Genetic Background and Environment Influence the Effects of Mutations in *pykF* and Help Reveal Mechanisms Underlying Their Benefit

A Dissertation

Presented to

the Faculty of the Department of Biology and Biochemistry

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By Fen Peng

August 2015

Genetic Background and Environment Influence the Effects of Mutations in *pykF* and Help Reveal Mechanisms Underlying Their Benefit

Fen Peng

APPROVED:

Dr. Tim Cooper, Chair
Department of Biology and Biochemistry

Dr. Renwick Dobson
University of Canterbury, New Zealand

Dr. Elizabeth Ostrowski
Department of Biology and Biochemistry

Dr. Hye-Jeong Yeo
Department of Biology and Biochemistry

Dr. Dan Wells, Dean
College of Natural Sciences and Mathematics

Acknowledgements

First and foremost I would like to express my deepest and most genuine appreciation to my advisor, Dr. Tim Cooper. I would like to thank you, Dr. Cooper, for all the time and effort you have invested in me. I owe a special debt of gratitude to you; words cannot describe my happiness and fortune to be given the opportunity to complete my graduate studies under your guidance.

I would like to thank my committee member, Dr. Renwick Dobson, for his help with my dissertation, especially the biochemistry part. I am sincerely grateful for your unusual patience and assistance on academic, as well as personal, affairs during my stay in New Zealand. I had a wonderful time there. I would also like to acknowledge and express special appreciation to the other members, current and former, of my committee: Dr. Elizabeth Ostrowski, Dr. Hye-Jeong Yeo, and Dr. Azevedo Ricardo for their time and advice, without which this dissertation would not be possible.

Thanks to our collaborators from New Zealand, Katherine Donovan, and Dr. Sarah Kessans for their hard work as well as brilliant communication and teamwork skills. It was my honor and pleasure to work with them. Thanks to all of my lab members: Yinhua Wang, Kelly Phillips, Andrea Wünsche, Rachel Staples, Dustin Phipps, Kristina Duan, Ann Tate, Kedar Karkare, and Gerardo Castillo for providing me such valuable discussions, advice, and generous help to my project's development and progress. Special

thanks go to Kristina Duan, Kelly Phillips, and Andrea Wünsche for their time and contributions in proof reading the dissertation.

I also thank friends in the division, including: Hongan Long, Fan Li, Wei-ting Lin, Ata Kalirad, Huy Vu, Fei Yuan, etc. I appreciate them for making my life at UH enjoyable and incredibly worthwhile. I would like to extend this appreciation to my friends outside the department, Jie Huang, Xuan Zeng, and Trang Duong, for your friendship, love, and unending support.

Lastly, yet most importantly, I would like to thank my family: my parents and parents-in-law for their understanding and husband, Chuntang Fu, for always being there, supporting and loving me.

Genetic Background and Environment Influence the Effects of Mutations in *pykF* and Help Reveal Mechanisms Underlying Their Benefit

---

An Abstract of a Dissertation

Presented to

the Faculty of the Department of Biology and Biochemistry

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By Fen Peng

August 2015

Abstract

Resolving the relationship between genotypes and their effects remains a central challenge in the study of adaptation. Although parallel mutations, a signature of adaptation, have been observed a lot in natural and lab-evolved populations, it is unknown if they are equally adaptive, or even if they affect similar biological processes to cause phenotypic changes. Using eight independently occurring mutations in *pykF* identified from a long-term evolution experiment with *Escherichia coli*, I found the mutations confer similar benefits in the ancestral background, but variable effects in the background in which they were evolved. Differences in mutation × background interactions were found to be driven by different suites of mutations in each genetic background, rather than by different *pykF* mutations. Through biochemical and physiological studies with the *pykF* mutations in the ancestor, I found that although the mutations affect enzymes in a range of different ways, the net effect of these changes is to lead to changes in the same biological pathways, and thus to confer similar fitness effects.

An adaptive mutation may no longer be beneficial if the given genetic background or environment changes. Relatively few studies, however, have examined the combined effect of genetic and environmental context on fitness effects of a mutation. To do this, I measured fitness effect conferred by one *pykF* mutation in 23 divergent genetic backgrounds and five environments. I found the environment, genetic background, and interactions between them, all significantly affect fitness of the mutation, which makes it harder to predict evolutionary fate of new mutations. Nevertheless, I found that initial

fitness of a progenitor strain can be used to predict contribution of a mutation: a mutation

will contribute less when added to fitter progenitors.

Table of Contents

Chapter 1 Introduction

1.1 Adaptation

Adaptation is central to the study of evolutionary change. It can be defined as any trait that was selected because it caused an organism to be better able to survive and reproduce. How to identify adaptations has been a key question asked by evolutionary biologists. In the past century, research on adaptation has mostly focused on the selective pressures resulting in morphological changes. Convergence, the independent occurrence of similar morphological changes in different populations, has long been taken as the evidence of adaptation (Losos 2011). For instance, a streamlined body independently evolved in sharks, tunas, and dolphins strongly suggests that the convergent body shape is adaptive. However, the evolution of some adaptations might be contingent on previous genotypic changes, so that convergence is not expected to be ubiquitous. Moreover, convergence does not always indicate adaptation; it might occur by chance or as a correlated response to selection on another trait (Losos 2011). All of these factors make it hard to identify adaptations.

Mutations are the ultimate source of genetic variation that is acted on by natural selection to cause evolutionary change. Thus, the study of molecular level changes can be used to identify potentially adaptive traits, and even why a trait has been favored by natural selection by connecting the molecular change to the adaptive trait. As molecular techniques have developed and become widespread in many different biological systems,

studies have accumulated that identify the genetic basis of adaptation (reviewed in Dean and Thornton 2007, and Barrett and Hoekstra 2011). These studies follow two broad strategies. First, candidate adaptive alleles can be identified if they show a statistical signature of historical selection, especially if they occur in focal loci thought to affect selected phenotypes (Colosimo et al. 2005; Mullen and Hoekstra 2008; Storz and Kelly 2008). Second, in genetically tractable systems, direct experimental (*e.g.*, functional) tests can be carried out to test the hypothesized genetic bases of adaptation (Hoekstra et al. 2006; Barrett et al. 2008; Storz et al. 2010). Many studies have found that convergent phenotypes have evolved through similar changes at the gene level, indicating a genetic constraint in adaptation (reviewed in Gompel and Prud'homme 2009). However, it is unknown to what extent adaptive mutations in the same gene produce convergence in phenotypes. Moreover, very few studies have applied manipulative molecular experimental techniques to evolution, so that it is very rarely understood *how* adaptive mutations mediate beneficial phenotypic changes. Mapping the connections between genotypic change (mutations), biochemical mechanisms, and organism-level phenotypic effects (including fitness) is required to fully understand an adaptive mutation (Barrett and Hoekstra 2011).

1.2 Experimental evolution

One common focus of experimental evolution aims to understand the connections between genotype and phenotype. To facilitate this study and explore adaptation in a controlled, laboratory context, researchers rely on microorganisms such as bacteria and yeast, which present many distinct advantages: (1) they can be grown in defined and

controllable environments, so that selective histories can be both replicated and manipulated; (2) they reproduce quickly, and allow many generations in a relatively short period of time; (3) they can be frozen and revived, which allows direct comparison at different evolutionary stages; (4) many of them have small genomes, which facilitates whole genome-sequencing as a feasible approach for discovering mutations after selection; (5) their genomes can often be manipulated such that alleles can be introduced to genetic backgrounds of interest and their effects determined.

1.3 Mapping from genotype to phenotype

A major challenge in the study of adaptation is to identify the genetic basis of adaptations. New genome-scale sequencing techniques combined with experimental evolution approaches have made it easier to begin to determine the phenotypic effect of adaptive mutations. First, sequencing of replicate populations started from a common genotype and evolved in the same environmental conditions enables identification of mutations arising in parallel, a signature of their being selected. These mutations can be introduced to a reference background, for example the ancestor of an evolution experiment, and their effects on organismal phenotypes determined in the evolution environment (Cooper et al. 2001; Khan et al. 2011). Complicating this approach, however, the effect of one mutation may depend on the presence of other mutations fixed in the same population (epistasis). These interactions can mean that a mutation that is beneficial in one genetic context (*e.g.*, an evolved genotype) is neutral or even deleterious in another (*e.g.*, the ancestral genotype). Although epistatic interactions are pervasive (reviewed in Kouyos et al. 2007; Parera et al. 2009; Lalić and Elena 2012), they have usually been examined in studies that

3

have focused on randomly generated mutations, most of which have neutral or deleterious phenotypic effects. Relatively few studies have examined epistasis between adaptive mutations. One study examining the epistatic interactions between adaptive mutations was conducted by Khan *et al.* (2011). Investigating five adaptive mutations, they found that negative epistasis was common. This finding was also observed in several other studies (Chou et al. 2011; Rokyta et al. 2011; Wang et al. 2013). In addition, environmental context can also affect the effect of one mutation (pleiotropy) (Remold 2012). Thus, to map from genotype to phenotype, the genetic and environmental context where a mutation appears has to be carefully considered.

1.4 Mechanisms connecting genotype to phenotype

Genetic manipulations to determine the fitness effect of specific mutations can identify those that are adaptive, but to understand *why* a mutation is adaptive requires a knowledge of the lower-level biochemical and physiological mechanisms that connect genotype to phenotype. The complexity of this mapping is the ultimate reason why the mechanism of action of adaptive mutations is generally unknown (Dean and Thornton 2007). To gain insight into how adaptive mutations cause phenotypic effects, there are several aspects we can investigate, such as a mutation's effect on transcriptional regulation, protein activity and function, and metabolism. By integrating information from these investigations we hope to determine: how an adaptive mutation mediates its fitness advantage, and whether phenotypic convergence reflects constraints at genetic, functional, or physiological levels (Orr 2005). Together, this data will help us to predict if

future mutations would target a phenotype in the same way—in other words, will evolution be predictable.

1.5 Long-term evolution experiment

Experiments in this dissertation are based on clones isolated from a long-term evolution experiment with *Escherichia coli* (LTEE) started in 1988 by Lenski and colleagues (Lenski et al. 1991). In that ongoing experiment, 12 populations founded from a common ancestor have been evolved in a glucose-limited environment for more than 60,000 generations. Similar phenotypic changes among replicate populations were discovered, such as increased fitness (Lenski and Travisano 1994), increased cell size (Lenski et al. 1998), and gene (Cooper et al. 2003) and protein expression profiles (Pelosi et al. 2006), which suggest phenotypic parallelism (I used parallelism instead of convergence to refer to similar changes from independent populations started from the same founder genotype). Similarly, mutations in the same gene were also identified in multiple replicate populations (Woods et al. 2006). One of the genes that have been mutated in all of the 12 populations is *pykF*. This gene encodes pyruvate kinase, a key enzyme in glycolysis. Non-synonymous point mutations in *pykF* were identified in 10 of the 12 populations. Eight of these mutations are unique, they are: P70Q, P70T, D127N, I264F, A301S, A301T, G381A and T462I (Figure 1.1). Repeated fixation of *pykF* mutations in 12 independent populations is a strong indicator that mutations in this gene are adaptive.
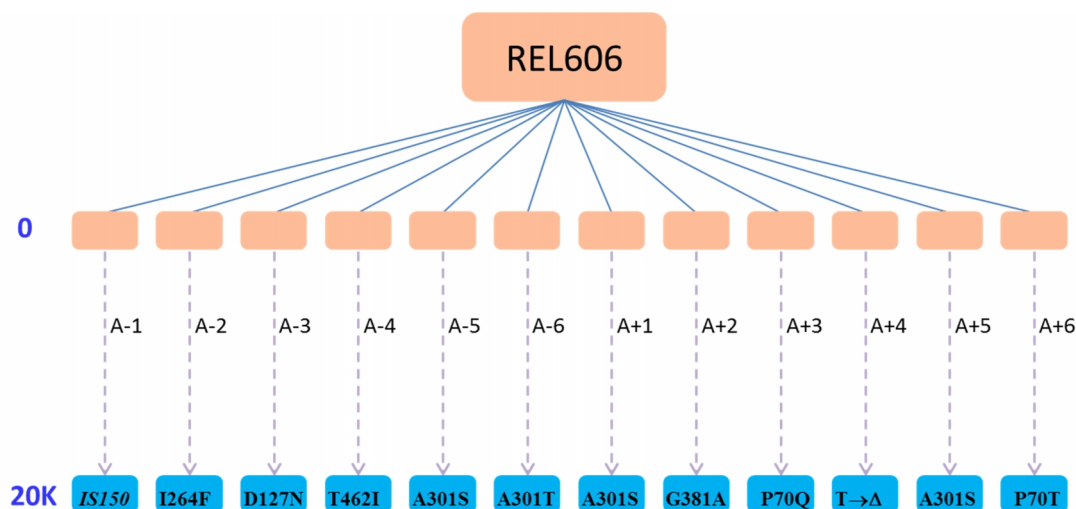
Figure 1.1. *pykF* mutations identified from the LTEE at 20,000 generations. 12 replicate populations (numbered A-1 to A+6) were found by the ancestor, REL606, and independently evolved in a glucose-limited environment. At 20,000 generations, one clone was isolated from each population and sequenced at specific candidate genes. Mutations in *pykF* were identified in all 12 isolated clones, and there is no more than one non-synonymous *pykF* mutation in any clone. Point mutations are found in 10 clones, and the other two have an IS*150* and a frameshift mutation. As shown, A301S mutation was identified in three independently evolved clones.

1.6 Pyruvate kinase

Pyruvate kinase regulates the last rate-limiting step in glycolysis, catalyzing the irreversible de-phosphorylation of phosphoenolpyruvate (PEP) to pyruvate, and the phosphorylation of ADP to ATP (Valentini et al. 2000) (Figure 1.2). It is a conserved protein and exists in the glycolysis pathway in both eukaryotes and prokaryotes (Hattori et al. 1995; Valentini et al. 2000). Pyruvate kinase has recently been identified as an important player in cancer progression (Christofk et al. 2008; Luo and Semenza 2012; Tamada et al. 2012).

In *E. coli*, there are two pyruvate kinase isoenzymes, pyruvate kinase type I (PykF) and pyruvate kinase type II (PykA). PykA is thought to contribute little to pyruvate kinase activity (Garrido-Pertierra and Cooper 1983; Ponce et al. 1995). Through X-ray crystallography, the protein structure of *E. coli* PykF has been resolved and carefully examined in previous studies (Mattevi et al. 1995; Zhu et al. 2010). PykF is a homotetramer, each subunit composed of three domains: domain A, domain B, and domain C. Two main mechanisms of PykF regulation have been characterized: (1) homotropical regulation by its substrate PEP (binding of PEP at one active site in the homotetramer facilitates binding of PEP at other active sites) and (2) allosteric regulation by fructose-1, 6-bisphosphate (FBP). Regulation by FBP, whose concentration is sensitive to the glucose flux in a cell, allows pyruvate kinase to be a crucial metabolic regulation point (Mattevi et al. 1995). Mapping *pykF* mutations occurring in the LTEE to the pyruvate kinase structure shows that they localize to the active site, perhaps affecting PEP binding or catalysis, or close to the allosteric regulation site, which might affect the FBP allosteric regulation. These mutations are expected to change the enzyme function in some way, thus are ideal to the study of how mutations act to change phenotypes.
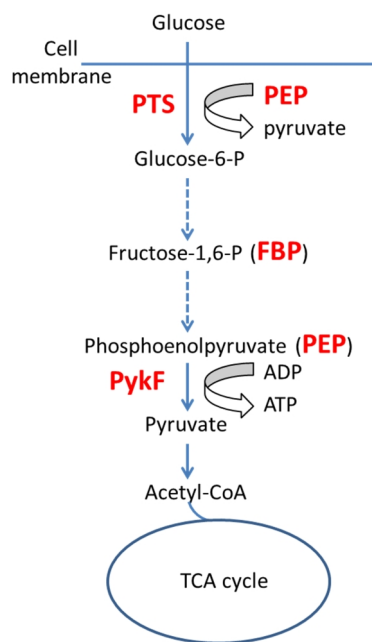
Figure 1.2. An overview of glycolysis in *E. coli*. Phosphotransferase system (PTS) is used to import a PTS resource, for example glucose, into the cell. In this process, PEP is dephosphorylated to pyruvate, and glucose is phosphorylated to glucose-6-phosphate (glucose-6-P). Downstream steps produce intermediates, including FBP and PEP. Pyruvate kinase uses PEP as a substrate and FBP as an allosteric regulator, to produce pyruvate. Pyruvate can be converted to acetyl-CoA, which feeds into the tricarboxylic acid (TCA) cycle.

1.7 Dissertation outline

In this dissertation, I aimed to explore the phenotypic effect of adaptive mutations in *pykF*, the mechanisms underlying that effect, and the influence of genetic and environmental context. Specifically, I address the following questions: (1) How similar are the phenotypic effects of different mutations occurring in the same gene (chapter 2)? (2) Do the effects of mutations change as the populations in which they occurred continue to evolve (chapter 2)? (3) What are the mechanisms underlying fitness effects conferred by *pykF* mutations (chapter 3)? (4) Does the effect of an adaptive mutation depend on its

8

genetic background and environment context, and their interaction? If so, are we able to predict the nature of this dependence (chapter 4)?

In chapter 2, I measured the effect of *pykF* mutations arising in each of 10 replicate evolved populations on fitness in the ancestral and their evolved genetic background. Two key *pykF* mutations were also measured for fitness effects in the alternative evolved backgrounds. I found that mutations in the same gene confer similar benefits, *i.e.*, are phenotypically parallel in the same ancestral background, but have variable effects in their evolved background, due to different epistatic effects, which were driven by different sets of mutations fixed in each evolved line.

In chapter 3, my aim was to unravel mechanisms underlying the similar fitness advantages conferred by the *pykF* mutations found in chapter 2. I found that mutations in *pykF* are likely to confer fitness benefits by shortening lag times following transfer of a strain to fresh medium. The shortened lag phase was found to be associated with increased concentrations of PEP in stationary phase, which in turn facilitates more rapid uptake of the limited glucose resource when cells are transferred to fresh medium. Although the evolved mutations changed the PykF enzyme in a range of different ways, the overall effect on cell metabolism and fitness revealed to be similar.

In chapter 4, I measured fitness effects of one *pykF* mutation across different natural isolate strains and different resource environments. I showed that fitness of a mutation strongly depends on the genetic background, the environment, and the interaction between the genetic background and the environment. I also showed that the initial

growth rate of a natural isolate can be used to qualitatively predict fitness of a new

mutation: the larger the initial growth rate of a recipient strain, the smaller the benefit of

the mutation.

1.8 References

Barrett, R. D. H., and H. E. Hoekstra. 2011. Molecular spandrels: tests of adaptation at the genetic level. Nat. Rev. Genet. 12:767–780.

Barrett, R. D. H., S. M. Rogers, and D. Schluter. 2008. Natural selection on a major armor gene in threespine stickleback. Science 322:255–257.

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

Christofk, H. R., M. G. Vander Heiden, M. H. Harris, A. Ramanathan, R. E. Gerszten, R. Wei, M. D. Fleming, S. L. Schreiber, and L. C. Cantley. 2008. The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. Nature 452:230–233.

Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarreal Jr., M. Dickson, J. Grimwood, J. Schmutz, R. M. Myers, D. Schluter, and D. M. Kingsley. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. Science 307:1928–1933.

Cooper, T. F., D. E. Rozen, and R. E. Lenski. 2003. Parallel changes in gene expression after 20, 000 generations of evolution in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 100:1072–1077.

Cooper, V. S., D. Schneider, M. Blot, and R. E. Lenski. 2001. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. J. Bacteriol. 183:2834–2841.

Dean, A. M., and J. W. Thornton. 2007. Mechanistic approaches to the study of evolution: the functional synthesis. Nat. Rev. Genet. 8:675–688.

Garrido-Pertierra, A., and R. A. Cooper. 1983. Evidence for two distinct pyruvate kinase genes in *Escherichia coli* K-12. FEBS Lett. 162:420–422.

Gompel, N., and B. Prud'homme. 2009. The causes of repeated genetic evolution. Dev. Biol. 332:36–47.

Hattori, J., B. R. Baum, S. G. Mchugh, S. D. Blakeley, D. T. Dennis, and B. L. Miki. 1995. Pyruvate kinase isozymes: ancient diversity retained in modern plant cells. Biochem. Syst. Ecol. 23:773–780.

Hoekstra, H. E., R. J. Hirschmann, R. A. Bundey, P. A. Insel, and J. P. Crossland. 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. Science 313:101–104.

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

Kouyos, R. D., O. K. Silander, and S. Bonhoeffer. 2007. Epistasis between deleterious mutations and the evolution of recombination. Trends Ecol. Evol. 22:308–315.

Lalić, J., and S. F. Elena. 2012. Magnitude and sign epistasis among deleterious mutations in a positive-sense plant RNA virus. Heredity. 109:71–77.

Lenski, R. E., J. A. Mongold, P. D. Sniegowski, M. Travisano, F. Vasi, P. J. Gerrish, and T. M. Schmidt. 1998. Evolution of competitive fitness in experimental populations of *E. coli* : what makes one genotype a better competitor than another? Antonie Van Leeuwenhoek 73:35–47.

Lenski, R. E., M. R. Rose, S. C. Simpson, and S. C. Tadler. 1991. Long-term experimental evolution in *Escherichia coli*. I. adaptation and divergence during 2,000 generations. Am. Nat. 138:1335–1341.

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

Losos, J. B. 2011. Convergence, adaptation, and constraint. Evolution 65:1827–1840.

Luo, W., and G. L. Semenza. 2012. Emerging roles of PKM2 in cell metabolism and cancer progression. Trends Endocrinol. Metab. 23:560–566.

Mattevi, A., G. Valentini, M. Rizzi, M. L. Speranza, M. Bolognesi, and A. Coda. 1995. Crystal structure of *Escherichia coli* pyruvate kinase type I: molecular basis of the allosteric transition. Structure 3:729–741.

Mullen, L. M., and H. E. Hoekstra. 2008. Natural selection along an environmental gradient: A classic cline in mouse pigmentation. Evolution 62:1555–1570.

Orr, H. A. 2005. The genetic theory of adaptation: a brief history. Nat. Rev. Genet. 6:119–127.

Parera, M., N. Perez-Alvarez, B. Clotet, and M. A. Martínez. 2009. Epistasis among deleterious mutations in the HIV-1 protease. J. Mol. Biol. 392:243–250.

Pelosi, L., L. Kühn, D. Guetta, J. Garin, J. Geiselmann, R. E. Lenski, and D. Schneider. 2006. Parallel changes in global protein profiles during long-term experimental evolution in *Escherichia coli*. Genetics 173:1851–1869.

Ponce, E., N. Flores, A. Martinez, F. Valle, and F. Bolívar. 1995. Cloning of the two pyruvate kinase isoenzyme structural genes from *Escherichia coli*: the relative roles of these enzymes in pyruvate biosynthesis. J. Bacteriol. 177:5719–5722.

Remold, S. 2012. Understanding specialism when the jack of all trades can be the master of all. Proc. R. Soc. B Biol. Sci. 279:4861–4869.

Rokyta, D. R., P. Joyce, S. B. Caudle, C. Miller, C. J. Beisel, and H. A. Wichman. 2011. Epistasis between beneficial mutations and the phenotype-to-fitness map for a ssDNA virus. PLoS Genet. 7:e1002075.

Storz, J. F., and J. K. Kelly. 2008. Effects of spatially varying selection on nucleotide diversity and linkage disequilibrium: insights from deer mouse globin genes. Genetics 180:367–379.

Storz, J. F., A. M. Runck, H. Moriyama, R. E. Weber, and A. Fago. 2010. Genetic differences in hemoglobin function between highland and lowland deer mice. J. Exp. Biol. 213:2565–2574.

Tamada, M., M. Suematsu, and H. Saya. 2012. Pyruvate kinase M2: Multiple faces for conferring benefits on cancer cells. Clin. Cancer Res. 18:5554–5561.

Valentini, G., L. Chiarelli, R. Fortin, M. L. Speranza, A. Galizzi, and A. Mattevi. 2000. The allosteric regulation of pyruvate kinase. J. Biol. Chem. 275:18145–18152.

Wang, Y., C. D. Arenas, D. M. Stoebel, and T. F. Cooper. 2013. Genetic background affects epistatic interactions between two beneficial mutations. Biol. Lett. 20120328.

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

Zhu, T., M. F. Bailey, L. M. Angley, T. F. Cooper, and R. C. J. Dobson. 2010. The quaternary structure of pyruvate kinase type 1 from *Escherichia coli* at low nanomolar concentrations. Biochimie 92:116–120.

Chapter 2 Same mutation different effects: the influence of genetic background on adaptation

Parallel evolution, widely seen in natural and lab-evolved populations, illustrates the major role of natural selection in evolution, since similar evolutionary changes are unlikely to occur by chance alone. Gene-level parallel genetic changes have often been observed in replicated populations evolving in similar environments, however, this observation does not imply that these parallel mutations are equally adaptive or, even, that in all cases they affect the same basic biological process. For these reasons, understanding the extent to which genetic parallelism corresponds to phenotypic parallelism is important to determining the role of parallel genetic changes in phenotypic divergence. I aimed to test how similar the phenotypic effects of mutations in the same gene are, and whether these effects change as the populations in which they substitute continue to evolve. In this study, I focused on eight independently occurring *pykF* mutations, identified from 12 initially identical populations of *Escherichia coli* that evolved in a glucose-limited environment for more than 20,000 generations. Phenotypes, here measured as organism fitness, conferred by the eight *pykF* mutations were measured in both the ancestor and their own evolved background. I found that the eight mutations conferred similar benefits of ~10% in the ancestor; but had much more variable effects in their evolved backgrounds at 20,000 generations, ranging from 0% to 25%. Moreover, by also determining the effect of focal mutations in clones isolated from populations evolved for 30,000 and 50,000 generations, I found a tendency toward them conferring a declining benefit between 20,000 and 50,000 generations. Finally, by transplanting two

different *pykF* mutations, which had different effects in the evolved backgrounds in which they fixed, into all of the evolved backgrounds, I determined that differences in mutation × background interactions are more likely driven by different backgrounds, rather than by different *pykF* mutations. My findings indicate that mutations occurring in the same gene possess the potential to have parallel phenotypic effects, but this may change due to epistatic interactions between different mutations and backgrounds.

2.1 Introduction

Parallel evolution, defined as similar changes occurring in independent lineages evolving under common selection pressure, strongly indicates the action of natural selection, since it is unlikely to observe similar changes from a large pool of possibilities by chance alone. Indeed, several studies have successfully used parallel changes to identify adaptive changes from a set of all evolved changes in lab experiments (Wichman et al. 1999; Cooper et al. 2003; Pelosi et al. 2006; Huse et al. 2010; Gerstein et al. 2012) or from broader physiological/molecular analyses in natural populations (Nachman et al. 2003; Colosimo et al. 2005; Prud'homme et al. 2006).

Compelling examples of parallel phenotypic evolution in natural (Stewart et al. 1987; Losos et al. 1998; Rundle et al. 2000) and lab-evolved populations (Ferea et al. 1999; Cooper et al. 2001, 2003; Pelosi et al. 2006; Huse et al. 2010) have been accumulating. For example, in a long-term evolution experiment with *E. coli*, phenotypic parallelism was found at many levels: the twelve populations show similar changes in most or all populations in cell size (Lenski et al. 1998), growth rate (Lenski and Travisano 1994),

negative DNA supercoiling (Crozat et al. 2005), and gene (Cooper et al. 2003) and protein expression profiles (Pelosi et al. 2006). Recent studies have shown that parallel phenotypic changes often have a parallel genetic basis (reviewed in Gompel and Prud'homme 2009). Genetic parallelism has been observed in many independent lab-evolved populations, such as viruses (Bull et al. 1997; Wichman et al. 1999), bacteria (Woods et al. 2006; Ostrowski et al. 2008; Crozat et al. 2010), and yeast (Gerstein et al. 2012). However, defining genetic parallelism is difficult, since it can be characterized in different ways, such as by mutations affecting identical functional groups, pathways, genes, amino acids or nucleotides. In most studies with bacteria, genetic parallelism was most significant at the gene level (Schluter et al. 2004; Herring et al. 2006; Woods et al. 2006; Tenaillon et al. 2012; Le Gac et al. 2013; Martin and Orgogozo 2013; Vogwill et al. 2014). One example is a recent study (Tenaillon et al. 2012) with 115 lab-evolved *E. coli* populations that, after 2,000 generations of selection at elevated temperature, found genetic parallelism mostly at the gene level, although a few identical nucleotide and amino acid changes were observed.

Genetic parallelism is a signature of adaption, but that does not imply that parallel genetic changes are equally adaptive. Understanding the extent to which genetic parallelism corresponds to phenotypic parallelism is important to determining the role of parallel genetic changes in evolutionary divergence: if parallel mutations are not equally adaptive, distinct suites of subsequent mutations may occur, contributing to an escalating cycle of divergence. In addition, whether different mutations in the same gene have different effects is clearly relevant to the likelihood that multiple mutations in a gene will be

beneficial, and can imply the existence of distinct physiological ways in which adaptation may occur via a single gene. For two reasons, gene-level genetic parallelism may not lead to phenotypic parallelism. First, the effects on the phenotype might be different between different mutations of the same gene. This could occur, for example, if mutations affect different functional regions of the same enzyme. Second, even if mutations in the same gene have similar effects in one genetic background, these effects might, due to epistatic interactions, differ in the actual backgrounds in which the mutations occur.

Lenski *et al.* (1991) evolved 12 populations in a glucose-limited environment for 20,000 generations. By sequencing candidate and random genes in a clone isolated from each population it was found that genetic changes were concentrated in relatively few genes (Woods et al. 2006). One of these genes is *pykF*, which is mutated in all 12 clones. Mutations in this gene are also common in evolution experiments selecting the same ancestral strain in different environments (Phillips et al. 2015) and selecting different ancestral strains in a similar glucose-limited environment (Moore and Woods 2006).

In this study, I measured fitness effects conferred by *pykF* mutations in the ancestor and their own evolved background to assess the extent to which genetic parallelism corresponds to phenotypic parallelism, and how epistatic interactions affect phenotypic parallelism. I found that mutations occurring in the same gene possess the potential to have parallel phenotypic effects, but this may change due to epistatic interactions between different mutations and backgrounds; and the different epistatic interactions are more likely driven by different backgrounds.

2.2 Results

2.2.1 Genetic parallelism has the potential to produce phenotypic parallelism

pykF mutations were identified in all 12 evolved *E. coli* populations by 20,000 generations. Ten are unique—one mutation occurred in three independently evolved populations—and eight are non-synonymous point mutations, the other two mutations are an IS*150* insertion and a frameshift mutation (Table 2.1). In order to examine if different mutations occurring in the same gene have the same effect in the same genetic background, I constructed a series of strains with evolved non-synonymous mutations in *pykF* separately added to their common ancestral background, and measured the fitness of constructed strains relative to the ancestor. I found no detectable fitness difference among them; all *pykF* mutations conferred a ~10% benefit (one-way ANOVA: $F_{7, 32} = 1.67$, $P = 0.15$; Figure 2.1). Similar fitness effects conferred by mutations in *pykF* clearly indicate that gene-level parallel genetic changes can have the same phenotypic effect. The fitness effect conferred by a deletion allele (Del), which has an equivalent fitness effect as the original evolved IS*150* mutation (Khan et al. 2011), was measured in the ancestor as well (Figure 2.1). Interestingly, fitness effects of the point mutations were significantly different from that of the deletion allele, which contributed less benefit (~6%) in the ancestor (See Figure 2.1 for Dunnett's test and Tukey's HSD tests following one-way ANOVA for the deletion allele and the eight point mutations).

Table 2.1. *pykF* mutations identified from clones at 20,000 generations.

Population	Allele	Position (in nucleotides)	Mutation	Amino acid change
20K A-1	IS*150*	683	::IS*150*	insertion
20K A-2	I264F	790	A→T	Ile → Phe
20K A-3	D127N	379	G→A	Asp → Asn
20K A-4	T462I	1385	C→T	Thr → Ile
20K A-5	A301S	901	G→T	Ala → Ser
20K A-6	A301T	901	G→A	Ala → Thr
20K A+1	A301S	901	G→T	Ala → Ser
20K A+2	G381A	1142	G→C	Gly → Ala
20K A+3	P70Q	209	C→A	Pro → Gln
20K A+3	synonymous[*]	507	A→G	synonymous
20K A+4	frame-shift	483	T→Δ	frame-shift
20K A+5	A301S	901	G→T	Ala → Ser
20K A+6	P70T	208	C→A	Pro → Thr

Data from Woods *et al.* (2006).
* I introduced the P70Q and the synonymous mutation together into ancestor. The constructed strain has the same fitness as the ancestral strain containing only P70Q (two-tailed *t* test: $t_5 = 0.90$, $P = 0.41$).
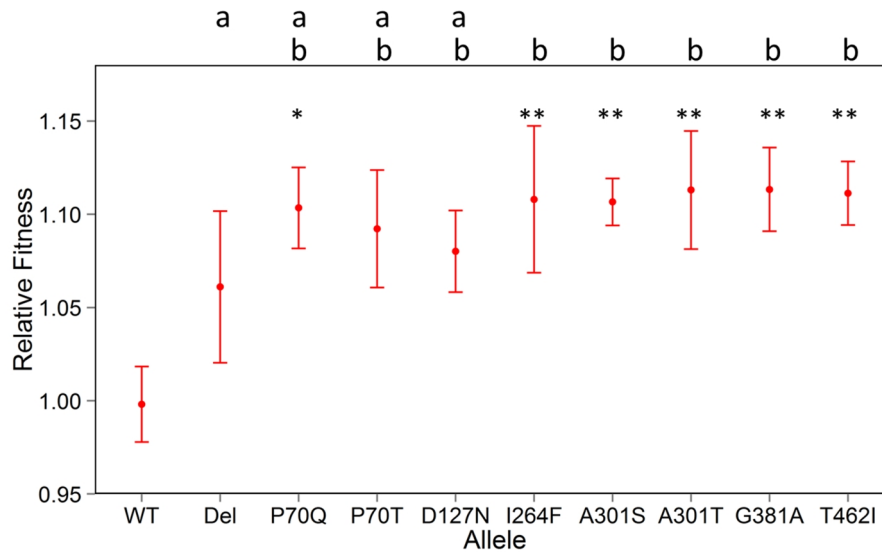
Figure 2.1. *pykF* mutations isolated from independently evolved glucose-selected populations were added to the ancestor and their fitness measured relative to the ancestor. WT indicates a control fitness estimate competing the ancestor against a derivative of itself containing a neutral screenable marker. This marked strain was used as the reference in all competitions. Points and error bars represent the mean and 95% confidence intervals of four independent fitness estimates (except that the A301S mutation effect was estimated with 12 replicates). One-way ANOVA found no detectable fitness difference ($F_{7, 32} = 1.67$, $P = 0.15$) among point mutations, but did identify a significant difference between effects if the Del mutation was included in the analysis ($F_{8, 35} = 3.71$, $P < 0.01$). Points sharing common letters indicate no significant difference was detected between the alleles (Tukey's HSD test). Asterisks reflect results of Dunnett's test comparing fitness conferred by point mutations against the Del mutation: ** $P < 0.01$, * $P < 0.05$.

2.2.2 Epistatic interactions with different genetic backgrounds change the fitness effects of pykF mutations

2.2.2.1 Effects of pykF mutations on fitness at 20,000 generations

In spite of having indistinguishable fitness effects in the ancestor, the *pykF* mutations may have different effects in their own evolved clones due to the epistatic interactions with different genetic backgrounds. To test this, I measured the fitness conferred by each *pykF* mutation in the evolved background in which it was identified. To do this, I

reverted the evolved *pykF* allele in each evolved clone and competed each reverted strain against its corresponding progenitor clone. Fitness effects were variable, ranging from 0% to 25% (one-way ANOVA: $F_{9, 30} = 27.89$, $P < 0.001$; Figure 2.2), indicating a variety of epistatic interactions between each mutation and its evolved background. Of all mutations in *pykF*, P70T is the only mutation that becomes neutral in its evolved background. Comparison of fitness effects of mutations in the ancestor and their own evolved background showed that of the 11 mutation-background combinations, six (Del, P70Q, D127N, A301S-2, A301S-3 and T462I) had higher fitness effects in the evolved background than in the ancestor (positive epistasis); three (I264F, A301T and G381A) had the same fitness in their evolved background and ancestor (no epistasis); and two (P70T and A301S-1) had smaller fitness effects in their evolved background than ancestor (negative epistasis) (Figure 2.2). These results indicate that interactions between *pykF* mutations and their evolved background (epistasis) are very common, and apparently complicated, involving positive, negative and no epistasis.
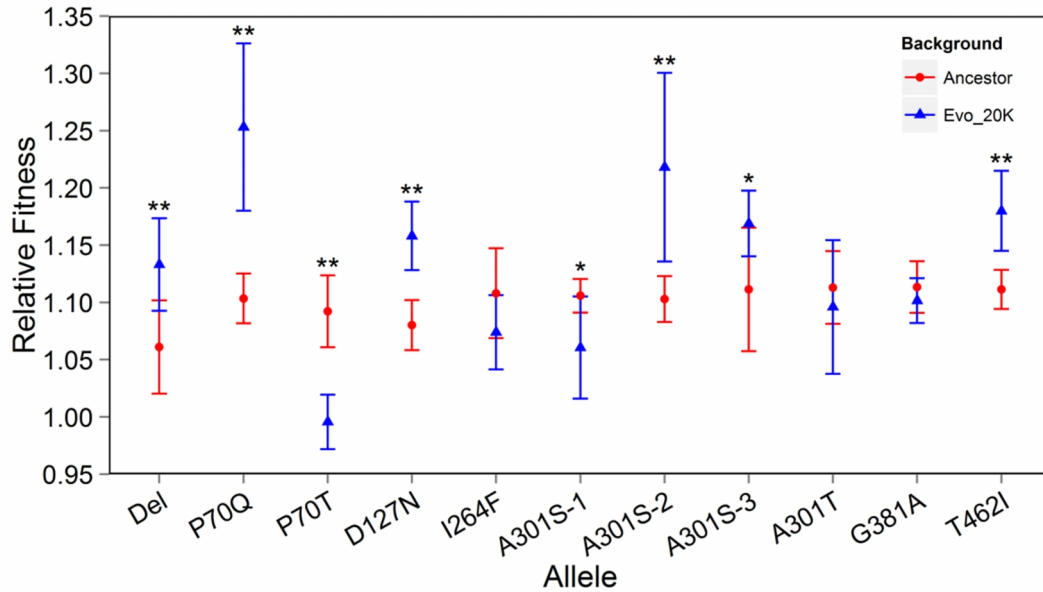
Figure 2.2. Fitness effects conferred by *pykF* mutations were measured in the ancestor (red) and in their own evolved backgrounds at 20,000 generations (blue). The A301S mutation occurred independently in three evolved clones (isolated from populations A-5, A+1 and A+5), and thus was measured in the three backgrounds, marked as A301S-1 (A-5), A301S-2 (A+1), and A301S-3 (A+5). Points and error bars represent the mean and 95% confidence intervals of independent fitness estimates (n = 4). Significant differences in fitness between ancestor and evolved background are indicated by asterisks: ** $P < 0.001$ and * $P < 0.05$ by *t*-tests. Eight pair-wise comparisons were significantly different. Of these, six remained so after sequential Bonferroni correction.

The eight *pykF* point mutations are distributed throughout the gene sequence. When mapped to the pyruvate kinase I structure, however, they cluster to three important regions: three mutations in the active site, three mutations at the tetrameric interfaces, and two mutations close to the allosteric regulation site (Figure 2.3). To test if different regions of the enzyme are responsible for the different fitness effects of mutations observed in the evolved background at 20,000 generations, I grouped mutations by enzyme functional region, and ran nested ANOVA. No detectable difference was found

among the three regions ($F_{2, 30} = 1.201$, $P = 0.315$), suggesting that the mutated region does not explain the different effects by *pykF* mutations in their own evolved background.
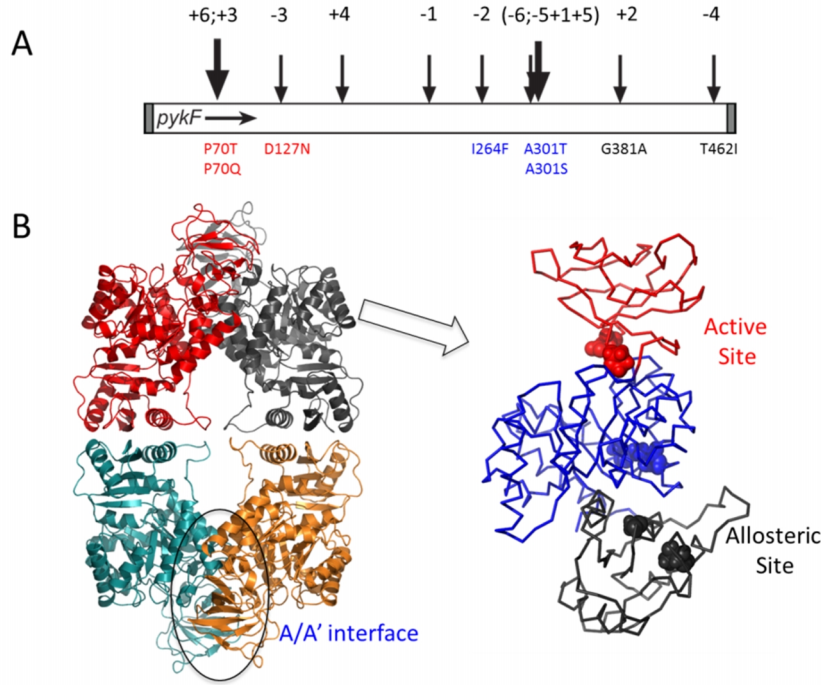
Figure 2.3. Position of *pykF* mutations identified in the 12 evolved populations mapped in primary and tertiary/quaternary protein space. **A,** Mutations are mapped to the protein sequence. The 12 populations are named as A-1 to A-6, and A+1 to A+6 (see 2.4 Materials and Methods), the number above the sequence shows the affected population (see Table 2.1); population +4 substituted a frameshift mutation, and population -1 substituted the IS*150* insertion mutation, which were not used in this study. **B,** Mutations are mapped to the enzyme structure. Pyruvate kinase type I encoded by *pykF* gene, is a homotetramer (left); each subunit has three domains: domain A (blue), domain B (red), and domain C (black) (right). Mutated residues are highlighted on the backbone model of the subunit structure: mutations P70Q, P70T, and D127N are located in the active site; I264F, A301S, and A301T are near the subunit interface (A/A' interface); G381A and T462I are in the allosteric regulation site. (This figure was created by Katherine Donovan and Renwick Dobson.)

23

2.2.2.2 Effects of pykF mutations on fitness over evolutionary time

As a population evolves, new mutations will arise and increase in frequency, thus the genetic background in which any given mutation finds itself will continuously change. An important question concerns how an earlier fixed mutation influences later adaptation. One hypothesis is earlier fixed mutations will play a more and more important role over evolution due to the dependence of later mutations on them. This predicts that the fitness effect of these early mutations will increase over time (Naumenko et al. 2012; Shah et al. 2015). To test this, I measured effects of the *pykF* mutations on fitness at 0, 20,000, 30,000, and 50,000 generations. As at 20,000 generations, mutations confer different effects in their evolved backgrounds at 30,000 (one-way ANOVA: $F_{5, 18} = 58.65$, $P <$ 0.001) and 50,000 generations (one-way ANOVA: $F_{5, 18} = 7.53$, $P < 0.001$) (Figure 2.4A). Technical constraints prevented construction of missing genotype-mutation combinations. To better examine how fitness conferred by mutations changes over evolution time, I focused on mutations whose effects were measured at all four time points (highlighted in color in Figure 2.4A). There is a fitness increase on average from ancestor to 20,000 generations (*t* test: $P < 0.001$), but a trend of decrease from 20,000 to 50,000 generations (*t* test between 20,000 and 30,000 generations: $P = 0.080$; between 30,000 and 50,000 generations: $P < 0.001$) (Figure 2.4B). One caveat here is that it is generally not known when the mutations occurred during the evolution time. Nevertheless, there is a tendency for fitness conferred by the mutations to decrease during the course of long-term evolution. The results observed in generations beyond 20,000 contradict my original

hypothesis, suggesting instead that the observed *pykF* mutations diminish in importance as evolution continues further and further.
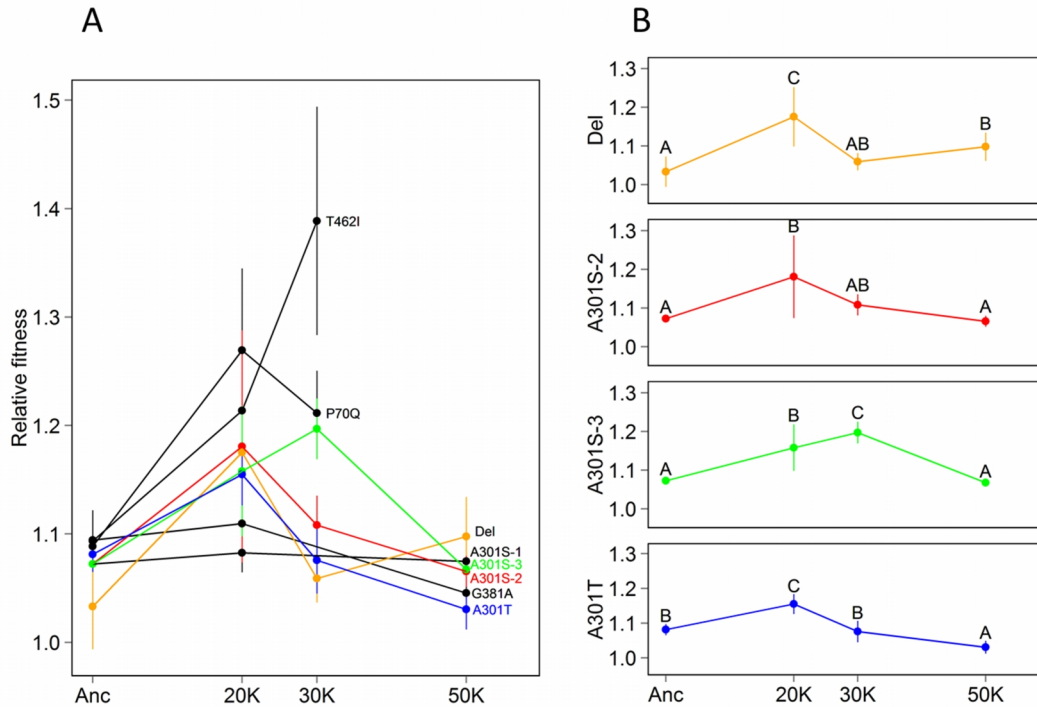
Figure 2.4. Fitness effects of mutations over evolutionary time. Points and error bars represent the mean and 95% confidence intervals of independent fitness estimates (n = 3 or 4). On x axis, "Anc", "20K", "30K", and '50K', indicate generations at 0, 20,000, 30,000 and 50000, respectively. **A,** All data. **B**, The individual fitness effect of four mutations assessed at all time points; common letters indicate no significant difference on fitness effect conferred by the same allele was detected between different generations (Tukey's HSD test).

2.2.3 Different genetic backgrounds drive the majority of differences in the epistatic interactions between different mutations and genetic backgrounds

Interactions exist between most of the *pykF* mutations (8 out of 11) and their genetic backgrounds at 20,000 generations. In principle, this interaction could be driven by the different mutations, by the different genetic backgrounds, or by some combination of

both. Considering mutations that have different effects in the ancestor and their own evolved background, finding that different mutations have similar effects in a given evolved background would support the possibility that the genetic background drives genetic interactions (Figure 2.5 A). By contrast, finding that they have similar effects in their own and in alternative evolved backgrounds would support the possibility that the mutation drives interaction effects (Figure 2.5B).
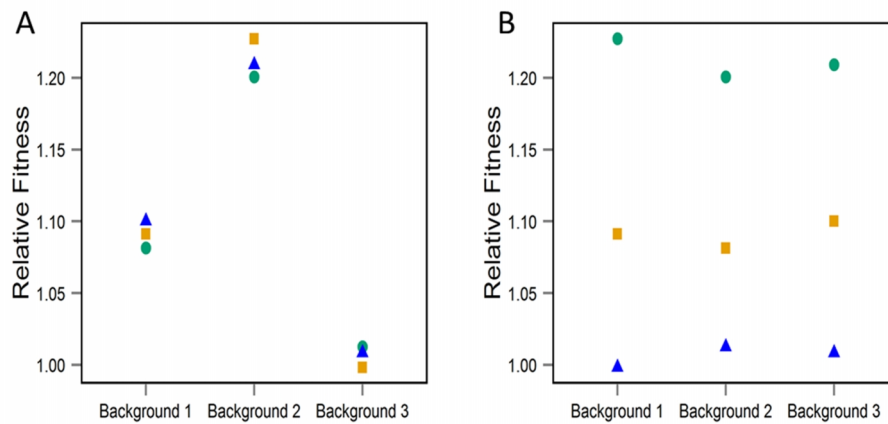
Figure 2.5. Mutation × background interactions. Three mutations are represented as different symbols and the panels show their fitness effects in three evolved genetic backgrounds. **A,** if interaction effects are driven by the genetic background, the fitness effect of mutations will differ between backgrounds, but not between one another in the same background. **B,** if interaction effects are driven by the different mutations, the fitness effect of a given mutation will not differ between genetic backgrounds, but different mutations will confer different effects in the same background.

To distinguish whether the different mutations or different genetic backgrounds drive the *pykF* mutation × background interactions, I selected two focal *pykF* mutations: Del and A301S, and moved them into the alternative 20,000-generation evolved backgrounds to construct two sets of evolved strains: one carrying the A301S mutation, and the other

carrying the Del mutation. Effects of the original, Del and A301S mutations in each background were estimated relative to the corresponding background carrying the wild type *pykF* allele (Figure 2.6). I found that the fitness effects of the same mutation across different backgrounds were significantly different (ANOVA for Del: $F_{10, 22} = 27.12$, $P < 0.001$; ANOVA for A301S: $F_{10, 22} = 20.28$, $P < 0.001$). However, different mutations have similar fitness effects when added to the same evolved background (See Table 2.2 for results of one-way ANOVAs). Genetic background explains most (~ 88%) of the variation in fitness whereas the variation explained by different *pykF* alleles is negligible (~2%) compared to statistical noise (~8%) (Table 2.3). These results support that the differences in epistatic interactions are more likely driven by different backgrounds, rather than different mutations.
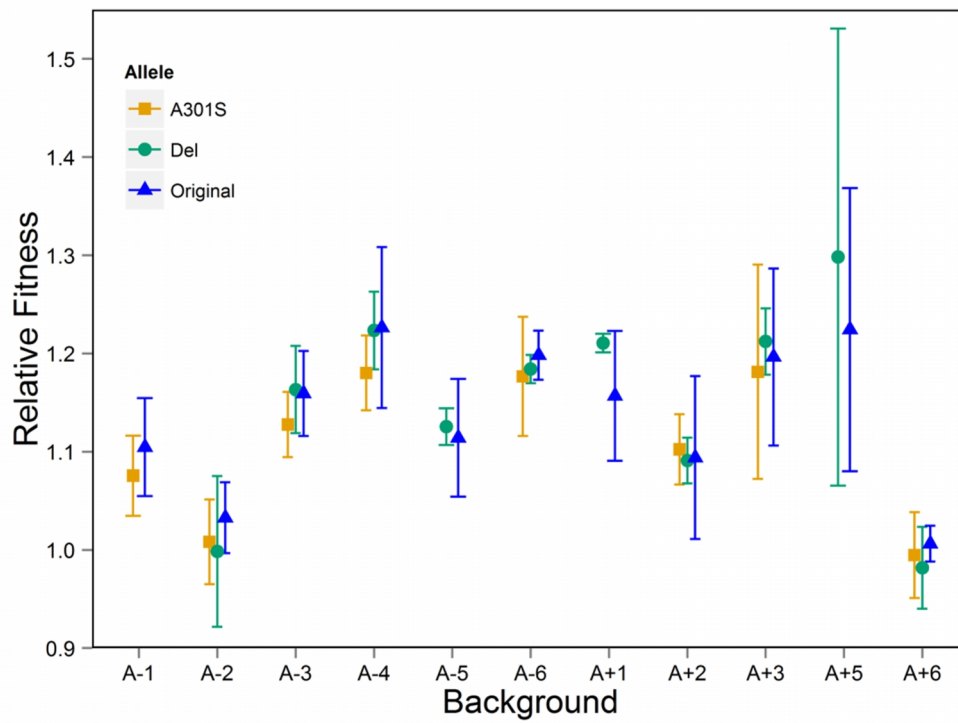
Figure 2.6. Fitness effects conferred by two focal mutations (Del and A301S) across the evolved backgrounds. "Original" indicates the *pykF* mutation identified from that evolved background. Points and error bars represent the mean and 95% confidence intervals of three independent fitness measurements. Del is the original allele in background A-1, and A301S allele is the original allele in backgrounds A-5, A+1, and A+5, thus only two points were plotted in these backgrounds.

Table 2.2. One-way ANOVA testing for differences in fitness effects of different mutations in the same genetic background.

Background	ANOVA (*P* value)
A-1	0.125
A-2	0.227
A-3	0.070
A-4	0.085
A-5	0.478
A-6	0.300
A+1	0.070
A+2	0.807
A+3	0.566
A+5	0.310
A+6	0.202

Table 2.3. Two-way ANOVA testing for fitness effects of the same mutations across the evolved backgrounds.

Factor	Df	SS	MS	*F*	*P*	variation explained
Background	10	0.4282	0.0428	46.98	< 0.0001	87.55%
Mutation	1	0.0086	0.0086	9.46	0.0036	1.76%
Background × Mutation	10	0.0122	0.0012	1.34	0.2410	2.49%
Residuals	44	0.0401	0.0009			8.20%

2.3 Discussion

An important question in evolutionary biology concerns the extent to which similar, but not identical, genetic changes might drive phenotypic parallelism (Wood et al. 2005). Consequences of adaptation, such as rate of adaptation, have been well studied, at least in controlled lab conditions, as well as in some natural experiments like longitudinal sampling of stable chronic infections (Orr and Coyne 1992; Orr 2005; Kawecki et al. 2012). Rapid developments in sequencing technology have enabled fast identification of genetic changes occurring during adaptation. There are still, however, few examples documenting the genetic basis of adaptation and, especially, the mapping between genotype and phenotype. Repeated fixation of mutations in the same gene is a signature of adaptation, which has been identified in many studies with natural and lab-evolved populations. They are the target of study on adaptation and provide a good opportunity to study the similarities and differences on phenotypic effects that genetic changes in a single gene can have. For instance, Ostrowski *et al.* (2008) found that independent lineages evolving in the same environment tend to have early mutations fixed in one of a few target genes, and the direct and correlated fitness effects of mutations in the same gene tended to be similar. However, that study could not rule out that other mutations might accompany the identified mutations, and the similarity of effects of mutations in the same gene need to be further explored at higher resolution. For example, while two adaptive mutations occurring in the gene *fis* conferred similar effects in the ancestor (Crozat et al. 2010), in another study, two different evolved alleles of the *topA* gene conferred different immediate effects on fitness and different interactions with other

mutations so that they altered potential for future adaptation (Woods et al. 2011). The small number of alleles examined makes it impossible to draw any general conclusion about the effects of selected mutations in the same gene on phenotypes.

The eight *pykF* mutations I examine here are distributed throughout the protein structure and might be expected to have different phenotypic effects since they affect different functional regions of the pyruvate kinase enzyme (Mattevi et al. 1995; Muñoz and Ponce 2003). However, I found that they had similar fitness effects when measured in the same background. This finding is consistent with studies that found mutations in the same gene have similar fitness effects (Ostrowski et al. 2008; Crozat et al. 2010; Gerstein et al. 2012). The parallelism pattern found in the effects of point mutations introduced to the ancestor may further constrain the evolutionary paths, and increase the predictability of adaptation. In addition to point mutations, one IS*150* mutation was identified in one of the isolated clones. In fact, there are many ways to produce loss-of-function alleles (deletion, transposition, point mutations to create early stop codons, and frame shift mutations). Thus loss-of-function alleles are more likely to occur as genetic materials acted on by natural selection. Indeed, adaptive loss-of-function alleles are commonly found in adaptive evolution experiments (Schneider and Lenski 2004; Chou et al. 2009; Gerstein et al. 2012). In this study, a deletion allele previously found to be equivalent to the evolved IS*150* insertion allele, conferred a lower benefit (~6%) in the ancestor than did the point mutations (~10%). The difference in fitness effects of the deletion and point mutation alleles indicates that these point mutations are not simply loss of function mutations. This conclusion is supported by the biochemical study of the evolved PykF

enzymes that demonstrated that they all retain at least partial enzymatic function. Indeed, some have considerably higher activity than the wild type enzyme (Pers. Comm. Katherine Donovan and Renwick Dobson).

Although mutations in *pykF* confer similar benefits in the ancestral genetic background, their effects were much more variable when compared over the different evolved genetic backgrounds in which they were isolated (20,000 generations). The interactions between mutations and their evolved background (epistasis) were very common, but were not consistent: positive, negative and no epistasis were all observed. This finding is different from predictions of a previously found pattern of diminishing returns epistasis common to many adaptive mutations interactions (Maisnier-Patin et al. 2005; Chou et al. 2011; Khan et al. 2011). This pattern predicts that *pykF* mutations will confer larger benefit in the less fit ancestor than in the higher fitness evolved strains. Other work predicts the relationship may be more complicated, but still expect some kind of general pattern. For example, an opposite trend from diminishing returns epistasis (Chou et al. 2009).

Moreover, the effect of epistasis changes in influencing the effect of *pykF* mutations as the population evolves newly fixed mutations. By measuring fitness conferred by focal mutations in genetic backgrounds isolated at 20,000, 30,000 and 50,000 generations, I found a tendency for fitness benefit to decrease between 20,000 and 50,000 generations, indicating less importance of the *pykF* mutations in later generations. This is different from what have been reported in previous studies that there is an increasing or conservative importance of early mutations as populations evolve (Naumenko et al. 2012;

Pollock et al. 2012; Ashenberg et al. 2013; Risso et al. 2014; Shah et al. 2015). However, those studies mostly focused on mutations in the same protein/gene, and it is generally agreed that within-gene epistasis is more synergistic compared to between-gene epistasis , which was the focus of my investigation (Lehner 2011). The declining importance of *pykF* mutations over evolution time might be explained by newly fixed redundant mutations.

One complication of interpreting the role of epistasis in influencing the effect of the *pykF* mutations is that it is generally not known exactly when and on which genetic background they occurred during the evolution of the different populations. For example, it was shown that the *pykF* mutation was fixed at around 1,500 generations in the A-1 population (Barrick et al. 2009). While the distinction between the fitness effect of the mutations in the backgrounds on which they arose vs. backgrounds prevailing at arbitrary subsequent time points is clearly relevant to some questions, for example the effect of previous mutations in making *pykF* mutations more or less likely to fix in a population, it is not relevant to questions focusing generally on the changing effect of early mutations as additional mutations are acquired by a population.

I found that differences in the mutation × background interaction are more likely driven by different backgrounds, rather than different mutations. The different genetic backgrounds, formed by different suites of mutations fixed in each strain, seem to play a more important role in influencing genetic and phenotypic trajectories of evolving populations. Though more research is needed to determine the mechanistic basis of the

epistatic effects between gene and genetic backgrounds, the variation in its magnitude suggests that epistasis has been shaped idiosyncratically by particular mutations occurring in the different evolving populations, rather than by some general process of the kind inferred from studies finding that the fitness of a strain, not its specific genotype, determines its interaction with new mutations (Chou et al. 2011; Khan et al. 2011; Kryazhimskiy et al. 2014). Overall, my results suggest that gene-level parallel genetic changes are consistent with predictability of adaptive evolution, but mutations at other loci complicate this predictability.

My study supports the view that it is reasonable to consider mutations in the same gene to be tentatively equivalent, due to the observation that they have similar effects in the same genetic background, no matter whether the background is ancestral or evolved. This would contribute to practical use: for example, sequencing of isolates from chronic infections where mutations in the same gene in different patients are sometimes found (Huse et al. 2010; Lieberman et al. 2011). My results indicate that mutations in the same gene might be functionally similar, suggesting that a general way may be possible to treat diseases that only have mutations in one gene. But for patients who have other mutations as well, the genetic context might complicate treatment, and should be individually considered.

2.4 Materials and Methods

Strains and mutations

REL606, the ancestor of the *E. coli* long term evolution experiment (LTEE) (Lenski et al. 1991), was used as the reference strain in this study. The LTEE consists of 12 populations: six populations were identical to REL606, which cannot grow on arabinose, named A-1 to A-6 and six populations were started from a mutant of REL606 that was selected for a mutation in *araA* that enabled growth on arabinose, named A+1 to A+6. The *araA* mutation is neutral in the glucose evolution and competition environments used here. The replicate populations were transferred daily (~6.64 generations per day) in a glucose-limited environment for more than 60,000 generations. One clone from each population was isolated and sequenced for mutated genes found following comprehensive study and sequencing of the A-1 population, and it was found that *pykF* mutations occurred in all 12 clones (Woods et al. 2006) (Table 2.1). Ten unique *pykF* mutations were identified, of which eight are point mutations. They are P70Q, P70T, D127N, I264F, A301S, A301T, G381A, T462I. The A301S mutation was identified independently in three strains (A-5, A+1, A+5). The other two *pykF* mutations are IS*150* (in practice I used a deletion allele, which is equivalent to the IS*150* insertion allele (Khan et al. 2011), to evaluate the effect of the insertion mutation, indicated as 'Del') and a frame shift mutation.

Genetic manipulation

Four series of strains were constructed: (1) the *pykF* mutations were separately added to the ancestral strain REL606 (Figure 2.7A); (2) the wild type allele was added to the evolved clones isolated from each population at 20,000 generations, replacing the original *pykF* allele (Figure 2.7B); (3) the Del allele was added to the evolved background from 20,000 generations to replace the original *pykF* allele (Figure 2.7C); (4) the A301S allele was added to the evolved background from 20,000 generations to replace the original *pykF* allele (Figure 2.7D). I also attempted to replace evolved *pykF* alleles with the wild type allele in clones isolated from all populations at 30,000 and 50,000 generations. I was unable to obtain several of these mutation-background combinations, and others failed quality control tests carried out to determine the presence of secondary mutations arising during the construction process.
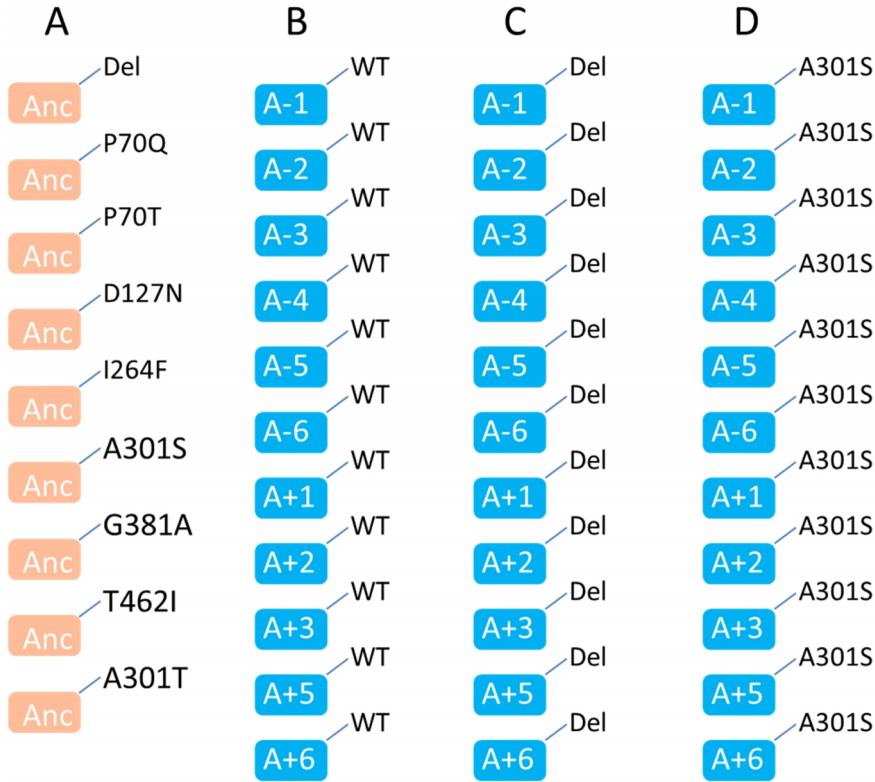
Figure 2.7. Strain constructions. Light orange and blue rectangles represent ancestral and evolved backgrounds, respectively. **A**, ancestral background, to which each evolved *pykF* mutation was individually added. **B**, **C**, **D**, evolved backgrounds, to which WT, Del, and A301S mutations were added; text in the blue rectangles represent each evolved genetic background.

Allelic replacement (Philippe et al. 2004) was used to exchange alleles on the chromosome. The approach has been described previously (Khan et al. 2011). Briefly, ~800 bp PCR products centered on each of the *pykF* mutations were cloned into to plasmid pCR2.1 using the TA cloning kit (Invitrogen), and transformed into TOP10F′ *E. coli* cells (Invitrogen). Plasmid pCR2.1 derivatives were digested with *XbaI* and *SacI* (NEB) to liberate the cloned fragment, which was ligated into the similarly digested

suicide vector, pDS132 (Philippe et al. 2004). Ligated plasmids were used to transform MFD*pir*. Conjugation was carried out between MFD*pir* cells containing pDS132 plasmid with a specific *pykF* allele and recipient cells (Ferrières et al. 2010). After each conjugation and subsequent two-step recombination process, I obtained at least one recipient strain pair where one strain had the replacement allele and the other was a control strain that retained its original allele and therefore should have fitness identical to the original recipient progenitor strain. Genotypes with replacement alleles were only used when their paired control strain had the same fitness as the progenitor.

Measurement of fitness effects

The fitness effects were estimated by head-to-head competitions of two strains using a neutral genetic marker in Davis-Mingioli medium supplemented with 25 µg/mL glucose (Lenski et al. 1991). This is the same medium used in the original LTEE experiment. All of the competitions were performed in 10ml of medium contained in 50 mL Erlenmeyer flasks (the evolution environment), instead of easily manipulated 96-well or 24-well blocks, because I found the fitness effects were dependent on culture vessel. *araA* markers were used for all strains except 20K A-4. For that I was not able to make a neutral *araA*$^+$ derivative of 20K A-4 and instead constructed a neutral *lacZ* marker. Competitions were carried out by growing two competitors overnight in LB liquid culture, transferring them to the evolution environment, and the next day transferring again at a diluted ratio of 1:100. After two daily transfers in this preconditioning environment, the two competitors were mixed at 1:1 ratio to start the competition. The cell density of both

competitors was obtained by plating appropriate amount of mixture to tetrazolium arabinose (TA) plates, on which Ara^+ and Ara^- cells form white and red colonies, respectively (for strains with *lacZ* markers, I used LB + Xgal + IPTG plates to distinguish the two competitors). After two transfer cycles of competition, an appropriate amount of mixture was plated again on TA (or LB + Xgal + IPTG) plates to obtain final cell densities of each competitor. The relative fitness was calculated as the ratio of the Malthusian parameter of one competitor to another (Lenski et al. 1991).

Statistical analyses

Plots and statistical analyses were all done in R version 3.1.3 (R Core Team 2015). All fitness competitions were performed as complete experimental blocks. One-way ANOVA analysis was performed to test for differences in fitness effects of *pykF* mutations in the ancestor and across different evolved backgrounds. Nested ANOVA was performed to test if effects from the three enzyme functional regions could explain different effects of mutations found in the evolved background. Two-way ANOVA was performed to test for the effects of the same mutations and genetic backgrounds on fitness. In all analysis, mutation and genetic background were treated as fixed effects.

2.5 References

Ashenberg, O., L. I. Gong, and J. D. Bloom. 2013. Mutational effects on stability are largely conserved during protein evolution. Proc. Natl. Acad. Sci. USA 110:21071–21076.

Barrick, J. E., D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, D. Schneider, R. E. Lenski, and J. F. Kim. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature 461:1243–1247.

Bull, J. J., M. R. Badgett, H. A. Wichman, J. P. Huelsenbeck, D. M. Hillis, A. Gulati, C. Ho, and I. J. Molineux. 1997. Exceptional convergent evolution in a virus. Genetics 147:1497–1507.

Chou, H. H., J. Berthet, and C. J. Marx. 2009. Fast growth increases the selective advantage of a mutation arising recurrently during evolution under metal limitation. PLoS Genet. 5:e1000652.

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

Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarreal Jr., M. Dickson, J. Grimwood, J. Schmutz, R. M. Myers, D. Schluter, and D. M. Kingsley. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. Science 307:1928–1933.

Cooper, T. F., D. E. Rozen, and R. E. Lenski. 2003. Parallel changes in gene expression after 20, 000 generations of evolution in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 100:1072–1077.

Cooper, V. S., D. Schneider, M. Blot, and R. E. Lenski. 2001. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. J. Bacteriol. 183:2834–2841.

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

Crozat, E., C. Winkworth, J. Gaffé, P. F. Hallin, M. A. Riley, R. E. Lenski, and D. Schneider. 2010. Parallel genetic and phenotypic evolution of DNA superhelicity in experimental populations of *Escherichia coli*. Mol. Biol. Evol. 27:2113–2128.

Ferea, T. L., D. Botstein, P. O. Brown, and R. F. Rosenzweig. 1999. Systematic changes in gene expression patterns following adaptive evolution in yeast. Proc. Natl. Acad. Sci. USA 96:9721–9726.

Ferrières, L., G. Hémery, T. Nham, A. M. Guérout, D. Mazel, C. Beloin, and J. M. Ghigo. 2010. Silent mischief: bacteriophage Mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. J. Bacteriol. 192:6418–6427.

Gerstein, A. C., D. S. Lo, and S. P. Otto. 2012. Parallel genetic changes and nonparallel gene-environment interactions characterize the evolution of drug resistance in yeast. Genetics 192:241–252.

Gompel, N., and B. Prud'homme. 2009. The causes of repeated genetic evolution. Dev. Biol. 332:36–47.

Herring, C. D., A. Raghunathan, C. Honisch, T. Patel, M. K. Applebee, A. R. Joyce, T. J. Albert, F. R. Blattner, D. van den Boom, C. R. Cantor, and B. Ø. Palsson. 2006. Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. Nat. Genet. 38:1406–1412.

Huse, H. K., T. Kwon, J. E. A. Zlosnik, D. P. Speert, E. M. Marcotte, and M. Whiteley. 2010. Parallel evolution in *Pseudomonas aeruginosa* over 39,000 generations *in vivo*. MBio 1:e00199–10.

Kawecki, T. J., R. E. Lenski, D. Ebert, B. Hollis, I. Olivieri, and M. C. Whitlock. 2012. Experimental evolution. Trends Ecol. Evol. 27:547–560.

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

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

Le Gac, M., T. F. Cooper, S. Cruveiller, C. Médigue, and D. Schneider. 2013. Evolutionary history and genetic parallelism affect correlated responses to evolution. Mol. Ecol. 22:3292–3303.

Lehner, B. 2011. Molecular mechanisms of epistasis within and between genes. Trends Genet. 27:323–331.

Lenski, R. E., J. A. Mongold, P. D. Sniegowski, M. Travisano, F. Vasi, P. J. Gerrish, and T. M. Schmidt. 1998. Evolution of competitive fitness in experimental populations of *E. coli* : what makes one genotype a better competitor than another? Antonie Van Leeuwenhoek 73:35–47.

Lenski, R. E., M. R. Rose, S. C. Simpson, and S. C. Tadler. 1991. Long-term experimental evolution in *Escherichia coli*. I. adaptation and divergence during 2,000 generations. Am. Nat. 138:1335–1341.

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

Lieberman, T. D., J. B. Michel, M. Aingaran, G. Potter-Bynoe, D. Roux, M. R. Davis Jr, D. Skurnik, N. Leiby, J. J. LiPuma, J. B. Goldberg, A. J. McAdam, G. P. Priebe, and R. Kishony. 2011. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. Nat. Genet. 43:1275–1280.

Losos, J. B., T. R. Jackman, A. Larson, K. de Queiroz, and L. Rodriguez-Schettino. 1998. Contingency and determinism in replicated adaptive radiations of island lizards. Science 279:2115–2118.

Maisnier-Patin, S., J. R. Roth, A. Fredriksson, T. Nyström, O. G. Berg, and D. I. Andersson. 2005. Genomic buffering mitigates the effects of deleterious mutations in bacteria. Nat. Genet. 37:1376–1379.

Martin, A., and V. Orgogozo. 2013. The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. Evolution 67:1235–1250.

Mattevi, A., G. Valentini, M. Rizzi, M. L. Speranza, M. Bolognesi, and A. Coda. 1995. Crystal structure of *Escherichia coli* pyruvate kinase type I: molecular basis of the allosteric transition. Structure 3:729–741.

Moore, F. B.-G., and R. Woods. 2006. Tempo and constraint of adaptive evolution in *Escherichia coli* (Enterobacteriaceae , Enterobacteriales). Biol. J. Linn. Soc. 88:403–411.

Muñoz, M. E., and E. Ponce. 2003. Pyruvate kinase: current status of regulatory and functional properties. Comp. Biochem. Physiol. Part B Biochem. Mol. Biol. 135:197–218.

Nachman, M. W., H. E. Hoekstra, and S. L. D'Agostino. 2003. The genetic basis of adaptive melanism in pocket mice. Proc. Natl. Acad. Sci. USA 100:5268–5273.

Naumenko, S. A., A. S. Kondrashov, and G. A. Bazykin. 2012. Fitness conferred by replaced amino acids declines with time. Biol. Lett. 8:825–828.

Orr, H. A. 2005. The genetic theory of adaptation: a brief history. Nat. Rev. Genet. 6:119–127.

Orr, H. A., and J. A. Coyne. 1992. The genetics of adaptation : a reassessment. Am. Nat. 140:725–742.

Ostrowski, E. A., R. J. Woods, and R. E. Lenski. 2008. The genetic basis of parallel and divergent phenotypic responses in evolving populations of *Escherichia coli*. Proc. Biol. Sci. 275:277–284.

Pelosi, L., L. Kühn, D. Guetta, J. Garin, J. Geiselmann, R. E. Lenski, and D. Schneider. 2006. Parallel changes in global protein profiles during long-term experimental evolution in *Escherichia coli*. Genetics 173:1851–1869.

Philippe, N., J. P. Alcaraz, E. Coursange, J. Geiselmann, and D. Schneider. 2004. Improvement of pCVD442, a suicide plasmid for gene allele exchange in bacteria. Plasmid 51:246–255.

Phillips, K., G. Castillo, A. Wünsche, and T. F. Cooper. 2015. Adaptation of *E. coli* to glucose promotes evolvability in lactose. Submitt. to Evol.

Pollock, D. D., G. Thiltgen, and R. A. Goldstein. 2012. Amino acid coevolution induces an evolutionary Stokes shift. Proc. Natl. Acad. Sci. USA 109:E1352–E1359.

Prud'homme, B., N. Gompel, A. Rokas, V. A. Kassner, T. M. Williams, S. D. Yeh, J. R. True, and S. B. Carroll. 2006. Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. Nature 440:1050–1053.

R Core Team. 2015. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria.

Risso, V. A., F. Manssour-Triedo, A. Delgado-Delgado, R. Arco, A. Barroso-delJesus, A. Ingles-Prieto, R. Godoy-Ruiz, J. A. Gavira, E. A. Gaucher, B. Ibarra-Molero, and J. M. Sanchez-Ruiz. 2014. Mutational studies on resurrected ancestral proteins reveal conservation of site-specific amino acid preferences throughout evolutionary history. Mol. Biol. Evol. 32:440–455.

Rundle, H. D., L. Nagel, J. W. Boughman, and D. Schluter. 2000. Natural selection and parallel speciation in sympatric sticklebacks. Science 287:306–308.

Schluter, D., E. A. Clifford, M. Nemethy, and J. S. McKinnon. 2004. Parallel evolution and inheritance of quantitative traits. Am. Nat. 163:809–822.

Schneider, D., and R. E. Lenski. 2004. Dynamics of insertion sequence elements during experimental evolution of bacteria. Res. Microbiol. 155:319–327.

Shah, P., D. M. McCandlish, and J. B. Plotkin. 2015. Contingency and entrenchment in protein evolution under purifying selection. Proc. Natl. Acad. Sci. USA 112:E3226–E3235.

Stewart, C. B., J. W. Schilling, and A. C. Wilson. 1987. Adaptive evolution in the stomach lysozymes of foregut fermenters. Nature 330:401–404.

Tenaillon, O., A. Rodríguez-Verdugo, R. L. Gaut, P. McDonald, A. F. Bennett, A. D. Long, and B. S. Gaut. 2012. The molecular diversity of adaptive convergence. Science 335:457–461.

Vogwill, T., M. Kojadinovic, V. Furio, and R. C. MacLean. 2014. Testing the role of genetic background in parallel evolution using the comparative experimental evolution of antibiotic resistance. Mol. Biol. Evol. 31:3314–3323.

Wichman, H. A., M. R. Badgett, L. A. Scott, C. M. Boulianne, and J. J. Bull. 1999. Different trajectories of parallel evolution during viral adaptation. Science 285:422–424.

Wood, T. E., J. M. Burke, and L. H. Rieseberg. 2005. Parallel genotypic adaptation: when evolution repeats itself. Genetica 123:157–170.

Woods, R. J., J. E. Barrick, T. F. Cooper, U. Shrestha, M. R. Kauth, and R. E. Lenski. 2011. Second-order selection for evolvability in a large *Escherichia coli* population. Science 331:1433–1436.

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

Chapter 3 Mechanistic basis underlying phenotypic parallelism caused by parallel genetic changes

Even when the fitness effect of an adaptive mutation can be estimated, it has usually not been determined *how* the mutation causes the fitness effect that it does. I have previously shown that independently selected adaptive mutations in *pykF* conferred similar fitness effects when measured in a defined background. It is unknown, however, if these parallel mutations affect fitness through the same underlying mechanism. Although this may seem likely, biochemical studies of the evolved PykF enzymes reveal a wide array of changes in enzyme kinetics that include increases and decreases in activity and stability relative to the wild type enzyme, so that it cannot be assumed that they have the same mechanism of action. To examine the basis of the benefit conferred by mutations in *pykF* I propose and test a mechanistic hypothesis whereby mutations in PykF act to increase stationary phase concentrations of PEP, which in turn facilitates more rapid uptake of glucose when cells are transferred to fresh medium. Consistent with predictions of this hypothesis: (1) *pykF* mutations tended to confer greater benefits in substrates requiring PEP for uptake than in alternative substrates, (2) stationary phase PEP levels are lower in the wild type strain than in strains containing any of the different evolved PykF enzymes, and (3) the *pykF* mutations confer most of their benefit through shortening the lag time inherent in each daily growth cycle, not by increasing cell growth rate. My results suggest that, while parallel mutations change the *in vitro* biochemical properties of PykF in a range of different ways, the net effect of these changes is to confer similar adaptive changes on cell metabolism.

3.1 Introduction

Parallel genetic changes are a signature of positive selection (Woods et al. 2006; Barrick et al. 2009), but typically do not speak to why the gene became the target of selection. To understand this, we need to have the knowledge of the mechanisms underlying the fitness benefit, *i.e.*, how mutations in a gene act to change the phenotypes (Orr 2005). In addition, studying beneficial mutations as identified through natural selection provides access to very specific mutant alleles that are unlikely to be found through the kind of random mutational screens usually used to create a mutant library. For this reason the spectrum of phenotypes among beneficial mutations may be quite different – for example, being enriched with variants with new functional attributes – than that of random mutants (Carroll & Marx 2013). My previous studies indicated that parallel mutations in *pykF* conferred similar fitness benefits when measured in a focal genetic background. Here, I examine the mechanisms underlying the fitness advantages, and the extent to which parallelism at the level of organism fitness extends to changes in protein function and physiology.

Pyruvate kinase I (PykF) is involved in glycolysis, a well-characterized metabolic pathway (Muñoz and Ponce 2003), providing a good opportunity to study how changes in its function or activity confer benefits to cells. PykF catalyzes the transfer of one phosphate group from phosphoenolpyruvate (PEP) to ADP, generating ATP and pyruvate. PEP is not only the substrate of PykF, but is also used by the phosphotransferase system (PTS) to import glucose into cells (Postma et al. 1993) (See Figure 1.2, page 7). One hypothesis has been proposed for the fitness benefit of *pykF* mutations that were selected

in the Lenski Long-Term Evolution Experiment: by decreasing the enzyme activity, mutations in *pykF* slow the conversion of PEP to pyruvate, and thus make more PEP available to drive the import of limiting resource, glucose, into cells (Schneider et al. 2000; Woods et al. 2006).

A recent study (Xu et al. 2012) demonstrated that a sufficient intra-cellular levels of PEP is necessary for *E. coli* cells to quickly respond to fluctuations in glucose concentration: when cells were given new glucose-containing media after a period of starvation, a higher PEP concentration was associated with more rapid import of glucose and shorter lag times before resumption of exponential growth. In that study, PEP accumulation was found to be associated with the ultrasensitive allosteric regulation of PEP carboxylase by fructose-1, 6-bisphosphate (FBP), a chemical intermediate in the early part of the glycolysis pathway. PEP carboxylase, like PykF, converts PEP to intermediates that feed the tricarboxylic acid (TCA) cycle. When glucose was exhausted from the medium, FBP levels quickly dropped causing a rapid decrease in the activity of PEP carboxylase. This decrease in activity stopped the consumption of PEP, conserving an intra-cellular store that could be used to prime initial uptake of glucose when glucose was reintroduced to the culture medium. PykF is also allosterically activated by FBP (Zhu et al. 2010). I hypothesize that this dependence is affected by the evolved mutations to reduce PykF enzyme activity at the low FBP prevailing during glucose starvation. A lower PykF activity will preserve its substrate, PEP, allowing more to be available to facilitate glucose import following transfer to fresh medium. More rapid initial glucose import will

help cells quickly adapt to the fresh environment, shortening their lag times and contributing to a fitness advantage.

In this study, I tested three key predictions of the hypothesis outlined above: (1) that the benefit of *pykF* mutations would be greater in resources that depend on PEP for their transport into cells, (2) that evolved mutations in *pykF* lead to higher levels of intra-cellular PEP at stationary phase, and (3) that the benefits of evolved PykF mutations occur during the lag phase or transition from lag to log growth phases, not to an increase in maximum cell growth rate. In all cases, my results were consistent with these predictions.

3.2 Results

3.2.1 pykF mutations affect PTS – a sugar transport system

PEP is not only the substrate of PykF, but is also used as a source of energy to drive the phosphotransferase system (PTS), a system required for the transfer of some carbon resources into the cell. If mutations in *pykF* affect stationary phase levels of PEP, I expect them to have a general effect on the initial translocation of PTS resources and therefore for *pykF* mutants to generally have higher fitness in PTS than non-PTS resource environments. I note that *pykF* mutations may well have pleiotropic effects that cause fitness to vary across environments, so this prediction does not imply that the effects of the mutation will be beneficial in all PTS environments. To test if mutations in *pykF* confer fitness advantages by affecting the PTS pathway, I measured the effect of one of the mutations (*pykF::IS150*) in five PTS and six non-PTS resources in the genetic

background where it occurred and was fixed (Figure 3.1). I found the *pykF*::IS*150* mutation has a pleiotropic effect across environments: it is neutral in most of the PTS environments (mean ± SEM: 0.981 ± 0.015), but has different deleterious effect in most of the non-PTS environments (mean ± SEM: 0.877 ± 0.022). I ran a two-way ANOVA with environment nested within resource type (PTS or non-PTS), and found a significant difference between PTS and non-PTS resources in affecting fitness of the mutation ($F_{1,\,32}$ = 302.9, $P < 0.001$). This result is consistent with my hypothesis that *pykF* mutations tended to confer greater fitness in environments requiring PEP for uptake than in alternative substrates.
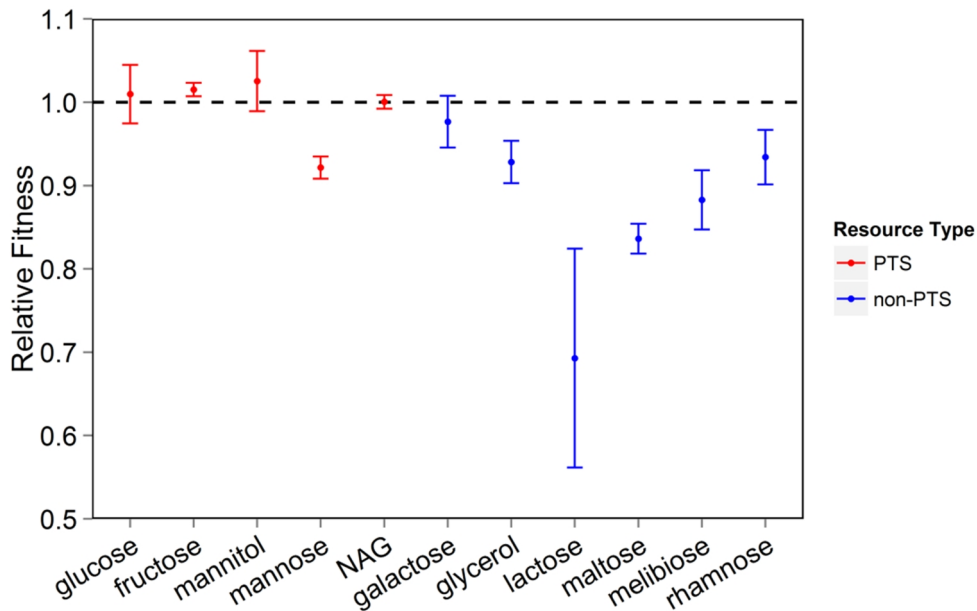
Figure 3.1. Fitness effect of *pykF*::IS*150* mutation in 11 resources. Five are PTS resources (red): glucose, fructose, mannitol, mannose, and NAG; six are non-PTS resources (blue): galactose, glycerol, lactose, maltose, melibiose, and rhamnose. The neutral effect of *pykF*::IS*150* in glucose seems contradictory with the previous 6% benefit shown in chapter 2, because the fitness assays here were performed in 96-well plates (see Materials and methods) rather than flasks.

3.2.2 Shorter lag times by pykF mutations might contribute to most of the fitness benefits

My hypothesis for the benefit of *pykF* mutations predicts that this benefit results from a shortening of population lag times, not an increase in growth rate. In the original evolution experiment, *E. coli* populations were transferred to fresh glucose-limited media every 24 hours. This regime results in a distinct growth cycle, of which two components are of interest to us. One is the lag time, the time taken for bacteria to begin growth following transfer to fresh media. The other is the log growth phase, the period of growth that can be quantified by determining the maximum growth rate. Fitness advantages can be caused by shorter lag times, a greater maximum growth rate, or both.

First, I determined growth curves of the ancestor strains carrying each of the three alleles (WT, Del, A301S) to obtain preliminary estimates of the times that strains spend in lag phase and log phase during the regular daily growth cycle. Due to their high throughput and precision, nowadays researchers prefer to estimate growth curves using a plate reader to measure optical density (OD) to track population growth. However, in my experimental setup the glucose concentration is too low to be detected by using this method. Therefore, to obtain better, though lower throughput, measures of growth curves, I measured them for representative genotypes in the exact same environment used in the evolution experiment. I did this by using serial dilutions and standard plate counts at different time points (Figure 3.2 upper panel). Logistic regressions were employed to fit the data to obtain lag times for strains encoding each of three different evolved *pykF* alleles. Unfortunately, I got poor goodness of fit measures for my data. Therefore, I used a simple and robust way to determine the lag phase: considering the time taken for a

population's cell density to double as the lag phase (Buchanan and Solberg 1972). The lag times for the three strains with WT, Del, and A301S allele are estimated to be close to two hours. Due to the high variation between replicates, I was not able to tell the difference in the lag times between strains with different alleles (ANOVA: $F_{2,15} = 0.16$, $P = 0.85$).

To directly measure when *pykF* mutations start to have advantage over the wild type allele in a 24-h growth cycle, I ran separate fitness assays competing the ancestor strain against otherwise isogenic derivative strains encoding each of the mutations, and measured fitness effects conferred by each mutation at different time points throughout a 24 hour competition (Figure 3.2 lower panel). A significant fitness benefit conferred by point mutations was detected after 6h competition (*t*-test: A301S, $P = 0.003$; G381A, $P = 0.001$; P70T, $P < 0.001$). Previous results show that lag times of the strains are ~2 hours, however, I was not able to obtain fitness measurements at this time point. Therefore, it is impossible to conclude directly from the results that shorter lag times contribute to their fitness advantage. As mentioned previously, the mutations can either shorten lag time or increase growth rate to cause benefit to cells. The log growth phase was not ended until ~8 hours (Figure 3.2 upper panel), but the fitness effects of the point mutations did not change from 5h to 24h (ANOVA: A301S, $F_{5,22} = 0.67$, $P = 0.65$; G381AA, $F_{5,22} = 0.59$, $P = 0.71$; P70T, $F_{5,22} = 0.44$, $P = 0.82$;). That means the log growth phase is very unlikely to contribute to the fitness benefit, thus most of the fitness benefits of *pykF* mutations are likely to be conferred by shortening lag phase or perhaps by accelerating the transition between lag phase and maximum growth rate.
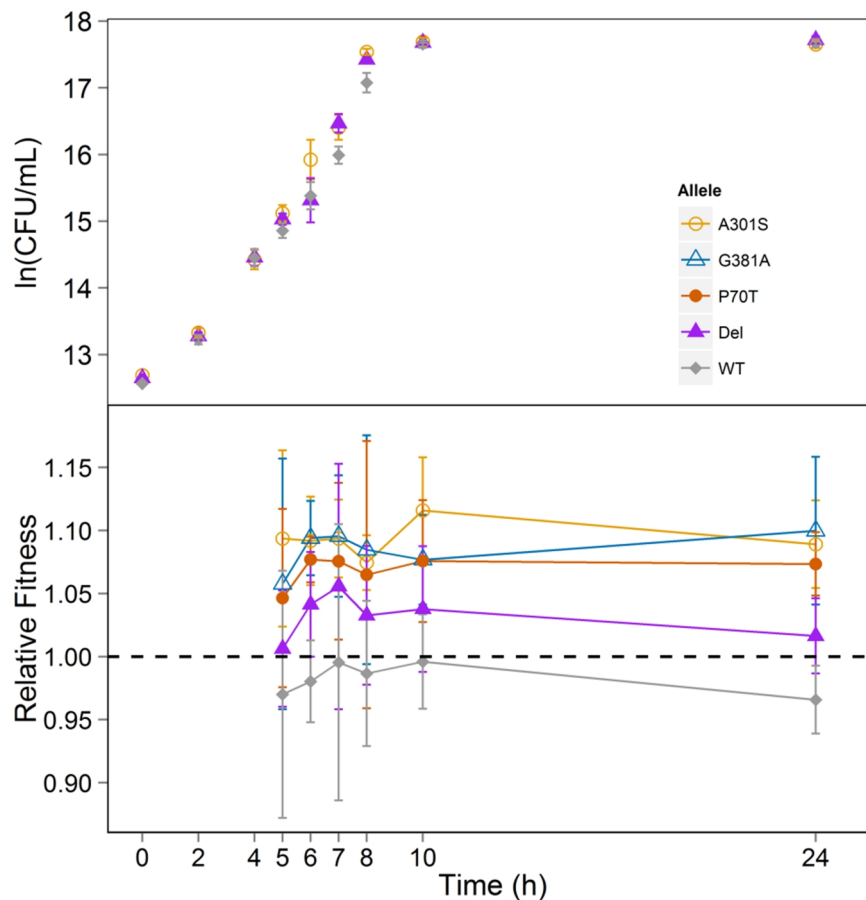
Figure 3.2. Growth and fitness measurements of the ancestral strain and derivatives encoding one of four different *pykF* alleles over 24 hours. Upper panel shows the growth measurement for ancestral strains with WT, Del, and A301S allele, and error bars indicate 95% confidence intervals of independent CFU estimates (n=6). Lower panel shows the fitness effects of *pykF* alleles at different time points. Symbols and error bars represent the mean and 95% confidence intervals of four independent measurements.

3.2.3 Mutational benefits might be due to accumulation of PEP

Finally, I tested the prediction that mutations in *pykF* cause the enzyme to be less active in stationary phase, leading to an increase in PEP. PEP is the substrate of pyruvate kinase and also drives glucose import into the cell through the PTS. An increase in PEP levels at the time cells are transferred to fresh media is hypothesized to be the basis of

the fitness advantages characterized in the results presented above. I found that PEP levels in a strain with the WT PykF enzyme was lower than in strains containing any of the evolved PykF enzymes (Figure 3.3). In five cases these differences were significant at $P < 0.1$ (Dunnett's test).
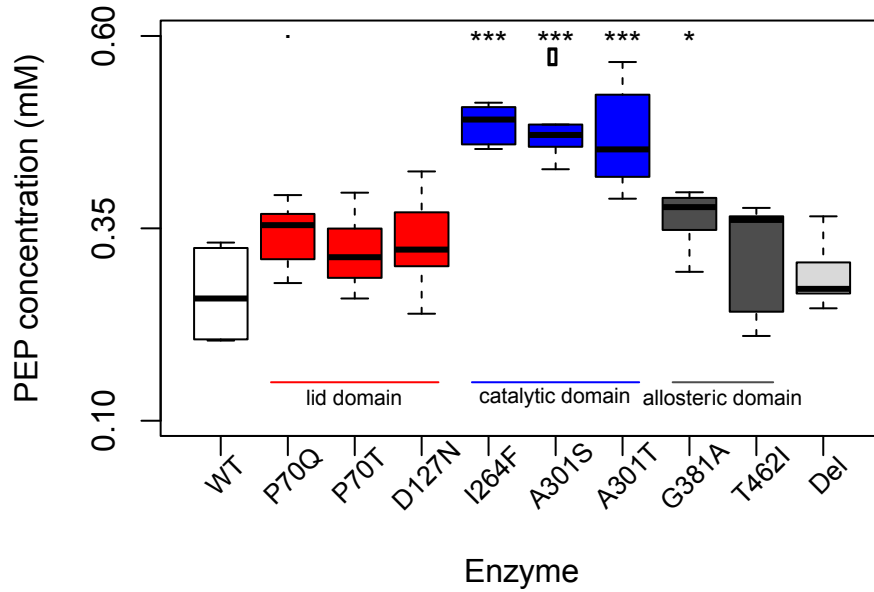
Figure 3.3. Box and whisker plot indicating PEP concentration of otherwise isogenic strains expressing the indicated PykF enzyme ('Del' is the deletion allele). Strains were sampled following 24 hours of growth in the evolution environment. Boxes indicate first and third quartiles, the central line indicates the median, and whiskers indicate 95% confidence intervals of the median. Symbols at top of panel reflect results of Dunnett's test comparing strains with evolved enzymes against the WT enzyme strain: "***" <0.001, "*" 0.05, " " 0.1. Results are from six independent replicate measurements of each strain. Measurements of PEP concentration carried out by Sarah Kessens and Renwick Dobson.

3.2.4 Mutated enzymes might behave similarly as a whole

The results presented above support the hypothesis that the benefit of evolved *pykF* mutations is due to their effect on increasing PEP levels at the time cells are transferred to new media. Ideally, I would like to propose a biochemical mechanism underlying this metabolic change. To this end, our collaborators, Katherine Donovan and Renwick Dobson, measured the enzyme kinetics for purified mutated enzymes as well as for the ancestral enzyme. They found that the mutated enzymes demonstrate a wide variety of values of $k_{cat}/S_{0.5}$ (catalytic efficiency), nH (the Hill coefficient, used to provide a quantitative measure of cooperative measure of cooperativity of ligand binding), and thermal stability, suggesting mutations change the enzyme biochemical attributes in different ways (Table 3.1). I consider two possibilities to reconcile this variation with the observed uniformity of effects on fitness, growth phase, and PEP concentrations. First, it may be that many of the kinetic parameters are physiologically irrelevant, either because of differences in *in vivo* and *in vitro* conditions that impact kinetic parameter estimation or because some parameters only become relevant in conditions outside physiologically relevant ranges (*e.g.*, differences in enzyme stability apparent at temperatures over 60 °C may have no relevance during growth in 37 °C as in our whole-organism assay environments). Second, different attributes of the enzymes might integrate in a way that results in a similar net biochemical effect. The first of these possibilities is difficult to evaluate because the *in vivo* conditions relevant to PykF activity are poorly known and probably impossible to exactly replicate.

For the second possibility, I carried out principal component analysis (PCA) on the three attributes of enzyme to explore similarities between mutated enzymes (Figure 3.4). The first component (PC1) explained 62.6%, and the second (PC2) 29.9%, of the variance, respectively. PC1 was mainly a combination of $k_{cat}/S_{0.5}$ and nH, while PC2 was mainly a reflection of thermal stability. Mutations affecting the same functional group such as G381A and T462I, and P70Q and P70T, tend to group together, however, the other mutants were widely distributed in the PCA score plot. To better estimate contributions of each attribute to the enzyme, I ranked the mutated enzymes in order of each enzyme attribute (Table 3.2). It showed that for each enzyme, not all three enzyme attributes have the higher or lower ranks, but often each enzyme has a higher rank (six of them have rank 1 or 2) for one attribute and a lower rank for another. Since the three attributes ($k_{cat}/S_{0.5}$, nH and thermal stability) affect the enzyme in the same direction, the lower value found in one of the attributes, but higher value in another, suggest that the point mutations might change the enzyme in different ways, with some affect $k_{cat}/S_{0.5}$ more, some nH more, and some thermal stability more, but the integrated effect on the enzyme might be similar, and possibly cause the similar phenotypes I have seen.

Table 3.1. Enzymatic attributes for WT and mutated enzymes. Enzyme activity will increase as the $k_{cat}/S_{0.5}$, nH, and thermal stability increases.

Enzyme	$k_{cat}/S_{0.5}$ (s^{-1}/ mM)	nH	Thermal stability (ºC)
WT	2955.56	1.51 ± 0.03	61.7
P70Q	1445.45	2.88 ± 0.15	61.5
P70T	1250	3.06 ± 0.45	61.8
D127N	2766.67	1.22 ± 0.31	60.5
I264F	21.13	2.83 ± 0.16	56.8
A301S	1500	2.27 ± 0.06	62
A301T	50.5	2.70 ± 0.19	65
G381A	235.81	2.74 ± 0.08	54.5
T462I	201.27	2.93 ± 0.10	57.3

Table 3.2. Rank of each enzyme in the order of decreasing each of the three attributes.

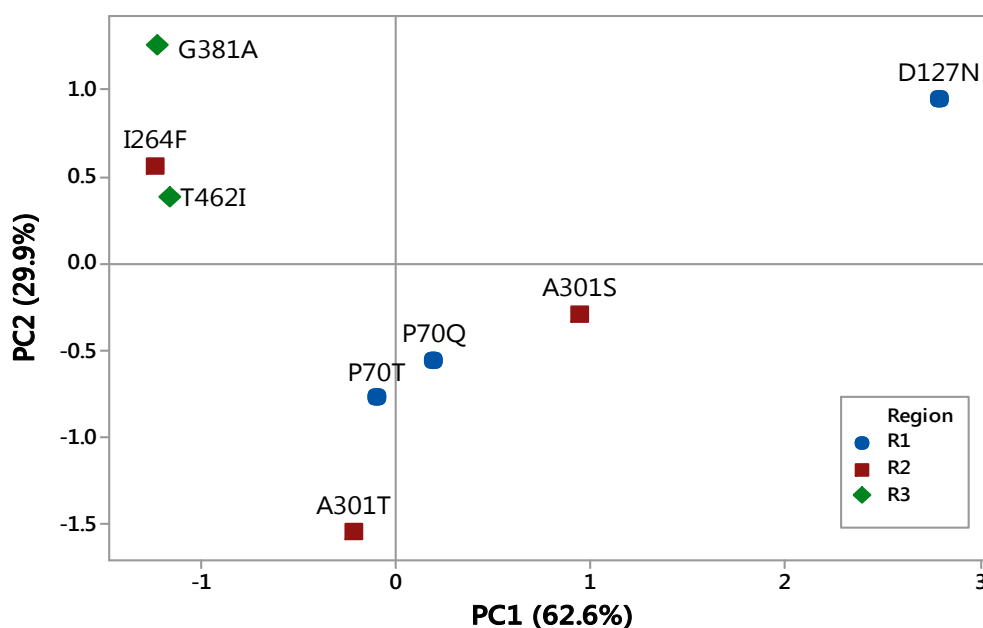| Enzyme	$k_{cat}/S_{0.5}$ (s^{-1}/ mM)	nH	Thermal stability (ºC)
WT	1	8	4
P70Q	4	3	1
P70T	5	1	5
D127N	2	9	7
I264F	9	4	6
A301S	3	7	2
A301T	8	6	1
G381A	6	5	9
T462I	7	2	8

Figure 3.4. Principal component analyses of the eight mutated enzymes. The eight mutations affect three regions of enzyme: R1, active site, include mutations P70T, P70Q, and D127N; R2, the tetrameric interfaces, include mutations I264F, A301S, and A301T; R3, close to the allosteric regulation site, include mutations G381A and T462I.

3.3 Discussion

This chapter described the mechanistic basis underlying beneficial effects of *pykF* mutations. I found that point mutations in *pykF* cause an increase in stationary phase concentrations of PEP. Such an increase might drive a rapid uptake of glucose following cell transfer to fresh medium, and shorten lag time of *pykF* mutant strains. Although the point mutations change the PykF enzyme in different ways, the overall change might integrate in a way that results in a similar net biochemical effect, and affects the similar changes on cell metabolism and physiological pathways to contribute to fitness advantages, suggesting that mutations in the same gene impose constraints on adaptation.

Tight regulation by FBP on mutant PykF might be the reason why PEP accumulates in the stationary phase. Mutant enzymes are hypothesized to require more FBP to catch up with WT enzyme activity, however, in stationary phase (starvation) FBP decreases to a very low level (Xu et al. 2012), thus mutant enzymes would have low activity, and PEP would be built up. Future experiments to directly test the hypothesis of tight regulation by FBP on mutant enzymes could involve modification of the pyruvate kinase to make FBP insensitive and test its effects on physiology, metabolism, and fitness. Furthermore, to clearly understand how mutations change PykF, more attributes of enzymes besides the three shown in results are needed and assessed in conditions close to the *in vivo* environment.

Previous studies connecting genotype to phenotype mainly focused on one phenotypic trait changed by adaptive mutations, such as gene expression (Cooper et al. 2003; Conrad et al. 2010; Chou and Marx 2012). This study, however, aims to bridge the gap between genotype and phenotype, broadly examining the mechanistic basis of adaptive mutations at multiple levels: from mutations to enzymes, physiological pathways, and fitness effects. As systems-level biology develops, more mechanistic studies are focusing on genome-scale changes of specific genetic perturbations, such as gene and protein expression profiles (Pelosi et al. 2006; Murakami et al. 2015), protein-protein interaction networks (Kaçar and Gaucher 2013), and metabolic networks (McCloskey et al. 2013). Integrating all such information is necessary to build a solid understanding of mechanic insights into the adaptive evolution.

3.4 Materials and Methods

Growth measurements

I used serial dilutions and standard plate counts at different time points to track the growth of the ancestral strain (REL606) carrying different *pykF* alleles: WT, Del, and A301S. Before measuring growth rate, all strains were taken from -80 degree freezer stocks, grown overnight in LB liquid medium, and then grown in the 50 mL Erlenmeyer flasks containing 10 mL Davis-Mingioli medium supplemented with 25 µg/mL of glucose for two 24-hour growth cycles, to allow the acclimation to the experimental environment. Preconditioned cultures were diluted 1:100 into 10 mL of fresh media in each flask to allow growth for another 24 hours. During this period, 10 µL of culture was taken at time points: 0, 2h, 4h, 5h, 6h, 7h, 8h, 10h, and 24h, and appropriately diluted before plating to obtain 200~400 colonies on each plate. Cell density at each time point for each strain can be calculated as the product of the cell count and the dilution factor.

Measurement of fitness effects

Fitness assays to determine fitness effect of *pykF* alleles at different time points were carried out as described previously (2.4-*Measurement of fitness effects*). REL606 is used as the reference strain for all competitions. The other competitor in each competition is the ancestral strain carrying each of the different alleles (WT, Del, P70T, A301S, and G381A). In 2.4-*Measurement of fitness effects*, only cell ratios at the beginning and the end of the competition are taken into account. In contrast, this experiment tracked cell ratio of the two competitors at different time points over a 24-hour growth cycle. The

time points used here are: 0, 2h, 4h, 5h, 6h, 7h, 8h, 10h, and 24h. Large variation of fitness effects of 2h and 4h provides little information on comparison effect of *pykF* mutations over time, thus data from the two time points were removed from analysis (see Figure 3.2).

The fitness effects conferred by *pykF*::IS*150* mutation in the genetic background where it occurred and was fixed were measured relative to its progenitor containing the opposite Ara marker in Davis-Mingioli medium supplemented with 25 μg/mL of each of the five PTS resources (glucose, fructose, mannitol, mannose, and NAG) and the six non-PTS resources (galactose, glycerol, lactose, maltose, melibiose, rhamnose). Fitness assays to determine fitness effect of *pykF*::IS*150* were carried out similarly as described previously (2.4-*Measurement of fitness effects*), except that here the competitions were performed in 96-well plates instead of flasks.

Statistical analyses

Plots and statistical analyses were all done in R version 3.1.3 (R Core Team 2015), except for Figure 3.4 which was created in Minitab. All fitness competitions were performed as complete experimental blocks. One-way ANOVA analysis was performed to test for differences in fitness effects of each *pykF* point mutations over time. Nested ANOVA was performed to test for differences of PTS and non-PTS resources in affecting fitness effects of the *pykF*::IS*150* mutation.

3.5 References

Barrick, J. E., D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, D. Schneider, R. E. Lenski, and J. F. Kim. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature 461:1243–1247.

Buchanan, R. L., and M. Solberg. 1972. Interaction of sodium nitrite, oxygen and pH on growth of *Staphylococcus aureus*. J. Food Sci. 37:81–85.

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

Chou, H. H., and C. J. Marx. 2012. Optimization of gene expression through divergent mutational paths. Cell Rep. 1:133–140.

Conrad, T. M., M. Frazier, A. R. Joyce, B. K. Cho, E. M. Knight, N. E. Lewis, R. Landick, and B. Ø. Palsson. 2010. RNA polymerase mutants found through adaptive evolution reprogram *Escherichia coli* for optimal growth in minimal media. Proc. Natl. Acad. Sci. USA 107:20500–20505.

Cooper, T. F., D. E. Rozen, and R. E. Lenski. 2003. Parallel changes in gene expression after 20, 000 generations of evolution in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 100:1072–1077.

Kaçar, B., and E. A. Gaucher. 2013. Experimental evolution of protein-protein interaction networks. Biochem. J. 453:311–319.

McCloskey, D., B. Ø. Palsson, and A. M. Feist. 2013. Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*. Mol. Syst. Biol. 9:661.

Muñoz, M. E., and E. Ponce. 2003. Pyruvate kinase: current status of regulatory and functional properties. Comp. Biochem. Physiol. Part B Biochem. Mol. Biol. 135:197–218.

Murakami, Y., Y. Matsumoto, S. Tsuru, B. W. Ying, and T. Yomo. 2015. Global coordination in adaptation to gene rewiring. Nucleic Acids Res. 43:1304–1316.

Orr, H. A. 2005. The genetic theory of adaptation: a brief history. Nat. Rev. Genet. 6:119–127.

Pelosi, L., L. Kühn, D. Guetta, J. Garin, J. Geiselmann, R. E. Lenski, and D. Schneider. 2006. Parallel changes in global protein profiles during long-term experimental evolution in *Escherichia coli*. Genetics 173:1851–1869.

Postma, P. W., J. W. Lengeler, and G. R. Jacobson. 1993. Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria. Microbiol. Rev. 57:543–594.

R Core Team. 2015. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria.

Schneider, D., E. Duperchy, E. Coursange, R. E. Lenski, and M. Blot. 2000. Long-term experimental evolution in *Escherichia coli*. IX. characterization of insertion sequence-mediated mutations and rearrangements. Genetics 156:477–488.

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

Xu, Y. F., D. Amador-Noguez, M. L. Reaves, X. J. Feng, and J. D. Rabinowitz. 2012. Ultrasensitive regulation of anapleurosis via allosteric activation of PEP carboxylase. Nat. Chem. Biol. 8:562–568.

Zhu, T., M. F. Bailey, L. M. Angley, T. F. Cooper, and R. C. J. Dobson. 2010. The quaternary structure of pyruvate kinase type 1 from *Escherichia coli* at low nanomolar concentrations. Biochimie 92:116–120.

Chapter 4 The role of environment and genetic background on fitness of a transferred mutation

The effect of a mutation may depend on its genetic background (epistasis), the environment (pleiotropy), or both. However, few empirical studies have investigated the joint influence of both factors in the same experiment. Those that have, have usually focused on randomly generated mutations, which may behave differently than beneficial mutations selected because they improve organism adaptation in a specific genetic and environmental context. I measured the fitness effect conferred by the beneficial *pykF* deletion mutation (Δ*pykF*) in 23 natural isolates of *Escherichia coli* in five different resource environments. I found that the genetic background, the environment, and the interaction between the genetic background and the environment, all have significant effects on the fitness effect of the mutation, with genetic background being the most important in determining fitness. In agreement with the diminishing returns epistasis found in previous experiments, a negative relationship between the maximum growth rate of the recipient strains and the fitness effect of Δ*pykF* were found in all five environments, although only in one environment was this relationship significant. My results suggest that the effect of beneficial mutations will be difficult to predict, depending on specific interactions with genotype and environment. Nevertheless, I find support for a growing body of research predicting a qualitative relationship between mutation effects and initial fitness of a progenitor strain, predicting that the contribution of a mutation will be smaller when added to fitter progenitors.

4.1 Introduction

The dependence of a mutation's effects on the genetic background, known as epistasis (G × G), is important to many evolutionary theories, such as adaptation (Chou et al. 2011; Khan et al. 2011; Vogwill et al. 2014), maintenance of sex (Kondrashov and Kondrashov 2001; Kouyos et al. 2007), and speciation (Brideau et al. 2006; Anderson et al. 2010). Previous studies have shown that epistasis is pervasive, however, these studies mainly focused on randomly generated mutations which mostly have neutral or deleterious effects (reviewed in Kouyos et al. 2007; Parera et al. 2009; Lalić and Elena 2012). Recent studies that examined epistatic interactions involving adaptive mutations have typically found evidence for a tendency toward negative epistasis, whereby combined mutation effects tend to be smaller than expected from the sum of their individual effects (MacLean et al. 2010; Chou et al. 2011; Khan et al. 2011; Rokyta et al. 2011; Wang et al. 2013).

Environmental conditions have also been shown to influence mutation effects, a phenomenon known as pleiotropy (G × E). Mutations might have smaller or larger effects in different environments (magnitude pleiotropy), or have effects that differ in sign (antagonistic pleiotropy, *e.g.*, a mutation being beneficial in a selected environment, but being deleterious in alternative environments) (Remold 2012). The different types of pleiotropy can greatly affect whether a mutation will be favored by selection. Moreover, epistasis can also depend on the environment (Kishony and Leibler 2003; Remold and Lenski 2004; Lalic and Elena 2012). This kind of interaction, epistatic pleiotropy (G × G × E), also has important evolutionary consequences. Organisms could escape from a

deleterious epistatic effect in one environment if epistasis confers a benefit in alternative environments, thus allowing populations to follow evolutionary paths that are otherwise inaccessible (Lindsey et al. 2013). In short, the extent to which the phenotypic effect of a particular mutation is determined by the genetic background, environment, and the interaction between them, is critical to predicting its evolutionary fate. Relatively few studies, however, have examined the combined effect of epistasis and pleiotropy for beneficial mutations. One study that did found that fitness effects of beneficial mutations identified from adaptation of a lab evolved strain of *E. coli* were greatly influenced by the genetic backgrounds and external environments (Flynn et al. 2013). In that case, however, the genetic backgrounds in which beneficial mutations were tested were evolved from the same ancestor and differed by at most four mutations, thus epistasis is restricted due to the small number of potentially different mutation interactions. It remains unclear how interactions between a focal adaptive mutation and generally divergent genetic backgrounds, will be affected by different external environments.

In this study I investigate the fitness of a *pykF* deletion mutation (Δ*pykF,* same as Del, which was used in chapter 2 and 3) that was added to 23 divergent strains of *E. coli* and measured in five resource environments. The Δ*pykF* mutation is equivalent to a beneficial *pykF*::IS*150* allele that was isolated from one of the evolved populations in the long-term evolution experiment (described in previous chapters) (Khan et al. 2011). The Δ*pykF* mutation confers a ~6% benefit when measured in the ancestral background in the evolution environment: minimal medium supplemented with glucose (chapter 2). Since beneficial mutations were selected in a specific genetic and environmental context, they

might be more strain and environment specific than are the random mutations studied in most experiments following the effects of genotype and environment on phenotypes. The five resources used as sole carbon sources in the environments I use in this study share a common uptake mechanism, the phosphotransferase system (PTS). A study on the effect of presumably similar environments will provide insight into mechanisms of the interaction between given mutations and environments, and help predict the evolutionary fate of these mutations.

My experimental design also allows me to test the generality of the observation that the effect of beneficial mutations generally declines in proportion to the initial fitness of a genotype it is added to (MacLean et al. 2010; Chou et al. 2011; Khan et al. 2011). To test this, I measured the maximum growth rate of the 23 divergent genotypes in each of the five resource environments, and estimated the correlation between each genotype's maximum growth rate and the fitness change conferred by adding Δ*pykF*.

4.2 Results

4.2.1 Gene-by-environment interactions, gene-by-genotype interactions, and gene-by-genotype-by-environment interactions

First, I sought to determine whether the Δ*pykF* mutation has consistent effects over different environments. I measured effects of the Δ*pykF* mutation on fitness when it was added to 23 *E. coli* strains in five PTS resource environments: glucose, fructose, mannose, mannitol, and *N*-acetyl-glucosamine (NAG). Differences in the effect of the Δ*pykF* mutation on fitness in PTS and non-PTS resources provide evidence that it might confer a fitness advantage by affecting PTS mediated nutrient uptake (described in chapter 3).

Therefore, even in divergent genetic backgrounds, I expected the Δ*pykF* mutation to be beneficial in resources that use the PTS transport system. The reasons that I choose five PTS resource as the study environments are: (1) estimates of pleiotropy will be conservative putting a lower bound on the influence of environment on mutation effects and epistatic interactions; (2) the five resources are common sugars that generally exist in hosts (such as bird, marsupial, human) in which *E. coli* was found, therefore the results have some ecological relevance (Gordon and Cowling 2003; Walk et al. 2009); (3) the Δ*pykF* mutation is expected to be beneficial in PTS resources, thus I can use the resources to test the diminishing returns epistasis. Determining the mean fitness effect of the Δ*pykF* mutation in the 23 genotypes for each environment, I found an overall significant beneficial effect in four of five resources. Only in mannose was there no mean fitness effect (Figure 4.1).
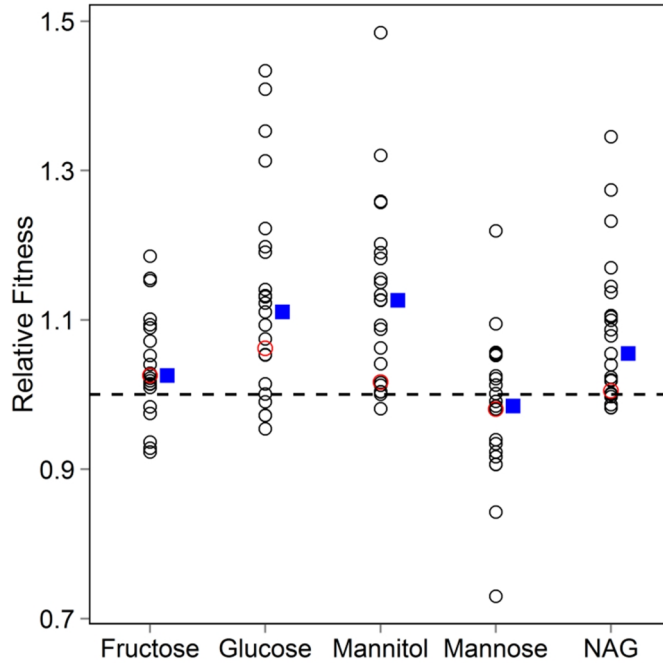
Figure 4.1. The Δ*pykF* mutation was added to the 23 genetic backgrounds and their fitness measured relative to the original progenitor strain in each of five PTS resource environments. Hollow symbols represent the mean fitness of the Δ*pykF* mutation in each strain based on three or four independent estimates. Red hollows indicate the fitness of Δ*pykF* in the genetic background on which it arose and fixed (See 4.4 Materials and Methods). Blue squares represent the grand mean fitness in each environment.

Next, I determined the frequency and form of interactions between the Δ*pykF* mutation and the natural isolate genotypes. One-way ANOVAs showed that the genetic background is important. The effect of genotype is significant in all five PTS resource environments (Table 4.1). Using a two-way ANOVA, I also examined the combined contributions of genotype and environment to fitness effects of the mutation (Table 4.2). I found a significant influence on the fitness effect of the Δ*pykF* mutation of the genetic background (G × G), the environment (G × E), and their interaction (G × G × E), which

explained 44%, 19%, 23% of the total variation in fitness, respectively. Two possible explanations can be brought forward to explain the significant G × G × E effect: changes in the (1) magnitude and/or (2) rank order of the fitness effects acquired by a background by adding the Δ*pykF* mutation across environments. To test if rank order changes across the environments, I plotted mean fitness of mutation in each genetic background over the five resource environments (Figure 4.2). Crossed lines generally seen between environments clearly indicate that changes in the rank order is one main factor accounting for the significant G × G × E effect. The rank order change in the G × G × E effect makes it harder to predict the evolutionary fate of a newly occurring mutation, since genetic background acquiring this mutation best fits in one environment, but might not be so in alternative environments.

Table 4.1. Results of one-way ANOVA testing the effect of genotype in each PTS environment.

Environment	Df	MS (Genotype)	MS	F	P
Fructose	22, 67	0.018	0.001	13.646	< 0.0001
Glucose	22, 68	0.074	0.005	16.149	< 0.0001
Mannitol	22, 67	0.060	0.004	17.260	< 0.0001
Mannose	22, 69	0.037	0.005	7.220	< 0.0001
NAG	22, 69	0.039	0.002	18.850	< 0.0001

Table 4.2. Results of two-way ANOVA testing the effect of genotype and resource environment on fitness.

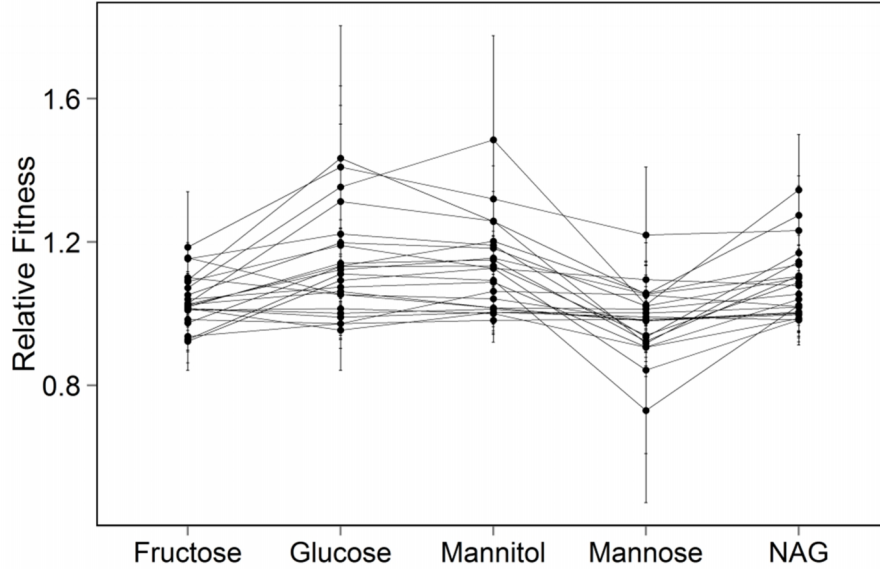| Factor	Df	SS	MS	*F*	*P*	Variation explained
Genotype	22	3.295	0.150	45.105	< 0.0001	43.6%
Environment	4	1.408	0.352	105.993	< 0.0001	18.6%
Genotype ×	88	1.727	0.020	5.909	< 0.0001	22.8%
Residuals	340	1.129	0.003			14.9%

Figure 4.2. Fitness effects of the Δ*pykF* mutation in the 23 genetic backgrounds in each of the five resource environments. Points and error bars represent the mean and 95% confidence intervals of three or four independent fitness estimates. Lines connect the same strains across environments.

To examine the effect of environment in more detail, I determined the proportion of beneficial, neutral, and deleterious mutations across environments (Figure 4.3). Seven strains (30% of total) act as generalists, in that the effect of the $\Delta pykF$ mutation does not depend on the environment (one-way ANOVA: $P > 0.05$; Table 4.3). Of these, six strains have no fitness change on adding the mutation in all environments, only one, BL21, shows consistent fitness benefits of the mutation, but with marginal non-significance (one-way ANOVA: $P = 0.054$; Table 4.3). For the genotypes in which the mutation effects were not consistent across environments, the mutation confers significant benefits to seven (30% of total) genotypes in at least four environments, and to 13 (57% of total) genotypes in three or more environments, indicating most of the genotypes being generally good backgrounds for the $\Delta pykF$ mutation to occur in. These results suggest that it is more likely to have no fitness change of adding one mutation to a generalist than to a non-generalist strain. For the non-generalists, while the effect of adding the $\Delta pykF$ mutation was variable across environments, it was usually beneficial. This phenomenon, called magnitude pleiotropy, reflects a change in the size of the mutation's effect across environments. Antagonistic pleiotropy, where genotypes are beneficial in some environments, but neutral or deleterious in others, was only slightly detected.
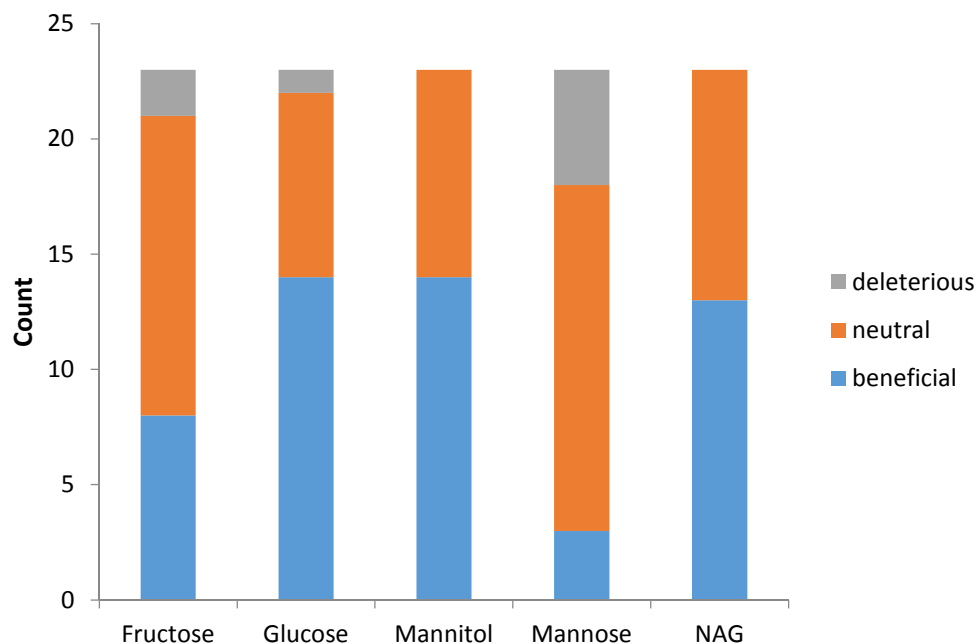
Figure 4.3. Number of genotypes in which Δ*pykF* has beneficial, neutral, or deleterious fitness effect in each PTS environment. Genotypes that are classified into three categories based on *t*-tests: beneficial (fitness effect of Δ*pykF* is significantly larger than zero), neutral (fitness effect of Δ*pykF* is not significantly different from zero), and deleterious (fitness effect of Δ*pykF* is significantly less than zero). Differences from a zero fitness effect were identified using a two-tailed *t*-test with significance assessed at $P < 0.05$.

Table 4.3. Results of one-way ANOVAs testing the effect of resource environment for each genotype.

Genotype	MS(Environment)	MS	F	P
TA141	0.002	0.012	0.618	0.657
B706	0.003	0.015	0.725	0.589
TC941	0.014	0.051	1.051	0.414
H442	0.009	0.028	1.240	0.336
TW10509	0.009	0.018	1.817	0.178
TA271	0.026	0.042	2.297	0.107
BL21	0.124	0.142	3.040	0.053
E1118	0.031	0.017	6.894	< 0.01
H260	0.033	0.019	5.927	< 0.01
TA144	0.081	0.049	6.251	< 0.01
B175	0.148	0.012	47.165	< 0.01
B921	0.083	0.010	31.068	< 0.01
E101	0.424	0.108	14.682	< 0.01
E560	0.210	0.012	65.488	< 0.01
FBGM1	0.618	0.209	10.375	< 0.01
FBGM12	0.053	0.013	14.882	< 0.01
FBGM18	0.222	0.096	8.704	< 0.01
FBGM4	0.092	0.032	10.011	< 0.01
FBGM9	0.387	0.094	15.478	< 0.01
M114	0.131	0.017	28.975	< 0.01
MG1655	0.043	0.012	13.759	< 0.01
R424	0.355	0.110	11.273	< 0.01
TA280	0.035	0.011	12.274	< 0.01

4.2.2 Effect of adding a mutation depends on the fitness of the recipient genotype.

Previous work has found that the fitness of a genotype determines the fitness increase conferred by a new beneficial mutation; the fitter a genotype, the less it benefits from a new beneficial mutation—a pattern referred to as diminishing returns epistasis (Chou et al. 2011; Khan et al. 2011; Kryazhimskiy et al. 2014). This relationship has, however, only been made comparing closely related strains and measuring fitness in the environment in which the focal transferred mutations were initially selected.

To test if the relationship extends to more divergent genetic backgrounds and different environments, I measured the maximum growth rate of the 23 genotypes in five resource environments, and estimated the correlation between each genotype's maximum growth rate and the fitness change by adding Δ*pykF* (Figure 4.4). In mannose the correlation was significant at $P < 0.05$. Although no significant relationship was found in the other four resource environments, they all showed a negative correlation between maximum growth rate and fitness change, and a one-tailed sign test showed that the tendency toward negative relationships is itself significant ($P = 0.031$). This negative relationship found in natural isolates is in agreement with the observations from previous experiments with lab strains (Chou et al. 2011; Khan et al. 2011). It indicates that adding a beneficial mutation to a genotype that already has high fitness does not contribute as much fitness increase as it would if the mutation was added to an initially less fit genotype.
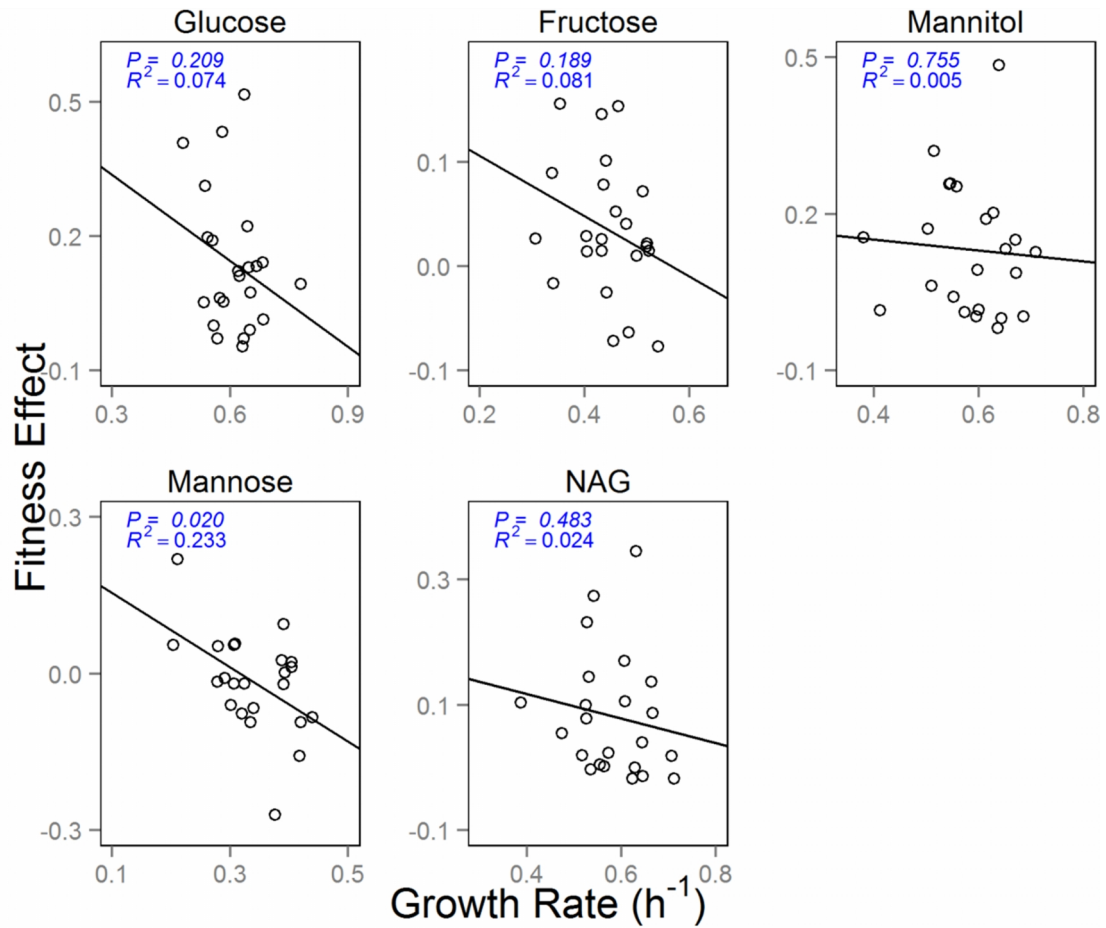
Figure 4.4. Relationships between maximum growth rate and relative fitness of adding the $\Delta pykF$ mutation to the 23 genotypes in five PTS resource environments. Hollows represent the mean of maximum growth rate of the 23 genotypes ($3 \leq n \leq 32$), and the mean of fitness of the $\Delta pykF$ mutation in each genotype ($n = 3$ or 4). Lines show linear regression with P value and R^2 calculated in a Pearson correlation shown for each graph. Spearman non-parametric correlations were also performed, and they did not affect conclusions of significance.

4.3 Discussion

Despite many studies of individual effects of genetic background and environmental conditions on fitness, my work represents one of the few studies to investigate the combined effect of background and environment in affecting fitness, and to my

knowledge the first study focusing on a beneficial mutation added to divergent genetic backgrounds. My experiments show that in natural isolate strains, the fitness effect of adding a new mutation depends not only on the genetic background (Lunzer et al. 2005; Weinreich et al. 2005; Lozovsky et al. 2009; Chou et al. 2011) and environmental (Remold and Lenski 2001; Ostrowski et al. 2005; Remold 2012) context, but also their interactions (Kishony and Leibler 2003; Remold and Lenski 2004; Agrawal and Whitlock 2010; Lalic and Elena 2012; Flynn et al. 2013). This finding suggests that knowledge of both epistasis and pleiotropy is required to predict the evolutionary fate of mutations.

Compared to the environment and interactions with environment, the genetic background explains a larger total variation of fitness. However, the effect of environment on fitness might be underestimated due to a bias in choosing resource environments: the resource environments used here all employ the same sugar uptake mechanism—PTS, the system that I predict the $\Delta pykF$ mutations will affect. Effects of the $\Delta pykF$ mutation in such resource environments will be more similar than in environments that are more different from each other. I expect more importance of environmental effect on fitness, if a wide range of environmental conditions is included.

As observed in previous studies (Ostrowski et al. 2005, 2008), I found limited antagonistic pleiotropy: the $\Delta pykF$ mutation that is beneficial in each genotype in the environment in which it was initially selected (glucose), also tends to be beneficial in alternative environments (magnitude pleiotropy). Again, one possible reason for this phenomenon is the biased choice of environmental resources where the $\Delta pykF$ mutation

was tested in. However, *Flynn et al.* (2013) also found limited antagonistic pleiotropy after assessing the effect of a combination of five beneficial mutations in 1,920 environments. All beneficial mutations considered in these studies (Ostrowski et al. 2005, 2008; Flynn et al. 2013), including my study, are early adaptive mutations in the evolution experiment, and it was considered that the limitation of antagonistic pleiotropy is likely to cause early adaptation to expand niche breadth with limited cost (Flynn et al. 2013).

My results indicate that the initial absolute fitness of a genotype in a defined environment may have some influence on the fate of a transferred mutation: the larger the initial fitness is, the less benefit a new mutation will contribute. However, although overall there is a negative relationship between the maximum growth rate of the recipient genotype and fitness change conferred on adding the Δ*pykF* mutation, only in mannose was this relationship significant. Clearly, initial growth rate is not the only factor that affects fitness change of a transferred mutation, and the complexity of natural isolates (*e.g.*, divergent genetic backgrounds, and different ecological histories) act to add noise to the negative relationship (diminishing returns epistasis) that was found in previous studies (Chou et al. 2011; Khan et al. 2011; Kryazhimskiy et al. 2014).

In summary, the combination of pleiotropy and epistasis can strongly influence the fate of a mutation in a genotype to adapt to a new environment. We need to caution that, due to the adaptation to different organisms or environments, conclusions drawn from lab strains might not hold, or, at least, hold less strongly, when applied to natural isolate strains. For

example, the relatively weak negative relationship between maximum growth rate of a progenitor strain and fitness change of adding a new mutation in natural isolates compares to a strong and significant relationship in lab strains.

4.4 Materials and Methods

Strains

The 23 strains used in this study include: (1) TC941, the genetic background where *pykF*::IS*150* (equivalent to Δ*pykF*) was identified as part of the Lenski evolution experiment (Schneider et al. 2000); (2) MG1655 and BL21, commonly used *E. coli* lab strains; 3) 20 strains obtained from Francis Moore (University of Ohio, Akron) and the Michigan State University Shiga-toxin producing *E. coli* (STEC) strain collection, which were collected from different organisms and different environments, and reflect a diverse set of strains. The Δ*pykF* mutation, a 787-bp deletion of 1413-bp *pykF* gene, was introduced to all strains by Yinhua Wang using an approach described previously (Khan et al. 2011). An *araA⁻* mutation was also introduced into these strains by Yinhua Wang. This mutation introduces a neutral phenotypic marker used to distinguish strains in fitness competitions.

Measurement of fitness effects

The fitness effects conferred by the Δ*pykF* mutation in each genetic background were measured relative to its progenitor containing the opposite Ara marker in five PTS

resource environments: fructose, glucose, mannitol, mannose, and NAG (Lenski et al. 1991). For each resource, fitness assays were carried out by head-to-head competitions of the two competitors in Davis-Mingioli medium supplemented with 25 µg/mL of that resource. Fitness assays were carried out as described previously (2.4-*Measurement of fitness effects*), except that all of the competitions in this study were performed in 96-well plates. In some strains and some environments, the *araA* marker was not neutral. In these cases, fitness effect conferred by the $\Delta pykF$ mutation was corrected by the *araA* marker cost or benefit.

Growth rate measurements

Before measuring growth rate, all strains were taken from -80 degree freezer stocks, grown overnight in LB liquid culture, and then grown in Davis-Mingioli medium supplemented with 500 µg/mL of one resource (*i.e.*, fructose, glucose, mannitol, mannose, or NAG) for two 24-hour growth cycles, to allow acclimation to the experimental environment. Preconditioned cultures were diluted 1:100 into 200 µL of fresh media in each well of a 96-well microtiter plate, and OD_{450} was measured at 5-minute intervals over 24 hours in a VersaMax plate reader (Molecular Dynamics, CA). All incubations were at 37 ºC. I had difficulty in obtaining smooth growth curves for some strains (specifically, B175, B706, B921, E101, E1118, H442, MG1655, and TA280) in all or most of the five resource environments. For these strains, to avoid bias, I ran sufficient growth assays to obtain at least as many smooth growth curves as obtained for other strains. The number of replicates for these strains is usually more than six. For the strains

that always give good growth curves, I had at least four individual measurements. Strain maximum growth rates were estimated by applying a custom R script to the collected growth data.

Statistical analyses

Plots and statistical analyses were all done in R version 3.1.3 (R Core Team 2015). Fitness competitions were performed as complete experimental blocks. Two blocks were used to obtain four individual fitness measurements, with two replicates in one block. I found no individual block effect or effect of interactions between block and environment, genotype, or environment and genotype. One-way ANOVA analysis was performed to test for differences in fitness effects of the Δ*pykF* mutation in each strain or in each resource environment. Two-way ANOVA was performed to test for the effects of genetic background and environment on fitness. In all analyses, genetic background and resource environment were treated as fixed effects.

4.5 References

Agrawal, A. F., and M. C. Whitlock. 2010. Environmental duress and epistasis: how does stress affect the strength of selection on new mutations? Trends Ecol. Evol. 25:450–458.

Anderson, J. B., J. Funt, D. A. Thompson, S. Prabhu, A. Socha, C. Sirjusingh, J. R. Dettman, L. Parreiras, D. S. Guttman, A. Regev, and L. M. Kohn. 2010. Determinants of divergent adaptation and Dobzhansky-Muller interaction in experimental yeast populations. Curr. Biol. 20:1383–1388.

Brideau, N. J., H. A. Flores, J. Wang, S. Maheshwari, X. Wang, and D. A. Barbash. 2006. Two Dobzhansky-Muller genes interact to cause hydbrid lethality in *Drosophila*. Science 314:1292–1295.

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

Flynn, K. M., T. F. Cooper, F. B.-G. Moore, and V. S. Cooper. 2013. The environment affects epistatic interactions to alter the topology of an empirical fitness landscape. PLoS Genet. 9:e1003426.

Gordon, D. M., and A. Cowling. 2003. The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: Host and geographic effects. Microbiology 149:3575–3586.

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

Kishony, R., and S. Leibler. 2003. Environmental stresses can alleviate the average deleterious effect of mutations. J. Biol. 2:14.

Kondrashov, F. A., and A. S. Kondrashov. 2001. Multidimensional epistasis and the disadvantage of sex. Proc. Natl. Acad. Sci. USA 98:12089–12092.

Kouyos, R. D., O. K. Silander, and S. Bonhoeffer. 2007. Epistasis between deleterious mutations and the evolution of recombination. Trends Ecol. Evol. 22:308–315.

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

Lalic, J., and S. F. Elena. 2012. Epistasis between mutations is host-dependent for an RNA virus. Biol. Lett. 9:20120396.

Lalić, J., and S. F. Elena. 2012. Magnitude and sign epistasis among deleterious mutations in a positive-sense plant RNA virus. Heredity. 109:71–77.

Lenski, R. E., M. R. Rose, S. C. Simpson, and S. C. Tadler. 1991. Long-term experimental evolution in *Escherichia coli*. I. adaptation and divergence during 2,000 generations. Am. Nat. 138:1335–1341.

Lindsey, H. A., J. Gallie, S. Taylor, and B. Kerr. 2013. Evolutionary rescue from extinction is contingent on a lower rate of environmental change. Nature 494:463–467.

Lozovsky, E. R., T. Chookajorn, K. M. Brown, M. Imwong, P. J. Shaw, S. Kamchonwongpaisan, D. E. Neafsey, D. M. Weinreich, and D. L. Hartl. 2009. Stepwise

acquisition of pyrimethamine resistance in the malaria parasite. Proc. Natl. Acad. Sci. USA 106:12025–12030.

Lunzer, M., S. P. Miller, R. Felsheim, and A. M. Dean. 2005. The biochemical architecture of an ancient adaptive landscape. Science 310:499–501.

MacLean, R. C., G. G. Perron, and A. Gardner. 2010. Diminishing returns from beneficial mutations and pervasive epistasis shape the fitness landscape for rifampicin resistance in *Pseudomonas aeruginosa*. Genetics 186:1345–1354.

Ostrowski, E. A., D. E. Rozen, and R. E. Lenski. 2005. Pleiotropic effects of beneficial mutations in *Escherichia coli*. Evolution 59:2343–2352.

Ostrowski, E. A., R. J. Woods, and R. E. Lenski. 2008. The genetic basis of parallel and divergent phenotypic responses in evolving populations of *Escherichia coli*. Proc. Biol. Sci. 275:277–284.

Parera, M., N. Perez-Alvarez, B. Clotet, and M. A. Martínez. 2009. Epistasis among deleterious mutations in the HIV-1 protease. J. Mol. Biol. 392:243–250.

R Core Team. 2015. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria.

Remold, S. 2012. Understanding specialism when the jack of all trades can be the master of all. Proc. R. Soc. B Biol. Sci. 279:4861–4869.

Remold, S. K., and R. E. Lenski. 2001. Contribution of individual random mutations to genotype-by-environment interactions in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 98:11388–11393.

Remold, S. K., and R. E. Lenski. 2004. Pervasive joint influence of epistasis and plasticity on mutational effects in *Escherichia coli*. Nat. Genet. 36:423–426.

Rokyta, D. R., P. Joyce, S. B. Caudle, C. Miller, C. J. Beisel, and H. A. Wichman. 2011. Epistasis between beneficial mutations and the phenotype-to-fitness map for a ssDNA virus. PLoS Genet. 7:e1002075.

Schneider, D., E. Duperchy, E. Coursange, R. E. Lenski, and M. Blot. 2000. Long-term experimental evolution in *Escherichia coli*. IX. characterization of insertion sequence-mediated mutations and rearrangements. Genetics 156:477–488.

Vogwill, T., M. Kojadinovic, V. Furio, and R. C. MacLean. 2014. Testing the role of genetic background in parallel evolution using the comparative experimental evolution of antibiotic resistance. Mol. Biol. Evol. 31:3314–3323.

Walk, S. T., E. W. Alm, D. M. Gordon, J. L. Ram, G. A. Toranzos, J. M. Tiedje, and T. S. Whittam. 2009. Cryptic lineages of the genus *Escherichia*. Appl. Environ. Microbiol. 75:6534–6544.

Wang, Y., C. D. Arenas, D. M. Stoebel, and T. F. Cooper. 2013. Genetic background affects epistatic interactions between two beneficial mutations. Biol. Lett. 20120328.

Weinreich, D. M., N. F. Delaney, M. A. Depristo, and D. L. Hartl. 2005. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 312:111–114.