associated protein H-NS, also mentioned above for its role in temperature regulation. At low temperature, H-NS blocks methylation of both GATC sites (White-Ziegler *et al.*, 1998). Taken together, this is another example of multiple overlapping systems regulating virulence gene expression.

In *Yersinia*, the effect of DNA methylation has been observed but has not been fully characterized. Dam over-expression interferes with the regulation of several thermally-regulated phenotypes, suggesting a role for DNA methylation in the regulation of these systems. Dam over-expressing strains have altered motility, surface antigens, and biofilm production (Fälker *et al.*, 2007). Importantly, Dam over-expressing strains of *Y. pseudotuberculosis* and *Y. pestis* attenuate virulence. Experimentally, such over-expressing strains are effective vaccines (Julio *et al.*, 2001; Robinson *et al.*, 2005; Kubicek-Sutherland *et al.*, 2014). Conversely, a Dam over-expressing strain of *Y. enterocolitica* increases invasion into epithelial cells (Fälker *et al.*, 2005). This is similar to observations in *Salmonella typhimurium*, where DNA methylation over-expression attenuates virulence while invasion ability increases and strains with modified Dam activity can serve as vaccination strains (Heithoff *et al.*, 2001).

Hypothesis

Following the observation that modulating Dam expression overrides temperature signals in *Yersinia*, and given that a cell cycle-linked 2.5 generations of growth is necessary to restore motility after temperature downshifts, we hypothesize that differential DNA methylation plays a role in temperature regulation. This is further supported by our observation of differential patterns of methylation-sensitive restriction enzyme activity in DNA isolated from different temperatures (unpublished results). In the next chapter, we describe the differential temperature-responsive methylome of *Y. enterocolitica* where we identify several potential regulators responsible for temperature-regulated phenotypes. This is similar to several recently published studies which describe the methylome of enteric pathogens, also identifying regions of DNA methylation with regulatory functions (Fang *et al.*, 2012; Bendall *et al.*, 2013; Mou *et al.*, 2014; Pirone-Davies *et al.*, 2015; Atack *et al.*, 2015; Cota *et al.*, 2016). Further defining the role of DNA methylation on host-parasite interactions will add to this productive model and provide new insights to infection interaction.

References

Achtman, M., Morelli, G., Zhu, P.X., Wirth, T., Diehl, I., Kusecek, B., *et al.* (2004) Microevolution and history of the plague bacillus, Yersinia pestis. *Proc Natl Acad Sci USA* **101**: 17837–17842.

Achtman, M., Zurth, K., Morelli, C., Torrea, G., Guiyoule, A., and Carniel, E. (1999) *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci USA* **96**: 14043–14048.

Arber, W. (1965a) Host-controlled modification of bacteriophage. *Annual Reviews in Microbiology*.

Arber, W. (1965b) Host specificity of DNA produced by *Escherichia coli*: V. The role of methionine in the production of host specificity. *Journal of Molecular Biology* **11**: 247–256.

Arber, W., and Dussoix, D. (1962) Host specificity of DNA produced by *Escherichia coli*: I. Host controlled modification of bacteriophage λ. *Journal of Molecular Biology* **5**: 18–36.

Arber, W., and Linn, S. (1969) DNA Modification and Restriction. *Annu Rev Biochem* **38**: 467–500.

Arold, S.T., Leonard, P.G., Parkinson, G.N., and Ladbury, J.E. (2010) H-NS forms a superhelical protein scaffold for DNA condensation. *Proc Natl Acad Sci USA* **107**: 15728–15732.

Atack, J.M., Srikhanta, Y.N., Fox, K.L., Jurcisek, J.A., Brockman, K.L., Clark, T.A., *et al.* (2015) A biphasic epigenetic switch controls immunoevasion, virulence and niche adaptation in non-typeable *Haemophilus influenzae*. *Nature Communications* **6**: 1–12.

Axler-DiPerte, G.L., Hinchliffe, S.J., Wren, B.W., and Darwin, A.J. (2009) YtxR Acts as an Overriding Transcriptional Off Switch for the *Yersinia enterocolitica* Ysc-Yop Type 3 Secretion System. *J Bacteriol* **191**: 514–524.

Barras, F., Marinus, M.G. (1989) The Great GATC - DNA Methylation in *Escherichia coli*. *Trends in Genetics* **5**:139–143.

Ben-Gurion, R., and Shafferman, A. (1981) Essential virulence determinants of different *Yersinia* species are carried on a common plasmid. *Plasmid* **5**: 183–187.

Bendall, M.L., Luong, K., Wetmore, K.M., Blow, M., Korlach, J., Deutschbauer, A., and Malmstrom, R.R. (2013) Exploring the Roles of DNA Methylation in the Metal-Reducing Bacterium *Shewanella oneidensis* MR-1. *J Bacteriol* **195**: 4966–4974.

Blyn, L.B., Braaten, B.A., and Low, D.A. (1990) Regulation of *pap* pilin phase variation by a mechanism involving differential *Dam* methylation states. *EMBO J* **9**: 4045–4054.

Bottone, E.J. (1999) *Yersinia enterocolitica*: overview and epidemiologic correlates. *Microbes Infect* **1**: 323–333.

Böhme, K., Steinmann, R., Kortmann, J., Seekircher, S., Heroven, A.K., Berger, E., *et al.* (2012) Concerted actions of a thermo-labile regulator and a unique intergenic RNA thermosensor control *Yersinia* virulence. *PLoS Pathog* **8**: e1002518–23.

Campbell, J.L., and Kleckner, N. (1988) The rate of Dam-mediated DNA adenine methylation in *Escherichia coli*. *Gene* **74**: 189–190.

Campbell, J.L., and Kleckner, N. (1990) *E. coli oriC* and the *dnaA* gene promoter are sequestered from *dam* methyltransferase following the passage of the chromosomal replication fork. *Cell* **62**: 967–979.

Carniel, E. (2001) The *Yersinia* high-pathogenicity island: an iron-uptake island. *Microbes Infect* **3**: 561–569.

Carniel, E., and Mollaret, H.H. (1990) Yersiniosis. *Comp Immunol Microbiol Infect Dis* **13**: 51–58.

Casadesús, J., and Low, D.A. (2006) Epigenetic Gene Regulation in the Bacterial World. *Microbiol Mol Biol Rev* **70**: 830–856.

Chen, S., Thompson, K.M., and Francis, M.S. (2016) Environmental Regulation of *Yersinia* Pathophysiology. *Front Cell Infect Microbiol* **6**: 1424–27.

Collier, J., McAdams, H.H., and Shapiro, L. (2007) A DNA methylation ratchet governs progression through a bacterial cell cycle. *Proc Natl Acad Sci USA* **104**: 17111–17116.

Cornelis, G.R., Biot, T., de Rourvroit, C.L., Michiels, T., Mulder, B., Sluiters, C., *et al.* (1989) The Yersinia-Yop Regulon. *Molecular Microbiology* **3**: 1455–1459.

Cornelis, G.R., Boland, A., Boyd, A.P., Geuijen, C., Iriarte, M., Neyt, C., *et al.* (1998) The virulence plasmid of *Yersinia*, an antihost genome. *Microbiol Mol Biol Rev* **62**: 1315–1352.

Cornelis, G.R., Laroche, Y., Balligand, G., Sory, M.P., and Wauters, G. (1987) *Yersinia enterocolitica*, a Primary Model for Bacterial Invasiveness. *Rev Infect Dis* **9**: 64–87.

Cornelis, G.R., Sluiters, C., Delor, I., Geib, D., Kaniga, K., de Rourvroit, C.L., *et al.* (1991) *ymoA*, a *Yersinia enterocolitica* Chromosomal Gene Modulating the Expression of Virulence Functions. *Molecular Microbiology* **5**: 1023–1034.

Correnti, J., Munster, V., Chan, T., and van der Woude, M. (2002) Dam-dependent phase variation of Ag43 in *Escherichia coli* is altered in a *seqA* mutant. *Molecular Microbiology* **44**: 521–532.

Cota, I., Bunk, B., Spröer, C., Overmann, J., König, C., and Casadesús, J. (2016) OxyR-dependent formation of DNA methylation patterns in $OpvAB^{OFF}$ and $OpvAB^{ON}$ cell lineages of *Salmonella enterica*. *Nucleic Acids Res* **44**: 3595–3609.

Dame, R.T., Wyman, C., and Goosen, N. (2001) Structural basis for preferential binding of H-NS to curved DNA. *Biochimie* **83**: 231–234.

Diekmann, S. (1987) DNA Methylation Can Enhance or Induce DNA Curvature. *EMBO J* **6**: 4213–4217.

Dorman, C.J. (1991) DNA supercoiling and environmental regulation of gene expression in pathogenic bacteria. *Infection and Immunity* **59**: 745–749.

Dussoix, D., and Arber, W. (1962) Host specificity of DNA produced by *Escherichia coli*: II. Control over acceptance of DNA from infecting phage λ. *Journal of Molecular Biology* **5**: 37–49.

Ellison, D.W., and Miller, V.L. (2006) H-NS Represses *inv* Transcription in *Yersinia enterocolitica* through Competition with RovA and Interaction with YmoA. *J Bacteriol* **188**: 5101–5112.

Ellison, D.W., Young, B., Nelson, K., and Miller, V.L. (2003) YmoA Negatively Regulates Expression of Invasin from *Yersinia enterocolitica*. *J Bacteriol* **185**: 7153–7159.

Falconi, M., Colonna, B., Prosseda, G., Micheli, G., and Gualerzi, C.O. (1998) Thermoregulation of *Shigella* and *Escherichia coli* EIEC pathogenicity. A temperature-dependent structural transition of DNA modulates accessibility of *virF* promoter to transcriptional repressor H-NS. *EMBO J* **17**: 7033–7043.

Falconi, M., Prosseda, G., Giangrossi, M., Beghetto, E., and Colonna, B. (2001) Involvement of FIS in the H-NS-mediated regulation of *virF* gene of Shigella and enteroinvasive *Escherichia coli*. *Molecular Microbiology* **42**: 439–452.

Fang, G., Munera, D., Friedman, D.I., Mandlik, A., Chao, M.C., Banerjee, O., *et al.* (2012) Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nature Biotechnology* **30**: 1–11.

Fälker, S., Schilling, J., Schmidt, M.A., and Heusipp, G. (2007) Overproduction of DNA Adenine Methyltransferase Alters Motility, Invasion, and the Lipopolysaccharide O-Antigen Composition of *Yersinia enterocolitica*. *Infection and Immunity* **75**: 4990–4997.

Fälker, S., Schmidt, M.A., and Heusipp, G. (2005) DNA methylation in *Yersinia enterocolitica*: role of the DNA adenine methyltransferase in mismatch repair and regulation of virulence factors. *Microbiology* **151**: 2291–2299.

Han, Y., Qiu, J., Guo, Z., Gao, H., Song, Y., Zhou, D., and Yang, R. (2007) Comparative transcriptomics in *Yersinia pestis*: a global view of environmental modulation of gene expression. *BMC Microbiol* **7**: 96.

Heithoff, D.M., Enioutina, E.Y., Daynes, R.A., Sinsheimer, R.L., Low, D.A., and Mahan, M.J. (2001) *Salmonella* DNA adenine methylase mutants confer cross-protective immunity. *Infection and Immunity* **69**: 6725–6730.

Herbst, K., Bujara, M., Heroven, A.K., Opitz, W., Weichert, M., Zimmermann, A., and Dersch, P. (2009) Intrinsic Thermal Sensing Controls Proteolysis of *Yersinia* Virulence Regulator RovA. *PLoS Pathog* **5**: e1000435–16.

Hernday, A., Krabbe, M., Braaten, B., and Low, D.A. (2002) Self-perpetuating epigenetic pili switches in bacteria. *Proc Natl Acad Sci USA* **99**: 16470–16476.

Heroven, A.K., Nagel, G., Tran, H.J., Parr, S., and Dersch, P. (2004) RovA is autoregulated and antagonizes H-NS-mediated silencing of invasin and *rovA* expression in *Yersinia pseudotuberculosis*. *Molecular Microbiology* **53**: 871–888.

Heusipp, G., Fälker, S., and Alexander Schmidt, M. (2007) DNA adenine methylation and bacterial pathogenesis. *International Journal of Medical Microbiology* **297**: 1–7.

Hoe, N.P., and Goguen, J.D. (1993) Temperature Sensing in *Yersinia pestis*: Translation of the LcrF Activator Protein Is Thermally Regulated. *J Bacteriol* **175**: 7901–7909.

Holtzendorff, J., Hung, D., Brende, P., Reisenauer, A., Viollier, P.H., McAdams, H.H., and Shapiro, L. (2004) Oscillating global regulators control the genetic circuit driving a bacterial cell cycle. *Science* **304**: 983–987.

Huovinen, E., Sihvonen, L.M., Virtanen, M.J., Haukka, K., Siitonen, A., and Kuusi, M. (2010) Symptoms and sources of *Yersinia enterocolitica*-infection: a case-control study. *BMC Infect Dis* **10**: 122.

Hurme, R., and Rhen, M. (1998) Temperature sensing in bacterial gene regulation — what it all boils down to. *Molecular Microbiology* **30**: 1–6.

Isberg, R.R., and Leong, J.M. (1990) Multiple β_1 chain integrins are receptors for invasin, a protein that promotes bacterial penetration into mammalian cells. *Cell* **60**: 861–871.

Julio, S.M., Heithoff, D.M., Provenzano, D., Klose, K.E., Sinsheimer, R.L., Low, D.A., and Mahan, M.J. (2001) DNA Adenine Methylase Is Essential for Viability and Plays a Role in the Pathogenesis of *Yersinia pseudotuberculosis* and *Vibrio cholerae*. *Infection and Immunity* **69**: 7610–7615.

Kapatral, V., and Minnich, S.A. (1995) Co-ordinate, temperature-sensitive regulation of the three *Yersinia enterocolitica* flagellin genes. *Molecular Microbiology* **17**: 49–56.

Kapatral, V., Olson, J.W., Pepe, J.C., Miller, V.L., and Minnich, S.A. (1996) Temperature-dependent regulation of *Yersinia enterocolitica* class III flagellar genes. *Molecular Microbiology* **19**: 1061–1071.

Koczura, R., and Kaznowski, A. (2003a) The *Yersinia* high-pathogenicity island and iron-uptake systems in clinical isolates of *Escherichia coli*. *Journal of Medical Microbiology* **52**: 637–642.

Koczura, R., and Kaznowski, A. (2003b) Occurrence of the *Yersinia* high-pathogenicity island and iron uptake systems in clinical isolates of Klebsiella pneumoniae. *Microb Pathog* **35**: 197–202.

Kortmann, J., and Narberhaus, F. (2012) Bacterial RNA thermometers: molecular zippers and switches. *Nature Methods* **10**: 255–265.

Kubicek-Sutherland, J.Z., Heithoff, D.M., Ersoy, S.C., Shimp, W.R., and Mahan, M.J. (2014) Immunization with a DNA adenine methylase over-producing *Yersinia pseudotuberculosis* vaccine confers robust cross-protection against heterologous pathogenic serotypes. *Vaccine* **32**: 1451–1459.

Laub, M.T., Chen, S.L., Shapiro, L., and McAdams, H.H. (2002) Genes directly controlled by CtrA, a master regulator of the *Caulobacter* cell cycle. *Proc Natl Acad Sci USA* **99**: 4632–4637.

Lawrenz, M.B., and Miller, V.L. (2007) Comparative Analysis of the Regulation of *rovA* from the Pathogenic *Yersiniae*. *J Bacteriol* **189**: 5963–5975.

Li, Y., Hu, Y., Francis, M.S., and Chen, S. (2015) RcsB positively regulates the *Yersinia* Ysc-Yop type III secretion system by activating expression of the master transcriptional regulator LcrF. *Environ Microbiol* **17**: 1219–1233.

Liu, J., Thanikkal, E.J., Obi, I.R., and Francis, M.S. (2012) Elevated CpxR~P levels repress the Ysc–Yop type III secretion system of *Yersinia pseudotuberculosis*. *Research in Microbiology* **163**: 518–530.

Lu, M., Campbell, J.L., Boye, E., and Kleckner, N. (1994) SeqA: a negative modulator of replication initiation in *E. coli*. *Cell* **77**: 413–426.

Løbner-Olesen, A., Marinus, M.G., and Hansen, F.G. (2003) Role of SeqA and Dam in *Escherichia coli* gene expression: a global/microarray analysis. *Proc Natl Acad Sci USA* **100**: 4672–4677.

Madrid, C., Nieto, J.M., and Juarez, A. (2002) Role of the Hha/YmoA family of proteins in the thermoregulation of the expression of virulence factors. *International Journal of Medical Microbiology* **291**: 425–432.

Marczynski, G.T., and Shapiro, L. (2002) Control of chromosome replication in *Caulobacter crescentus*. *Annu Rev Microbiol* **56**: 625–656.

Marinus, M.G., and Casadesús, J. (2009) Roles of DNA adenine methylation in host–pathogen interactions: mismatch repair, transcriptional regulation, and more. *FEMS Microbiol Rev* **33**: 488–503.

Maurelli, A.T., and Sansonetti, P.J. (1988) Identification of a chromosomal gene controlling temperature-regulated expression of *Shigella* virulence. *Proc Natl Acad Sci USA* **85**: 2820–2824.

McNally, A., Thomson, N.R., Reuter, S., and Wren, B.W. (2016) “Add, stir and reduce”: *Yersinia* spp. as model bacteria for pathogen evolution. *Nature Methods* **14**: 177–190.

Miller, H.K., Kwuan, L., Schwiesow, L., Bernick, D.L., Mettert, E., Ramirez, H.A., *et al.* (2014) IscR Is Essential for *Yersinia pseudotuberculosis* Type III Secretion and Virulence. *PLoS Pathog* **10**: e1004194–21.

Minnich, S.A., and Rohde, H.N. (2007) A rationale for repression and/or loss of motility by pathogenic *Yersinia* in the mammalian host. *Adv Exp Med Biol* **603**: 298–311.

Mohapatra, S.S., Fioravanti, A., and Biondi, E.G. (2014) DNA methylation in Caulobacter and other Alphaproteobacteria during cell cycle progression. *Trends in Microbiology* **22**: 528–535.

Mou, K.T., Muppirala, U.K., Severin, A.J., Clark, T.A., Boitano, M., and Plummer, P.J. (2014) A comparative analysis of methylome profiles of *Campylobacter jejuni* sheep abortion isolate and gastroenteric strains using PacBio data. *Front Microbiol* **5**: 782.

Narberhaus, F., Waldminghaus, T., and Chowdhury, S. (2006) RNA thermometers. *FEMS Microbiol Rev* **30**: 3–16.

Nickerson, C.A., and Achberger, E.C. (1995) Role of Curved DNA in Binding of *Escherichia coli* RNA Polymerase to Promoters. *J Bacteriol* **177**: 5756–5761.

Nieto, J.M., Madrid, C., Miquelay, E., Parra, J.L., Rodriguez, S., and Juarez, A. (2002) Evidence for Direct Protein-Protein Interaction between Members of the Enterobacterial Hha/YmoA and H-NS Families of Proteins. *J Bacteriol* **184**: 629–635.

Palmer, B.R., and Marinus, M.G. (1994) The Dam and Dcm Strains of *Escherichia coli* - a Review. *Gene* **143**: 1–12.

Pepe, J.C., Badger, J.L., and Miller, V.L. (1994) Growth phase and low pH affect the thermal regulation of the *Yersinia enterocolitica inv* gene. *Molecular Microbiology* **11**: 123–135.

Perry, R.D., and Fetherston, J.D. (1997) *Yersinia pestis*--etiologic agent of plague. *Clinical microbiology reviews*.

Pirone-Davies, C., Hoffmann, M., Roberts, R.J., Muruvanda, T., Timme, R.E., Strain, E., *et al.* (2015) Genome-Wide Methylation Patterns in *Salmonella enterica* Subsp. *enterica* Serovars. *PLoS ONE* **10**: e0123639–13.

Prosseda, G., Falconi, M., Giangrossi, M., Gualerzi, C.O., Micheli, G., and Colonna, B. (2004) The virF promoter in Shigella: more than just a curved DNA stretch. *Molecular Microbiology* **51**: 523–537.

Prosseda, G., Mazzola, A., Di Martino, M.L., Tielker, D., Micheli, G., and Colonna, B. (2010) A Temperature-Induced Narrow DNA Curvature Range Sustains the Maximum Activity of a Bacterial Promoter in Vitro. *Biochemistry* **49**: 2778–2785.

Quade, N., Mendonca, C., Herbst, K., Heroven, A.K., Ritter, C., Heinz, D.W., and Dersch, P. (2012) Structural Basis for Intrinsic Thermosensing by the Master Virulence Regulator RovA of *Yersinia*. *Journal of Biological Chemistry* **287**: 35796–35803.

Reisenauer, A., and Shapiro, L. (2002) DNA methylation affects the cell cycle transcription of the CtrA global regulator in *Caulobacter*. *EMBO J* **21**: 4969–4977.

Reisenauer, A., Kahng, L.S., McCollum, S., and Shapiro, L. (1999) Bacterial DNA methylation: a cell cycle regulator? *J Bacteriol* **181**: 5135–5139.

Reuter, S., Connor, T.R., Barquist, L., Walker, D., Feltwell, T., Harris, S.R., *et al.* (2014) Parallel independent evolution of pathogenicity within the genus *Yersinia*. *Proc Natl Acad Sci USA* **111**: 6768–6773.

Righetti, F., Nuss, A.M., Twittenhoff, C., Beele, S., Urban, K., Will, S., *et al.* (2016) Temperature-responsive in vitro RNA structurome of *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci USA* 201523004.

Roberts, D., Hoopes, B.C., McClure, W.R., and Kleckner, N. (1985) IS10 transposition is regulated by DNA adenine methylation. *Cell* **43**: 117–130.

Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A.S., Bickle, T.A., Bitinaite, J., *et al.* (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res* **31**: 1805–1812.

Robinson, V.L., Oyston, P.C.F., and Titball, R.W. (2005) A *dam* mutant of *Yersinia pestis* is attenuated and induces protection against plague. *FEMS Microbiology Letters* **252**: 251–256.

Rohde, J.R., Fox, J.M., and Minnich, S.A. (1994) Thermoregulation in *Yersinia enterocolitica* is coincident with changes in DNA supercoiling. *Molecular Microbiology* **12**: 187–199.

Rohde, J.R., Luan, X.S., Rohde, H.N., Fox, J.M., and Minnich, S.A. (1999) The *Yersinia enterocolitica* pYV virulence plasmid contains multiple intrinsic DNA bends which melt at 37°C. *J Bacteriol* **181**: 4198–4204.

Ross, W., Gosink, K.K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., *et al.* (1993) A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* **262**: 1407–1413.

Steinmann, R., and Dersch, P. (2013) Thermosensing to adjust bacterial virulence in a fluctuating environment. *Future Microbiology* **8**: 85–105.

Straley, S.C., and Perry, R.D. (1995) Environmental modulation of gene expression and pathogenesis in *Yersinia*. *Trends in Microbiology* **3**: 310–317.

Sulakvelidze, A. (2000) *Yersiniae* other than *Y. enterocolitica, Y. pseudotuberculosis,* and *Y. pestis*: the ignored species. *Microbes Infect* **2**: 497–513.

Tauxe, R.V., Vandepitte, J., WAUTERS, G., Martin, S.M., Goossens, V., De Mol, P., *et al.* (1987) *Yersinia enterocolitica* infections and pork: the missing link. *The Lancet* **1**: 1129–1132.

Tobe, T., Yoshikawa, M., Mizuno, T., and Sasakawa, C. (1993) Transcriptional Control of the Invasion Regulatory Gene *virB* of *Shigella flexneri* - Activation by VirF and Repression by H-NS. *J Bacteriol* **175**: 6142–6149.

Travers, A.A., and Muskhelishvili, G. (2005) DNA supercoiling — a global transcriptional regulator for enterobacterial growth? *Nature Reviews Microbiology* **3**: 157–169.

Travers, A.A., Muskhelishvili, G., and Thompson, J.M.T. (2012) DNA information: from digital code to analogue structure. *Phil Trans R Soc A* **370**: 2960–2986.

Urig, S., Gowher, H., Hermann, A., Beck, C., Fatemi, M., Humeny, A., and Jeltsch, A. (2002) The *Escherichia coli* Dam DNA Methyltransferase Modifies DNA in a Highly Processive Reaction. *Journal of Molecular Biology* **319**: 1085–1096.

What's in that pork? We found antibiotic-resistant bacteria--and traces of a veterinary drug. (2013) What's in that pork? We found antibiotic-resistant bacteria--and traces of a veterinary drug. *Consum Rep* **78**: 44–46.

White-Ziegler, C.A., Angus Hill, M.L., Braaten, B.A., van der Woude, M.W., and Low, D.A. (1998) Thermoregulation of *Escherichia coli* pap transcription: H-NS is a temperature-dependent DNA methylation blocking factor. *Molecular Microbiology* **28**: 1121–1137.

Wilson, G. (1991) Restriction and Modification Systems. *Annu Rev Genet* **25**: 585–627.

Wion, D., and Casadesús, J. (2006) N6-methyl-adenine: an epigenetic signal for DNA–protein interactions. *Nature Reviews Microbiology* **4**: 183–192.

CHAPTER 2

Comparison of *Yersinia enterocolitica* DNA methylation at ambient and host temperatures

Article submitted to *Microbial Genomics*

Abstract

Pathogenic bacteria recognize environmental cues to vary gene expression for host adaptation. Moving from ambient to host temperature, *Yersinia enterocolitica* responds by immediately repressing flagella synthesis and inducing the virulence plasmid (pYV)-encoded type III secretion system. In contrast, shifting from host to ambient temperature requires 2.5 generations to restore motility suggesting a link to the cell cycle. We hypothesized that differential DNA methylation contributes to temperature-regulated gene expression. We tested this hypothesis by comparing single-molecule real-time (SMRT) sequencing of *Y. enterocolitica* DNA from cells growing exponentially at 22°C and 37°C. The inter-pulse duration ratio rather than the traditional QV scoring was the kinetic metric to compare DNA from cells grown at each temperature. All 565 *Yen*I restriction sites were fully methylated at both temperatures. Among the 27,118 DNA adenine methylase (Dam) sites, 42 had differential methylation patterns while 17 remained

unmethylated regardless of temperature. A subset of the differentially methylated Dam sites localized to promoter regions of predicted regulatory genes including LysR-type and PadR-like transcriptional regulators, and a cyclic-di-GMP phosphodiesterase. The unmethylated Dam sites localized with a bias to the replication terminus, suggesting they were protected from Dam methylase. No cytosine methylation was detected at Dcm sites.

Author Summary

Organisms sense and respond to their environment, in part, by epigenetic variation mediated by DNA methylation. Pathogenic bacteria vary gene expression to allow survival and activate virulence systems in response to host temperature. *Yersinia enterocolitica*, a facultative intracellular pathogen, respond by immediately repressing flagella synthesis and inducing the virulence plasmid-encoded type III secretion system. In this work, we examined the locations of DNA methylation throughout the *Y. enterocolitica* genome. While most methylation target sites were fully methylated, we identified sites with disparate temperature-dependent methylation. Several of these sites were within promoter regions of predicted regulatory genes. Differences in DNA methylation in promoter sequences are often responsible for variations in transcription. Identification of these differences in methylation provide likely candidates for regulators responsible for temperature-dependent phenotypes.

Introduction

The adaptation of facultative bacterial pathogens to their mammalian host environment requires significant global changes in gene regulation to establish infection. Host cues utilized for this transition vary among pathogens, but for many bacteria, host temperature is a key environmental signal (Maurelli, 1989). Temperature sensing is especially prominent in the pathogenic *Yersinia*. *Y. enterocolitica*, a Gram-negative enteropathogen, shows significant phenotypic changes between the narrow range of 30°C and 37°C (host temperature). Changes include a temperature-dependent requirement for calcium ion (2.5 mM at 37°C) (Michiels and Cornelis, 1991; Straley *et al.*, 1993), modification of LPS acylation to circumvent Toll-like receptor (TLR) 4 stimulation (Rebeil *et al.*, 2004; Montminy *et al.*, 2006), metabolic differences such as urease and acetoin production (de Koning-Ward and Robins-Browne, 1997), and the reciprocal temperature-controlled regulation between two type-III secretion systems (TTSS). The latter includes the immediate repression of flagellum biosynthesis and concomitant induction of the pYV (virulence plasmid) TTSS at 37°C (Rohde JR *et al.*, 1994; Minnich and Rohde HN, 2007). The reciprocal temperature regulation of these two TTSS may be required because of the substrate reciprocity of their exported proteins (Minnich and Rohde HN, 2007). Co-expression of flagella and the virulence TTSS at 37°C would result in injection of flagellin into host cells by the pYV-encoded TTSS. Flagellins are potent cytokine inducers of TLRs 5 and Ipaf and their injection into host cells could effectively countermand the pYV-TTSS effectors, termed *Yersinia*

outer proteins (Yops). The Yops collectively act to suppress the host innate immune system. Conversely, the flagellar TTSS exports Yops into the host extracellular milieu, rather than direct injection into host cells, diluting their effect. Thus, the immediate cessation of flagellin expression and concomitant induction of the Yops may be essential during the initial stages of host infection.

In contrast to the *Y. enterocolitica* rapid response to host temperature, acclimation to 25°C after a temperature downshift (37°C to 25°C) is much slower. *Y. enterocolitica* adapted to 37°C and downshifted to 25°C requires ~2.5 generations before flagellins (*fleABC*) are expressed (Rohde JR *et al.*, 1994). The relative timing varies between four and 10 hrs, depending on growth rate (rich vs minimal medium), but the 2.5 generation requirement is consistent. This suggests restoration of a low temperature phenotype is cell-cycle dependent. A model of temperature-dependent differential DNA methylation could link temperature-regulated genes to the cell cycle if expression of a key regulatory gene was sensitive to DNA methylation. This is because two generations of DNA replication are required for a DNA site to go from a fully methylated to an unmethylated state.

Bacterial DNA methylation occurs at adenine and cytosine bases providing epigenetic information in the form of N6-methyladenine (N6mA), N4-methylcytosine (N4mC), and 5-methylcytosine (5mC). DNA methylation is a mechanism to discriminate self from non-self (restriction sensitive), direct mismatch repair, excise non-methylated, i.e. non-self, strands of DNA (Wilson, 1991), and activate transpositions (Roberts D *et al.*, 1985). Recent studies expand the role of bacterial DNA methylation to include regulation of cell cycle progression (Collier *et al.*, 2007;

Collier, 2009) and modulation of gene expression (Low *et al.*, 2001; Casadesús and Low, 2006; Srikhanta *et al.*, 2010). For example, DNA replication forks leave a wake of transient hemi-methylated sites on the newly synthesized DNA strand. Hemi-methylation of gene promoters can either activate or repress expression by promoting or inhibiting the binding of specific transcription factors. Transient hemi-methylation states associated with the passing of DNA replication forks, account for the link between *Caulobacter crescentus* developmental gene regulation and its cell cycle (Collier *et al.*, 2007) and the frequency of *E. coli Tn*10 transposition (Campbell and Kleckner, 1988). Thus, modulation of DNA methylation provides a nondestructive and reversible means of DNA modification. This biphasic epigenetic switch is one mechanism organisms use to sense and adapt to their environment (Atack *et al.*, 2015).

Temperature modulations of DNA supercoiling, histone-like protein DNA binding, and intrinsic DNA bends also contribute to *Y. enterocolitica* host adaptation. DNA supercoiling levels naturally respond to changes in temperature to control expression of flagella and virulence factors or can be artificially manipulated using gyrase inhibitors or novobiocin-resistant mutants (Rohde JR *et al.*, 1994). Intrinsic DNA bends associated with poly-A and T tracts are sensitive to temperature and effectively melt at 37°C. Changes in DNA structure affect binding of histone-like proteins, promoter function, and methylation (Dorman, 1991). Coupling temperature-regulation to DNA methylation is also evident in *Yersinia*. Overproduction of DNA adenine methylase (Dam) in *Y. pseudotuberculosis* overrides the low temperature repression of Yop expression, but not the low Ca^{2+}

requirement for Yop secretion (Julio *et al.*, 2001). Dam methylates the N6 position of adenine in 5'-GATC-3' sequences. In *Y. enterocolitica*, overproduction of Dam alters motility, invasion, and results in increased amounts of rough lipopolysaccharide (LPS) lacking O-antigen side chains (Fälker *et al.*, 2007). Together, these results suggest that Dam sites may show variation in DNA methylation in different environmental conditions including temperature.

Until recently, high-throughput analysis of genomic epigenetic markers was limited to analyzing 5mC via bisulfite conversion followed by sequencing, such as Sanger, pyrosequencing, or whole genome amplification (Tost and Gut, 2007; Zilberman and Henikoff, 2007). The limitation was significant since many studies indicate bacterial DNA methylation centers on N6mA as the modified base (Reisenauer *et al.*, 1999; Wion and Casadesús, 2006; Marinus and Casadesús, 2009). With the advent of single-molecule real-time (SMRT) sequencing, a method for sequencing the methylome of bacteria to identify site modifications of adenine exists (Fang *et al.*, 2012; Murray *et al.*, 2012; Davis BM *et al.*, 2013). SMRT sequencing allows analysis of the native prepared DNA without the requirement of a whole genome amplification, therefore revealing the DNA as the organism has modified it *in vivo* (Flusberg *et al.*, 2010).

In this study, we tested the hypothesis that the *Y. enterocolitica* genome has temperature-dependent differences in DNA methylation. We compared the methylome of *Y. enterocolitica* DNA isolated from cells grown at 22°C and 37°C, providing the first systematic analysis of the epigenetic modifications in this important pathogen. We identified Dam sites with disparate methylation patterns

dependent upon growth temperature, and characterized them in relation to promoter regions or coding sequences.

Materials and Methods

Bacterial strains, culture media, and DNA techniques.

Y. enterocolitica strain 8081v (R^-M^+) (Kinder *et al.*, 1993) was grown to late exponential phase in Luria-Bertani broth at either 22°C or 37°C. The presence of the virulence plasmid (pYV) was verified both by PCR and plating on Congo Red LB plates in calcium-chelating conditions (CR-MOX). Genomic DNA was isolated using Sigma (St. Louis) GenElute Bacterial Genomic DNA Kit according to the manufacturer's protocol.

SMRT sequencing.

SMRT sequencing analyzes the rate at which nucleotides are incorporated on the complimentary replicating strand and compares these kinetics to an *in silico* control. The resulting inter-pulse duration (IPD ratio) provides a kinetic profile that identifies modified bases (Schadt *et al.*, 2013). Libraries of *Y. enterocolitica* 8081v cultures grown at each temperature (22°C and 37°C) were prepared for SMRT sequencing via circular consensus sequencing using a library construction protocol described previously (Travers KJ *et al.*, 2010). Briefly, native chromosomal and plasmid DNA preparations were sheared to an average size of 500bp via adaptive

focus acoustics (Corvaris, Woburn, MA, USA), end-repaired, and A-tailed and hairpin adapters with a single T-overhang were ligated. Primers were annealed and sequenced on the Pacific Biosciences RS instrument using C2 chemistry. These libraries were sequenced to a mean coverage depth of 228 (22°C) and 178 (37°C) across the chromosome, and 149 and 68 respectively across the virulence plasmid; average read length was 2656 and 2682, respectively. Reads were mapped to the reference genome (RefSeq NC_008791 and NC_008800 for the chromosome and virulence plasmid, respectively) using BLASR (Chaisson and Tesler, 2012). Base modification and motif detection were performed using the Modification and Motif Detection protocol in the software program SMRTPipe v.1.3.3. Positions with coverage of >25 and Score (QV) of >40 were considered modified. The Score (QV) equals -10*log(p-value), where the p-value was determined from a t-test between the sample and the *in silico* control derived from whole genome amplified (WGA) (i.e. unmethylated) samples (http://pacb.com/applications/base_modification/index.html). Due to the lower read depth coverage on the virulence plasmid, sites of adenine methylation with coverage >10 and an IPD ratio above 3 were considered modified, and was further confirmed because each point of modification was located at known methylation sites.

Results

DNA sequence analysis confirmed four genes encoding potential methyltransferase enzymes.

The NEB REBASE database identifies 4 potential methyltransferase enzymes (Roberts RJ *et al.*, 2009) which were confirmed by the genomic sequencing of *Y. enterocolitica* strain 8081. The previously reported restriction modification (R-M) *yenI* system (Kinder *et al.*, 1993) was present at loci YE1808. The *yenI* operon encoded a restriction enzyme and a methyltransferase. The *Yen*I restriction enzyme is *Pst*I-like, recognizing the sequence 5'-CTGCAG-3'. Three 'orphan' methyltransferase enzymes, i.e., lacking a corresponding restriction enzyme, were also identified: YE3972 and YE2361 encoded DNA adenine methyltransferases (Dam) that recognize 5'-GATC-3'. Previous studies show phenotypic changes following *dam* overexpression in *Yersinia* involve overexpression of loci YE3972 (Fälker *et al.*, 2005; Fälker *et al.*, 2006; Fälker *et al.*, 2007). A third 'orphan' methylase identified as YE2362, was a putative DNA cytosine methyltransferase (Dcm). The coding sequence for this putative Dcm overlapped the 3'-end of the YE2361 Dam methylase.

Inter-pulse duration analysis identified methylated nucleotides.

Our sequencing had increased read-depth coverage of DNA isolated from cells grown at 22°C (222) compared to the sequencing read-depth coverage from cells grown at 37°C (178). This difference is depicted in Figure 2.2 A and B. The

software package to analyze DNA genomic sequences by Pacific Biosciences includes a MEME-chip analysis to identify DNA base modifications. A benefit of SMRT sequencing is that two data types exist in which to determine these base modifications. A quality value (QV) score can be determined by calculating -10*log (p-value) compared to a whole genome amplified (unmethylated) *in silico* control. However, QV scoring has a strong dependence on sequence coverage (number of reads over a given position), so reliability of this value increases as coverage increases. Base modifications can also be determined using the inter-pulse duration (IPD) ratio when sequenced samples differ in their coverage. Comparing the IPD ratio at each base across the genome using this method alleviated bias due to sequence coverage differences (Fig. 2.3). Therefore, we used the IPD ratio, vs QV scoring, as our kinetic metric to compare DNA prepared from cells grown at each temperature.

Throughout, the following symbols depict DNA methylation states: ■■-denotes fully methylated sites; □■-denotes hemimethylated sites; and □□-denotes unmethylated sites. In addition, patterns that varied with temperature are depicted as methylation condition separated by an arrow indicating the shift from 22°C to 37°C, e.g. □□→□■.

***Y. enterocolitica* DNA had no evidence of Dcm (5mC) methylation but complete methylation of all *Yen*I restriction sites at both 22°C and 37°C.**

There was no evidence of N4mC or 5mC base modification (Figs. 2.2 and 2.3), despite adequate sequence read-depth coverage. Therefore, we concluded that the Dcm methylase (YE2362) was not active in the LB-growth conditions at the two temperatures tested. In contrast, IPD analysis of all 565 *Yen*I (5'-CTGCAG-3') restriction endonuclease sites showed each adenine was methylated (N6mA) at both temperatures on both DNA strands (Table 2.1). This complete methylation was expected because these sequences are targets for DNA restriction. These sites are present throughout the genome so identification of adenine methylation at these sites served as an internal control for identifying N6mA at non-*Yen*I sites. The genes and their chromosome positions are depicted in Fig. 2.1.

***Y. enterocolitica* DNA had different Dam methylation patterns at 22°C and 37°C.**

Genomic DNA sequence analysis identified 27,118 5'-GATC-3' Dam methylase sites in the *Y. enterocolitica* genome. The chromosome contained 26,664 sites and the pYV (virulence plasmid) contained 454 sites. IPD analysis showed that the majority of these Dam sites were fully methylated on both DNA strands at both temperatures analyzed (Table 2.1). Importantly, a subset of 42 Dam sites had different temperature-dependent methylation patterns (Table 2.2). In addition to these 42 sites, 17 Dam sites remained unmethylated at both temperatures

(□□→□□, Table 2.3). Temperature-dependent Dam methylation patterns were found throughout the genome in both regulatory and gene open reading frame sequences. The genes identified were classified as coding for, putative regulators, ribosomal RNAs, membrane-associated proteins, metabolic proteins, virulence proteins, or for hypothetical proteins of no known function. We found no evidence of cytosine methylation in DNA from cells grown in these experimental conditions. Finally, all *Yen*I restriction endonuclease sites were fully methylated on both DNA strands at both temperatures. The specifics of the differential methylation are in Tables 2.2 and 2.3 and outlined in the following paragraphs.

Two Dam sites were unmethylated at 22°C and methylated at 37°C (□□→■■) localized in the 5'-regulatory regions of YE0914 and YE4070, at 36 and 100 bp respectively 5' from the predicted AUG start codons (Table 2.2). YE0914 is identified by NCBI as a hypothetical protein but EMBL's Pfam sequence analysis identified it as a LysR-type transcription regulator (LTTR) (Finn *et al.*, 2016). LTTRs contain a helix-turn-helix DNA-binding domain and are among the most abundant type of transcriptional regulators within prokaryotes (Henikoff *et al.*, 1988; Maddocks and Oyston, 2008). YE4070 codes for a putative outer membrane oligogalacturonate-specific porin, KdgM.

Four Dam sites were fully methylated at 22°C and unmethylated at 37°C (■■→□□). These included the YE0335 (*hemN*) gene coding for coproporphyrinogen III oxidase with the Dam site 75 bp 5' from the start codon. Two (■■→□□) sites were within ribosomal RNA genes, both a 16s rRNA and 23s

rRNA, and one was within the coding region of YE1322, a putative RTX-family protein, associated with type I secretion system pore-forming toxins. YE1322, with a predicted open reading frame of 6333 bp, contains 17 additional Dam sites, four of which had full methylation at 22°C and hemimethylation at 37°C (■■→■□).

Of the seven Dam sites unmethylated at 22°C and hemimethylated at 37°C (□□→□■), five were in probable regulatory regions and two were within gene coding regions. Of the five Dam sites in intergenic regulatory regions, three were in genes identified as possible regulatory proteins. The first regulatory gene identified with this category was YE1259, with a Dam site 51 bp 5' from the AUG start codon. YE1259 was identified as a PadR-like transcriptional regulator. The second regulatory gene in this category was YE3423, with a Dam site 85 bp 5' from the AUG start codon. YE3423 codes for an ArsR-family transcriptional regulator. Of note, this Dam sites is positioned between two divergently-transcribed genes, placing it 66 bp 5' from the AUG start codon of YE3424. YE3424 codes for a putative zinc metallopeptidase. A BLAST search of the YE3424 predicted amino acid sequence had 70% identity and 83% similarity to enterohemorrhagic *E. coli* (EHEC) metallopeptidase, SprT, a type III secretion system effector. The third regulator gene in this category of Dam methylation (□□→□■) was YE0316 with a Dam site 206 bp 5' from the AUG start codon. YE0316 codes for putative DNA-binding protein with 53% similarity to sigma-70 (*fecI*). Of note, this Dam sites is also positioned 172 bp 5' from the AUG start codon of divergently transcribed YE0315. YE0315 codes for a membrane transport protein with high similarity to *Salmonella*

typhimurium TonB. Two additional genes showed this pattern of methylation (□□→□■): YE0981, encoding a hypothetical protein with a Dam site 74 bp 5' from the start codon; and YE1098 with the Dam site 51 bp 5' from the start codon. This gene encodes GutA, also referred to as SrlA, a glucitol/sorbitol-specific IIC2 component, a subunit of the phoshotransferase system (Meadow *et al.*, 1990).

Of the 24 Dam sites fully methylated at 22°C and hemi-methylated at 37°C (■■→■□), 22 were located within structural gene coding sequences (Table 2.2). Of the two sites in regulatory regions, one ■■→■□ Dam site is positioned 288 bp 5' from the AUG start codon of YE2225, a predicted cyclic-di-GMP phosphodiesterase with a conserved EAL domain. This enzyme inactivates cyclic-di-GMP, a common bacterial secondary messenger (Romling *et al.*, 2013). Interestingly, within the promoter of YE2225, three additional Dam sites had differential methylation patterns. The three Dam sites most proximal to the coding region, 38, 55, and 73 bp 5' from the AUG start codon, remained unmethylated at each temperature. It is noteworthy that this regulatory region, from 38 to 288 5' from the start AUG codon, contains four Dam sites that show atypical methylation patterns (Tables 2.2 and 2.3). Statistically, only one Dam site was predicted over a span of 256 bp. The second Dam site in a potential regulatory region is on the virulence plasmid, 86 bp from the 5' start of YEP0064, a putative pseudogene. This pattern of methylation (■■→■□) was also prominent in rRNA genes. We identified five Dam sites in 16S and 23S rRNA genes, which were position-specific at the 3'-ends. This included four of the seven 16S ribosomal RNA genes (YEr007, YEr010,

YEr018, and YEr022) with this conserved pattern of atypical methylation (Tables 2.2 and 2.3).

Seventeen Dam sites were identified to be unmethylated at each temperature (□□→□□). Ten were located in probable regulatory regions (Table 2.3). As described above, three unmethylated Dam sites were located 5' to the start of YE2225, the putative cyclic-di-GMP phosphodiesterases described above. Two were at 131- and 144 bp 5' from the start codon of YE0983, a hypothetical protein. The remaining unmethylated Dam sites were 5' to the AUG start codon of an oligogalacturonate lyase, a bifunctional transaminase, a putative transporter, a putative glycosyl transferase, and YE2151, a TetR-like transcriptional regulator similar to NemR (Table 2.3).

Discussion

The most significant finding of this work identified two subsets of Dam methylation sites with temperature-dependent Dam methylation patterns or Dam sites that remained unmethylated regardless of temperature. To our knowledge, this is the first systematic analysis of temperature-dependent DNA methylation of a facultative bacterial pathogen. Importantly, IPD ratio, rather than the traditional QV scoring, was the kinetic metric used to compare DNA so that DNA sequence coverage differences between cells grown a different temperatures was not an issue. Both categories of Dam sites were localized either within regulatory regions (5' from the predicted AUG start site) or within gene open-reading frames. Among

the genes with differential methylation were those encoding regulatory, virulence, metabolic, membrane-associated proteins, and both 23s and 16s rRNA genes. Importantly, we found no evidence of Dcm methylation at predicted Dcm sites (5'-CCWGG-3') on the chromosome nor on the pYV. To determine if the temperature-dependent methylation patterns identified correlated with gene expression we capitalized on the *Y. enterocolitica* comprehensive transcriptome analyzes conducted by Bent *et al.* (2015). This study is a comprehensive RNA-seq analysis of *Y. enterocolitica* 8081v transcripts prepared from cultures grown at 25°C in LB broth and from 37°C cultures grown in (i) conditioned RPMI, (ii) in surface contact with mouse macrophages, or (iii) internalized by mouse macrophages.

The LTTR at loci YE0914 has a Dam site in the predicted regulatory region that is unmethylated at 22°C and fully methylated at 37°C (□□→■■) exemplifying a gene that would require two rounds of DNA replication to restore the unmethylated state after a temperature downshift. This gene is highly conserved among the enteric bacteria with 88% identity to an LTTR in *Serratia marscesens* and 76% identity to an LTTR in *Shigella*. To our knowledge, YE0914, nor its enteric homologs, have been characterized. LTTRs are among the most abundant types of transcriptional regulators present in bacteria, and control a diverse subset of genes, including motility, metabolism, and virulence (Maddocks and Oyston, 2008). The *E. coli* LTTR, OxyR, activates one of its many targets, the phage Mu *mom* gene, only when three Dam sites upstream of the promoter are methylated (Bölker and Kahnmann, 1989). Therefore, there is precedence for LTTR expression correlating

with promoter methylation. Transcription of YE0914 in the RNA-seq transcriptomic analysis by Bent *et al.* (2015) does not show temperature-dependent regulation between *Y. enterocolitica* cells grown in LB at 26°C compared to conditioned RPMI at 37°C (Bent *et al.*, 2015). However, the p-value in this study for YE0914 at these two temperatures is very high (0.65-0.93), suggesting significant heterogeneity in expression within sampled populations. Heterogenic expression is also indicative of epigenetic mechanisms whereby bacteria modify gene expression to "bet hedge" within otherwise clonal populations (Casadesús and Low, 2013).

Among the regulatory genes identified with unmethylated Dam site(s) (□□→□□) upstream of the translational start was the TetR-like transcriptional regulator YE2151. This gene has a predicted 64% amino acid similarity (52% identity) to *E. coli* NemR (Gray *et al.*, 2013). NemR activates several stress-response genes required for survival when cells are exposed to reactive oxygen species such as hypochlorus acid (Hassett and Cohen, 1989). The transcriptomic analysis by Bent *et al.* (2015) shows significantly increased expression of YE2151 at elevated temperature. However, based on our methylation analysis, methylation did not appear to play a role in regulation since the unmethylated state of the Dam site identified did not change with temperature. It is noteworthy that this stress-response gene is localized near the terminus of DNA replication. Genes expressed during stress or stationary phase are positionally-biased to be near the terminus and show a requirement for reduced supercoiling for expression (Muskhelishvili and Travers AA, 2013; Travers AA and Muskhelishvili, 2015). Previous studies by our

laboratory have likewise demonstrated that temperature regulation in *Y. enterocolitica* correlates to changes in DNA supercoiling. DNA methylation can affect histone-like protein binding which in turn dictate regional DNA domain confirmation. As such, lack of methylation at the chromosome terminus may influence gene expression indirectly through Dam site protection.

Also showing differential regulation in its upstream regulatory region, was the regulatory gene, YE1259 (□□→■□), coding for a PadR-like transcriptional regulator. PadR family transcriptional regulators are involved in *Vibrio* virulence. For example, AphA is involved in the regulation of motility and virulence in *Vibrio parahaemolyticus* (Wang *et al.*, 2013). In *V. parahaemolyticus,* both an LTTR (AphB) and AphA co-regulate acetoin production and motility (Kovacikova *et al.*, 2004; Kovacikova *et al.*, 2005). Notably these phenotypes are temperature regulated in *Y. enterocolitica,* where acetoin and flagella are only produced at low temperature (Chester and Stotzky, 1976). The transcriptomic analysis by Bent *et. al.* (2015) however, reveals little variation in temperature expression of YE1259. YE1098, coding for GutA, also showed the same pattern of methylation as YE1259 (□□→■□). van der Woude *et al.* (van der Woude *et al.*, 1998) reported that the Dam site 44 bp from the *E. coli gutABD* transcription start site is unmethylated. GutR and CRP compete for binding at this Dam site, and GutR binding prevents methylation. However, changing the Dam site in *E. coli* did not alter expression, suggesting that methylation does not play a direct role in *gutABD* regulation. In *Y. enterocolitica* we noted the presence of a putative CRP-binding site overlapping

this Dam site suggesting similar regulation. The fact that we saw a transition in methylation pattern with growth temperature suggests methylation of this operon may be sensitive to other environmental conditions. Indeed, Bent *et al.* (2015) show significant activation of *gutA* at 37°C.

The most extensive atypical Dam methylation was seen in YE2225. Over a span of 288 bp upstream of the translational start site are four Dam sites, three of which remained unmethylated at both temperatures and one, the most distal from the AUG start site, went from fully methylated at 25°C to hemimethylated at 37°C (■■→■□). YE2225 encodes a putative cyclic-di-GMP phosphodiesterase (PDE) with a conserved EAL domain required for hydrolysis of the common bacterial signaling messenger di-c-GMP (Romling *et al.*, 2013). Transcriptomic analysis by Bent *et al.* (2015) shows this gene undergoes a steady upregulation when exposed to mouse macrophages grown in RPMI at 37°C reaching a ~5.5-fold peak increase at 60 min. Cyclic-di-GMP plays an essential role during environmental transitions of bacteria. Phenotypes governed by cyclic-di-GMP include virulence gene expression and the transition from the motile to sessile states during biofilm formation (Bobrov *et al.*, 2011) . For the related *Y. pestis*, cyclic-di-GMP regulates exopolysaccharide production during biofilm formation in the flea vector. Hydrolysis of cyclic-di-GMP is essential in the transition from flea ambient temperature to mammalian host temperature. Bobrov *et al.* (2011) show mutational inactivation of HmsP, the *Y. pestis* cyclic-di-GMP phosphodiesterase, results in a 1000-fold reduction in virulence. Our results suggest methylation in concert with higher temperature may

alter cyclic-di-GMP levels essential for host infection. Interestingly in *V. cholera*, both AphA and AphB repress the activation of the EAL-containing *acgA*. AcgA further shows effects on motility, virulence, and biofilm formation (Kovacikova *et al.*, 2005).

YE0316 (□□→■□) is a homolog of FecI. FecI is a sigma factor which regulates expression of the ferric di-citrate uptake system clustered chromosomally in the iron regulatory region of *E. coli* (Buchanan, 2005). The *fecI* gene is temperature-regulated in *E. coli*, with higher expression at 37°C, and controlled by histone-like nucleoid structuring (H-NS) protein (White-Ziegler and Davis TR, 2009). This temperature regulation is true for YE0316; expression is increased over eight-fold at 120 mins following 37°C exposure (Bent *et al.*, 2015). Notably, H-NS binding is affected by DNA methylation (White-Ziegler *et al.*, 1998).

Several locations of disparate methylation patterns were within or near ribosomal RNA elements. Of the seven 16S rRNA genes present on the chromosome, four showed temperature variation in methylation. Similarly, of the seven 23S rRNA elements, five showed temperature variation in methylation. The positions for each of these Dam sites were localized to the 3’ end of the coding region. Furthermore, the ribosomal RNAs showing this methylation pattern were localized near the origin of replication. Genomic position does effect expression levels (Muskhelishvili and Travers AA, 2013; Gerganova *et al.*, 2015).

Finally, we identified 17 Dam sites which remain unmethylated. This number is consistent with unmethylated sites determined for uropathogenic *E. coli* and *C.*

crescentus (Blyn *et al.*, 1990; van der Woude *et al.*, 1998; Fang *et al.*, 2012; Kozdon *et al.*, 2013). This suggests these sites are permanently protected by either binding by proteins, local DNA domain structure, or both. It is noteworthy that 14 of these 17 unmethylated sites are clustered in 10 genes positioned near the origin (five sites) or terminus of DNA replication (12 sites).

The lack of cytosine methylation detected in our analysis is consistent with the transcriptome analysis of Bent *et al.* (2015). Our interpretation of their data indicates either no or minimal expression of YE2361 (*dam)* and YE2362 (*dcm*) at all growth and temperature conditions monitored in their experiments. Conversely, *yenI* (YE1808) and *dam* (YE3972) are expressed constitutively with little temperature variation.

Overproduction of Dam affects motility and invasion abilities in *Y. enterocolitica*, implying that temperature-dependent phenotypic differences may be regulated by temperature-responsive differential methylation (Low *et al.*, 2001). In this study, we showed there is a subset of Dam sites displaying temperature-dependent methylation patterns. Of the Dam sites in regulatory regions, some patterns correlate with temperature regulation when compared to the RNA-seq data of Bent *et al.* (2015). Surprisingly, we did not see differences in methylation of regulatory regions of key genes involved in TTSS temperature regulation suggesting if methylation is a component of motility and Yop expression, it is indirect. However, the mechanism underlying this regulation by Dam remains unclear, the genes identified here provide targets for further investigation.

Acknowledgements

We thank Derek Pouchnik and Mark Wildung (Washington State University genomics core laboratory) for DNA sequencing assistance, and Jonas Korlach (Pacific Biosciences) for assistance in DNA methylation data interpretation.

References

Atack, J.M., Srikhanta, Y.N., Fox, K.L., Jurcisek, J.A., Brockman, K.L., Clark, T.A., *et al.* (2015) A biphasic epigenetic switch controls immunoevasion, virulence and niche adaptation in non-typeable *Haemophilus influenzae*. *Nature Communications* **6**: 1–12.

Bent, Z.W., Poorey, K., Brazel, D.M., LaBauve, A.E., Sinha, A., Curtis, D.J., *et al.* (2015) Transcriptomic Analysis of *Yersinia enterocolitica* Biovar 1B Infecting Murine Macrophages Reveals New Mechanisms of Extracellular and Intracellular Survival. *Infection and Immunity* **83**: 2672–2685.

Blyn, L.B., Braaten, B.A., and Low, D.A. (1990) Regulation of *pap* pilin phase variation by a mechanism involving differential Dam methylation states. *EMBO J* **9**: 4045–4054.

Bobrov, A.G., Kirillina, O., Ryjenkov, D.A., Waters, C.M., Price, P.A., Fetherston, J.D., *et al.* (2011) Systematic analysis of cyclic di-GMP signaling enzymes and their role in biofilm formation and virulence in *Yersinia pestis*. *Molecular Microbiology* **79**: 533–551.

Bölker, M., and Kahnmann, R. (1989) The *Escherichia coli* Regulatory Protein OxyR Discriminates Between Methylated and Unmethylated States of the Phage Mu *mom* Promoter. *EMBO J* **8**: 2403–2410.

Buchanan, S.K. (2005) Bacterial metal detectors. *Molecular Microbiology* **58**: 1205–1209.

Campbell, J.L., and Kleckner, N. (1988) The rate of Dam-mediated DNA adenine methylation in *Escherichia coli*. *Gene* **74**: 189–190.

Casadesús, J., and Low, D.A. (2006) Epigenetic Gene Regulation in the Bacterial World. *Microbiol Mol Biol Rev* **70**: 830–856.

Casadesús, J., and Low, D.A. (2013) Programmed Heterogeneity: Epigenetic Mechanisms in Bacteria. *Journal of Biological Chemistry* **288**: 13929–13935.

Chaisson, M.J., and Tesler, G. (2012) Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics* **13**: 238.

Chester, B., and Stotzky, G. (1976) Temperature-dependent cultural and biochemical characteristics of rhamnose-positive *Yersinia enterocolitica*. *J Clin Microbiol* **3**: 119–127.

Collier, J. (2009) Epigenetic regulation of the bacterial cell cycle. *Current Opinion in Microbiology* **12**: 722–729.

Collier, J., McAdams, H.H., and Shapiro, L. (2007) A DNA methylation ratchet governs progression through a bacterial cell cycle. *Proc Natl Acad Sci USA* **104**: 17111–17116.

Davis, B.M., Chao, M.C., and Waldor, M.K. (2013) Entering the era of bacterial epigenomics with single molecule real time DNA sequencing. *Current Opinion in Microbiology* 1–7.

de Koning-Ward, T.F., and Robins-Browne, R.M. (1997) A novel mechanism of urease regulation in *Yersinia enterocolitica*. *FEMS Microbiology Letters* **147**: 221–226.

Dorman, C.J. (1991) DNA supercoiling and environmental regulation of gene expression in pathogenic bacteria. *Infection and Immunity* **59**: 745–749.

Fang, G., Munera, D., Friedman, D.I., Mandlik, A., Chao, M.C., Banerjee, O., *et al.* (2012) Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nature Biotechnology* **30**: 1–11.

Fälker, S., Schilling, J., Schmidt, M.A., and Heusipp, G. (2007) Overproduction of DNA Adenine Methyltransferase Alters Motility, Invasion, and the Lipopolysaccharide O-Antigen Composition of *Yersinia enterocolitica*. *Infection and Immunity* **75**: 4990–4997.

Fälker, S., Schmidt, M.A., and Heusipp, G. (2005) DNA methylation in *Yersinia enterocolitica*: role of the DNA adenine methyltransferase in mismatch repair and regulation of virulence factors. *Microbiology* **151**: 2291–2299.

Fälker, S., Schmidt, M.A., and Heusipp, G. (2006) Altered Ca^{2+} Regulation of Yop Secretion in *Yersinia enterocolitica* after DNA Adenine Methyltransferase Overproduction Is Mediated by Clp-Dependent Degradation of LcrG. *J Bacteriol* **188**: 7072–7081.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., *et al.* (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**: D279–D285.

Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., *et al.* (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature Methods* **7**: 461–465.

Gerganova, V., Berger, M., Zaldastanishvili, E., Sobetzko, P., Lafon, C., Mourez, M., *et al.* (2015) Chromosomal position shift of a regulatory gene alters the bacterial phenotype. *Nucleic Acids Res* **43**: 8215–8226.

Gray, M.J., Wholey, W.Y., Parker, B.W., Kim, M., and Jakob, U. (2013) NemR Is a Bleach-sensing Transcription Factor. *Journal of Biological Chemistry* **288**: 13789–13798.

Hassett, D.J., and Cohen, M.S. (1989) Bacterial adaptation to oxidative stress: implications for pathogenesis and interaction with phagocytic cells. *FASEB J* **3**: 2574–2582.

Henikoff, S., Haughn, G.W., Calvo, J.M., and Wallace, J.C. (1988) A Large Family of Bacterial Activator Proteins. *Proc Natl Acad Sci USA* **85**: 6602–6606.

Julio, S.M., Heithoff, D.M., Provenzano, D., Klose, K.E., Sinsheimer, R.L., Low, D.A., and Mahan, M.J. (2001) DNA Adenine Methylase Is Essential for Viability and Plays a Role in the Pathogenesis of *Yersinia pseudotuberculosis* and *Vibrio cholerae*. *Infection and Immunity* **69**: 7610–7615.

Kinder, S.A., Badger, J.L., Bryant, G.O., Pepe, J.C., and Miller, V.L. (1993) Cloning of the YenI restriction-endonuclease and methyltransferase from *Yersinia enterocolitica* serotype O:8 and construction of a transformable R(-)M(+) mutant. *Gene* **136**: 271–275.

Kovacikova, G., Lin, W., and Skorupski, K. (2004) *Vibrio cholerae* AphA uses a novel mechanism for virulence gene activation that involves interaction with the LysR-type regulator AphB at the *tcpPH* promoter. *Molecular Microbiology* **53**: 129–142.

Kovacikova, G., Lin, W., and Skorupski, K. (2005) Dual regulation of genes involved in acetoin biosynthesis and motility/biofilm formation by the virulence activator AphA and the acetate-responsive LysR-type regulator AlsR in *Vibrio cholerae*. *Molecular Microbiology* **57**: 420–433.

Kozdon, J.B., Melfi, M.D., Luong, K., Clark, T.A., Boitano, M., Wang, S., *et al.* (2013) Global methylation state at base-pair resolution of the *Caulobacter* genome throughout the cell cycle. *Proc Natl Acad Sci USA* **110**: E4658–67.

Low, D.A., Weyand, N.J., and Mahan, M.J. (2001) Roles of DNA Adenine Methylation in Regulating Bacterial Gene Expression and Virulence. *Infection and Immunity* **69**: 7197–7204.

Maddocks, S.E., and Oyston, P.C.F. (2008) Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology* **154**: 3609–3623.

Marinus, M.G., and Casadesús, J. (2009) Roles of DNA adenine methylation in host–pathogen interactions: mismatch repair, transcriptional regulation, and more. *FEMS Microbiol Rev* **33**: 488–503.

Maurelli, A.T. (1989) Temperature Regulation of Virulence Genes in Pathogenic Bacteria - a General Strategy for Human Pathogens. *Microb Pathog* **7**: 1–10.

Meadow, N.D., Fox, D.K., and Roseman, S. (1990) The bacterial phosphoenol-pyruvate: glycose phosphotransferase system. *Annu Rev Biochem* **59**: 497–542.

Michiels, T., and Cornelis, G.R. (1991) Secretion of Hybrid Proteins by the *Yersinia* Yop Export System. *J Bacteriol* **173**: 1677–1685.

Minnich, S.A., and Rohde, H.N. (2007) A rationale for repression and/or loss of motility by pathogenic *yersinia* in the mammalian host. *Adv Exp Med Biol* **603**: 298–311.

Montminy, S.W., Khan, N., McGrath, S., Walkowicz, M.J., Sharp, F., Conlon, J.E., *et al.* (2006) Virulence factors of *Yersinia pestis* are overcome by a strong lipopolysaccharide response. *Nat Immunol* **7**: 1066–1073.

Murray, I.A., Clark, T.A., Morgan, R.D., Boitano, M., Anton, B.P., Luong, K., *et al.* (2012) The methylomes of six bacteria. *Nucleic Acids Res* **40**: 11450–11462.

Muskhelishvili, G., and Travers, A.A. (2013) Integration of syntactic and semantic properties of the DNA code reveals chromosomes as thermodynamic machines converting energy into information. *Cell Mol Life Sci* **70**: 4555–4567.

Rebeil, R., Ernst, R.K., Gowen, B.B., Miller, S.I., and Hinnebusch, B.J. (2004) Variation in lipid A structure in the pathogenic *yersiniae*. *Molecular Microbiology* **52**: 1363–1373.

Reisenauer, A., Kahng, L.S., McCollum, S., and Shapiro, L. (1999) Bacterial DNA methylation: a cell cycle regulator? *J Bacteriol* **181**: 5135–5139.

Roberts, D., Hoopes, B.C., McClure, W.R., and Kleckner, N. (1985) IS10 transposition is regulated by DNA adenine methylation. *Cell* **43**: 117–130.

Roberts, R.J., Vincze, T., Posfai, J., and Macelis, D. (2009) REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res* **38**: D234–D236.

Rohde, J.R., Fox, J.M., and Minnich, S.A. (1994) Thermoregulation in *Yersinia enterocolitica* is coincident with changes in DNA supercoiling. *Molecular Microbiology* **12**: 187–199.

Romling, U., Galperin, M.Y., and Gomelsky, M. (2013) Cyclic di-GMP: the First 25 Years of a Universal Bacterial Second Messenger. *Microbiol Mol Biol Rev* **77**: 1–52.

Schadt, E.E., Banerjee, O., Fang, G., Feng, Z., Wong, W.H., Zhang, X., *et al.* (2013) Modeling kinetic rate variation in third generation DNA sequencing data to detect putative modifications to DNA bases. *Genome Research* **23**: 129–141.

Srikhanta, Y.N., Fox, K.L., and Jennings, M.P. (2010) The phasevarion: phase variation of type III DNA methyltransferases controls coordinated switching in multiple genes. *Nature Reviews Microbiology* **8**: 196–206.

Straley, S.C., Plano, G.V., Skrzypek, E., Haddix, P.L., and Fields, K.A. (1993) Regulation by Ca^{2+} in the *Yersinia* low-Ca^{2+} response. *Molecular Microbiology* **8**: 1005–1010.

Thomson, N.R., Howard, S., Wren, B.W., Holden, M.T.G., Crossman, L., Challis, G.L., *et al.* (2006) The Complete Genome Sequence and Comparative Genome Analysis of the High Pathogenicity *Yersinia enterocolitica* Strain 8081. *PLoS Genet* **2**: e206–13.

Tost, J., and Gut, I.G. (2007) DNA methylation analysis by pyrosequencing. *Nat Protoc* **2**: 2265–2275.

Travers, A.A., and Muskhelishvili, G. (2015) DNA structure and function. *FEBS J* **282**: 2279–2295.

Travers, K.J., Chin, C.S., Rank, D.R., Eid, J.S., and Turner, S.W. (2010) A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Res* **38**: e159–e159.

van der Woude, M., Hale, W.B., and Low, D.A. (1998) Formation of DNA methylation patterns: nonmethylated GATC sequences in *gut* and *pap* operons. *J Bacteriol* **180**: 5913–5920.

Wang, L., Ling, Y., Jiang, H., Qiu, Y., Qiu, J., Chen, H., *et al.* (2013) AphA is required for biofilm formation, motility, and virulence in pandemic *Vibrio parahaemolyticus*. *International Journal of Food Microbiology* **160**: 245–251.

White-Ziegler, C.A., and Davis, T.R. (2009) Genome-Wide Identification of H-NS-Controlled, Temperature-Regulated Genes in *Escherichia coli* K-12. *J Bacteriol* **191**: 1106–1110.

White-Ziegler, C.A., Angus Hill, M.L., Braaten, B.A., van der Woude, M.W., and Low, D.A. (1998) Thermoregulation of *Escherichia coli pap* transcription: H-NS is a temperature-dependent DNA methylation blocking factor. *Molecular Microbiology* **28**: 1121–1137.

Wilson, G. (1991) Restriction and Modification Systems. *Annu Rev Genet* **25**: 585–627.

Wion, D., and Casadesús, J. (2006) N6-methyl-adenine: an epigenetic signal for DNA–protein interactions. *Nature Reviews Microbiology* **4**: 183–192.

Zilberman, D., and Henikoff, S. (2007) Genome-wide analysis of DNA methylation patterns. *Development* **134**: 3959–3965.

Table 2.1. Comparison of *Yen*I, Dam, and Dcm methylation sites at 22°C and 37°C.

MOTIF [a]	TOTAL # of Motifs	22°		37°	
		# of Atypical Methylation Detected	# of Fully Methylation Detected	# of Atypical Methylation Detected	# of Fully Methylation Detected
5'-G<u>A</u>TC-3' (Dam)					
Chromosome	26664	28	26636	50	26614
pYV Plasmid	454	7	447	7	447
5'-CTGC<u>A</u>G-3' (*Yen*I)					
Chromosome	554	0	554	0	554
pYV Plasmid	11	0	11	0	11
5'-<u>C</u>CWGG-3' (Dcm)					
Chromosome	7237	0	0	0	0
pYV Plasmid	88	0	0	0	0

[a]Methylated base is underlined.

Table 2.2. Dam sites showing temperature-dependent methylation differences at 22°C and 37°C.

Position	22° [a]	37° [a]	Name	Location [b]	Cat. [c]	Description
20357	■■	□□	YEr002	ORF	rR	23S rRNA
315457	■■	■□	YEr007	ORF, -160 from 3' end	rR	16S rRNA
315631	■□	■□		+16 from 3' end of 16S rRNA		
356036	■■	■□	YEr010	ORF, -158 from 3' end	rR	16S rRNA
356210	■■	■□		+16 from 3' end of 16S rRNA		
373662	□□	■□	YE0315	-172 from TSS	M	putative membrane transport proteins, similar to Salmonella TonB
			YE0316	-206 from TSS	MT	putative DNA-binding protein
399373	■■	□□	YE0335	-75 from TSS	MT	HemN, coproporphyrinogen III oxidase
623091	□□	■□	YE0546	ORF	M	putative glycosyltransferase
888993	□□	■□	YE0762	ORF	MT	CysD, sulfate adenylyltransferase subunit 2
1042835	□□	■■	YE0914	-36 from TSS	R	putative LysR-type transcriptional regulator
1104202	□□	■□	YE0981	-74 from TSS	H	hypothetical protein
1111932	■■	■□	YE0990	ORF	MT	AYP/GTP-binding protein
1113753	■■	■□	YE0992	ORF	H	hypothetical protein
1228189	□□	■□	YE1098	-182 from TSS	MT	GutA, pts system, glucitol/sorbitol-specific iic2 component
1403906	□□	■□	YE1259	-51 from TSS	R	putative PadR-like family transcriptional regulator
1468949	■■	■□	YE1322	ORF	V	putative RTX-family protein
1469615	■■	■□				
1471616	■■	□□				
1471736	■■	■□				
1471902	■■	■□				
1879399	■■	■□	YE1683	ORF	R	putative prophage encoded two-component system response regulator
1881251	■■	■□	YE1684	ORF	R	putative prophage encoded two-component system histidine kinase
2354334	■■	■□	YE2154	ORF. near 5'	MT	Rnt, ribonuclease T
2436018	■■	■□	YE2225	-288 from TSS	R	putative cyclic-di-GMP phosphodiesterase
2598581	■■	■□	YE2407	ORF	V	putative hemolysin
2602469	■■	■□	YE2409	+30 from 3' end of mviN	MT	MviN, putative membrane-associated protein
2852693	■■	■□	YE2635	ORF		putative metallo-beta-lactamase superfamily protein
3352496	■■	■□	YE3082	ORF	MT	RfbX, putative O-antigen transporter

3548081	■■	■□	YEr017	ORF	rR	23S rRNA
3549372	■□	□□				
3550740	■■	■□	YEr018	+17 from 3' end of 16S rRNA	rR	16S rRNA
3655022	■■	■□	YE3343	ORF. near 3'	MT	OutL, general secretion pathway protein L
3739389	□□	■□	YE3423	-85 from TSS	R	ArsR-family transcriptional regulator
			YE3424	-66 from TSS	M	putative zinc metallopeptidases
4243325	■■	□□	YEr022	+17 from 3' end of 16S rRNA	rR	16S rRNA
4441739	□□	■■	YE4070	-100 from TSS	MT	putative oligogalacturonate-specific porin protein
2443	■■	■□	YEP0004	ORF	V	YopQ, virulence plasmid protein
2654	■□	■□		ORF		
15967	■□	■■	YEP0018	ORF	MT	LcrD, low calcium response locus membrane protein d
21305	■□	■□	YEP0026	ORF	V	YscP, putative type III secretion protein
23409	■□	■■	YEP0029	ORF	V	YscS, putative type III secretion protein
24586	■■	■□	YEP0031	ORF	V	YscU, putative type III secretion protein
25258	■□	■■		ORF		
45942	■■	■□	YEP0064	-86 from TSS	H	putative pseudogene
48481	■□	■■	YEP0066	-158 from TSS	M	putative YadA invasin
66135	■■	■□	YEP0096	ORF	MT	putative plasmid copy control protein

[a]■■, fully methylated; ■□, hemimethylated; □□, unmethylated

[b]ORF, Open Reading Frame; TSS, Translational start site.

[c]H, Hypothetical; R, Regulator; rR, ribosomal RNA; M, membrane-associated; MT, metabolic; V, virulence

Table 2.3. . Dam sites unmethylated at both 22°C and 37°C.

Position	22° [a]	37° [a]	Name	Location [b]	Cat. [c]	Description
316977	□□	□□	YEr008	ORF	rR	23s rRNA
357556	□□	□□	YEr011	ORF	rR	23s rRNA
1105938	□□	□□	YE0983	-131 from TSS	H	hypothetical protein
1105951	□□	□□		-144 from TSS		
1998525	□□	□□	YE1808	ORF	MT	*Yen*I, methyltransferase-endonuclease
1998670	□□	□□				
2057696	□□	□□	YE1876	-106 from TSS	MT	OgI, oligogalacturonate lyase
2351648	□□	□□	YE2151	-23 from TSS	R	transcriptional repressor NemR
2435768	□□	□□	YE2225	-38 from TSS	R	putative cyclic-di-GMP phosphodiesterase.
2435785	□□	□□		-55 from TSS		
2435803	□□	□□		-73 from TSS		
2656973	□□	□□	YE2469	-369 from TSS	MT	ArgM, bifunctional succinylornithine transaminase/acetylornithine transaminase
2857854	□□	□□	YE2639	-318 from TSS	M	putative transporter protein
3351041	□□	□□	YE3081	ORF	MT	WbcD, putative 6-deoxy-D-Gul transferase
			YE3080	-191 from TSS	MT	WbcE, putative glycosyl transferases
4241957	□□	□□	YEr021	ORF	rR	23s rRNA
4243004	□□	□□	YEt079	ORF		tRNA-Ala
65753	□□	□□	YEP0094	ORF	MT	IS541 transposase
			YEP0095	-240 from TSS	MT	putative plasmid copy number

[a] □□, unmethylated

[b] ORF, Open Reading Frame; TSS, Translational start site.

[c] H, Hypothetical; R, Regulator; rR, ribosomal RNA; M, membrane-associated; MT, metabolic

Figure 2.1. *Y. enterocolitica* sequence analysis identifies 4 methyltransferases.

The figure was generated by the REBASE program provided by New England Biolabs. YE0519 is a putative restriction endonuclease with unidentified motif specificity. YE1681 is a putative orphan methyltransferase with unidentified motif specificity, 'orphan' referring to the absence of a corresponding restriction endonuclease. YE1808, encodes the *Yen*I restriction modification system recognizing motif 5'-CTGCAG-3' previously described (Thomson *et al.*, 2006). YE2361 is identified as a DNA adenine methyltransferase (Dam) recognizing 5'-GATC-3'. This ORF overlaps by 4bp with the downstream ORF, YE2362, which, according to NCBI, encodes a DNA cytosine methyltransferase (Dcm). YE3972 is an orphan Dam having recognition specificity to 5'-GATC-3' as reported previously (Fälker *et al.*, 2005; Fälker *et al.*, 2006; Fälker *et al.*, 2007).

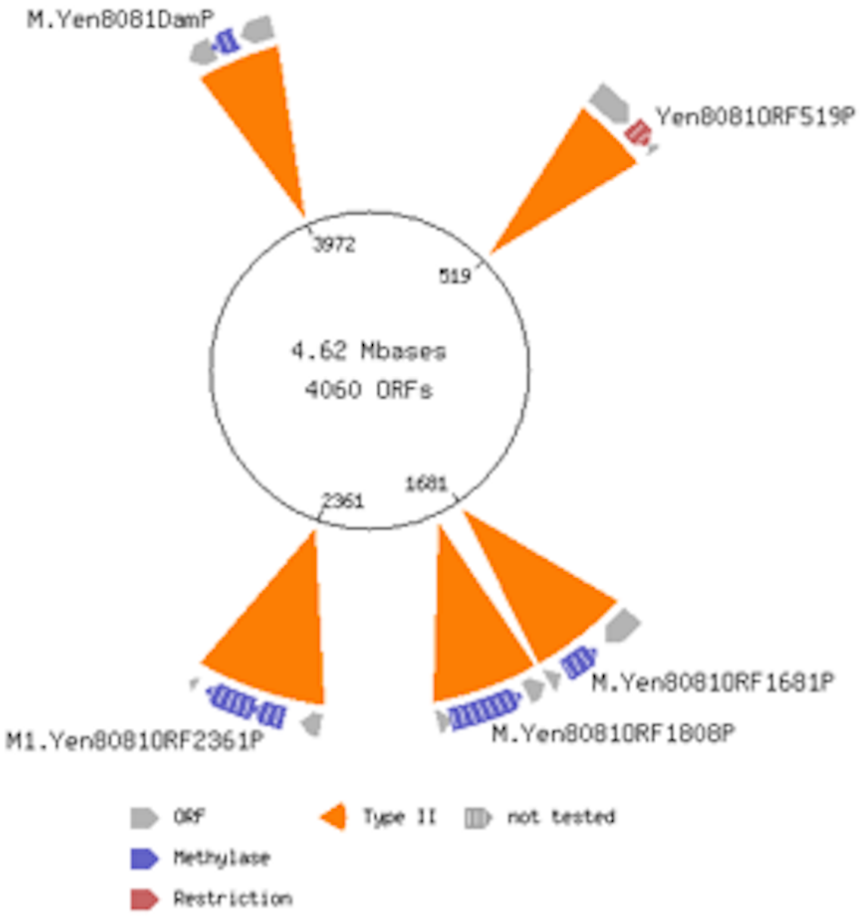

M.Yen8081DamP
Yen80810RF519P
3972
519
4.62 Mbases
4060 ORFs
2361
1681
M1.Yen80810RF2361P
M.Yen80810RF1681P
M.Yen80810RF1808P
ORF
Methylase
Restriction
Type II
not tested

Figure 2.2. Base modification score to detect methylation is influenced by coverage bias.

DNA methylation can be determined by an increased Quality Value (QV) score (-log*10(P-value) compared to unmodified *in silico* control) (Murray *et al.*, 2012). However, the scatter plot in panel (A) illustrates the coverage of the 37°C sample (x-axis) is significantly higher than that of the 22°C sample (y-axis), shown with a 1:1 correlation in the dotted red line. QV score has a strong dependence on coverage, increasing as coverage increases, thus Panel (B) reveals a similar trend when comparing scores as in Panel A, with scores skewing toward the 37°C sample. Therefore, this illustrates that QV scores alone provides a poor metric for comparison between each sample. (C,D) Scatter plot of sequencing coverage and QV score for all genomic positions at 22°C and 37°C respectively, again separated for each base. The threshold for detected base modification above background is indicated by the black line. Scatter plot distributions show significantly higher scores for many adenine (red) residues above the background, whereas no increased scores are evident for all other bases, thus indicating only adenine residues are methylated.

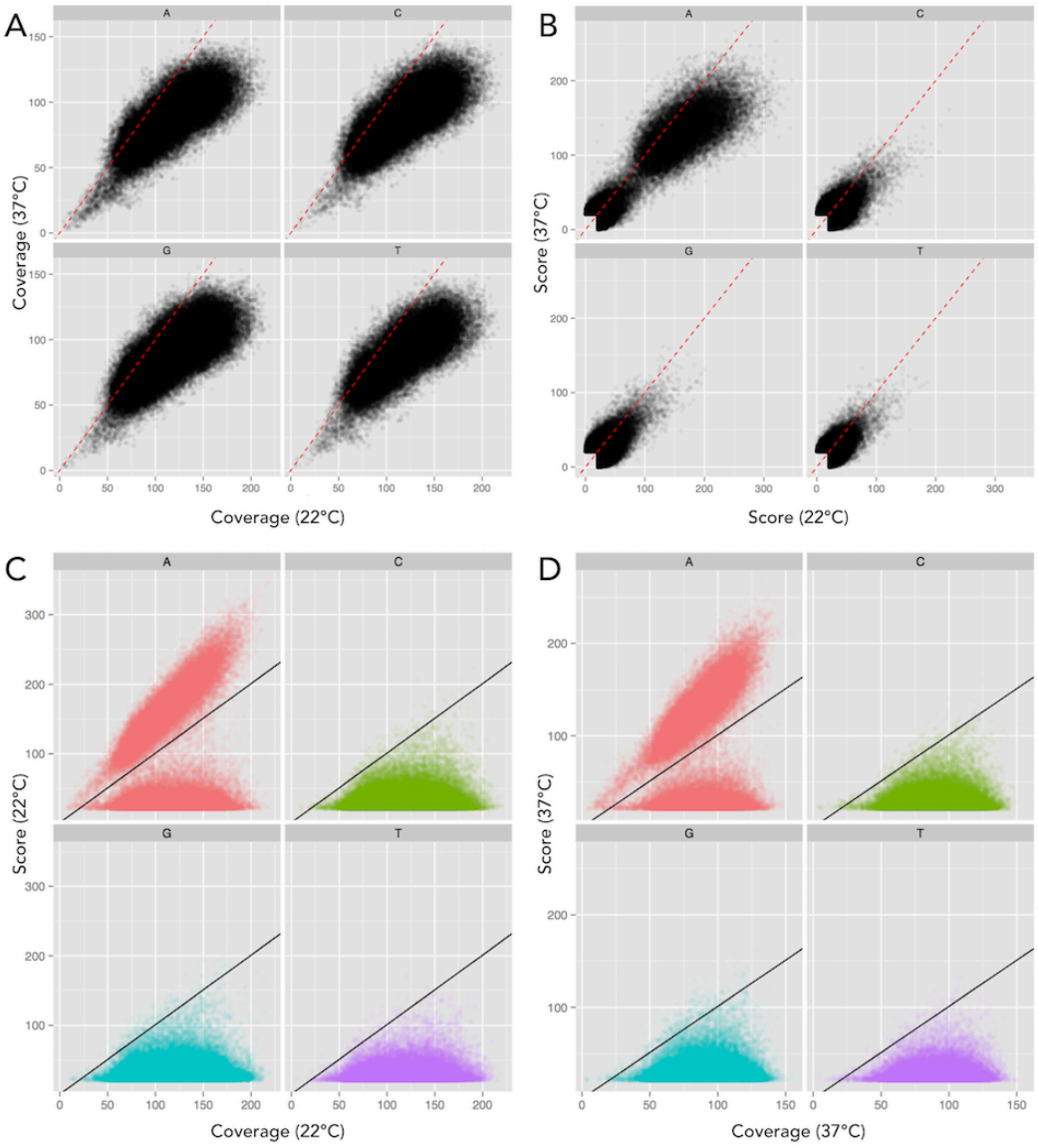
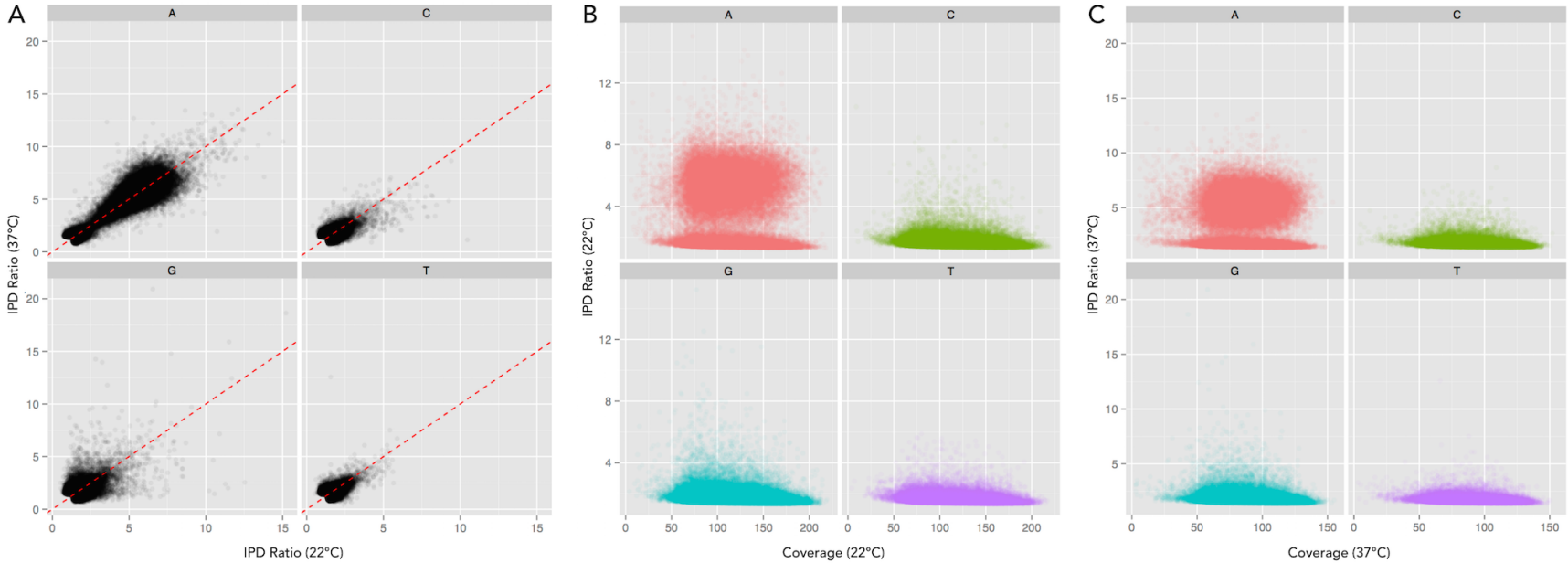

A
A
C
G
T
Coverage (37°C)
Coverage (22°C)
B
Score (37°C)
Score (22°C)
C
Score (22°C)
Coverage (22°C)
D
Score (37°C)
Coverage (37°C)

Figure 2.3. Interpulse duration (IPD) ratio provides an unbiased means of comparing samples with different coverage.

(A) Scatter plot comparison of the IPD ratios for all residues from each sample, 22°C on x-axis and 37°C on y-axis, showing no bias toward either sample (1:1 correlation in dotted red line). This permits an unbiased comparison of each sample, indiscriminate of coverage, thus providing a good metric for comparison of each sample. (B,C) Scatter plot of sequencing coverage and IPD ratio for all genomic positions at 22°C and 37°C respectively, separated for each base. Increased IPD ratio above baseline indicates DNA methylation. In each sample, many adenine (red) residues show an increased IPD ratio, revealing adenine methylation.


A
A
C
G
T
IPD Ratio (37°C)
IPD Ratio (22°C)
B
A
C
G
T
IPD Ratio (22°C)
Coverage (22°C)
C
A
C
G
T
IPD Ratio (37°C)
Coverage (37°C)

CHAPTER 3

LITERATURE REVIEW— Evolution of Citrate Utilization in *Escherichia coli*

History

How genetic information evolves to produce new phenotypes, new genes, and ultimately new species is a central question that lies at the root of basic biology. Due to the long time frames required for evolutionary changes to occur, much of evolution has remained historically inferential based on DNA and protein sequence comparisons, phenotypic similarities, and the fossil record. To observe 'real time' evolution, microorganisms such as bacteria, bacteriophages, and yeasts have been recently employed as an experimental system to empirically observe and quantify evolution in action. These microbiological experiments leverage rapid generation times, small size, and high population numbers, against much slower-growing larger multicellular organisms. Using bacteria, experimental organisms typically produce progeny in less than one hour, and a small flask may support a population of $>10^9$ organisms. Controlled conditions can be ensured using defined environments, aseptic technique, and employing isogenic strains. Because bacterial genomes are haploid any genetic changes rapidly manifest quantifiable phenotypic changes. Another major benefit of using microbial populations in evolution

experiments is their ease of preservation. Bacteria and yeast can be frozen and stored indefinitely for later analysis and comparison, providing in essence a “fossil record” that can be revived and assessed for evolutionary changes over time. Revived cultures can thus be compared to their ancestral parents throughout experimental time frames. With the advent of rapid and cost-effective whole genome sequencing methods, genetic changes can be identified at the nucleotide level.

The longest running evolution experiment was initiated on 24 February 1988 by Richard Lenski. It is now commonly known as the ‘LTEE', for Long-Term Evolution Experiment. The LTEE, as conceived, is remarkably simple in execution. 12 separate isogenic populations of *Escherichia coli* B are grown every day in minimal growth media. The environment is kept at uniformly stable conditions; 37°C with constant aeration. The LTEE medium contains a small amount of glucose (0.0025%) as a sole carbon source, enough glucose to support a 100-fold increase in population size or the equivalent of six or seven generations. Every day, 1% of the culture is transferred to a new flask containing fresh growth medium to allow for the growth of another six or seven generations. Each of the 12 cultures is preserved by freezing samples in glycerol every 500 generations. This experiment is still ongoing, and to date has produced over 65,000 generations. This experiment is an exceptional means of following and cataloging evolutionary changes under defined conditions.

Organisms in the LTEE reproduce in genetic isolation. The strain of *E. coli* B was selected because it does not harbor lysogenic bacteriophages, is not naturally

competent for DNA uptake, and lacks a fertility factor for mating (Lenski *et al.*, 1991). This experimental design, that eliminates lateral gene transfer, ensures that individual populations reproduce in isolation. Thus, any evolutionary changes providing fitness gains arise only by random spontaneous mutations.

The LTEE is designed as an open-ended experiment to compare the gained fitness of the *E. coli* B evolved strains compared to their ancestral parental strain. To facilitate these fitness comparisons, six of the 12 starting isogenic strains had a neutral mutation in arabinose operon added (Ara^-). The other six Ara^+ strains reduce tetrazolium on arabinose trypic soy agar (TA) indicator plates producing red colonies in contrast to the six Ara^- strains which remain white on these indicator plates. This phenotypic difference provides as easy visual screen when comparing numbers of colony forming units in competition analyses. As such, mean fitness is tested by introducing equal inocula of the evolved strain and its ancestral parent, with the reciprocal Ara phenotype, into the minimally defined glucose medium and allowing them to 'compete' over 24 h. A gain of fitness, reflected by generation time differences, is easily scored by comparing colony numbers of each strain (red vs. white) on the tetrazolium TA indicator plates.

Each of the 12 populations in the LTEE have shown considerable parallel evolution. As expected, each of the 12 populations increased in mean fitness under the experimental conditions (Lenski *et al.*, 1991; Lenski and Travisano, 1994). Cellular volume and surface area increased in all populations (Lenski and Travisano, 1994). Evolved populations showed a greater ability to import and metabolize glucose, allowing faster growth in the limited glucose environment compared to the

founding ancestral strain. Furthermore, many of the populations showed mutations in the same genes. Notably, 10 of the 12 developed point mutations in *topA*, a topoisomerase that increased DNA supercoiling (Crozat *et al.*, 2005). All 12 populations acquired a point mutation in *pykF*, a pyruvate kinase very rapidly within the first 2,000 generations (Woods *et al.*, 2006).

Divergent evolution has also been observed in the LTEE. Though many of the populations developed comparable mutations, these mutations typically did not occur at the same position in the gene in question. Several populations exhibited large chromosomal rearrangements such as inversions and deletions (Schneider and Lenski, 2004; Raeside *et al.*, 2014), and there was evidence of many insertions sequence (IS) elements moving throughout the genome (Papadopoulos *et al.*, 1999; Schneider and Lenski, 2004). Interestingly, several populations developed mutations in DNA repair mechanisms, thereby allowing greater genetic diversity (Sniegowski *et al.*, 1997; Barrick *et al.*, 2009) due to an ~1000 increase in mutation rate. Lenski *et al.* suggest this development of high mutation rates (*mut* strains) to be responsible for greater genetic diversity, thus giving natural selection increased opportunities to select for higher fitness clades in a constant experimental environment. The initial observed parallel evolution of the 12 experimental cultures was somewhat unexpected. In fact, Lenski's group first reported that these 12 populations, evolving in 'lockstep', appeared to negate Gould's evolutionary theory of contingency, the idea that if evolution were repeated, different outcomes would ensue. The single source for energy metabolism, glucose, easily obtained and metabolized, yielded similar phenotypic solutions to rapidly and efficiently procure

and utilize this finite energy source. However, serendipitously, the environment contained one additional resource yet to be tapped. The Davis-Mingolini (DM) minimal media recipe had historically contained citrate as an additional benign additive. Citrate (1.7 mM) was initially added to defined media by Bernard Davis because it augmented the bactericidal activity of penicillin (Davis, 1949). Use of citrate in defined *E. coli* minimal media has continued when it was later determined that citrate functioned as an iron chelator and enhanced *E. coli* iron uptake (Frost and Rosenberg, 1973). It is well known that *E. coli* cannot grow aerobically on citrate, but its use as a chelator is a standard component in minimal media recipes. Because *E. coli* can synthesize and metabolize citrate internally, the Lenski laboratory set up protocols to periodically check for external citrate utilization by extended incubation on DM plates [Blount dissertation].

Aerobic citrate utilization

In 2008, Blount *et al*. reported that one of the 12 *E. coli* LTEE populations had evolved the ability to grow on citrate aerobically (Blount *et al.*, 2008). This key innovation allowed this population to metabolize the citrate and proliferate to a significantly greater population density than the other 11 LTEE populations, still limited by growth solely on glucose. Acquisition of citrate utilization by this Ara-3 population emerged after 15 years of the LTEE, or ~31,000 generations. This observation was immediately hailed as evidence of of a novel metabolic evolutionary innovation. The NewScientist reported,

> *"A major evolutionary innovation has unfurled right in front of researchers' eyes. It's the first time evolution has been caught in the act of making such a rare and complex new trait." (https://www.newscientist.com/article/dn14094-bacteria-make-major-evolutionary-shift-in-the-lab/)*

Carl Zimmer, discussing the work in the *Scientific American* podcast *Science Talk*, added,

> "E. coli *couldn't eat citrate or at least they couldn't until something very odd happened in Lenski's lab. Citrate is in that broth that this scientist used to rear this E. coli, and one day Lenski and his students noticed something strange, which is basically that one of the flasks had gone very cloudy with a lot of bacteria; and when they realized what happened was, those* E. coli *were eating the citrate. They have determined that this is a genetic change that has happened, and so if you define* E. coli *as a species then what we may have here is the origin of [a] new species."* (http://www.scientificamerican.com/podcast/episode/everything-you-ever-wanted-to-know-08-10-09/)

Starting with a simple bacterium, Professor Lenski had observed the evolution of a complex phenotype within 15 years of starting his experiment.

Citrate Metabolism in *Escherichia coli*

The fact that *E. coli* cannot grow aerobically on citrate as the sole carbon source has long been a defining characteristic used to distinguish *E. coli* from other related enteric organisms within the *Enterbacteriacaea* family. Diagnostic laboratories ubiquitously use the IMViC series of diagnostic tests to identify clinical and environmental samples of enteric bacteria, especially within the coliform group. A coliform is defined as a Gram-negative, rod-shaped bacterium which produces acid and gas from lactose within 48 hours at 37°C. The IMViC series includes tests for indole production (I) from tryptophan, acid production from glucose metabolism indicated by the pH indicator methyl red (M), acetoin production (Vogus-Proskauer test) (Vi), and aerobic growth on Simmons citrate media (C). In laboratory jargon, with respect to the IMViC diagnostic panel, *E. coli* by definition is termed ++-- i.e., it is indole and methyl red positive and VP (acetoin) and citrate negative. In contrast, *Enterobacter aerogenes*, also lactose positive, is --++. This key diagnostic can rapidly discriminate samples that contain fecal contamination (*E. coli*) from non-fecal contamination (*Enterobacter*) and citrate utilization is a main component in this determination.

E. coli also possesses all the enzymes for a fully functioning tricarboxylic acid, a.k.a. the Kreb's or citric acid cycle (Lara and Stokes, 1952). Citrate is one of the chemicals utilized in this metabolic cycle by nearly all living organisms. Briefly, the cycle begins when acetyl Co-A derived from carbohydrate, lipid, and amino acid catabolism is condensed with oxaloacetate to form citrate using the enzyme citrate

lyase. A series of chemical transformations follows whereby two carboxylic acid groups are removed in the form of carbon dioxide along with one purine triphosphate (ATP or GTP) and three energy-rich electron carriers NADH to be used in oxidative phosphorylation. The four carbon product, succinate, generated by oxidation of citrate, is sequentially reduced back to oxaloacetate to keep the cycle moving.

Under anaerobic conditions, *E. coli* possess the ability to grow on citrate if another co-metabolite is supplied such as glucose, pyruvate or lactate (Vaughn *et al.*, 1950; Lara and Stokes, 1952; Lütgens and Gottschalk, 1980). The barrier preventing aerobic utilization of citrate is the inability to transport citrate across the cellular membrane into the cytoplasm. Anaerobically, this is mediated by CitT, a citrate/succinate antiporter (Pos *et al.*, 1998). Antiporters transport bidirectionally, so when citrate is imported into the cell, succinate is exported. Fermentation of citrate proceeds as citrate is cycled back through the TCA cycle, producing acetate and succinate, along with one ATP. It is noteworthy, then, that as CitT transports the usable carbon source into the cell, it also removes the metabolic byproduct of citrate fermentation. Le Chatelier's principle shows that systems, and in this case enzymes, do not proceed toward their end in the presence of their product. For the chemical transformations catabolizing citrate to continue to proceed, succinate must continually be removed from the system. Furthermore, citrate will not be imported, despite it's high concentration outside the cell, without the antiporter's necessary co-substrate. Therefore, in the case of citrate utilization, succinate must continually be produced by reduction of oxaloacetate to fumarate to succinate in

order to provide the necessary substrate to continue to import citrate into the cells. If citrate were the sole carbon source, only reducing power generated from acetate metabolism would be available resulting in a 'stalled' equilibrium with no net gain of energy or reducing power. It is for this reason anaerobic citrate use requires a co-substrate to supply the needed reducing power for growth.

Importantly, citrate cannot be used aerobically because *citT* is not expressed. However, several aerobic citrate-utilizing strains of *E. coli* have been isolated from livestock and sewage samples (Ishiguro *et al.*, 1978; Ishiguro *et al.*, 1979). All of these environmental strains harbor a plasmid-borne copy of a citrate transporter expressed aerobically. Furthermore, constitutive expression of the *citT* located on a high-copy plasmid sufficiently conferred a strong Cit^+ phenotype *E. coli* (Pos *et al.*, 1998). Barry Hall, in 1982, reported a spontaneous Cit^+ phenotype of *E. coli* emerging rapidly within 14 days. This mutant was recovered in a genetic selection to isolate an *E. coli* K12 strain capable of growing on phenylarabinose, the sole carbon source supplied in a minimal defined medium (Hall, 1982). Hall subsequently determined that the *E. coli* was actually growing on the citrate iron chelator, like the Lenski LTEE Cit^+ isolate, not the intended phenylarabinose substrate. Hall was able to map the Cit^+ mutations to the *cit* gene cluster on the *E. coli* chromosome using P1 transduction and he speculated this mutant had activated the anaerobic CitT transporter.

Nobel laureate Werner Arber, in the early 1990s, laid out three distinct possibilities for molecular changes at the genetic level that could facilitate evolution in bacteria: local sequence change, DNA rearrangements, and gene acquisition

(Arber, 1991; Arber, 1993). Local nucleotide substitutions include base insertions, substitutions, and deletions caused by errors during replication and effects from mutagenesis. DNA rearrangements occur frequently in bacterial genomes by homologous recombination of like sequences or by recombination of IS and transposon sequences. Such rearrangements provide additional mechanisms to delete, amplify, invert, or translocate large segments of the microbial genome. Gene acquisition via horizontal or lateral gene transfer via transformation of DNA, phage uptake, or conjugal mating is a mechanism whereby bacteria can incorporate a segment of DNA from an exogenous source, i.e. other organisms with unique genetic elements. The experimental setup of the LTEE provides a mechanism to observe the effects of all these molecular changes except for the latter. Therefore, the observed innovations in the LTEE should occur from local nucleotide changes or DNA rearrangements.

Several years after first reporting the isolation of the novel *E. coli* Cit$^+$ population in the LTEE, Lenski's lab published their genetic analysis in 2012, chronicling the genetic changes they had observed in this evolved strain (Blount *et al.*, 2012). In this analysis, Blount *et al*. (2012) reported that the Cit$^+$ strain showed amplification of the *citT* locus coupled with a promoter capture event of the upstream constitutive *rnk* promoter. This amplification and recombination event created the *rnk-citT* fusion allowing aerobic CitT expression. Replay experiments from frozen fossils have shown a recurring common theme: amplification of *citT* followed by a promoter capture of a neighboring gene or an acquisition of an insertion element upstream of *citT*, either of which allows aerobic expression. These

mutations account, in part, for the novel Cit$^+$ phenotype. A point mutation in *dctA* (dicarboxylic acid transporter for succinate, fumarate, and malate) at position -20 from the translational start is also a necessary mutation for efficient citrate utilization, as it allows recapture of succinate exported during citrate uptake (Quandt *et al.*, 2014).

Importantly, adaptation to aerobic citrate use follows three evolutionary stages as defined by Lenski: potentiation, actualization, and refinement (Blount *et al.*, 2012; Quandt *et al.*, 2014). Potentiation in their scenario involves neutral mutations that may not have a discernible phenotype but are necessary for future mutations leading to phenotypic change. Actualization involves a subsequent mutation(s) that generates the phenotype. Refinement is the further optimization of the phenotype once established in a population. The potentiating changes in the Cit$^+$ phenotype are not yet fully understood and are under active investigation by Lenski and others. The *citT* promoter capture represents actualization and the *dctA* mutation represents refinement. Recent work by Quandt et al. (Quandt *et al.*, 2015) suggest that a negative mutation in *gltA* (citrate lyase) was a potentiating mutation followed by a second restorative mutation in *gltA* occurring after the Cit$^+$ phenotype appeared. Taken together, the sequence of potentiating, actualization, and refinement mutations has led Lenski's group to postulate this rare innovative mutation was the result of 'historical contingency'. This theory suggests why the Cit$^+$ phenotype occurred only after 31,000 generations and why it has only been recovered in one of the 12 parallel cultures. In the following chapter we provide experiments that challenged this historical contingency argument. Instead we

hypothesized that this rare Cit^+ mutation was in actuality an artifact of the experimental design with regard to this mutation and that historical contingency was not a ‘true’ requirement.

References

Arber, W. (1991) Elements in microbial evolution. *Journal of Molecular Evolution* **33**: 4–12.

Arber, W. (1993) Evolution of prokaryotic genomes. *Gene* **135**: 49–56.

Barrick, J.E., Yu, D.S., Yoon, S.H., Jeong, H., Oh, T.K., Schneider, D., *et al.* (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* **461**: 1243–U74.

Blount, Z.D., Barrick, J.E., Davidson, C.J., and Lenski, R.E. (2012) Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* **489**: 513–518.

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

Crozat, E., Philippe, N., Lenski, R.E., Geiselmann, J., and Schneider, D. (2005) Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. *Genetics* **169**: 523–532.

Davis, B.D. (1949) The Isolation of Biochemically Deficient Mutants of Bacteria by Means of Penicillin. *Proc Natl Acad Sci USA* **35**: 1–10.

Frost, G.E., and Rosenberg, H. (1973) The inducible citrate-dependent iron transport system in *Escherichia coli* K12. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **330**: 90–101.

Hall, B.G. (1982) Chromosomal Mutation for Citrate Utilization by *Escherichia coli* K-12. *J Bacteriol* **151**: 269–273.

Ishiguro, N., Oka, C., and Sato, G. (1978) Isolation of Citrate-Positive Variants of *Escherichia coli* From Domestic Pigeons, Pigs, Cattle, and Horses. *Appl Environ Microbiol* **36**: 217–222.

Ishiguro, N., Oka, C., HANZAWA, Y., and Sato, G. (1979) Plasmids in *Escherichia coli* Controlling Citrate-Utilizing Ability. *Appl Environ Microbiol* **38**: 956–964.

Lara, F., and Stokes, J.L. (1952) Oxidation of Citrate by *Escherichia coli*. *J Bacteriol* **63**: 415–420.

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

Lenski, R.E., Rose, M.R., Simpson, S.C., and Tadler, S.C. (1991) Long-Term Experimental Evolution in *Escherichia coli*. I. Adaptation and Divergence During 2,000 Generations. *American Naturalist* **138**: 1315–1341.

Lütgens, M., and Gottschalk, G. (1980) Why a Co-Substrate Is Required for Anaerobic Growth of *Escherichia coli* on Citrate. *J Gen Microbiol* **119**: 63–70.

Papadopoulos, D., Schneider, D., Meier-Eiss, J., Arber, W., Lenski, R.E., and Blot, M. (1999) Genomic evolution during a 10,000-generation experiment with bacteria. *Proc Natl Acad Sci USA* **96**: 3807–3812.

Pos, K.M., Dimroth, P., and Bott, M. (1998) The *Escherichia coli* citrate carrier CitT: a member of a novel eubacterial transporter family related to the 2-oxoglutarate/malate translocator from spinach chloroplasts. *J Bacteriol* **180**: 4160–4165.

Quandt, E.M., Deatherage, D.E., Ellington, A.D., Georgiou, G., and Barrick, J.E. (2014) Recursive genomewide recombination and sequencing reveals a key refinement step in the evolution of a metabolic innovation in *Escherichia coli*. *Proc Natl Acad Sci USA* **111**: 2217–2222.

Quandt, E.M., Gollihar, J., Blount, Z.D., Ellington, A.D., Georgiou, G., and Barrick, J.E. (2015) Fine-tuning citrate synthase flux potentiates and refines metabolic innovation in the Lenski evolution experiment. *eLife* **4**: 2163.

Raeside, C., Gaffé, J., Deatherage, D.E., Tenaillon, O., Briska, A.M., Ptashkin, R.N., *et al.* (2014) Large chromosomal rearrangements during a long-term evolution experiment with *Escherichia coli*. *mBio* **5**: e01377–14.

Schneider, D., and Lenski, R.E. (2004) Dynamics of insertion sequence elements during experimental evolution of bacteria. *Research in Microbiology* **155**: 319–327.

Sniegowski, P.D., Gerrish, P.J., and Lenski, R.E. (1997) Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* **387**: 703–705.

Vaughn, R.H., Osborne, J.T., Wedding, G.T., Tabachnick, J., Beisel, C.G., and Braxton, T. (1950) The Utilization of Citrate by *Escherichia coli*. *J Bacteriol* **60**: 119–127.

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

CHAPTER 4

Rapid evolution of citrate utilization by *Escherichia coli* by direct selection requires *citT* and *dctA*

Work published in *Journal of Bacteriology*:

Van Hofwegen DJ, Hovde CJ, Minnich SA. (2016) Rapid Evolution of Citrate Utilization by *Escherichia coli* by Direct Selection Requires *citT* and *dctA*. *J Bacteriol* **198**:1022–1034.

Abstract

The isolation of aerobic citrate-utilizing *Escherichia coli* (Cit^+) in long term evolution experiments (LTEE) has been termed a rare, innovative, presumptive speciation event. We hypothesized that direct selection would rapidly yield the same class of *E. coli* Cit^+ mutants and follow the same genetic trajectory: potentiation, actualization, and refinement. This hypothesis was tested with wild-type *E. coli* B, K12, and three K12 derivatives: *E. coli ΔrpoS::kan* (impaired for stationary phase survival), *E. coli ΔcitT::kan* (deleted for the anaerobic citrate/succinate antiporter) and *E. coli ΔdctA::kan* (deleted for the aerobic succinate transporter). *E. coli* underwent adaptation to aerobic citrate metabolism that was readily and repeatedly achieved using minimal medium supplemented with

citrate (M9C), M9C with 0.005% glycerol, or M9C with 0.0025% glucose. Forty-six independent *E. coli* Cit$^+$ mutants were isolated from all *E. coli* derivatives except *E. coli ΔcitT::kan*. Potentiation/actualization mutations occurred within as few as 12 generations and refinement mutations occurred within 100 generations. Citrate utilization was confirmed using Simmons-, Christensen-, and LeMaster Richards citrate media and quantified by mass spectrometry. *E. coli* Cit$^+$ mutants grew in clumps and long incompletely divided chains, a phenotype that was reversible in rich media. Genomic DNA sequencing of four *E. coli* Cit$^+$ mutants revealed the required sequence of mutational events leading to a refined Cit$^+$ mutant. These events showed amplified *citT* and *dctA* loci followed by DNA rearrangements consistent with promotor capture events for *citT*. These mutations were equivalent to the amplification and promoter capture CitT-activating mutations identified in the LTEE.

Importance

E. coli cannot use citrate aerobically. Long term evolution experiments (LTEE) by Blount *et al.* (Blount *et al.*, 2012) found a single aerobic, citrate-utilizing *E. coli* after 33,000 generations (15 years). This is interpreted as a speciation event. Here we show why it probably was not a speciation event. Using similar media, 46 independent citrate utilizing mutants were isolated in as few as 12 to 100 generations. Genomic DNA sequencing revealed an amplification of the *citT* and *dctA* loci and DNA rearrangments to capture a promoter to express CitT,

aerobically. These are the same mutations identified by the LTEE, and we conclude that the rarity of this mutant was an artifact of the LTEE conditions, not a unique evolutionary event. No new genetic nformation (gene function) evolved.

Introduction

How genetic information evolves to generate new phenotypes/species is a central question in biology. Long term evolutionary experiments (LTEE) using microorganisms have been initiated by several groups, in part, to empirically observe this phenomenon (Adams and Rosenzweig 2014). LTEE using bacteria, bacteriophage, or yeast have distinct advantages that include high population numbers, rapid generation times, and the opportunity to freeze intermittent populations (frozen fossils) to track mutations over time. Coupled with whole genome sequencing, evolutionary changes can be genetically characterized to identify a mutation(s) required for a specific phenotypic change and frozen intermediates can be revived to replay and confirm the events. The most famous and meticulously documented LTEE are those, initiated in 1988, by Richard Lenski's laboratory (Lenski *et al.*, 1991). 12 parallel cultures of *Escherichia coli* REL606 (an *E. coli* B strain) have been growing aerobically in minimal salts medium with low glucose concentrations (0.0025%) for 27 years. Cultures are transferred daily into fresh medium. Frozen samples are preserved for each culture every 500 generations providing a tremendous resource to study long-term bacterial adaptation under controlled conditions.

To date, the Lenski LTEE cultures exceed 63,000 generations, equivalent to over one million years of human evolution. Importantly, these experiments negate Gould's theory of contingency, the idea that if evolution was replayed, different outcomes would arise (Pennisi, 2013). All 12 cultures are undergoing surprisingly similar genetic trajectories with one exception. After 15 years (33,000 generations), one of the 12 cultures increased in turbidity by utilizing citrate aerobically (Cit^+) (Blount *et al.*, 2008). The minimal medium in these experiments contains 1.7 mM citrate as a chelating agent. It is well known that wild-type *E. coli* cannot use citrate as a carbon source, aerobically, because it lacks a citrate transporter, but can use citrate under anaerobic conditions via expression of the CitT citrate/succinate antiporter (Koser, 1923) (Koser, 1924) (Vaughn *et al.*, 1950) (Lütgens and Gottschalk, 1980) (Garrity *et al.*, 2006). Genetic analysis of Lenski's Cit^+ strain show amplification of the *citT* locus, coupled with a promoter capture event of the upstream constitutive *rnk* promoter, allowing oxic CitT expression. Replay experiments from frozen fossils show a recurring common theme: amplification of *citT* followed by a promoter capture of a neighboring gene or an acquisition of an insertion element upstream of *citT*, either of which allow aerobic expression. In part, these mutations account for the novel Cit^+ phenotype. A point mutation in *dctA* (dicarboxylic acid transporter for succinate, fumarate, and malate) at -20 from the *dctA* translational start is also a necessary mutation for efficient citrate utilization as it allows recapture of succinate exported during citrate uptake (Quandt *et al.*, 2014). Importantly, adaptation to aerobic citrate use follows three evolutionary stages as defined by Lenski: potentiation, actualization, and refinement (Blount *et al.*, 2012;

Quandt *et al.*, 2014). Potentiation involves mutations that may not have a discernable phenotype, but prepare the cell for future mutations leading to phenotypic change. Actualization involves a subsequent mutation(s) that generates the phenotype. Refinement is the further optimization of the phenotype once established in a population. The potentiating changes in the Cit^+ phenotype are under active investigation by Lenski and others (Turner *et al.*, 2015) and not yet fully documented, but the *citT* promoter capture represents actualization and the *dctA* mutation represents refinement. Some authors attest this evolved *E. coli* Cit^+ is an exceedingly rare, innovative, gain-of-function mutation and argue for recognition of this *E. coli* variant as a newly evolved species (Lenski, 2011; Pennisi, 2013). This argument is made, in part, because citrate utilization is a key diagnostic to differentiate *E. coli* from other coliform species.

In this study, we challenged this interpretation that aerobic utilization of citrate by *E. coli* is necessarily a rare or innovative mutation warranting speciation. We hypothesized that the isolation of such mutants should be relatively easy, would follow the same genetic trajectory identified in the LTEE, and would utilize information present on the chromosome and not involve evolution of new information (novel gene function). As such, we predicted that the extremely long time required for *E. coli* to evolve to Cit^+ was due to the LTEE conditions and not potentiating genetic events requiring 33,000 generations. To test our hypothesis a direct selection was used. Our rationale was based on the serendipitous outcome of Hall's 1982 experiments to isolate *E. coli* K12 able to use phenyl-arabinose (Hall, 1982). After 14 days of aeration in minimal medium with phenyl-arabinose, cell

numbers increased. However, the cells were growing on the citrate chelator in the medium, like the Cit^+ from Lenski's LTEE, not the intended phenyl-arabinose substrate. Hall determines that at least two mutations are required for citrate utilization and, by transduction, maps the mutation to the citrate operon cluster of the *E. coli* chromosome. He also speculates that these mutations have activated the anaerobic citrate transporter. The adaptation was not documented by DNA analysis because genomic sequencing capabilities were not then available nor is it clear the Cit^+ phenotype had undergone refinement. In retrospect, Hall's experiment can be interpreted is an unintended 'direct selection' for Cit^+ mutants. The Lenski LTEE can be interpreted as an unintended genetic 'screen' for Cit^+ mutants. A direct selection is defined as conditions in which only the desired mutant can grow while a genetic screen is defined as conditions in which both the desired mutant and its parent can grow. We speculated that the difference in selective conditions could account for the days it took Hall versus the years it took Lenski to acquire the Cit^+ phenotype.

To test our hypothesis, we did direct and modified direct selection experiments. We used wild-type *E. coli* K12, three mutant derivatives of *E. coli* K12 as controls: *E. coli ΔrpoS::kan* (deleted for the σ^{38} stress response protein, impaired for stationary phase survival), *E. coli ΔcitT::kan* (deleted for the anaerobic citrate/succinate antiporter) and *E. coli ΔdctA::kan* (deleted for a di-carboxylic acid transporter), and two strains of *E. coli* B (B and REL606). Direct selections used three media variations to isolate Cit^+ mutants: (1) M9 minimal medium with citrate (6.8 mM) as the sole carbon source (M9C), (2) M9C amended with a low

concentration of glycerol (0.005%, M9CG50) containing either 6.8mM or 1.7mM citrate (M9LC50), and (3) M9C amended with a low concentration of glucose (0.0025%, M9C25) containing either 6.8mM or 1.7mM citrate (M9LC25). The latter medium, M9LC25, was equivalent to that used in the Lenski LTEE (Lenski *et al.*, 1991). Cultures were incubated aerobically and growth measured by turbidity and plate counts. Citrate utilization was detected using differential media and quantified by mass spectrometry of media filtrates during various stages of growth. Cit^+ mutants were phenotypically characterized by growth dynamics and microscopy; and genetically analyzed by transduction, and genomic and regional DNA sequencing. These experiments also provided the opportunity to compare the evolutionary trajectories of a direct selection versus a long term genetic screen.

Materials and Methods

Bacterial strains and growth conditions.

The bacterial strains used in this study are listed in Table 4.1. In order to use the *E. coli* K12 Keio mutant collection that provides kanamycin-marked mutations in every nonessential gene, systematic analyses of citrate utilization were done with this strain. The *rpoS*$^-$*, dctA*$^-$, and *citT*$^-$ controls were *E. coli* K12 kanamycin-marked mutations that allowed transduction experiments (see below). *E. coli* REL606 is the strain used in Lenski's LTEE. Stock cultures were maintained on Luria Bertani (LB) agar and grown in LB broth (Difco, MI). Direct selective citrate medium consisted of M9 mineral salts amended with 6.8 mM citrate and 50 μM thiamin (M9C). This 0.2%

citrate concentration (6.8 mM versus M9 chelator concentration of 1.7mM) is standard for a sole carbon source. It also has the added advantage of supporting more *E. coli* cell division than chelating concentrations so that initial mutants would be visually more apparent during selection. In addition, 6.8 mM citrate or higher is used in diagnostic media to differentiate *E. coli* from citrate-utilizing coliforms. Modified directive selective citrate media consisted of M9C amended with either 0.005% glycerol (M9CG50) or 0.0025% glucose (M9C25). To reproduce the LTEE condition these media were also used with the lower 1.7 mM citrate concentration (M9LC25 or M9LCG50, respectively). Colonies were isolated on LeMaster Richards minimal media supplemented with 6.8 mM citrate (Paliy and Gunasekera, 2007) and solidified with 1.5% Nobel agar (LRC), Simmons citrate (Becktin Dickenson, NJ) or Christensen citrate agars (Sigma, MO). These latter two media are selective differential media to diagnostically detect citrate utilization in enteric bacteria. In Simmons citrate agar, citrate is the sole carbon source (6.8 mM or 0.2%) and citrate-metabolizing organisms will cleave citrate to oxaloacetate and acetate. Oxaloacetate dehydrogenase then converts oxaloacetate to pyruvate and CO_2 and the CO_2 is converted to sodium carbonate resulting in an alkaline pH shift and conversion of the pH indicator, bromthymol blue, from green (neutral) to blue (alkaline). Wild-type *E. coli* does not grow on Simmons citrate agar or does so minimally, producing pinpoint yellow colonies (acidic) after one week incubation. Christensen citrate agar (0.3% citrate) also contains a low concentration of glucose (0.01%) and the pH indicator phenol red. *E. coli,* citrate negative, grows on the limited glucose to produce visible colonies in 18-24 h but are colorless to slightly

yellow (acidic). Citrate positive organisms will likewise grow on the glucose first and then convert to citrate metabolism, producing cerise (alkaline) colonies by the same reaction pathway described for Simmons citrate agar. All bacteria were stored at -80°C in LB or M9C containing 5% glycerol. Mutants were verified to be *E. coli* by comparing growth on Sorbitol MacConkey agar (Difco, MI) with 4-methyl-umbelliferyl-β-D-glucuronide (SMAC-MUG) (FENG and HARTMAN, 1982) or Eosin Methylene Blue (EMB) agar (Difco). Bacterial strains harboring kanamycin resistance (Kn^R) were grown where noted with 50 µg/ml kanamycin (Sigma, MO).

Selection for *E. coli* Cit^+ mutants.

Three protocols were used to select *E. coli* Cit^+ mutants from four *E. coli* K12 strains (wild-type, *ΔrpoS::kan, ΔcitT::kan, ΔdctA::kn*) and two *E. coli* B strains (B and REL606). In the first method, each *E. coli* K12 strain was grown overnight at 37°C with aeration in 5 ml of LB broth, after which cells were concentrated by centrifugation, and cell pellets were washed twice with M9 salts. Cells were re-suspended in 50 ml M9C broth in 250 ml Erlynmeyer flasks to a final concentration of 5.0-7.0 x 10^7 CFU/ml and incubated at 37°C with aeration at 160 rpm. Growth was visually monitored daily for increased turbidity and measured by absorption at OD_{600}. An aliquot of cells was removed weekly from each flask, quantified by triplicate plate counts on LB agar, and cryopreserved. During extended incubations, flasks were re-hydrated with sterile distilled water weekly to compensate for evaporation.

The second two methods were modified direct selections using 50 ml M9C in 250 ml Erlenmeyer flasks amended with either 0.005% glycerol (M9CG50) or 0.0025% glucose (M9C25). Both carbon concentrations supported ~six generations of growth (OD_{600} = 0.03). Flasks were incubated at 37°C with constant shaking for one week and then diluted (1:100) into fresh media. This cycle was repeated until culture turbidity at OD_{600} increased to >0.6. Aliquots were cryopreserved weekly. These experiments were also conducted with 1.7 mM citrate to determine if the concentration of citrate affected mutant selection. These low citrate media are designated M9LC25 (glucose) and M9LCG50 (glycerol).

In all variations of Cit^+ mutant selection, once the absorbance reached an OD_{600} >0.07 (1.7mM citrate-containing media) or >0.6 (6.8mM citrate-containing media), an aliquot of the culture was diluted (1:100) into 10mL of M9C medium and incubated with aeration until an absorbance (OD_{600} >0.07 or >0.6) was again observed. Individual colonies were isolated by streak dilution on LRC agar or diluted and plated on Simmons- and Christensen citrate agars. Large colonies arising after two to three days incubation on LRC agar were likewise tested on Simmons- and Christensen citrate agars. Blue colonies on Simmons citrate agar (strong citrate use) were re-purified by 12 h growth in LB broth and plated on fresh Simmons citrate agar. Large isolated colonies were grown in M9C broth for cryopreservation. All Cit^+ mutants were verified to be *E. coli* by patching isolated colonies sequentially onto SMAC-MUG, Simmons citrate agar, and Christensen citrate agars.

P1(*vir*) bacteriophage transduction experiments.

Transduction experiments were conducted with P1(*vir*) bacteriophage as described by Lennox (LENNOX, 1955) with a multiplicity of infection (m.o.i.) of 0.1:1. Donor cell lysates were preserved at 4°C over chloroform. Phage titers were quantified by plaque counts done in triplicate by 10-fold dilutions in four ml LB soft agar (0.4%) seeded with 100 µl of *E. coli* K12 wild-type and over-layed on LB agar plates or by dropping 10 µl of ten-fold dilutions on LB soft agar overlays seeded with 100 µl of *E. coli* K12.

Mass spectrometry of culture filtrates for metabolite identification.

To verify citrate utilization, putative Cit^+ isolates were inoculated into five ml of M9C broth. Cultures were incubated until exponential phase, and diluted (1:100) into 50 ml M9C in 250ml shake flasks. Growth was monitored in six h intervals by turbidity at OD_{600}. Additionally, one ml aliquots for each time point were centrifuged to remove cells, filter-sterilized using 0.2 µm filters, and a subset of these samples, correlating to the various stages of growth curves, analyzed by mass spectrometry. The concentration of citrate, succinate, acetate, fumarate, malate, glutamine, and indole were measured using reverse-phase chromatographic separation combined with multiple reaction monitoring methods for all of the analytes except acetate where a single ion reaction monitoring method was used. The liquid chromatographic separation was done using a Waters Acquity UPLC equipped with a Waters Acquity UPLC BEH C18 column with 1.7 µm particle size packing. The

column dimensions were 2.1 mm diameter by 50 mm in length and was maintained at 28°C. A binary solvent was used where Solvent A was 99.9% water with 0.1% formic acid and Solvent B was 99.9% acetonitrile with 0.1% formic acid. The solvent flow was 100 ul/min throughout the course of the separation. The solvent mixing program began with the ratio of 100% A/0% B. The concentration of solvent B was increased linearly to 3% over three min following injection. The gradient rate was then increased and the %B reached 80% at 14 min. The ratio of A and B were held constant for one min, the concentration of A was returned to 100% over the next 2.5 min, and the column was allowed to equilibrate at 100% A for 2.5 min before the next injection. A ten µl injection volume was used for all standards and samples. The column eluent flow was passed through the flow cell of an Acquity Photodiode Array Detector and then to the ESI sprayer of a Waters Xevo TQ MS triple quadrupole mass spectrometer. The mass spectrometry methods were programmed to acquire data in negative ion mode for all of the analytes except indole where the data was collected in positive ion mode. The negative ion mode capillary voltage was held at 2.5 kV and the positive ion mode capillary voltage was held at 3.4 kV. The desolvation gas flow was 400 L/hr, the desolvation temperature was 300ºC, and the source temperature was 150ºC. The organic acids eluted first and data was acquired in the first five min of the chromatographic run. At five min the method switched to positive ion mode and the data for indole was acquired within the next 10 min. The settings for precursor masses, product ion masses, cone voltages, collision voltages, and dwell times are listed in the Table 4.2. All standards and samples were analyzed in triplicate. The data was integrated and

analyzed using the TargetLynx quantitation software from Waters Corp. An external calibration was used and the standard data were fitted to quadratic curves. The samples, except the "media only" control, were brought to 1% formic acid by dilution of 99 µl of clarified culture media with 1 µl of formic acid prior to analysis. The "media only" sample was diluted 10-fold and brought to 1% formic acid before analysis.

Whole Genome sequencing.

Chromosomal DNA was prepared from *E. coli* Cit^+ mutants designated DV130, DV133, DV133T, and DV159 (Table 4.1) using the GenElute Bacterial Genomic DNA kit (Sigma, MO) or a bacterial genomic DNA extraction protocol (Ausubel *et al.*, 1987). Genomic DNA samples (15 µg) were sequenced using the Pacific Biosciences RS II sequencing platform at the Washington State University genomics core lab or the Icahn School of Medicine at Mount Sinai (NY) genomics core sequencing facilities. These sequences were deposited in the National Center for Biotechnology Information Sequence Read Archive. Sequences were assembled *de novo* via HGAP3 (PacBio) and compared to the genome of *E. coli* K12 MG1655.

Microscopic Analysis.

Digital micrographs of cells were taken using a Nikon Microphot FXA phase microscope at 400x magnification fitted with a Photometrics Cool SNAP cf camera.

Bacterial colonies were photographed using a Nikon dissecting microscope at 20x magnification with the same camera.

Results

Multiple independent *E. coli* K12 Cit$^+$ phenotypes were isolated by direct selection using M9C.

A direct selection in M9C broth containing 0.2 % citrate as the sole carbon source was performed. This experiment is analogous to that performed by Hall (Hall, 1982) in which *E. coli* is subjected to starvation conditions. Six independent *E. coli* Cit$^+$ isolates were recovered (Table 4.1 and Fig. 4.1). Experiments consisted of four flasks inoculated with *E. coli* wild-type, *E. coli ΔrpoS::kan*, *E. coli ΔcitT::kan,* or *E. coli ΔdctA::kan*. Bacterial growth was visually monitored daily for increased turbidity, quantified weekly by plate counts on LB agar, and aliquots were cryopreserved weekly. Two Cit$^+$ mutants arose from *E. coli* wild-type, three Cit$^+$ mutants arose from *E. coli ΔrpoS::kan,* and one arose from *E. coli ΔdctA::kan*. Representative growth curves from two experiments are shown in Fig. 4.1. During weeks of declining bacterial numbers, presumably subpopulations underwent potentiating mutations until the Cit$^+$ phenotype was actualized as evidenced by exponential growth. All Cit$^+$ *E. coli* were recovered within 40 days of the beginning of the selection incubations, with one exception from *E. coli ΔdctA::kan* that arose

after 134 days (Table 4.1). Recovery of *E. coli* Cit$^+$ did not occur among any of the *E. coli ΔcitT::kan* (Table 4.1).

Because of the potential for contamination in the extended incubation of these cultures, all Cit$^+$ isolates were meticulously confirmed to be *E. coli* by scoring 100 colonies from LB agar plates for specific *E. coli* phenotypes on SMAC-MUG agar for the *E. coli* wild-type strains, and SMAC-MUG Kn agar for the *E. coli ΔrpoS::kan* strain. All hundred colonies at each time point were sorbitol and MUG positive (fluoresced under UV light), reactions confirmatory for *E. coli*, and the *E. coli ΔrpoS::kan* strain was positive for kanamycin resistance (data not shown). This confirmed these cells matched the phenotype of the starting culture. Kanamycin was not used in the medium during the Cit$^+$ selection experiments. *E. coli* Cit$^+$ from direct selection flasks were plated on LRC agar and a single large colony from each experiment was used for further characterization.

Multiple independent *E. coli* K12, *E. coli* B, and *E. coli* REL606 Cit$^+$ phenotypes were isolated using a modified direct selection in M9C medium supplemented with glycerol or glucose.

The direct selection of Cit$^+$ mutants described above was repeated with the four *E. coli* K12 test strains using M9C medium supplemented with either 0.005% glycerol or 0.0025% glucose, equivalent low-level carbon concentrations used in the LTEE by Lenski (Lenski *et al.*, 1991). The rationale for these experiments was to provide a limiting usable carbon source to support ~six generations of growth and

thereby increase the likelihood of mutations among replicating cells, in contrast to the limited cell division in the 'starvation' direct selection described above. Glycerol was used to determine if catabolite repression or growth rate affected selection of Cit^+ mutants. Fourteen *E. coli* Cit^+ isolates (Table 4.1) were recovered from these experiments: five out of six from glycerol containing flasks with *E. coli* K12 wild-type or *E. coli ΔrpoS::kan,* and nine out of nine from glucose containing flasks with *E. coli* K12 wild-type. After inoculation (9×10^5 to 7×10^6 CFU/ml), the turbidity of the cultures rose to an OD_{600} of 0.02 to 0.03 overnight, or the equivalent of ~six generations based on plate count (data not shown). These cultures were maintained for seven days in stationary phase after which cells were inoculated into fresh media at a 1:100 dilution, similar to Lenski's LTEE but with a purposed extended time in stationary phase to enrich for the desired Cit^+ phenotype. Representative experiments are shown in Fig. 4.2, panels A and B. Following weekly transfers into fresh media, subpopulations presumably underwent potentiating mutations until the *E. coli* Cit^+ phenotype was actualized as evidenced by exponential growth of *E. coli* wild-type and *E. coli ΔrpoS::kan.* All but one *E. coli* Cit^+ were recovered within two to five weekly transfers, representing 12 to 30 generations. In general, adaptation to Cit^+ was slower among cells in M9C supplemented with glucose than cells in M9C supplemented with glycerol (first mutants obtained after the fifth versus second transfer; Table 4.1). Within the 63 days, recovery of *E. coli* Cit^+ did not occur among any of the flasks containing *E. coli ΔcitT::kan* or *E. coli ΔdctA::kan*. We concluded that direct selection of *E. coli* K12 Cit^+ readily occurred when M9C medium was supplemented with low levels of glycerol or glucose, similar to direct selection in

M9C medium. Furthermore, lack of citrate-dependent growth in either the *E. coli ΔcitT::kan* or *E. coli ΔdctA::kan* experiments showed both of these genes were necessary for adaptation to citrate utilization within this timeframe.

To show isolation of potentiated/actualized Cit$^+$ mutants were strain- and citrate concentration-independent, experiments were done using *E. coli* K12, *E. coli* B, and *E. coli* REL606 in media with 1.7mM citrate. Over the nine-week course, 15 of 16 *E. coli* K12; four of 12 *E. coli* B, and six of 12 *E. coli* REL606 containing flasks gave rise to Cit$^+$ mutants (Table 4.1). Representative experiments are shown in Fig. 4.2, panels C and D. As expected, reducing the citrate level by one fourth reduced *E. coli* K12 maximum OD_{600} by the same factor. Although *E. coli* B and *E. coli* REL606 followed the same pattern, the maximum OD_{600} attained was 0.08. We did not carry *E. coli* B or *E. coli* REL606 to refinement because they are both defective in growth on succinate (data not shown), which also explains the lower level of growth. However, we predict that refined mutants would be obtained with longer selections, as shown with Cit$^+$ strain SO191, derived from the phenotypically equivalent *E. coli K12 ΔdctA::kan* after 4 months. We concluded from these experiments that isolation of Cit$^+$ mutants were rapidly and readily attained regardless of citrate concentration and strain lineage.

All independently isolated *E. coli* Cit$^+$ phenotypes had long lag phases before re-initiation of growth in M9C.

Cit$^+$ cells from all three selection protocols displayed a long lag phase when transferred (1:100 dilution) from turbid primary selection media into fresh M9C broth. This lag phase varied from four to seven days. Additionally, when cells were plated on Simmons citrate agar, small yellow colonies only appeared after 2-3 days. Colonies plated on Christensen citrate agar had a weak, delayed cerise reaction after 24-36 h compared to the unselected *E. coli* strains that all appear as white colonies. Cells from secondary turbid M9C broth cultures were streak-purified on LRC agar and consistently resulted in several larger colonies amidst a background of pinpoint colonies, after 2-3 days of incubation. These large colony *E. coli* Cit$^+$ derivatives displayed more rapid growth in M9C broth and this shorter lag phase (24-36 h) was stable on subsequent transfers into fresh medium (Fig. 4.4). Importantly, when these large colony *E. coli* Cit$^+$ isolates were grown on Simmons citrate agar, blue colonies, indicative of strong citrate metabolism, appeared within 36-48 h. When grown on Christensen citrate agar a strong positive cerise reaction was observed within 12 h. Collectively, these strong positive reactions on citrate differential media and shortened lag phases indicated refinement of the *E. coli* Cit$^+$ phenotype had occurred.

The long lag phase before growth re-initiation of primary *E. coli* Cit$^+$ isolates could be shortened to 12 to 18 h by using filter-sterilized M9C spent media from turbid *E. coli* Cit$^+$ cultures, re-adjusted for M9 salts and citrate. Serial dilutions of spent media showed a growth-promoting dose-dependent response (data not

shown). This suggested quorum sensing could be involved. Although auto-inducer 2 (AI-2) was detected in these spent media supernates, as assayed with *Vibrio harveyi* test strains (Greenberg *et al.*, 1979) (Bassler *et al.*, 1997), addition of purified AI-2 to these cultures had no growth promoting effect and we did not pursue this further (data not shown).

Surprisingly, Cit^+ isolates were easily obtained using the *E. coli ΔrpoS::kan* strain. To show that lack of RpoS or other mutations present in this strain (Table 4.1) was not a contributing factor for citrate utilization we took strain DV133 and made a P1(*vir*) bacteriophage lysate on this isolate. This lysate was used to transduce *E. coli* K12 wild-type at an m.o.i. of 0.1:1. Transduced cells were diluted in 50 ml M9C broth and incubated at 37°C with aeration. Because citrate binds free calcium and therefore prevents P1(*vir*) phage infection (Wall and Harriman, 1974), concern for secondary infection of transductants by transducing or lytic phage in this mixture was alleviated. After four days, the transduced culture displayed turbid growth with an OD_{600} = 0.7 and a sample from this flask was streaked directly onto Simmons citrate agar. Blue colonies appeared after two days. These colonies were sorbitol, MUG, and lactose positive and Kn^S indicating these transductants were not contaminated by the DV133 donor (lactose negative on EMB agar and Kn^R). This *E. coli* K12 wild-type Cit^+ transductant was designated strain DV133T (Table 4.1).

We also identified a method of isolating refined *E. coli* Cit^+ on solid media. Actualized *E. coli* Cit^+ (long lag-phase, pinpoint yellow colonies on Simmons citrate agar) were incubated with maintained humidity for several weeks. During this extended incubation small blue papillae arose within these yellow colonies (Fig.

4.3). Using a dissecting scope for viewing colonies these papillae were picked and streaked onto fresh Simmons citrate agar. Large blue colonies appeared in two to three days and were separated from background yellow pinpoint colonies. Different papillae within the same colony or from different colonies represented independent transitions from actualized Cit$^+$ to the refined stage of citrate metabolism. This transition was not observed with un-potentiated *E. coli* K12 wild-type or *E. coli ΔrpoS::kan* cultures (data not shown). Using this technique additional *E. coli* Cit$^+$ refined phenotypes were isolated.

E. coli Cit$^+$ phenotype growth correlated with citrate depletion in M9C broth.

Growth curves were conducted with two Cit$^+$ refined isolates, DV133 and DV159. Filtered supernates obtained from these growth curve experiments were subjected to mass spectrometry at timed intervals to detect citrate and other metabolite concentrations. As shown in Fig. 4.4, bacterial growth correlated with depletion of citrate in the culture medium. Low levels of succinate were detected throughout these growth experiments and acetate levels peaked in late exponential phase and dropped to undetectable levels in the stationary phases (data not shown). We did not detect any of the other organic acids tested. Although Kovac's reagent gave a weak positive indole reaction when added to stationary phase culture samples, we did not detect this analyte by mass spectrometry.

E. coli Cit$^+$ phenotypes from M9C broth were clumped or incompletely divided and colonies derived directly from the broth were not clonal.

Microscopic examination of *E. coli* Cit$^+$ cells from M9C medium showed predominantly elongated pre-divisional cells, often in clumps or 'rafts', estimated to contain 20-100 cells (Fig. 4.5, panels A and C). Also, large blue colonies on Simmons Citrate agar from this broth were not clonal due to the deposition of these clumped cells (data not shown). However, this clumping phenotype was reversible if cells were grown in LB broth (Fig. 4.5, panels B and D). Therefore, to obtain stable, clonal, large blue colonies on Simmons citrate agar, presumptive Cit$^+$ isolates required cycling growth through LB broth to mid-exponential phase, to reverse clumping and assure clonal isolation. We found this purification step essential in the preparation of DNA for genomic sequencing.

Genetic and genomic DNA analyses of _E. coli_ Cit$^+$ phenotypes showed amplification of _citT_ followed by promoter capture and _dctA_ amplification.

Genomic DNA was prepared from four *E. coli* Cit$^+$ isolates DV130, DV133, DV133T, and DV159 (Table 4.1) and was sequenced using PacBio RS II sequencing technology. Accession numbers are in Table 4.1. *E. coli* DV159 is an independent wild-type Cit$^+$ isolate from an M9C direct selection. The other three strains represent staged isolates to Cit$^+$ refinement and were used to identify genetic changes at each stage. *E. coli* DV130 (*ΔrpoS::kan* background) displayed weak Cit$^+$ activity evidenced by a long lag phase when placed in M9C and yellow colonies

when grown on Simmons citrate agar. *E. coli* DV133 was derived from strain DV130 as a large colony isolate from LRC agar. This refined Cit$^+$ strain had a shorter lag phase in M9C compared to its DV130 parental strain and yielded blue colonies on Simmons citrate agar. Strain DV133T is the *E. coli* K12 wild-type that was transduced with P1(*vir*) grown on strain DV133 and selected for Cit$^+$ in M9C broth. Like strain DV133, the DV133T transductant showed a strong Cit$^+$ phenotype upon its initial isolation indicating the essential mutation for strong Cit$^+$ utilization had been transduced. We reasoned the genomic sequence of this strain in the wild-type background would facilitate mutation(s) identification.

The read-depth profile of strain DV130's genome showed four-fold coverage of *citT* and two-fold coverage of *dctA* above the normalized coverage of 100X for the rest of the chromosome, suggesting these regions had undergone amplification (Fig. 4.6A). However, duplications in these regions were polished out by the software in the final assembly process. Analysis of individual 'sub-reads' for *citT* showed tandem duplications (data not shown). The compiled genomic DNA sequence also showed a single base pair deletion in *dctA,* 26 nt from the start of translation, five nt from the mutation identified by Quandt *et al. (*2014) required for the refined Cit$^+$ phenotype (data not shown). The read-depth profile of strain DV133 likewise shows increased read depths for both *citT* and *dctA,* both at two-fold coverage suggesting a partial reduction in *citT* amplification compared to the DV130 parental strain (Fig. 4.6A). Importantly, the DNA sequence of *citT* shows *insI*1, an IS30 transposase, has inserted 5' to the *citT* gene, a presumptive promoter capture (Fig. 4.6B). This insertion is not present in any of the DV130 sub-reads

indicating this insertion event correlated with the switch between the potentiated Cit$^+$ weak phenotype and the strong or refined Cit$^+$ phenotype in strain DV133. This same *insI*1 is present in DV133T (Fig. 4.6B) indicating this region was transduced into the *E. coli* wild-type and confirmed its role for the strong or refined Cit$^+$ phenotype. DV133T also shows an amplified *dctA* region of 140 kb generated by a recombination between *rhsA* and *rhsB*, genes that share homology and bracket *dctA* at 79 min on the *E. coli* chromosome (Fig. 4.6C). There are no *dctA* point mutations in the gene or regulatory sequence.

To further verify that both *citT* and *dctA* had undergone gene amplification, we transduced DV133T with P1(*vir*) grown on strains JW0604-1 (*ΔcitT::kan*) and JW3496 (*ΔdctA::kan*). DV133T transduced with phage grown on strain JW0604-1 were plated on Christensen citrate Kn agar and were scored for Cit$^+$. We found six of 20 colonies were Kn^R and Cit$^+$. DV133T transduced with phage grown on JW3496 showed three of ten colonies were Kn^R and Cit$^+$. These results are only compatible with a gene duplication for both *citT* and *dctA* (Anderson and Roth, 1981) (Sonti and Roth, 1989).

We concluded that potentiation/actualization of Cit$^+$ first requires amplification of *citT* as shown in strain DV130. This would potentiate access to citrate by a gene dosage effect. Refinement requires a promoter-capture that leads to higher aerobic expression of *citT,* as shown for DV130-derviative strains DV133 and DV133T*,* and amplification of *dctA* by recombination at a well-documented 'hot-spot' for chromosome duplications (Lin *et al.*, 1984) (Petes and Hill, 1988).

Sequence analysis showed strain DV159 *citT* has undergone a duplication and deletion to generate a fusion of *citT* with the neighboring *uspG* gene (Fig. 4.6B). The *usp*G gene is stress-induced and expressed at extremely high levels during carbon starvation (Bochkareva *et al.*, 2002) (Kvint *et al.*, 2003). Sequence analysis also showed a similar amplification of a large 140kb region of the chromosome containing *dctA* consistent with recombination between *rhsA* and *rhsB*, similarly as determined in strain DV133T (Fig. 4.6C). Point mutation(s) in *dctA* and its regulatory region were not detected in DV159. The genomic rearrangement identified in strain DV159 is analogous to the original Cit$^+$ strain identified in the LTEE (Blount *et al.*, 2012).

Discussion

The most important finding of this work was that *E. coli* underwent rapid adaptation to aerobic citrate metabolism that was readily and repeatedly achieved using direct or modified direct selections. The genetic trajectory of this adaptation and the classes of mutations identified followed the same pattern of genetic events characterized in LTEE and centered on *citT* and *dctA* expression. Importantly, potentiated/actualized *E. coli* Cit$^+$ were obtained in as few as 12 generations and refined phenotypes in less than 100 generations. Phenotypic and genetic analyses of these *E. coli* Cit$^+$ provided insight into the mechanism for the adaptations and suggested why the LTEE took 33,000 generations to reach this phenotype. Also, this study provided a unique opportunity to compare the results of a direct selection

with a long-term genetic screen. Finally, because this adaptation did not generate any new genetic information and only required expanded expressions of two existing transporters (*citT* and *dctA*), generation of *E. coli* Cit$^+$ phenotypes in our estimation do not warrant consideration as a speciation event. In fact, mutations in these two loci are sufficient for the LTEE Cit$^+$ phenotype (Quandt *et al.*, 2014).

We obtained 46 independent *E. coli* Cit$^+$ phenotypes that followed a similar staged trajectory from potentiation/actualization to refinement as defined by Lenski (Blount *et al.*, 2012) (Quandt *et al.*, 2014). Emerging *E. coli* Cit$^+$ displayed long lag phases and weak utilization of citrate, analogous to the LTEE mutants. Potentiation, in the LTEE, is undescribed due to the 33,000 generations of mutations that need to be analyzed. Because our *E. coli* Cit$^+$ were recovered after a minimum of 12 generations, potentiation only requires *citT* amplification, as determined for strain DV130. Increased expression of the CitT transporter allowed minimal but sufficient access to citrate and a low level of cell division. Gene amplifications are the most common mutations identified (Reams *et al.*, 2012) and set the stage for subsequent promoter capture recombinatory events. Thus, Cit$^+$ cells amplified for *citT*, such as weakly Cit$^+$ DV130, transitioned by *citT* promoter capture and *dctA* amplification to its strongly Cit$^+$ derivative, DV133. This combination was confirmed to be the only required changes by transducing this mutation into *E. coli* K12 wild-type to create DV133T.

A simple model emerges that explains Cit$^+$ mutants derived by either direct selection or by LTEE. Cells are potentiated by *citT* amplification (>4X) and actualized by subsequent chromosome remodeling to capture a promoter that

allows aerobic *citT* expression. The copy number of *citT* required for growth before this capture is high, but afterwards is decreased (2X) to reduce the fitness cost of maintaining numerous *citT* gene copies. Finally, Cit^+ refinement results when the DctA transporter is amplified (2X) to import succinate. The same mechanism, gene amplification, promoter capture, and subsequent reduction in gene copy number, is identified by Roth to explain the recovery of cryptic β-galactosidase expression by *E. coli* in the famous Cairn's directed-mutagenesis experiments (Cairns *et al.*, 1988) (Reams and Roth, 2015). This mechanism also explains similar adaptation events in *Salmonella* (Kugelberg *et al.*, 2006) (Kugelberg *et al.*, 2010).

We do not, as yet, fully understand the dynamics of delayed onset of growth by Cit^+ isolates after initial selection. Because spent media from *E. coli* Cit^+ isolates over-rode this lag phase in a dose-dependent manner, quorum sensing may play a subtle role in this process. However, a more likely explanation is the loss of C4-dicarboxylates (succinate) that accompanies citrate uptake by CitT. Potentiated *E. coli* would be growing on the equivalent of a two-carbon substrate. Delayed growth may reflect a needed threshold level of succinate in the medium before citrate can be efficiently metabolized. Also, since *E. coli* K12 can grow on succinate, this molecule may support emergence of Cit^+-dependent Cit^--competing phenotypes in the population as they scavenge this metabolite produced from Cit^+ cells. Cit^+ cells grow in M9C as large 'rafts' of cell complexes. We speculate these 'rafts' include succinate-scavenging Cit^- cells because blue colonies isolated from M9C broth on Simmons citrate agar yielded varied phenotypes when re-purified. For this reason, we found it necessary to cycle initial Cit^+ isolates through LB broth, to disperse the

'rafts' and recover clonal isolates. In summary, it appeared loss of succinate is a growth-limiting step at this adaptive stage.

Refinement of the *E. coli* Cit$^+$ phenotypes in our experiments required the *dctA* C4-dicarboxylate (succinate) transporter. In all experiments, the *E. coli ΔdctA::kan* strains did not produce a Cit$^+$ phenotype with one exception, strain SO191. All genomic sequences of refined *E. coli* Cit$^+$ isolates had amplification of *dctA* or a mutation in its regulatory region. In *E. coli*, expression of *dctA* is inhibited by glucose (catabolite repression) and is normally not expressed until stationary phase (Davies *et al.*, 1999). DctA regulates its own expression and the deletion mutation, identified in a DV130 population, centers between the -10 promoter sequence and the translational ATG start site. This may de-repress its expression like the *dctA* point mutation identified in the LTEE (Quandt *et al.*, 2014). Expression of *dctA* is also dependent on the two-component regulatory DcuSR system (Witan, Monzel, *et al.*, 2012; Witan, Bauer, *et al.*, 2012). DctA directly interacts with DcuSR and forms a functional tripartite transporter/sensor that localizes at the cell poles and pre-division sites (Scheu *et al.*, 2014). DctA polar localization does not occur in the absence *dcuS* expression or in a *dcuS* deletion. *E. coli* possess three additional succinate transporters DcuA, DcuB, and DcuC, which are also under DcuSR regulation (Zientz *et al.*, 1996; Janausch *et al.*, 2001). These three transporters are expressed under anaerobic conditions like *citT*. We predict recapture of succinate may occur via activation of any of these anaerobic succinate transporters and mutations in their regulation provide three more potential routes to refinement of a Cit$^+$ phenotype. Such mutants may be a subset among our uncharacterized Cit$^+$

isolates such as strain SO191. This latter strain was the only *E. coli ΔdctA::kan* that became Cit^+, occurring remarkably after 134 days in stationary phase. *E. coli* K12 *ΔdctA::kan* cannot utilize succinate under aerobic conditions. Preliminary characterization of the Cit^+ SO191 showed retained Kn^R and the ability to grow on succinate, consistent with aerobic expression of another dicarboxylic acid transporter.

One unanticipated result was the repeated successful isolation of *E. coli ΔrpoS::kan* Cit^+ mutants. This strain was incorporated into our experiments as an intended negative control. We assumed the RpoS stress response would be required for stress-induced mutations in starving cells. RpoS-dependent stress-induced point mutations and gene amplification occur by repair of double-stranded breaks and double-stranded ends. The former of these occurs by the RpoS-controlled error-prone polymerase DinB causing nucleotide insertions, deletions and base-substitutions, and the latter by double-stranded end-initiated replication (Hersh *et al.*, 2004). RpoS-mediated stress response point mutations in chromosomes of starving *E. coli* have been reported to account for half the base substitutions and frameshift mutations and thus, RpoS is considered to be essential for adaptation under starvation (Galhardo *et al.*, 2007). However, our results show that only gene amplification and recombination were required for generation of Cit^+ mutants.

Our results explain why the LTEE led to a single Cit^+ isolate. By design, the LTEE are open-ended experiments to track the evolution of 12 parallel *E. coli* cultures under controlled conditions. Aerobic citrate utilization was not a specified

aim. The highly delayed evolution (33,000 generations) of Cit$^+$, understandably gives the impression that this is a rare, innovative evolutionary event. Our studies highlight why it is not. The LTEE experimental design and the use of *E.coli* REL606 account for the delayed emergence of a Cit+ isolate. First, potentiated and actualized cells were routinely diluted away by the LTEE daily 1:100 dilution transfers. Cells with amplified *citT* had only a brief advantage to use citrate once glucose was expended and were likely lost with each daily dilution/transfer into fresh glucose-containing media due to the fitness cost of multiple gene copies. This dynamic is described for *Salmonella* gene amplifications (Reams *et al.*, 2010). Gene amplification likely came to a steady state and did not reach a level that permitted the next step, the rarer promoter fusion event. Second, if actualized Cit$^+$ cells did arise, they have a long lag phase and would have difficulty out-competing the well-adapted *E. coli* specialized for growth in the LTEE low glucose concentrations. Our more rapid recovery of Cit$^+$ mutants in glycerol compared to glucose, support the idea that slower cell division reduces the fitness cost of gene amplifications. Again, daily transfers to fresh glucose specifically selects against slowly dividing mutants. Third, refined Cit$^+$ mutants depend on DctA to re-capture succinate. Normal expression of this transporter is repressed by glucose (catabolite repression), shows auto-repression, and requires DcuRS, all conditions reached in stationary phase. Thus, access to succinate occurred only briefly between LTEE transfers. Fourth, the *E. coli* strain used, REL606, has a *dcuS* five base-pair deletion that prevents *dctA* expression (Yoon *et al.*, 2012; Quandt *et al.*, 2014; Turner *et al.*, 2015). This strain is defective in the very pathway required for competitive citrate

utilization. This defect also explains why our short-term direct selection yield actualized and not refined *E. coli* REL606 Cit$^+$ cells. Thus, before citrate can be metabolized efficiently by this strain, an extra mutation to suppress this defect was required. Nonetheless, we predict that Cit$^+$ mutations, even in REL606, may have occurred sooner and repeatedly in the LTEE protocol, if the *dcuS* in this strain was repaired because then, only gene amplifications (*citT* and *dctA*) would be required, as we found.

Three of our results support this reasoning about the *dcuS* defect: i) we rapidly isolated Cit$^+$ mutants in *E. coli* REL606 and another *E. coli* B strain, ii) our characterized *E. coli* K12 Cit$^+$ mutants only required *dctA* duplication, a more common process than a *dctA*-specific point mutation required to suppress a *dcuS* defect, iii) we isolated a Cit$^+$ mutant in *E. coli* K12 *ΔdctA::kan*, a strain phenotypically equivalent to the REL606 *dcuS* mutation (DctA-negative) in 134 days. The pathway to Cit$^+$ with a defect in *dctA* expression does not require 33,000 generations of potentiating mutations. Interestingly, the time required for the LTEE Cit$^+$ mutation to occur is predictable. For both amplification/promoter capture (*citT*) and a point mutation (*dctA*) to occur in the same cell under non-selective condition is on the order of one in 10^{14}. This is almost exactly the number of cells Lenski screened to find Cit$^+$ and the frequency Hall predicted for his Cit$^+$ mutation (Hall, 1982; Pennisi, 2013).

Our experiments also presented a unique opportunity to compare the results of a direct selection to a long-term genetic screen. It is proposed that a strong (direct) selection might drive an evolving population to a “quick fix rendering a

better solution less accessible" (Barrick and Lenski, 2013). Our results suggest the converse is true. The *E. coli* Cit$^+$ genomes sequenced from our direct selection experiments are equivalent to the amplification and promoter capture mutations identified in the LTEE (Blount *et al.*, 2012). The requirement for *dctA* activation is also common to both experiments. We conclude there is no significant difference in the *E. coli* solutions to aerobic citrate use between these experiments. Isolation of a Cit$^+$ mutant in a *dctA*-deletion strain is indicative that a direct selection led to an additional solution that has, as of now, not been identified in the LTEE. Because *E. coli* REL606 Cit$^+$ has only been identified in one of the 12 parallel cultures of the LTEE, it may well be that a weak selection inhibits access to a 'citrate-use solution'. Frozen fossils from the 12 LTEE cultures could be used to test this possibility. They could be revived and the experiments replayed from any generational landmark with longer incubations extended into deeper stationary phase, the selection conditions described by us, to determine if the pathway to citrate use has been preserved or lost among all 12 *E. coli* cultures.

In summary, *E. coli* can rapidly mutate to a Cit$^+$ phenotype in a relatively short timeframe if put under direct selection. This indicates that the 33,000 generations to potentiate the evolutionary resources for the Cit$^+$ phenotype is not a direct requirement but merely reflects experimental conditions. As such, Cit$^+$ mutants exemplify the adaptation capability of microorganisms but as of yet, the LTEE has not substantiated evolution in the broader sense by generation of new genetic information, i.e. a gene with a new function. Interestingly, our findings parallel the conclusions of bacterial starvation studies by Zinsser and Kolter (Zinser,

2004) in which *E. coli* adaptations are dominated by changes in the regulation of pre-existing gene activities rather than the generation of new gene activities, de novo. The LTEE isolation of Cit^+ mutants has become a textbook example of the power of long term evolution to generate new species. But based on our results, *E. coli* arrives at the same solution to access citrate in days vs. years, as originally shown by Hall (Hall, 1982). In either case, genes involved in the process maintain their same function but show expanded expression by deregulation. Because of this, we argue that this is not speciation any more than any other regulatory mutant of *E. coli.* A more accurate, albeit controversial, interpretation of the LTEE is that *E. coli*'s capacity to evolve is more limited than currently assumed.

Acknowledgements

We thank Derek Pouchnik and Mark Wildung (Washington State University genomics core laboratory), Robert Sebra (Icahn School of Medicine core genomics laboratory) for providing sequencing services, Jonas Korlach (Pacific Biosciences) for assistance in genomic DNA sequence analysis and identification of gene duplications, and Dr. Lee Deobald (University of Idaho) for mass spectrometry assistance. We also thank Dr. Patricia L. Hartzell (University of Idaho) and Dr. Patrick J. Hrdlicka (University of Idaho) for valuable discussions during this investigation.

Funding Information: This work was supported by the University of Idaho Agricultural Experiment Station Hatch projects under award numbers IDA01467 (CJH) and IDA01406 (SAM), and the National Institute of Health, under award number P20GM103408 (CJH, SAM). The funders had no role in study design, data collection, interpretation, or the decision to submit the work for publication.

References

Adams J, Rosenzweig F. (2014) Experimental microbial evolution: history and conceptual underpinnings. *Genomics* **104**:393–398.

Anderson, P., and Roth, J. (1981) Spontaneous tandem genetic duplications in *Salmonella typhimurium* arise by unequal recombination between rRNA (*rrn*) cistrons. *Proc Natl Acad Sci USA* **78**: 3113–3117.

Ausubel, F.M., Brent, R., Klingston, R.E., and Moore, D.D. (1987) *Current Protocols in Molecular Biology,* vol. 1 Wiley. New York.

Barrick, J.E., and Lenski, R.E. (2013) Genome dynamics during experimental evolution. *Nat Rev Genet* **14**: 827–839.

Bassler, B.L., Greenberg, E.P., and Stevens, A.M. (1997) Cross-species induction of luminescence in the quorum-sensing bacterium *Vibrio harveyi*. *J Bacteriol* **179**: 4043–4045.

Blount, Z.D., Barrick, J.E., Davidson, C.J., and Lenski, R.E. (2012) Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* **489**: 513–518.

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

Bochkareva, E.S., Girshovich, A.S., and Bibi, E. (2002) Identification and characterization of the *Escherichia coli* stress protein UP12, a putative *in vivo* substrate of GroEL. *European Journal of Biochemistry* **269**: 3032–3040.

Cairns, J., Overbaugh, J., and Miller, S. (1988) The Origin of Mutants. *Nature* **335**: 142–145.

Davies, S.J., Golby, P., Omrani, D., Broad, S.A., Harrington, V.L., Guest, J.R., *et al.* (1999) Inactivation and regulation of the aerobic C-4-dicarboxylate transport (*dctA*) gene of *Escherichia coli*. *J Bacteriol* **181**: 5624–5635.

Feng, P., and Hartman, P.A. (1982) Fluorogenic Assays for Immediate Confirmation of *Escherichia coli*. *Appl Environ Microbiol* **43**: 1320–1329.

Galhardo, R.S., Hastings, P.J., and Rosenberg, S.M. (2007) Mutation as a Stress Response and the Regulation of Evolvability. *Critical Reviews in Biochemistry and Molecular Biology* **42**: 399–435.

Garrity, G., Staley, J.T., Boone, D.R., and De Vos, P. (2006) *Bergey's Manual® of Systematic Bacteriology: Volume Two: The Proteobacteria*.

Greenberg, E.P., Hastings, J.W., and Ulitzur, S. (1979) Induction of luciferase synthesis in *Beneckea harveyi* by other marine bacteria. *Arch Microbiol* **120**: 87–91.

Hall, B.G. (1982) Chromosomal Mutation for Citrate Utilization by *Escherichia coli* K-12. *J Bacteriol* **151**: 269–273.

Hersh, M.N., Ponder, R.G., Hastings, P.J., and Rosenberg, S.M. (2004) Adaptive mutation and amplification in *Escherichia coli*: two pathways of genome adaptation under stress. *Research in Microbiology* **155**: 352–359.

Janausch, I., Kim, O., and Unden, G. (2001) DctA- and Dcu-independent transport of succinate in *Escherichia coli*: contribution of diffusion and of alternative carriers. *Arch Microbiol* **176**: 224–230.

Koser, S.A. (1923) Utilization of the salts of organic acids by the colon-aerogenes group. *J Bacteriol* **8**: 493–520.

Koser, S.A. (1924) Correlation of citrate utilization by members of the colon-aerogenes group with other differential characteristics and with habitat. *J Bacteriol* **9**: 59–77.

Kugelberg, E., Kofoid, E., Andersson, D.I., Lu, Y., Mellor, J., Roth, F.P., and Roth, J.R. (2010) The Tandem Inversion Duplication in *Salmonella enterica*: Selection Drives Unstable Precursors to Final Mutation Types. *Genetics* **185**: 65–80.

Kugelberg, E., Kofoid, E., Reams, A.B., Andersson, D.I., and Roth, J.R. (2006) Multiple pathways of selected gene amplification during adaptive mutation. *Proc Natl Acad Sci USA* **103**: 17319–17324.

Kvint, K., Nachin, L., Diez, A., and Nyström, T. (2003) The bacterial universal stress protein: function and regulation. *Current Opinion in Microbiology* **6**: 140–145.

Lennox, E.S. (1955) Transduction of Linked Genetic Characters of the Host by Bacteriophage P1. *Virology* **1**: 190–206.

Lenski, R.E., Rose, M.R., Simpson, S.C., and Tadler, S.C. (1991) Long-Term Experimental Evolution in *Escherichia coli*. I. Adaptation and Divergence During 2,000 Generations. *American Naturalist* **138**: 1315–1341.

Lenski, R.E. (2011) Evolution in action: a 50,000-generation salute to Charles Darwin. *Microbe* **6**:30–33.

Lin, R.J., Capage, M., and Hill, C.W. (1984) A Repetitive DNA Sequence, *rhs*, Responsible for Duplications Within the *Escherichia coli* K-12 Chromosome. *Journal of Molecular Biology* **177**: 1–18.

Lütgens, M., and Gottschalk, G. (1980) Why a Co-Substrate Is Required for Anaerobic Growth of *Escherichia coli* on Citrate. *J Gen Microbiol* **119**: 63–70.

Paliy, O., and Gunasekera, T.S. (2007) Growth of *E. coli* BL21 in minimal media with different gluconeogenic carbon sources and salt contents. *Appl Microbiol Biotechnol* **73**: 1169–1172.

Pennisi, E. (2013) The Man Who Bottled Evolution. *Science* **342**: 790–793.

Petes, T.D., and Hill, C.W. (1988) Recombination Between Repeated Genes in Microorganisms. *Annu Rev Genet* **22**: 147–168.

Quandt, E.M., Deatherage, D.E., Ellington, A.D., Georgiou, G., and Barrick, J.E. (2014) Recursive genomewide recombination and sequencing reveals a key refinement step in the evolution of a metabolic innovation in *Escherichia coli*. *Proc Natl Acad Sci USA* **111**: 2217–2222.

Reams, A.B., and Roth, J.R. (2015) Mechanisms of Gene Duplication and Amplification. *Cold Spring Harb Perspect Biol* **7**: a016592–26.

Reams, A.B., Kofoid, E., Kugelberg, E., and Roth, J.R. (2012) Multiple Pathways of Duplication Formation with and Without Recombination (RecA) in *Salmonella enterica*. *Genetics* **192**: 397–415.

Reams, A.B., Kofoid, E., Savageau, M., and Roth, J.R. (2010) Duplication Frequency in a Population of *Salmonella enterica* Rapidly Approaches Steady State With or Without Recombination. *Genetics* **184**: 1077–1094.

Scheu, P.D., Steinmetz, P.A., Dempwolff, F., Graumann, P.L., and Unden, G. (2014) Polar Localization of a Tripartite Complex of the Two-Component System DcuS/DcuR and the Transporter DctA in *Escherichia coli* Depends on the Sensor Kinase DcuS. *PLoS ONE* **9**: e115534–23.

Sonti, R.V., and Roth, J.R. (1989) Role of Gene Duplications in the Adaptation of *Salmonella typhimurium* to Growth on Limiting Carbon-Sources. *Genetics* **123**: 19–28.

Turner, C.B., Blount, Z.D., Mitchell, D.H., and Lenski, R.E. (2015) Evolution and coexistence in response to a key innovation in a long-term evolution experiment with *Escherichia coli*.

Vaughn, R.H., Osborne, J.T., Wedding, G.T., Tabachnick, J., Beisel, C.G., and Braxton, T. (1950) The Utilization of Citrate by *Escherichia coli*. *J Bacteriol* **60**: 119–127.

Wall, J.D., and Harriman, P.D. (1974) Phage P1 Mutants with Altered Transducing Abilities for *Escherichia coli*. *Virology* **59**: 532–544.

Witan, J., Bauer, J., Wittig, I., Steinmetz, P.A., Erker, W., and Unden, G. (2012) Interaction of the *Escherichia coli* transporter DctA with the sensor kinase DcuS: presence of functional DctA/DcuS sensor units. *Molecular Microbiology* **85**: 846–861.

Witan, J., Monzel, C., Scheu, P.D., and Unden, G. (2012) The sensor kinase DcuS of *Escherichia coli*: two stimulus input sites and a merged signal pathway in the DctA/DcuS sensor unit. *Biological Chemistry* **393**: 1–8.

Yoon, S.H., Han, M.-J., Jeong, H., Lee, C.H., Xia, X.-X., Lee, D.-H., *et al.* (2012) Comparative multi-omics systems analysis of *Escherichia coli* strains B and K-12. *Genome Biol* **13**: R37.

Zientz, E., Six, S., and Unden, G. (1996) Identification of a third secondary carrier (DcuC) for anaerobic C-4-dicarboxylate transport in *Escherichia coli*: Roles of the three Dcu carriers in uptake and exchange. *J Bacteriol* **178**: 7241–7247.

Zinser, E.R. (2004) *Escherichia coli* evolution during stationary phase. *Research in Microbiology* **155**: 328–336.

Table 4.1. *E. coli* strains used in this study

TABLE 1. ***E. coli*** **strains used in this study**

Strain	**Genotype/description**	Reference
MG1655	(wt) *rph-1*	E. Top, Univ. Idaho collection
JW0604-1	F-, *Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), ΔcitT750::kan, λ-, rph-1, Δ(rhaD-rhaB)568, hsdR514*	Keio collection, Yale Univ.
JW3496-1	F-, *Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ-, ΔdctA783::kan, rph-1, Δ(rhaD-rhaB)568, hsdR514*	Keio collection, Yale Univ.
JW5437-1	F-, *Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ-, ΔrpoS746::kan, rph-1, Δ(rhaD-rhaB)568, hsdR514*	Keio collection, Yale Univ.
E. coli B	(wt) F-	Coli Genetic Stock Center, Yale Univ.
REL606	F-, *tsx-467(Am), araA230, Δlon, rpsL227(strR), ΔhsdR, [mal+]K-12(λS)*	Coli Genetic Stock Center, Yale Univ.
Cit+ Mutants	**Parent/description**	Reference
	Direct selection in M9C	
DV159	MG1655 (wt), Cit+ at 30 days	This study
DV160	MG1655 (wt), Cit+ <40 days	This study
DV133T	MG1655 (wt), phage transduction from DV133	This study
SO191	JW3496-1 (*ΔdctA*), Cit+ at 134 days	This study
DV130	JW5437-1 *(ΔrpoS),* Cit+ at 43 days	This study
DV133	JW5437-1 *(ΔrpoS),* refined from DV130 via several cycles on SC agar	This study
DV162	JW5437-1 *(ΔrpoS),* Cit+ <40 days	This study
DV268	JW5437-1 *(ΔrpoS),* Cit+ <40 days	This study
	Modified direct selection in M9C25 (glucose-supported)	
DV247	MG1655 (wt), Cit+ at 31 days	This study
DV290	MG1655 (wt), Cit+ at 35 days	This study
DV291	MG1655 (wt), Cit+ at 35 days	This study
DV309	MG1655 (wt), Cit+ at 42 days	This study
DV312	MG1655 (wt), Cit+ at 42 days	This study
DV313	MG1655 (wt), Cit+ at 42 days	This study
DV314	MG1655 (wt), Cit+ at 42 days	This study
DV315	MG1655 (wt), Cit+ at 42 days	This study
DV409	MG1655 (wt), Cit+ at 77 days	This study
DV344	JW5437-1 *(ΔrpoS),* Cit+ at 68 days	This study
	Modified direct selection in M9CG50 (glycerol-supported)	
DV179	MG1655 (wt), Cit+ at 28 days	This study
DV215	MG1655 (wt), Cit+ at 21 days	This study
DV172	JW5437-1 *(ΔrpoS),* Cit+ at 19 days	This study
DV216	JW5437-1 *(ΔrpoS),* Cit+ at 21 days	This study
DV351	JW5437-1 *(ΔrpoS),* Cit+ at 35 days, and refined from papilliated colony	This study
	Modified direct selection in M9LC25 (glucose-supported, 1.7 mM citrate (low-citrate))	
DV564	MG1655 (wt), Cit+ at 40 days	This study
DV596	MG1655 (wt), Cit+ at 56 days	This study
DV600	MG1655 (wt), Cit+ at 56 days	This study
DV592	MG1655 (wt), Cit+ at 63 days	This study
DV593	MG1655 (wt), Cit+ at 63 days	This study
DV594	MG1655 (wt), Cit+ at 63 days	This study
DV595	MG1655 (wt), Cit+ at 63 days	This study
DV599	MG1655 (wt), Cit+ at 63 days	This study
DV601	MG1655 (wt), Cit+ at 63 days	This study
DV602	MG1655 (wt), Cit+ at 63 days	This study
DV603	MG1655 (wt), Cit+ at 63 days	This study
DV604	REL606, Cit+ at 63 days	This study
	Modified direct selection in M9LCG50 (glycerol-supported, 1.7 mM citrate (low-citrate))	
DV545	MG1655 (wt), Cit+ at 35 days	This study
DV546	MG1655 (wt), Cit+ at 35 days	This study
DV571	MG1655 (wt), Cit+ at 40 days	This study
DV572	MG1655 (wt), Cit+ at 41 days	This study
DV605	*E. coli* B (wt), Cit+ at 63 days	This study
DV608	*E. coli* B (wt), Cit+ at 63 days	This study
DV609	*E. coli* B (wt), Cit+ at 63 days	This study
DV610	*E. coli* B (wt), Cit+ at 63 days	This study
DV611	REL606, Cit+ at 63 days	This study
DV613	REL606, Cit+ at 63 days	This study
DV614	REL606, Cit+ at 63 days	This study
DV615	REL606, Cit+ at 63 days	This study
DV616	REL606, Cit+ at 63 days	This study
Phage	**Genotype/description**	Reference
P1*(vir)*	nonlysogenic transducing phage	P. Hartzell

Table 4.2. Mass spectrometry settings for precursor masses, product ion masses, cone voltages, collision voltages, and dwell times

TABLE 2.	Mass spectroscopy settings for presursor masses, product ion masses, cone voltages, collision voltages, and dwell times				
Analyte	Precursor m/z	Product m/z	Cone V	Collision V	Dwell Time (sec)
Acetate	58.8	SIR	28	N/A	0.11
Lactate1	88.74	42.87	14	10	0.11
Lactate2	88.74	44.88	14	10	0.11
Fumarate1	114.72	26.96	26	10	0.11
Fumarate2	114.72	70.93	26	8	0.11
Succinate1	116.87	72.94	26	10	0.11
Succinate2	116.87	98.32	26	11	0.11
Malate1	132.86	132.86	70.28	30	0.11
Malate2	132.86	132.86	70.28	30	0.11
Glutamine1	144.91	83.97	36	14	0.11
Glutamine2	114.91	109	36	12	0.11
Citrate1	190.87	190.87	86.9	28	0.11
Citrate2	190.87	190.87	110.95	28	0.11
Indole1	117.83	117.83	64.88	48	0.745
Indole2	117.83	117.83	90.95	48	0.745

SIR, single ion reaction; NA, not applicable.

Figure 4.1. Direct selection of *E. coli* in minimal M9C yielded Cit$^+$ mutants for both *E. coli* wild-type and *E. coli Δrpos::kan* strains but not *E. coli ΔcitT::kan*.

Two separate representative experiments are shown. *E. coli* wild-type (○), *E. coli Δrpos::kan* (◊) and *E. coli ΔcitT::kan* (□)were inoculated into individual 250 ml flasks containing 50 ml of M9 minimal medium containing citrate as the sole carbon source. *E. coli* wild-type and *E. coli ΔcitT::kan* decreased by one log until after day 23 when the wild-type strain (designated DV159) reinitiated exponential growth (panel A). A parallel experiment is shown in panel B. *E. coli Δrpos::kan* decreased by almost two logs until after day 37 when it reinitiated exponential growth. This strain was designated DV130.

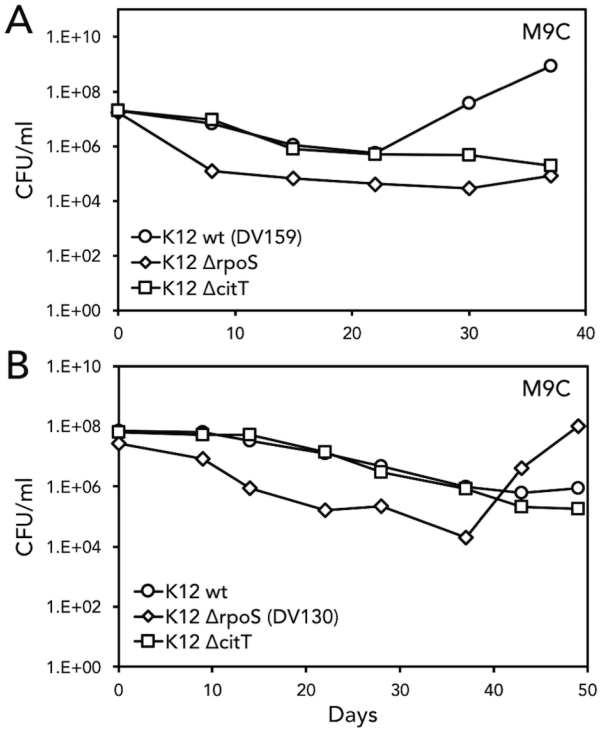


Figure 4.2. Modified direct selection of *E. coli* yielded Cit$^+$ mutants independent of strain and citrate concentration.

Mutants arose for *E. coli* K12, *E. coli* K12 *Δrpos::kan, E. coli* B, *E. coli* REL606, but not *E. coli ΔcitT::kan* or *E. coli ΔdctA::kan.* Four separate representative experiments are shown. Each flask was inoculated with ~5.0 x 10^5 CFU/ml and every seven days, cultures were diluted 1:100 into fresh medium. OD600 was measured before transfer. In panels A and B, *E. coli* K12 wild-type (○), *E. coli ΔrpoS::kan* (◊) *E. coli ΔdctA::kan* (*Δ*) and *E. coli ΔcitT::kan* (□) were inoculated into individual 250 ml flasks containing 50 ml of M9 containing 6.8 mM citrate and either 0.0025% glucose, M9C25 (Panel A) or 0.005% glycerol, M9CG50 (Panel B). This level of glucose or glycerol supported six generations of growth. In M9C25 (Panel A), Cit$^+$ mutants in *E. coli* K12 and *E. coli ΔrpoS::kan* arose after the forth and ninth transfers, respectively. In M9CG50 (Panel B), Cit$^+$ mutants arose for *E. coli* K12 and *E. coli ΔrpoS::kan* after the third transfer. Citrate-supported growth was not recovered for either *E. coli ΔdctA::kan* or *E. coli ΔcitT::kan* strains. Similarly, *E. coli* K12 wild-type (○), *E. coli* B (□), and *E. coli* REL606 (◊) were inoculated into M9 containing 1.7 mM citrate and either 0.0025% glucose, M9LC25 (Panel C) or 0.005% glycerol, M9LCG50 (Panel D). Cit$^+$ mutants arose in *E. coli* K12 after five or four transfers, in *E. coli* B after four or five transfers, and in *E. coli* REL606 after eight and five transfers (Panel C and D, respectively).

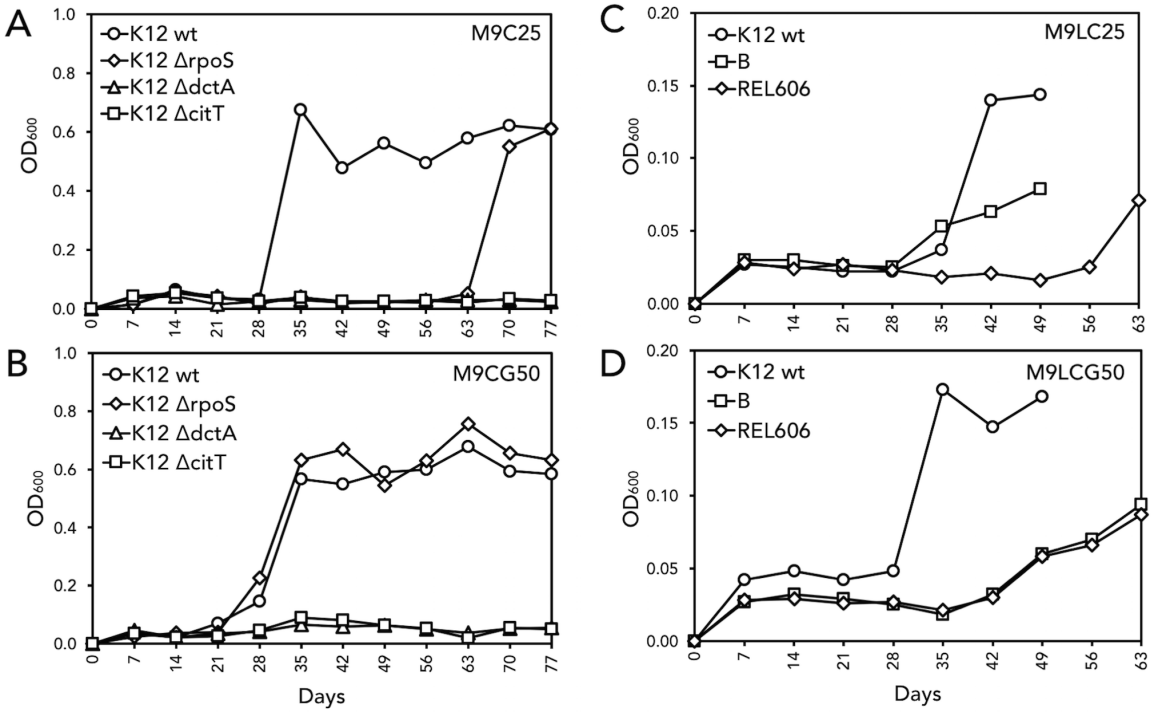

A
K12 wt
K12 ΔrpoS
K12 ΔdctA
K12 ΔcitT
M9C25
OD600
B
M9CG50
Days
C
K12 wt
B
REL606
M9LC25
D
M9LCG50

Figure 4.3. Extended incubation of actualized *E. coli* Cit$^+$ on Simmons citrate agar yielded refined *E. coli* Cit$^+$ papillae.

Primary direct selection M9C broth that displayed growth were plated on Simmons citrate agar. Small yellow colonies appeared after 3-4 days incubation and small blue papillae (irregular darkened regions) appeared as shown after 10-20 days. Each blue papillae represents an independent mutation to refinement.

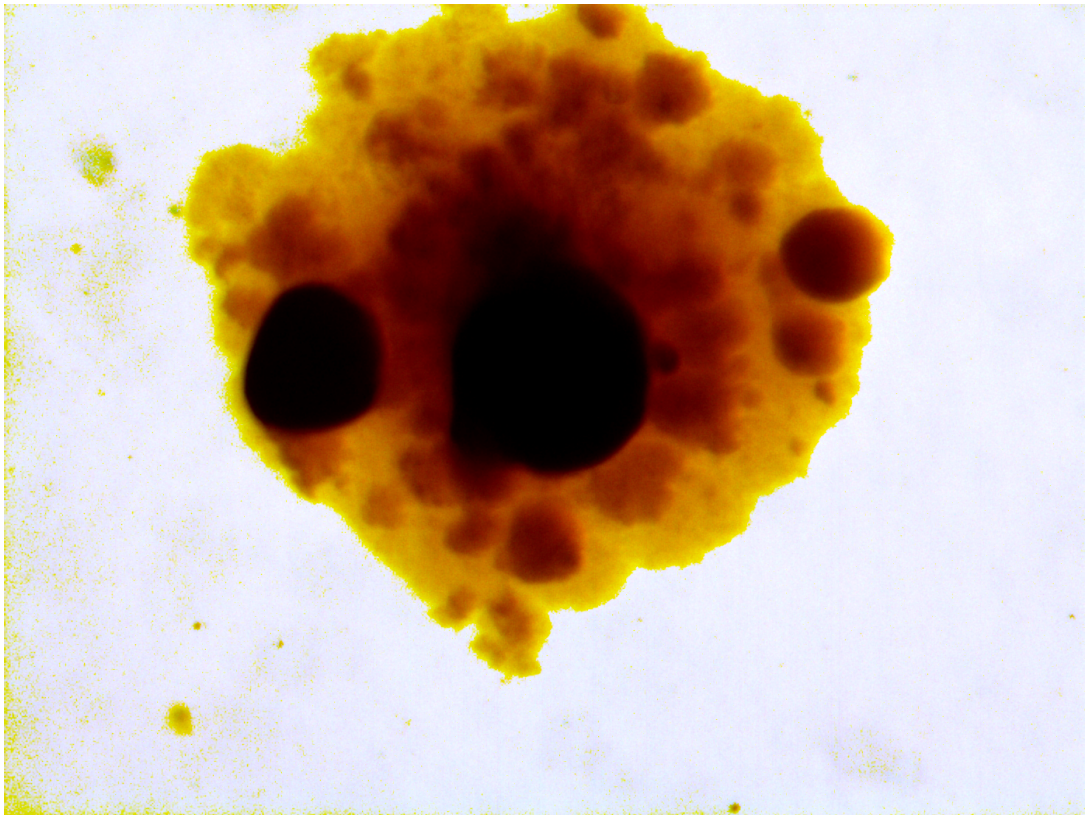


Figure 4.4. Mass spectrometry showed citrate depletion correlated with *E. coli* growth.

Two *E. coli* Cit$^+$ mutants derived from *E. coli* wild-type isolates, DV133T (Panel A) and DV159 (Panel B) were inoculated into M9C and the OD_{600} was measured every 6 h. Triplicate media samples at each time point were filter-sterilized and a subset of samples corresponding to time zero, late-lag, early-, mid-, late-exponential, and stationary phases were analyzed by mass spectrometry for citric acid. Bars represent ±S.E.

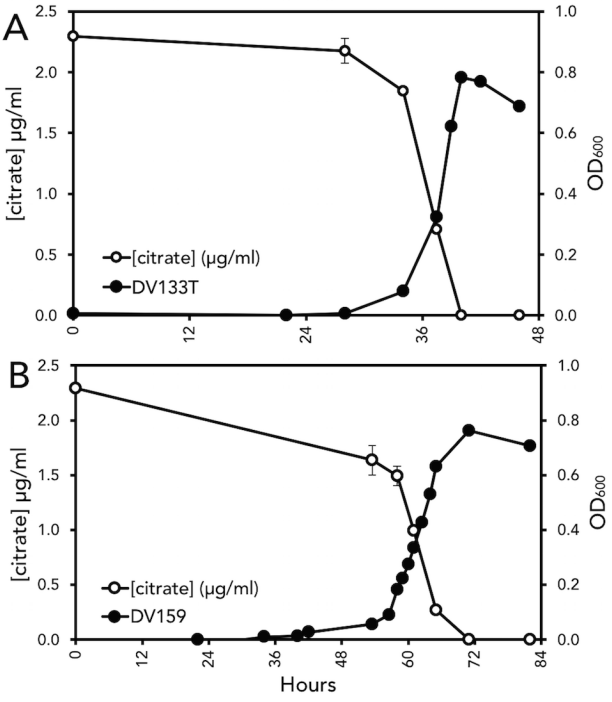


Figure 4.5. *E. coli* Cit$^+$ mutants grew in clumps in M9C broth but not in LB broth.

Phase contrast microscopy comparing growth of *E. coli* Cit$^+$ mutants DV133 and DV159 (Panels A and B, and Panels C and D, respectively) in M9C and LB broth (Panels A and C and B and D, respectively). Cells in M9C were clumped whereas growth in LB broth was dominated by single and pre-divisional cells. These pictures are representative of 10 microscopic fields; magnification = 400X.

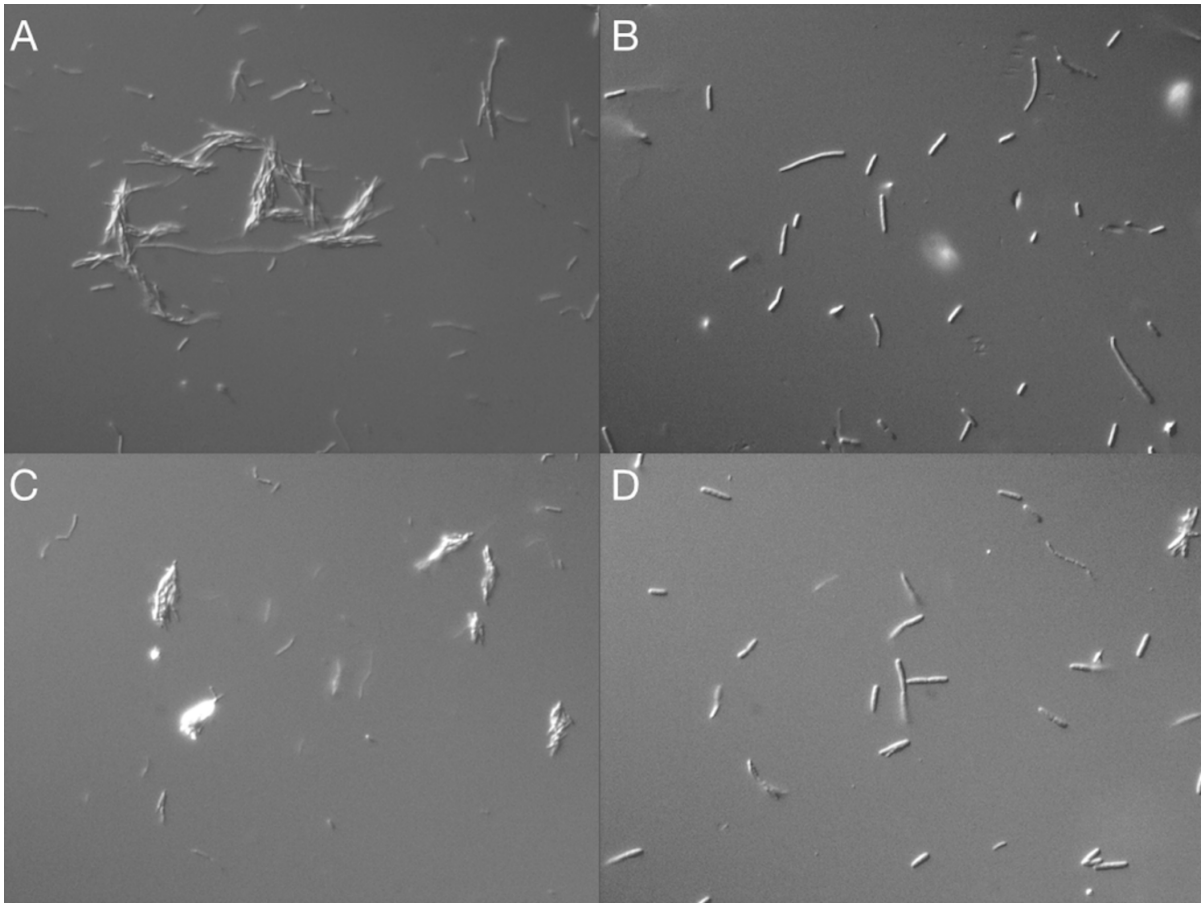
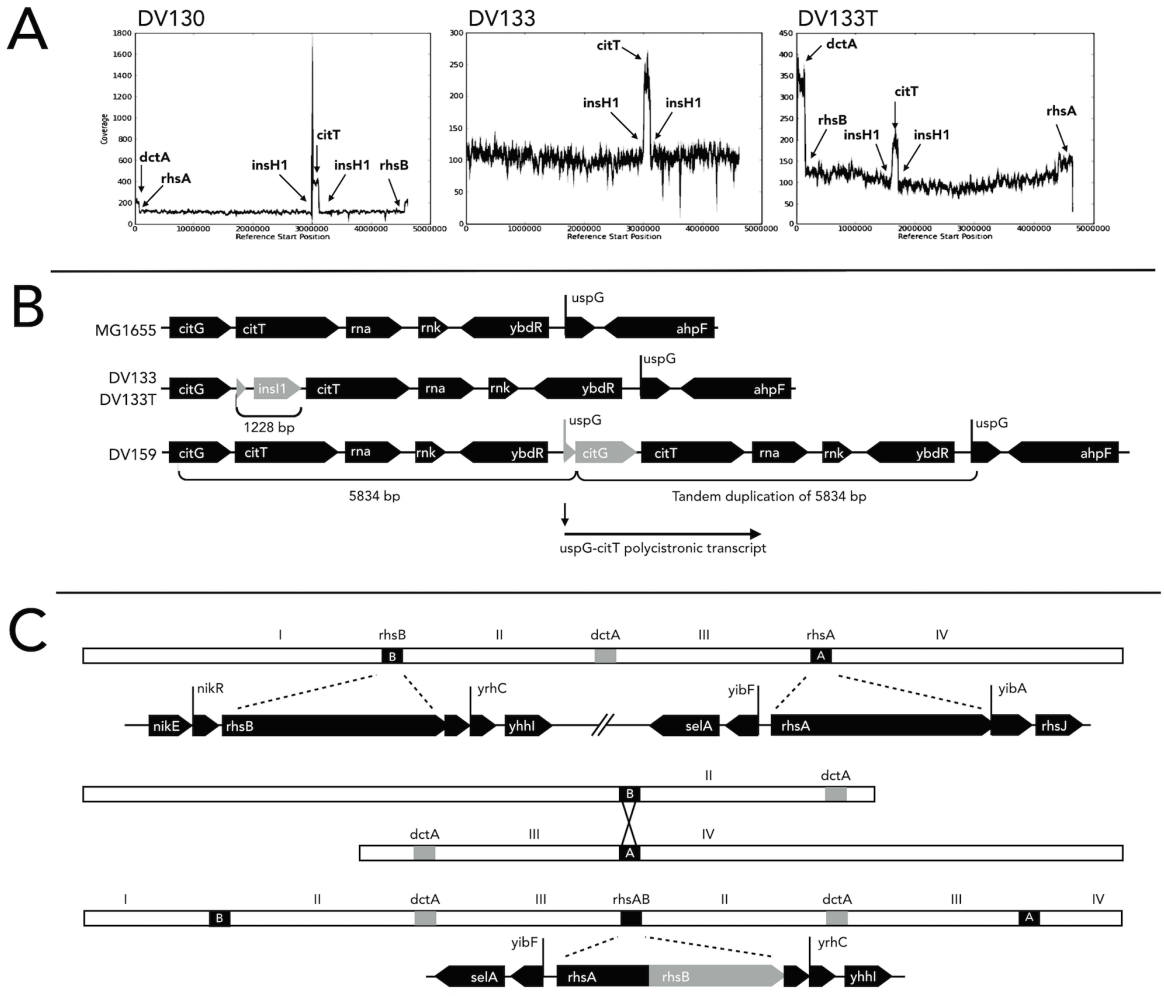


Figure 4.6. Genomic DNA sequence analysis of *E. coli* Cit$^+$ phenotypes showed gene amplifications associated with *dctA* and *citT* regions and promoter captures.

In Panel A, the genomic read depth profiles of *E. coli* Cit$^+$ isolates DV130, DV133 and DV133T are shown. The weakly Cit$^+$ phenotype of DV130 shows a four-fold increased coverage of the *citT* region and a two-fold increased coverage of *dctA* compared to the ~100X coverage of the rest of the chromosome. Strain DV133 was a strong Cit$^+$ phenotype derived from DV130. It showed a two-fold reduction in *citT* coverage compared to the parental DV130 strain, and the insertion of *insl*1 5' to *citT* (Panel B). When *E. coli* K12 wild-type was transduced with a lysate made on DV133, the same *insl1-citT* fusion was identified confirming this promoter capture was responsible for the transduced strong Cit$^+$ phenotype (Panel B). The genomic sequence of Cit$^+$ strain DV159 showed a *citT* duplication and deletion event that fused the promoter of *uspG* to *citG*, another promoter capture event (Panel B). All read-depth profiles show a two-fold increased coverage for the *dctA* region compared to the ~100X coverage for the chromosome. Sequence analysis of DV133, DV133T, and DV159 show a recombination between *rhsA* and *rhsB* based on the differences in positions of flanking genes *yibF* and *yrhC* (Panel C). The genomic organization of the *E. coli* wild-type chromosome is represented by the top picture in Panel C. The recombination of sister chromosomes to generate a large 140 kb duplication of this region is shown in the center picture (Panel C). The

determined gene map for both DV133 and DV159 is shown in the bottom picture with the *dctA* duplication (Panel C).



Chapter 5

Epilogue

Following the acceptance of our manuscript by the *Journal of Bacteriology*, we were notified by Dr. Thomas Silhavy, Editor in Chief, that our article was to be featured with an accompanying commentary, a practice that the journal reserves for manuscripts with broad and significant impact. Dr. Silhavy requested that one of our paper's anonymous reviewers write the commentary. This turned out to be John Roth, a pioneer in microbial genetics and member of the National Academy, along with a postdoctoral fellow in his laboratory, Sophie Maisnier-Patin. These two authors agreed with our overall interpretation of how *E. coli* acquires the Cit^+ phenotype in Lenski's LTEE (Roth and Maisnier-Patin, 2016). Twenty days after these two papers went on-line, Richard Lenski and Zachary Blount published a criticism of our paper on Lenski's blog site. Because our paper generated such positive and negative responses, this epilogue is written to address these criticisms.

Importantly, Roth and Maisnier-Patin agreed with our interpretation that generating *E. coli* Cit^+ mutants did not require the accumulation of neutral potentiating mutations (Roth and Maisnier-Patin, 2016). In their commentary, Roth and Maisnier-Patin stated our work highlights a need to reinterpret the concept that

'historical contingency' is a contributing factor in the evolution of aerobic citrate utilization in the LTEE (Blount *et al.*, 2008). Their commentary rightly identified the crux of the problem, i.e. that *E. coli* faces a 'bottleneck' during daily serial transfers that is not easily overcome unless the time for Cit^+ selection is extended as we show. The recognition that this 'bottleneck' is the key mechanism delaying a Cit^+ phenotype in the LTEE is why reinterpretation of the LTEE is critical.

The hypothesis that historical contingency is necessary for organisms to acquire new functions was first proposed by Stephen Jay Gould (Gould, 1989). Gould suggested that the evolution of innovative traits is predicated by a series of random, neutral-acting mutations, and thus evolution is unpredictable and unrepeatable. Lenski's LTEE, with its 12 parallel populations, is poised to test this fundamental concept. As such, the series of events leading to aerobic citrate utilization, by their interpretation, is contingent on acquiring unpredictable neutral but necessary potentiating mutations. These necessary random genetic events explain the long delay in acquiring the later actualizing mutation leading to a stable Cit^+ phenotype. As Blount explains:

> *"Perhaps the most intriguing and important question about the evolution of Cit^+ is also the most obvious. The enormous citrate resource had been there from the LTEE's beginning. So why did the Cit^+ trait evolve only once, and then only after such a long time? One plausible explanation is that Cit^+ was a historically contingent trait. Historically contingent traits require particular, non-*

> *guaranteed antecedent states, which is to say a particular history, to evolve. Their origins are therefore complex, and require multiple mutational steps (Blount, 2016b).”*

If, however, it could be demonstrated that the genetic events leading to citrate utilization occurred rapidly and repeatedly, it would rule out historical contingency as a contributing factor. Indeed, Blount agrees with this criterion:

> *“If evolution is highly contingent in the unpredictability sense, then one might expect evolutionary outcomes to be fundamentally un-repeatable even from the same starting point (Blount, 2016a).”*

The fact that the LTEE did not produce Cit^+ mutants in the first 31,000 generations, and that only one of the 12 populations produced a Cit^+ variant does give the appearance that historical contingency is a key factor in a population’s evolutionary history. However, the numerous independent Cit^+ mutants rapidly isolated in our experiments show historical contingency is not a prerequisite for Cit^+ adaptation (Van Hofwegen *et al.*, 2016). Roth agreed that the genetic events leading to aerobic citrate utilization were simple genomic duplications followed by a genomic deletion and promoter capture at the deletion join point. These same genomic rearrangements were observed in the LTEE populations (Raeside *et al.*, 2014). Changes in gene copy number are among the most common events for the microbial genome (Anderson and Roth, 1977; Sonti and Roth, 1989), so why the

extreme delay in the appearance of Cit$^+$ *E. coli* in the LTEE? Roth and Maisnier-Patin state:

> *"Thus, the delay may not reflect the necessity of a nonselective potentiation event (the "historical contingency") but may rather be due to selective events with a low probability of serial transfer."*

We illustrated that historical contingency is not involved in the manifestation of the Cit$^+$ phenotype by two methods. First, we could isolate Cit$^+$ mutants in as few as 12 generations, insufficient time for historical contingency to come into play. Second, the genetic loci for a strong Cit$^+$ phenotype could be directly transduced by P1 (*vir*) bacteriophage into the wild-type *E. coli* Cit$^-$ genetic background, showing a 'potentiating' genetic background is not required for the phenotype. The bottleneck for aerobic citrate utilization, we argue and Roth agrees, is not easily overcome in daily serial transfers because the initial stages of gene amplification confer a fitness cost, a selective disadvantage with transfer to fresh glucose each day. As Roth explains:

> *"These steps are easy to visualize when selection is imposed in a single long-term batch culture or during serial transfer with extended citrate selection. The process is more difficult with daily serial dilutions."*

The emergence of aerobic citrate utilization in our extended selection periods is something that Blount has argued cannot occur if the trait is historically contingent on nonselective potentiated mutations.

> *"Historically contingent traits require particular, non-guaranteed antecedent states, which is to say a particular history, to evolve. Their origins are therefore complex, and require multiple mutational steps. Some of these steps may be neutral, not uniquely beneficial, or possibly even mildly detrimental. Because the required steps are not uniquely favored, cumulative selection cannot predictably and rapidly facilitate their accumulation (Blount, 2016b)."*

Twenty days after our manuscript and the Roth and Maisnier-Patin commentary were posted online, Lenski and Blount responded on Lenski's blog site (https://telliamedrevisited.wordpress.com/2016/02/20/on-the-evolution-of-citrate-use/). Jeremy Fox, on his blog, hailed this response by Lenski and Blount as a "model example of post-publication review." (https://dynamicecology.wordpress.com/2016/03/04/friday-links-79/). In this blog response they agreed with our overall results stating:

> *"The actual science that was done and reported looks fine and interesting, though we have a few quibbles with some details that we will overlook for now."*

However, the majority of their response highlighted problems they had with our publication:

> *"The problem, then, is not with the experiments and data. Rather, the problem is that the results are wrapped in interpretations that are, in our view, unscientific and unbecoming."*

Claiming our interpretations are "unscientific and unbecoming" are pejorative comments. First, it implies that the peer review process, that included reviews by two members of the National Academy (Silhavy and Roth), could not discern that our conclusions were unscientific! Second, what is unbecoming about our scientific interpretation? Is it that we disagree and disproved their interpretations? This argument is a text book example of the 'No True Scotsman' fallacy.

Apart from the assertion that our interpretations (discussion section in Chapter 5) are "unscientific and unbecoming", I welcome criticism and clarification to my work. Lenski and Blount take issue with several aspects of our publication centered on the following points; "no new genetic information", "rapid evolution", "speciation event", and "historical contingency". It is these criticisms I would like to address.

Lenski and Blount take issue with the semantics regarding genetic information, i.e., our interpretation of 'no new genetic information' within the system. Lenski states:

> *"The authors assert repeatedly (last sentence of their Importance statement, and first and last paragraphs of their Discussion) that "no new genetic information evolved."*

First, Lenski's response misses the point by not taking this statement in its full context. We qualified "no new genetic information" stating this to mean 'novel gene function' for each instance:

Importance Section:

> *"No new genetic information (novel gene function) evolved."*

Introduction:

> *"We hypothesized that the isolation of such mutants should be relatively easy, would follow the same genetic trajectory as that identified in the LTEE, and would utilize information present on the chromosome and not involve evolution of new information (novel gene function)."*

Discussion:

> *"Finally, because this adaptation did not generate any new genetic information and required expanded expression of only two existing transporters (citT and dctA), generation of E. coli Cit^{+} phenotypes in our estimation does not warrant consideration as a speciation event."*

…

> *"As such, Cit$^+$ mutants exemplify the adaptation capability of microorganisms but, as of yet, the LTEE has not substantiated evolution in the broader sense by generation of new genetic information, i.e., a gene with a new function."*

Lenski's comments miss the point we were trying to make. We specifically characterized new information as novel gene function. Genetic information, from this perspective, is prescriptive, referring to the meaning of the coding sequence, i.e. the biological function of the encoded protein or regulatory region. The term information can be defined in terms of complexity or probability of a given sequence, however, this use of information theory fails to recognize the prescriptive nature of functional open reading frames in the genetic sequence (Barbieri, 2016). Biologically meaningful genetic information must specify some function. We hypothesized that the acquisition of aerobic citrate utilization would occur rapidly and use the existing biological information present in the genome. Lenski offers François Jacob's interpretation of natural selection:

> *"That's how evolution works—it's not as though new genes and functions somehow appear out of thin air. As the bacterial geneticist and Nobel laureate François Jacob wrote (Jacob, 1977): "[N]atural selection does not work as an engineer works. It works like a tinkerer—a tinkerer who does not know exactly what he is going to produce but uses whatever he finds around him, whether it be*

> *pieces of string, fragments of wood, or old cardboards; in short, it works like a tinkerer who uses everything at his disposal to produce some kind of workable object."*

However, that is exactly the point of our experiments. *E. coli* can rearrange its genome to rapidly develop the ability to use citrate aerobically, i.e., to 'tinker' with its chromosome. Moreover, we tested how much existing information was required for acquisition of this phenotype. By removing two of the genes responsible for conferring this phenotype, we tested if another solution could be reached by *E. coli*. Nevertheless, Lenski continues:

> *"The authors assert repeatedly (last sentence of their Importance statement, and first and last paragraphs of their Discussion) that "no new genetic information evolved." However, that statement flatly contradicts the fact that in their experiments, and ours, E. coli gained the new ability to grow on citrate in the presence of oxygen. We would further add (which we have not emphasized before) that these Cit$^+$ strains can grow on citrate as a sole carbon source—when E. coli grows anaerobically on citrate, it requires a second substrate for growth in order to use the citrate (a phenomenon called "co-metabolism")."*

The "new ability to grow on citrate in the presence of oxygen" was exactly the point of our hypothesis; that *E. coli* would rapidly evolve the ability to use citrate in oxic conditions. Our validation of this hypothesis (the isolation of 46 independent

Cit$^+$ mutants) illustrates that acquisition of this phenotype is not a novel innovation as Lenski *et al*. suggests. In the above statement, Lenski and Blount highlight that aerobic citrate utilization as a sole carbon source requires new information based on the fact that under anaerobic conditions, *E. coli* can only use citrate if it is provided a co-metabolite. This assertion, in my estimation, simply fails to recognize the difference between aerobic and anaerobic metabolism. *E. coli* requires a co-metabolite for anaerobic growth on citrate because of the need for reducing power (Lütgens and Gottschalk, 1980). Under anaerobic conditions, the tricarboxylic acid cycle (TCA) forward reactions are repressed because the electron transport chain has no oxygen to act as the final electron acceptor. The anaerobic CitT antiporter imports citrate, (six carbons) while exporting, succinate (four carbons). Citrate is then hydrolyzed to oxaloacetate (four carbons), and acetate (two carbons). Oxaloacetate is reduced back to succinate (running the TCA backwards), requiring reducing power in the form of NADH. This leaves only the acetate as the energy and reducing power source for the cell. The metabolism of acetate, under anaerobic conditions provides only enough reducing power to convert oxaloacetate back to succinate. This is a 'wash' on the energetic balance sheet, providing no net gain of energy nor reducing power. It is for this reason *E. coli* cannot grow anaerobically on citrate as a sole carbon source because it requires a co-metabolite to furnish additional reducing power to keep CitT functioning. The required co-metabolite, such as glucose, lactate, or pyruvate, provides a source of electrons, i.e. reducing power, to overcome this metabolic stalemate. Lütgens and Gottschalk (1980)

clearly show that under anaerobic conditions once this co-substrate is consumed, citrate metabolism ceases.

Aerobically, *E. coli* cannot grow on citrate because CitT is not expressed. *E. coli* cannot get this carbon source into the cell. The Cit^+ mutants in both the LTEE and our experiments over-ride this block in aerobic *citT* expression. Because both the TCA cycle and electron transport chain are active aerobically, the constraints on complete citrate oxidation are removed, as are the requirements for a co-metabolite. Overriding the requirement for co-metabolism is not new information, it is just the difference between anaerobic and aerobic metabolism. Indeed, if new information is required to oxidize citrate aerobically, this should have been evident in the genomic sequences of Cit^+ mutants compared to wild-type *E. coli*. We found no such additional information. Further, if additional mutations are required to circumvent this co-metabolism requirement, transduction of the Cit^+ phenotype should not have been feasible.

In the next critique, Lenski offers further clarification on information:

> *"The claim that "no new genetic information evolved" is based on the fact that the bacteria gained this new ability by rearranging existing structural and regulatory genetic elements. But that's like saying a new book—say, Darwin's Origin of Species when it first appeared in 1859—contains no new information, because the text has the same old letters and words that are found in other books."*

This analogy is nothing but a straw man argument. A more correct interpretation using this analogy would be the following: we started with Darwin's *Origin of Species,* i.e., the *E. coli* genome. After printing millions of copies, we found production errors where several pages had been duplicated and some of these duplications had been edited out leaving a 'cut and paste' error at the duplication junction. We started with the *Origin of Species* and ended with the *Origin of Species* with no new information. In other words, we didn't end up with a copy where Darwin's Origin of Species presents a new argument. It is simply the same book with a deletion, 'cut and paste' error, and a duplication. In short, no new information.

Lenski continues his criticism by taking issue with our experiments being called "rapid evolution":

> *"In the title of their paper and throughout, Van Hofwegen et al. emphasize that, in their experiments, E. coli evolved the ability to grow aerobically on citrate much faster than the 30,000 generations and ~15 years that it took in the LTEE. That's true, but it also obscures three points. First, we already demonstrated in replay experiments that, in the right genetic background and by plating on minimal-citrate agar, Cit$^+$ mutants sometimes arose in a matter of weeks (Blount et al., 2008)."*

In their 2008 PNAS paper they did do replay experiments, sampling from various generational time-points (See their Table 1) (Blount *et al.*, 2008). First, they

repeated the LTEE protocol using 72 replay populations. They isolated four Cit^+ variants. These four mutants only arose from starting populations taken from LTEE frozen fossils after 30,500 generations, no isolates from 5,000 – 30,000 LTEE generation start points. Additionally, these four isolates emerged between 750 and 3,700 generations of replay! 750 to 3,700 generations at 6.6 generations per day requires 113 to 561 days respectively for these mutants to appear under standard LTEE experimental conditions. Yet they criticize us for taking 19 days from generation zero!

In their second replay experiment, they took 68 clones from the same LTEE generational start points and plated directly onto minimal citrate agar plates and incubated them for long periods (59 days). They isolated five Cit^+ strains, again all after 32,000 generations starting clones i.e., 15 years of constant citrate exposure. In an expanded version of this ('Experiment 2') they isolated an additional eight Cit^+ clones; two from their 20,000 generation start point, two from 27,000 generation start point, and four after 31,000 generation start points. By their calculation they screened ~4.0×10^{13} cells in total. So in their combined replay experiments they acquired 17 additional Cit^+ mutants (again, compared to our isolation of 46 Cit^+ mutants with a maximum of 77 days and only requiring 12 to 100 generations). In summation, using their same protocol with extended time between transfers (1 week vs. daily), we get the same class of Cit^+ mutants in days, not decades, and starting with wild-type *E. coli*, not a strain that required, at a minimum, 20,000 generations of constant citrate exposure.

"Second, rapid evolution of citrate utilization—or any evolution of that function—was not a goal of the LTEE. So while it is interesting that Van Hofwegen et al. have identified genetic contexts and ecological conditions that accelerate the emergence of citrate utilization (as did Blount et al., 2008), that in no way undermines the slowness and rarity of the evolution of this function in the context of the LTEE (or, for that matter, the rarity of Cit$^+$ E. coli in nature and in the lab prior to our work)."

We fully acknowledge this and clearly stated this in our manuscript:

"By design, the LTEE are open-ended experiments to track the evolution of 12 parallel E. coli cultures under controlled conditions. Aerobic citrate utilization was not a specified aim. The highly delayed evolution (33,000 generations) of Cit$^+$ strains understandably gives the impression that this was a rare, innovative evolutionary event. Our studies highlight why it was not."

However, we would contend that our results do 'undermine' the overall interpretation for the long-delayed appearance of Cit$^+$ cells in the LTEE. Our experiments provide a mechanism explaining why it took so long for this trait to emerge in the LTEE, and the explanation is the basis of why Roth and Maisnier-Patin specifically state that the LTEE results need to be reinterpreted. Lenski and Blount imply we employed different genetic contexts and ecological conditions to accelerate Cit$^+$ mutant recovery. We used wild-type *E. coli* and the exact ecological

conditions (aeration, minimal medium, temperature) employed in the LTEE for the majority of our experiments. The only difference is we extended the time between culture transfers from one day to one week. As for "the rarity of Cit$^+$ *E. coli* in nature and in the lab prior to our work" we would simply use an evolutionary argument, there is no selective advantage for *E. coli* to require aerobic citrate metabolism. The primary natural habitat of *E. coli* is the lower intestine, a predominantly anaerobic environment, where the average generational time is two days (Savageau, 1983), conditions markedly different than our experiments or those of the LTEE. The caveat to this argument is there was one prior isolation of Cit$^+$ *E. coli* in the lab as reported by Barry Hall, the impetus for doing our study (Hall, 1982).

Finally, Lenski attempts to undermine the importance of our work based on examples of other direct selections:

> *"Third, the fastest time that Van Hofwegen et al. saw for the Cit$^+$ function to emerge was 19 days (from their Table 1), and in most cases it took a month or two. While that's a lot faster than 15 years, it's still much longer than typical "direct selections" used by microbiologists where a readily accessible mutation might confer, for example, resistance to an antibiotic after a day or two."*

Resistance to an antibiotic, such as streptomycin or rifampin, typically requires only a single mutation. This is genetically equivalent to acquiring bacteriophage resistance illustrated by Luria and Delbrück's seminal fluctuation tests (Luria and Delbrück, 1943). Phage resistance, they show, requires a single

mutation which is pre-existing in the population under direct selection. Luria and Delbrück calculate the mutation rate at 10^{-8} to 10^{-9} for phage resistance, the same rate for mutations conferring streptomycin or rifampin resistance. Notably, phage or antibiotic resistance appears in a day or two when more than 10^{8} cells are spread on selective plates. Lenski's example of direct selections yielding antibiotic resistance is a gross oversimplification when compared to acquiring Cit^{+} mutants. That is because we show isolating Cit^{+} mutants requires: (i) a large amplification of over 100 kbp; (ii) a subsequent deletion and promoter capture; and, (iii) for *E. coli* B used in the LTEE, a point mutation in the promoter of *dctA* to suppress the pre-existing mutation in *dcuS*. We calculated the probability of this occurring in the same cell at 1×10^{-14}. Barry Hall (1982), in his experiments also estimated this probability at $\sim 1 \times 10^{-14}$. In this specific criticism, Lenski and Blount imply the time to recover a phenotype dependent on a single mutation by direct selection (antibiotic resistance) should be equivalent to the time needed to recover a mutant that requires multiple mutations (Cit^{+}). The rarity of double mutations occurring in the same cell is obvious. This is illustrated by Joshua Lederberg's seminal experiments using conditional-lethal double and triple auxotrophs of *E. coli* to determine if recombination (mating) occurred between *E. coli* strains (Lederberg and Tatum, 1953). The multiple auxotrophic markers were employed for his direct selective crosses to negate the frequent recovery of revertants that occur in one to two days if single auxotrophs are used.

Finally, in perhaps the most revealing piece of criticism, Lenski attempts to establish motive, thus providing further grounds for dismissal:

> *"To say there's no new genetic information when a new function has evolved (or even when an existing function has improved) is a red herring that is promulgated by the opponents of evolutionary science."*

Our experiments were conducted because we hypothesized that potentiating mutations (historical contingency) was not a factor in deriving Cit^+ mutants. When we reviewed their 2008 paper, we also read the reference to Barry Hall's 1982 paper showing selection of Cit^+ *E. coli* in only 14 days, we suspected the conditions of the LTEE were a factor. Our hypothesis was correct. We are opposed to over-interpreted science, not evolution.

In summary, we think the LTEE are excellent experiments and are exactly the experiments that should be done to examine the evolutionary potential of bacterial cells. However, historical contingency for aerobic citrate utilization in the LTEE only follows from the accident of history that *E. coli* B was chosen as the founding strain, which, unbeknownst to Lenski, contains a frameshift mutation in *dcuS* (Yoon *et al.*, 2012; Turner *et al.*, 2015). This is further complicated by daily transfers that select against genome amplifications. As we state in our paper, the historical contingency or potentiating mutation(s) for Cit^+ is merely an artifact of the experimental design and the *E. coli* strain selected. Furthermore, Lenski *et al.* argue the potential for *E. coli* to evolve is open ended, however we interpret the overall lesson of the LTEE as a grand demonstration of genetic entropy and re-distribution of existing genetic information to adapt to a uniform environment. Indeed,

‘tinkering’ at best. To suggest that the acquisition of aerobic citrate utilization by degradation and rearrangement is grounds for speciation significantly over-interprets these results. Many pathogenic *E. coli* have acquired large segments of new genetic information by lateral gene transfer. For example, *E. coli* O157:H7 has almost a million extra base pairs of DNA compared to *E. coli* B or K12, yet we still classify it as *E. coli*. To think that a promoter capture and a suppressing point mutation allowing aerobic access to citrate is a step in speciation is arguably weak.

References

Anderson, R.P., and Roth, J.R. (1977) Tandem Genetic Duplications in Phage and Bacteria. *Annu Rev Microbiol* **31**: 473–505.

Barbieri, M. (2016) What is information? *Phil Trans R Soc A* **374**: 20150060–10.

Blount, Z.D. (2016a) History's Windings in a Flask: Microbial Experiments into Evolutionary Contingency. In *Chance in Evolution*. Pence, C., and Ramsey, G. (eds). Chance in Evolution.

Blount, Z.D. (2016b) A case study in evolutionary contingency. *Studies in History and Philosophy of Biological and Biomedical Sciences* 1–11. doi: 10.1016/j.shpsc.2015.12.007

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

Gould, S.J. (1989) *Wonderful life: New York*. NY: Norton.

Hall, B.G. (1982) Chromosomal Mutation for Citrate Utilization by *Escherichia coli* K-12. *J Bacteriol* **151**: 269–273.

Jacob, F. (1977) Evolution and Tinkering. *Science* **196**: 1161–1166.

Lederberg, J., and Tatum, E.L. (1953) Sex in bacteria; genetic studies, 1945-1952. *Science* **118**: 169–175.

Luria, S.E., and Delbruck, M. (1943) Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**: 491–511.

Lütgens, M., and Gottschalk, G. (1980) Why a Co-Substrate Is Required for Anaerobic Growth of *Escherichia coli* on Citrate. *J Gen Microbiol* **119**: 63–70.

Raeside, C., Gaffé, J., Deatherage, D.E., Tenaillon, O., Briska, A.M., Ptashkin, R.N., *et al.* (2014) Large chromosomal rearrangements during a long-term evolution experiment with *Escherichia coli*. *mBio* **5**: e01377–14.

Roth, J.R., and Maisnier-Patin, S. (2016) Reinterpreting Long-Term Evolution Experiments: Is Delayed Adaptation an Example of Historical Contingency or a Consequence of Intermittent Selection? *J Bacteriol* **198**:1009–1012.

Savageau, M.A. (1983) *Escherichia coli* habitats, cell types, and molecular mechanisms of gene control. *The American Naturalist* **122**: 732–744.

Sonti, R.V., and Roth, J.R. (1989) Role of Gene Duplications in the Adaptation of *Salmonella typhimurium* to Growth on Limiting Carbon-Sources. *Genetics* **123**: 19–28.

Turner, C.B., Blount, Z.D., Mitchell, D.H., and Lenski, R.E. (2015) Evolution and coexistence in response to a key innovation in a long-term evolution experiment with *Escherichia coli*. bioRxiv doi: 10.1101/020958

Van Hofwegen, D.J., Hovde, C.J., and Minnich, S.A. (2016) Rapid Evolution of Citrate Utilization by *Escherichia coli* by Direct Selection Requires *citT* and *dctA*. *J Bacteriol* **198**: 1022–1034.

Yoon, S.H., Han, M.-J., Jeong, H., Lee, C.H., Xia, X.-X., Lee, D.-H., *et al.* (2012) Comparative multi-omics systems analysis of *Escherichia coli* strains B and K-12. *Genome Biol* **13**: R37.