Experimental Tests of the Influence of Interactions Between Beneficial Mutations on Adaptation and Reproductive Isolation

---

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

Yinhua Wang

August 2015

Experimental Tests of the Influence of Interactions Between Beneficial Mutations on Adaptation and Reproductive Isolation

---

Yinhua Wang

APPROVED:

Dr. Tim Cooper, Chair
Department of Biology and Biochemistry

Dr. Gábor Balázsi
Stony Brook University

Dr. Rebecca Zufall
Department of Biology and Biochemistry

Dr. Ricardo Azevedo
Department of Biology and Biochemistry

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

ACKNOWLEDGEMENTS

My special thanks go to my adviser, Dr. Tim Cooper, for his guidance, support, and patience throughout this research. Without his help this dissertation would never have been possible. I feel truly honored to have had the chance to work with him. I am grateful to Dr. Rebecca Zufall, Dr. Ricardo Azevedo, Dr. Gábor Balázsi, and Dr. Dan Graur for their great efforts and valuable suggestions to this dissertation. I would like to thank our collaborators Ethan Knapp from Dr. Francisco Moore lab and Kenny Flynn from Dr. Vaughn Cooper lab for the sequencing analysis and biolog assay. I would also like to thank all my present and previous lab mates Fen Peng, Kelly Phillips, Andrea Wünsche, Rachel Staples, Dustin Phipps, Kristina Duan, Trang Duong, Rebecca Satterwhite, Carolina Anrenas, Daniel Stoebel, Duy Dinh, and Aisha Khan for their intelligent support and friendship. I also want to thank all my friends Fan Li, Wei-ting Lin, Huy Vu, Hongan Long, Bingjun Zhang, Fei Yuan, and Ata Kalirad for their encouragement and inspiration. They cheer me up and warm me up when I had difficulty in my study and in my life. Last, but certainly not least, my deepest gratitude goes to my husband Hongyu Guo, my daughter Shuyun, and our parents for their constant encouragement and total support in my attainment of this goal.

Experimental Tests of the Influence of Interactions Between Beneficial Mutations on Adaptation and Reproductive Isolation

---

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

Yinhua Wang

August 2015

ABSTRACT

Newly arising mutations create genetic variation which natural selection can act on to favor organisms better suited to the environment. The effect of any mutation can, however, not necessarily be determined independently, but may depend on complex interactions with the broader genetic background. This interaction has long been considered to be very important in understanding the structure and function of genetic pathways and the evolutionary dynamics of organisms. Despite the clear importance of such interactions, known as epistasis, the understanding of epistasis between beneficial mutations and overall genetic background is limited. In this dissertation, I present three studies to examine the basis and extent of epistatic interaction between beneficial mutations and their genetic backgrounds. First, I introduced each of four beneficial mutations selected in a long-term evolution experiment into a set of natural isolates of *Escherichia coli* and measured the fitness of the constructed strains. I found that the fitness effect of beneficial mutations was highly dependent on their genetic backgrounds, and that there was a negative relationship between the fitness effect of beneficial mutations and the initial fitness of the genetic background in which the mutation was introduced. These results suggest that the ability of a strain to adapt in environment is, at least in part, determined by the current fitness of a potential recipient strain. Second, in order to understand how the genetic background affects the epistatic interactions between focal beneficial mutations, I carried out experiments in which two beneficial mutations, in the genes *pykF* and *topA*, were introduced individually and in combination into seven natural isolates of *E. coli*. I found that the epistatic interaction between these two

mutations strongly depends on the genetic backgrounds and again the magnitude of epistasis tended to decrease as the fitness of the recipient strain increased. Finally, I tested how parallel and divergent adaptation affects reproductive isolation by examining the extent of epistasis between mutations obtained in different populations evolved either in the same or different environments. I found a prevailing negative deviation from expected fitness in recombinants from parents evolved in both the same and different environment, indicating that negative interactions between independently evolved mutations is one potential contributor to development of reproductive isolation. Overall, my work contributes to a growing body of work that demonstrates the importance of epistasis between beneficial mutations in major evolutionary processes.

Table of Contents

Chapter 1: General introduction

Overview

Adaptation, a fundamental question in evolutionary biology, is the process by which populations become better fitted to their environment through changes that increase survival or reproduction (Orr 2005). A primary goal of recent studies in experimental evolution is to explore the genetic basis of adaptation. Mutations are the ultimate source of all genetic variation within populations. Thus in order to fully understand adaptation through natural selection, it is essential to know how mutations function, what factors impact the fitness effect of mutations, and how mutations affect the evolution of organisms. Below, I review the definition of epistasis, which describes the effect of interactions between genes, and outline different forms epistatic interactions can take. I describe how epistasis can be examined and summarize evidence for its importance in basic evolutionary process.

What is epistasis?

When the phenotypic effect of a mutation is dependent on other mutations or its genetic background, the mutation is said to interact with them, a phenomena known as epistasis (Phillips 2008, De Visser et al. 2011, Macía et al. 2011). If epistasis influences the effect of a mutation, the effect of that mutation cannot be determined by itself, but only in the context of the genetic background in which it is present. The original definition of epistasis was proposed by William Bateson in 1907 (Bateson 1907), who

used this term to explain qualitative deviations from Mendelian inheritance. Specifically, he used epistasis to describe the observation that an allele at one locus masks the expression of an allele at another locus. Eye color determination in *Drosophila* is a classic example of this kind of epistasis (Mackenzie et al. 1999). The genes *scarlet*, *brown*, and *white* together play roles in determining eye pigmentation. The eye color of a mutant genotype at the *white* locus is white, indicating that *white* masks the genotypes at the other two loci and is, therefore, epistatic to *brown* and *scarlet*.

Later, Ronald Fisher (1918) suggested a derivative of epistasis, 'epistacy', to define the statistical deviation of the linear combination of multiple genetic effects for a trait. Now, epistasis usually refers to interactions between genes in which the effect of one gene at one locus depends on the other genes at another locus, or more broadly, the interactions between mutations and their genetic backgrounds in which the phenotypic effect of a mutation depends on the genetic background (Wolf 2000, Phillips 2008). For instance, if the effect of a mutation is beneficial in one genetic background, but deleterious in another, the sign of the fitness effect of this mutation depends on its genetic background. Such epistasis is called sign epistasis (Weinreich et al. 2005, Lalic and Elena 2012).

How to detect and characterize epistasis?

Reflecting the different ways of defining epistasis, there are different methods to detect and characterize it. According to Bateson's definition, it is easy to detect epistasis because phenotypes are qualitative. If epistasis is considered quantitatively and extended

to include any kind of genetic interaction, how to quantify epistasis becomes more complicated (Lehner 2011). Several models have been developed to statistically quantify the epistasis between different mutations. These models include additive, multiplicative, log, and minimum null models. In this dissertation, the null model I used to estimate epistasis is a multiplicative model because it best suits the *Escherichia coli*-based system employed in this study in which fitness is considered as the phenotype of interest. According to this model, the epistasis between mutations is defined as the difference between observed and expected fitness of some mutation combination, where expected fitness is calculated as the product of the fitness effects of individual mutations.

Epistasis can be estimated in both direction and magnitude of mutational effects on fitness. Direction was used to describe the force of selection on mutation and magnitude refers to the deviation between observed fitness and expected fitness of a genotype having mutations. If the difference is statistically smaller than 0, then the epistasis between the mutations is negative, which means that the combined effect is smaller than what would be expected from their separate effects, thereby decreasing the beneficial effects of beneficial mutations or enlarging the detrimental effect of deleterious mutations. If the difference is statistically greater than 0, the epistasis between mutations is positive, which means it reduces the detrimental effects of deleterious mutations and increase the beneficial effects of beneficial mutations.

The model bio-system used to study epistasis in experimental evolutionary studies

In contrast to comparative studies, experimental evolution enables researchers to directly observe evolution and quantitatively test theoretical predictions. Microbes, such as the bacterium *E. coli*, have many advantages that make them useful in experimental evolutionary studies (Elena and Lenski 2001). Firstly, *E. coli* are easily cultured in the laboratory and can be kept at a non-growing state by storing at -80°C, which allows direct comparison of evolved strains with their ancestor. In addition, short generation times and large population sizes can be achieved, so that evolutionary processes such as adaptation occur on an observable time scale. Moreover, *E. coli* can easily be genetically manipulated, facilitating construction of defined genotypes. Finally, replicate populations can be evolved from a common ancestor, so that multiple instances of evolution beginning from a single starting point can be obtained.

All these advantages of *E. coli,* in combination with development of whole genome sequencing methods, has allowed researchers to: (i) identify different beneficial mutations accumulated in hundreds of replicate populations over thousands of generations, (ii) directly measure the fitness effect of the beneficial mutations in mutant compared to the ancestor without mutations, and (iii) quantify the interactions between beneficial mutations and the genetic backgrounds by introducing specific focal mutations in different genetic backgrounds and estimate the direction and magnitude of epistasis (Phillips 2008, Wielgoss et al. 2011, Wielgoss et al. 2013).

The role and extent of epistasis between mutations

As direct experimental evidence of mutation interactions has accumulated (Elena and Lenski 1997, Sanders and Whitlock 2003, Remold and Lenski 2004, Sanjuan 2006, Sanjuan and Elena 2006, Pepin and Wichman 2007, Dworkin et al. 2009, Gerke et al. 2010, Chou et al. 2011, Khan et al. 2011, De Visser and Krug 2014), the prevalence and the importance of epistatic interactions has become widely accepted. It has been proposed that epistasis plays a very important role in several evolutionary processes, such as speciation (Dobzhansky 1937, Muller 1942, Coyne 1992, Brideau et al. 2006, Dettman et al. 2007, Dettman et al. 2010), the evolution and maintenance of sex (Muller 1942, Otto and Lenormand 2002, Azevedo et al. 2006, Kouyos et al. 2007), adaptation (Woods et al. 2011), drug resistance (Kryazhimskiy et al. 2011), and the evolution of ploidy (Omholt et al. 2000).

Negative epistasis between two or more genes that have functionally diverged in two isolated populations was proposed by Dobzhansky (1937) and Muller (1942) to cause hybrid dysfunction and eventually result in postzygotic reproductive isolation. Dobzhansky and Muller suggested that accumulation of different mutations in different lineages can promote speciation if they interact negatively to cause hybrids to have extreme low fitness, sterility, or to be inviable. Many studies have been conducted to test conflicts among interacting genes in closely related, but already established species (Axenovich et al. 1998, Fishman and Willis 2001, Brideau et al. 2006, Masly et al. 2006, Bikard et al. 2009, Cattani and Presgraves 2009, Mihola et al. 2009, Moyle et al. 2012)

and in independently evolved populations from common ancestor at initial stage of speciation (Dettman et al. 2010, Kwan and Rundle 2010). All these results provide evidence that negative epistasis involved in the hybrid incompatibility to cause reproductive isolation.

Negative epistasis between deleterious mutations, was suggested to explain the ubiquity of sexual reproduction and provide an advantage of sexual reproduction over asexual reproduction, known as mutational deterministic hypothesis (MDH) (Kondrashov 1988). In this hypothesis, the evolution of recombination is selected for by breaking down negative linkage disequilibrium generated by negative epistasis, reducing deleterious effects. High recombination rates are maintained because the breaking up of linkage disequilibria generated by negative epistasis enables more effective purging of deleterious mutations. If the rate at which deleterious mutations arise is high and their interaction affects fitness negatively, the ubiquity of different forms of genetic recombination, including sexual reproduction, can conceivably be explained. Although, more recently there are some theoretical and the experimental evidences challenging the ability of the MDH to explain the evolution of sex, the role of negative epistasis in this evolutionary process remains of wide interest (Kouyos et al. 2007).

It was observed that the fitness of populations adapting to a constant environment does not increase at a constant rate, but rapidly decelerates, though to a plateau or not remains unclear (Wiser et al. 2013, Good and Desai 2015). The negative epistasis between beneficial mutations was proposed to play an important role in the declining rate

of fitness increase as populations approach a fitness peak, known as diminishing returns epistasis (Chou et al. 2011, Khan et al. 2011, Rokyta et al. 2011, Pearson et al. 2012). In these cases, epistatic interactions contribute to the deceleration of adaptation by reducing the effect of later arising beneficial mutations.

Epistasis between mutations can also have important consequences for adaptive evolution by constraining the evolutionary pathways available to a population (Poelwijk et al. 2007, Woods et al. 2011, De Visser and Krug 2014). For example, Weinreich et al. (2006) examined the number of mutational trajectories of a particular beta-lactamase allele could be selected along to confer high-level resistance to a new antibiotic. By constructing all possible combinations of alleles corresponding to those trajectories, they found that, due to sign epistasis, only a small fraction of possible mutational trajectories were accessible to natural selection. A recent study found that epistasis contributes to contingency in antibiotic resistance enzyme protein evolution by amplifying the selective consequences of early arising mutations (Salverda et al. 2011).

Epistatic interactions between a mutation and its genetic background can impact the adaption of populations by affecting their ability to benefit from horizontally transferred mutations (Michener et al. 2014). Horizontal gene transfer (HGT) refers to the horizontal exchange of genetic materials between different organisms, even between distantly related species (Brown 2003) and has been suggested to play an important role in bacterial speciation (Ochman et al. 2000, Gogarten et al. 2002). Rapidly increasing amounts of genomic sequence data is revealing that horizontal gene transfer is quite

common in the evolutionary history and occurred in different organisms, ranging from prokaryotes to eukaryotes. Bacteria are probably frequently exposed to exogenous DNA fragments through transformation, transduction, and conjugation, but only a small proportion of such horizontally transferred DNA is likely to be maintained in the new host over evolutionary timescales (reviewed in (Popa and Dagan 2011)). Selective barriers are expected to be one of the key factors that constrain fixation of horizontally transferred genes in a new host. If the effect of a transferred allele is neutral or deleterious, the mutation is likely to be lost from the new population (Sorek et al. 2007). If the effect is beneficial, the mutation has the potential to spread rapidly in the new population (Treangen and Rocha 2011, Wiedenbeck and Cohan 2011, Acuna et al. 2012). Thus, the negative interaction between the horizontal transferred alleles and the host genetic background is likely to constrain the fixation of a horizontally transferred foreign gene, thereby promoting bacterial lineage divergence.

What we need to know more about epistasis between beneficial mutations?

Although many examples of epistatic interactions between beneficial mutations have been described, most recent studies focused on interactions between beneficial mutations arising within one gene, a single population or in replicate populations evolved from a common ancestor (Chou et al. 2011, Khan et al. 2011, Kryazhimskiy et al. 2011, Rokyta et al. 2011, Schenk et al. 2013, Kryazhimskiy et al. 2014). Thus there are relatively few mutational differences separating the different genotypes and, consequently, the chance for differential epistatic interactions to alter the effect of newly

occurring mutation is relatively low. As a result, only a small subset of the vast number of potential interactions relevant to the evolutionary success of mutations in natural environments is considered. In one demonstration that this omission may be important, higher-order epistasis—up to the order possible to be examined in lab studies, but then with no sign of a diminishment of influence—has been proposed to be evolutionarily important (Aragaki and Barber 2005, Weinreich et al. 2013, Fraïsse et al. 2014). Thus the overall genetic background-dependent nature of mutational effects, and how beneficial mutations selected in different environments interact with each other, needs to be further explored to better understand adaptive evolution in nature.

A better understanding of the interaction between beneficial mutations and diverse genetic backgrounds will allow us to address several important questions, such as whether the effect of beneficial mutation depends on genetic backgrounds, whether mutations selected in one genetic background tend to have similar interactions across different backgrounds, and if not, how these interactions depend on the genetic background and what pattern these dependencies may follow. Examining the interaction between beneficial mutations selected in different environments would allow us to directly test for the presence of Dobzhansky-Muller incompatibilities.

Focus of this dissertation

In this dissertation, I examined the extent and pattern of interactions occurring between beneficial mutations or between beneficial mutations and their genetic backgrounds. In particular, I address three topics. Chapter 2 demonstrates how the fitness

effect of beneficial mutations is affected by diverse genetic backgrounds by introducing each of four focal mutations into a set of natural isolates of *E. coli* and testing whether specific strain attributes—fitness, and the genetic and ecological distance between strains—can explain or even predict the fitness effect of beneficial mutations arising in a new genetic background. Chapter 3 illustrates how the overall genetic background affects the direction and magnitude of epistasis between two focal beneficial mutations, and the relationship between this epistasis and the initial fitness of genotype in which the beneficial mutations were introduced. Chapter 4 examines the interactions between beneficial mutations evolved in either the same or different environments to test for the role of adaptation in driving reproductive isolation at the initial stage of speciation.

Chapter 2: Benefit of mutations introduced into natural isolate strains of *Escherichia coli* depend on the recipient's growth rate, but notecological or genetic similarity

Summary

The effect of a given mutation can depend on interactions with the genetic background in which it is assessed. Studies in experimental microbial populations have demonstrated that such interactions are common among beneficial mutations and may generally follow a pattern that contributes to declining evolvability of more fit genotypes. In natural populations mutation-background interactions are also important because they can influence the spread of mutations that arise in one lineage and may then be transferred to others. We build on work that examined interactions between early beneficial mutations selected in a laboratory-evolved population of *Escherichia coli* to test the role of interactions between the same mutations and a diverse set of natural isolates of the same species. We find that the fitness effect of transferred mutations does not depend on the genetic or diet breadth similarity of recipient strains relative to the strain in which the mutations were originally selected. By contrast, the fitness effect of two mutations individually, as well as of the mutations considered together, was correlated to the initial fitness of the recipient strain. As in previous studies examining interactions between beneficial mutations, there was a pattern of diminishing returns whereby fit strains benefited proportionally less from transfer of the beneficial mutations.

Our results strengthen the view that the current fitness of a strain can be a major determinant of its ability to adapt.

Introduction

Mutations can interact with one another and with their broader genetic background to affect fitness, a phenomenon known as epistasis (De Visser et al. 2011). Such interactions play key roles in many aspects of developmental and evolutionary biology, including theories of speciation (Orr 1995, Schluter 2009), the evolution and maintenance of sex (Azevedo et al. 2006, De Visser and Elena 2007), adaptation (Kondrashov and Kondrashov 2001, Hayden et al. 2011, Woods et al. 2011) and evolutionary contingency (Weinreich et al. 2006, Salverda et al. 2011). Whereas, early experimental studies focused on interactions between deletion or other knockout mutations, advances in sequencing and genomic technologies now allow direct tests of interactions between spontaneously occurring beneficial mutations (MacLean et al. 2010, Chou et al. 2011, Khan et al. 2011, Rokyta et al. 2011, Wang et al. 2012, Kryazhimskiy et al. 2014). Studies that manipulate beneficial mutations have the potential to identify general patterns that may underlie some degree of predictability in evolutionary outcomes. For example, a common observation of studies measuring interactions between beneficial mutations is a pattern of diminishing returns epistasis, such that the marginal benefit of additional beneficial mutations declines with the fitness of the recipient genotype (MacLean et al. 2010, Chou et al. 2011, Khan et al. 2011, Rokyta et al. 2011, Wang et al. 2012, Kryazhimskiy et al. 2014). This pattern is consistent with the frequent observation

of decelerating fitness trajectories as populations adapt to constant environments (Kryazhimskiy et al. 2011, Wiser et al. 2013), and suggests that beneficial mutation interactions may generally follow a pattern of interaction that makes useful predictions regarding the form of fitness trajectories. In apparent contrast, some theoretical work predicts an excess of positive interactions along adaptive trajectories, highlighting the need for continued research in this area (Draghi and Plotkin 2013, Greene and Crona 2014).

Most studies that have directly examined interactions which affect beneficial mutations have focused on interactions between mutations arising within a single population or in replicate populations evolved from a common ancestor. In these cases there are relatively few mutational differences separating different genotypes. As a result, a vast complexity of mutation interaction relevant to natural populations is omitted, which may seriously limit our understanding of adaptive evolution. Indeed, introgression experiments indicate the importance of the broader genetic background in determining the effect of specific genetic regions (reviewed in Chandler et al. 2013); even the interaction between two focal mutations depends on the genetic background they are measured in (Wang et al. 2012).

A better understanding of the interaction between beneficial mutations and diverse genetic backgrounds will allow us to address whether mutations selected in one genetic background will tend to have similar effects across different backgrounds. This question is particularly important in considering the evolution of bacterial populations

where the horizontal transfer and integration of short DNA sequences—an average of ~50 bp in *E. coli* (Touchon et al. 2009)—means that interactions between single beneficial mutations and the broader genetic background into which they are transferred may be an important influence on the eventual fate of newly arising beneficial mutations. Indeed a new base may be introduced into a genome by recombination at a rate up to 100-fold greater than by mutation (Touchon et al. 2009). If the influence of interactions tends to be small, the effect of a beneficial mutation will be mostly independent of the particular genetic background, and it can spread broadly. If interactions are common, they may create a barrier preventing spread of potentially beneficial mutations between lineages. This kind of barrier has been proposed as a component of a bacterial species concept (Gevers et al. 2005). Alternatively, common mutation-background interactions may simply add a degree of stochasticity to the potential recipient range of transferred mutations, weakening the potential to predict outcomes from knowledge of overall background relatedness.

To the extent that mutations do have different effects across backgrounds, it is obviously of interest to identify attributes of those backgrounds that might explain, and even allow prediction of, those differences. At least three candidates have been presented: (i) as lineages diverge, they can develop unique co-adapted gene complex of epistatic interactions that might influence how they will interact with any new mutation (Orr 1995). Because closely related lineages are more likely to share these interactions, they may respond similarly to new mutations. (ii) Relatively distantly related lineages may evolve similar underlying genetic architectures by convergent evolution (Spor et al. 2009) and a

comparison of hundreds of bacterial metabolic networks suggests a correlation between metabolic network architecture and growth environment (Ruane and Donohue 2007, Kreimer et al. 2008). To the extent that this architecture can influence interactions between the genotype and new mutations, organisms that share selective environments, and thus have similar genetic architecture, may respond similarly to an introduced mutation, even if they are not genetically closely related. (iii) The relative benefit of a mutation transferred into different lineages is determined by the initial fitness of that lineage. This possibility stems from the frequent finding of negative interactions between beneficial mutations, perhaps the result of saturation in the possible improvement in the phenotypic processes that are being selected (MacLean 2010, Chou et al. 2011, Chou et al. 2014). If so, we might expect that potentially beneficial mutations will tend to confer larger benefits when transferred into less fit genotypes. Finally, interactions between mutations and the broader genetic background may be idiosyncratic, perhaps depending on relatively small numbers of large effect interactions or on interactions whose mechanistic basis changes between backgrounds. In that case it might be that no simple relationship can explain interaction effects, at least between a set of relatively closely related backgrounds.

We transferred beneficial mutations—in the genes or gene regions: *rbs*, *spoT*, *topA* and *pykF*—that fixed in a long-term lab evolved *E. coli* population to a diverse set of *E. coli* natural isolates (Khan et al. 2011). By measuring the fitness effect of each mutation × strain combination we determined the effect of global interactions between the focal mutations and recipient genetic backgrounds. We also examined if the fitness

effects of the beneficial mutations were correlated with phylogenetic relatedness and/or diet breadth similarity between the original and recipient genetic backgrounds. We found that the fitness of all mutations depended strongly on genetic background, but that this effect could not be explained by differences in the genetic relatedness of the recipients or similarity of their diet breadth profiles. It was explained by considering the initial fitness of the recipient strains, following a relationship of diminishing returns whereby more fit strains benefited least from transfer of the mutations.

Materials and Methods

Beneficial mutations and strain construction

Five beneficial mutations—occurring in the genes or gene regions: *rbs*, *topA, spoT, glmUS* and *pykF*—have been identified as the first to fix in a long-term evolving population of *E. coli* REL606 (Khan et al. 2011). We individually introduced four of these mutations into a series of natural isolate recipient strains, omitting the mutation upstream of *glmUS* because of the low transmission efficiency. Mutations were introduced into the chromosome of natural isolate strains using a suicide-vector based approach. As well as the four beneficial mutations, a mutation conferring an Ara- phenotype (*araA* D92G) the same mutation that distinguishes REL606 from its Ara+ derivative, REL607) was added into each natural isolate strain. In competition experiments this marker allowed us to distinguish progenitor from constructed strains on tetrazolium arabinose (TA) indicator media.

Bacterial strains

Natural isolate strains used as recipients for introduced beneficial mutations were chosen from a collection of 99 strains collected and sequenced as part of a Broad institute project and obtained from the Michigan State University STEC Center, and from strains obtained and used in a previous study by one of us (http://www.broadinstitute.org/) (Moore and Woods 2006). Genome sequences of strains were downloaded from the Broad Institute (http://www.broadinstitute.org/annotation/genome/escherichia_antibiotic_resistance/MultiDownloads.html) or obtained by de novo Illumina sequencing. Recipient strains used in this study are detailed in Table A2.1.

We sought to focus on intergenic epistasis by choosing recipient natural isolate strains that had the same amino acid sequence at the focal transferred gene as did the donor REL606 strain. For this reason the identity of potential recipient strains differed for constructions focusing on mutations in *spoT* and *topA*. The *rbs* and *pykF* mutations are large deletions that result in loss of function of target genes. In these cases, intra-gene interactions are not relevant, so these mutations were transferred into strains without regard to the original allele. The strains that were successfully constructed are a subset of the potential recipients because we were unable to successfully add either the focal beneficial mutation or the *araA* marker mutation into some target strains.

Fitness competition experiments

The fitness of constructed strains was measured using direct competition assays in the same medium used in the evolution experiment from which the clone containing the mutations manipulated in this work was isolated (Lenski et al. 1991, Khan et al. 2011). All competitions were carried out between strains with distinct Ara markers, which allows the two types to be distinguished on tetrazolium arabinose (TA) indicator media (Lenski et al. 1991). Except for control competitions, the two strains differed only by the presence or absence of an introduced mutation.

Genome and phylogenetic analysis

Core genes ('panorthologous' genes shared across all recipient strains) and accessory genes (genes shared amongst a subset of recipient strains) were identified using a previously described pipeline (Cooper et al. 2010). Briefly, the pipeline first identifies putative core genes using NCBI BLASTP (release 2.2.16) to analyze the total pool of genes for sequence similarity. Homologs were identified as those gene pairs that had BLAST hits in both directions within a bit score threshold scaled by the bit score of the self hit of the query gene. Pairs of homologous genes were grouped into families and panorthologs identified as genes from homolog families with exactly one gene from each genome. This analysis identified a total of 1648 panorthologs across the 27 genomes that were considered. A total of 7,017 accessory genes were identified as being present in at least one, but not all, strains. Adding a subsequent stringency filter to the panortholog gene set to require a limit of either <20 or <5 amino acid differences from the consensus sequence in the trimmed alignments of each gene across all genomes reduced the size of

the core genome to 1560 or 1442 genes, respectively, but did not qualitatively effect the outcome of any subsequent analysis. Likewise, an independent approach to identify core and accessory genome regions on the basis of shared DNA sequence windows, implemented in PANSEQ, resulted in relationships between strains that were largely unchanged from those determined by the whole gene analyses (Laing et al. 2010). Core and accessory genomes determined using Panseq were used for all analyses presented here.

Core and accessory genomes were used to build phylogenies with which to test for a phylogenetic signal in determining the effect of mutations introduced to the different recipient strains. PhyML was used to build a maximum likelihood tree of the core genome. For the accessory genome, a binary input file indicating the presence/absence of each accessory gene in each strain was analyzed using default parameters of PARS in PHYLIP (Felsenstein 2005).

Niche breadth

Niche breadth of strains was measured using Biolog PM1 plates. These plates allow estimation of the respiratory activity of each strain on each of 95 distinct substrates (Biolog, Hayward CA). Prior to growth in Biolog plates, each strain was grown overnight in lysogeny broth (LB) then concentrated by centrifugation and resuspended in phosphate buffered saline (PBS) to a defined optical density (OD). An aliquot of these cells was mixed with inoculating fluid (IF-0) containing a dye and transferred to each well of a PM1 plate. The plate was incubated at 37 C and its OD_{562} measured at 2 hr intervals until

12 hours, 4 hr intervals until 24 hours, and then at 48 hours. Respiration in each substrate was quantified as the area under the curve. A neighbor-joining tree was constructed using Biolog data using the program Neighbor in PHYLIP (Felsenstein 2005).

Growth rate analysis

The maximum growth rate of each progenitor strain was estimated by growing strains in 96 well plates containing DM medium supplemented with 500 μg/ml glucose and measuring changes in OD_{450} in a VersaMax plate reader (Molecular Dynamics, Sunnyvale CA). All strains were pre-conditioned in the assay media for two growth cycles prior to estimation of their growth rate. The higher concentration of glucose used in these assays compared to the competition assays was necessary for strains to achieve sufficient growth for reliable spectrophotometric detection. A custom R script was used to estimate the maximum growth rate of each strain.

Statistical analysis

All analyses were carried out in R (version 3.1.1). The nlme package was used to perform non-linear regressions. Recipient strains and introduced mutations were treated as fixed effects because our focus was on identifying and explaining mutation-genetic background interactions between the specific strains and mutations considered here. We use partial least square (PLS) regression to account for correlation between combinations of variables tested to explain the effect of mutations introduced into recipient strains. Whereas principal component regression seeks to determine orthogonal combinations of variables to maximize the amount of variation explained in those variables, PLS

determines combinations of variables that maximize the amount of variation in the response explained. PLS regression was implemented using the package pcr. Phylogenetic regression analyses were performed using the function pgls (phylogenetic general least square models) in the package Caper and tests for phylogenetic signal were performed using the function multiPhylosignal in the package Picante. The functions pd.calc and pd.bootstrap in the package Caper were used to compare the distance separating the strains used here to a distribution of distances between 1000 randomly chosen sets of the same number of strains from a set of 96 sequenced strains (all strains used here and additional strains sequenced by the Broad Institute) (Table A2.2 and Fig. A2.1). We use REL606 as the reference point for genetic and ecological distance regression analyses because it is the background in which mutations originally arose and were selected. All attributes of recipient strains are presented in tables A2.3, A2.4a, and A2.4b .

Results

Mutation effects depend strongly on genetic background

We added each of four beneficial mutations—in the gene or gene region *rbs*, *topA*, *spoT* and *pykF*— that fixed in a lab-evolved strain of *E. coli* into a series of recipient natural isolate strains of *E. coli* and measured their effect on strain fitness in the environment in which they were originally selected. Strains into which each mutation was transferred are presented in figure 2.1 and figure A2.1 (appendix) in the context of a broader phylogeny of *E. coli* obtained from a sequencing effort of diverse *E. coli* strains

(M&M for details). For all mutations, recipient strains represented a random sample with respect to a phylogeny constructed using 96 diverse strains (Fig. A2.2 and Table A2.2).

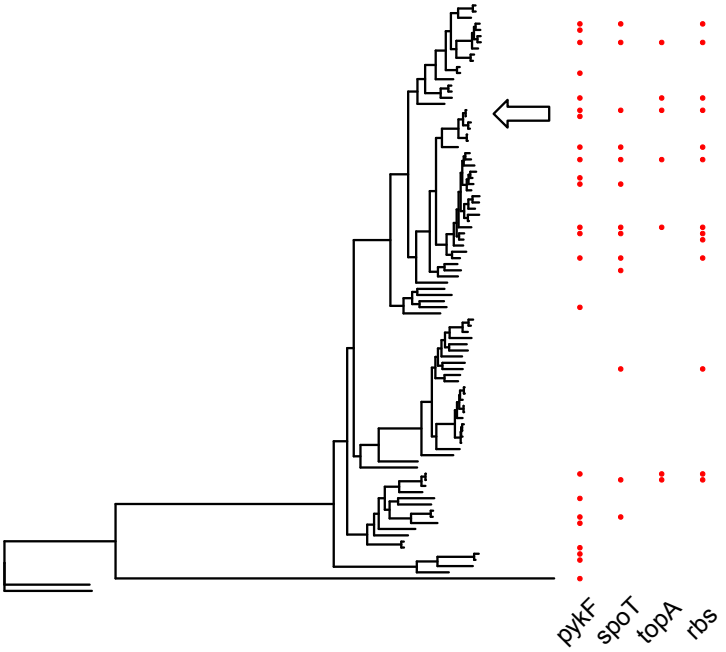

Figure 2.1 Phylogeny of 96 *E. coli* strains based on their shared core genome indicating strains to which the four beneficial mutations were transferred. Red dots indicate the recipients in which four beneficial mutations in *pykF*, *spoT*, *topA* and *rbs* were introduced. Arrow indicates strains REL606, the strain in which mutations were initially selected as part of a lab evolution experiment.

We initially asked if a given mutation had the same effect over different recipient strains; in other words, do mutation-by-genetic background interactions influence the fitness effect of these potentially beneficial mutations. For each of the four mutations, such interactions were statistically significant (Fig. 2.2 and Table 2.1). Moreover, the influence of these interactions on fitness clearly has the potential to be biologically meaningful. For example, considering the fitness effects conferred by the *pykF* mutation, which was introduced into 22 recipient strains, we see a range from ~0 to ~30%,

compared to an effect of ~10% in the strain in which it originally evolved. *spoT* mutation had significantly deleterious effects in most of recipient strains. It conferred a fitness benefit of ~10% in the strain in which it was originally selected, but was deleterious in eight, neutral in three and beneficial in only one recipient strains.

Table 2.1 Summary of mutation effects across recipient strains.

Mutation	**range**	**mean pair-wise difference in fitness**	**F**	**df***	**MS**	***P***
rbs	0.94 - 1.07	0.04	3.436	12,83	0.008	<0.001
topA	0.96 - 1.25	0.12	18.63	6,86	0.109	<0.001
spoT	0.84 - 1.09	0.10	11.56	12,121	0.078	<0.001
pykF	0.97 - 1.29	0.13	23.98	24,323	0.141	<0.001

*Degrees of freedom differ between mutations because they were introduced into overlapping but not identical sets of strains

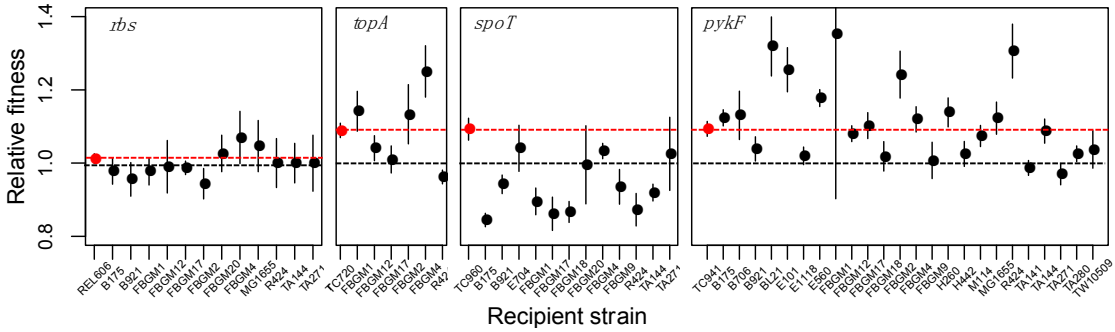

Figure 2.2 Fitness effects of transferred mutations in different strains. Points and error bars represent the mean and 95% CI of at least five independent fitness estimates. Strains are arranged alphabetically except that the strain in which the mutation was originally selected (red symbol) is at left. Black dash lines represent the fitness of recipients with mutation is the same level as corresponding progenitors and red dash lines represent the mean fitness effect of beneficial mutations in their original genetic backgrounds.

Basis of differences in mutation effects across strains

We sought to test candidate factors that might explain some of the variation in mutation effect across recipient backgrounds. Specifically, we test for a relationship between ecological, genetic, and growth attributes of recipient natural isolate strains, and the effect of each introduced mutation. Where appropriate, we do this taking into account the phylogeny of the recipient strains in order to reduce the chance of spurious relationships driven by closely related strains.

Mutation effects are not explained by phylogeny

The hypothesis that closely related strains are more likely to share genotype-by-mutation interactions and thus to respond relatively similarly to introduction of a new mutation leads to two testable predictions. First, that there will be some phylogenetic signal in the distribution of mutation fitness effects across recipient strains. Second, and more specifically, that there will be a negative relationship between the fitness effects of beneficial mutations and the genetic distance of recipient strains relative to the strain in which they were originally selected.

To test for a general influence of a phylogenetic signal on fitness effects we estimated Pagel's λ, a measure of phylogenetic signal in a response variable, considering the fitness effect of each mutation on phylogenies created based on both core and accessory genomes (Pagel 1999). Values of λ greater than 0 indicate some amount of phylogenetic signal such that the fitness effect of a transferred mutation is more similar in related strains than expected if by chance alone. In no case did the lower bound of the

maximum likelihood estimate of λ allows us to reject the null hypothesis of being no phylogenetic signal (core genome phylogeny: *pykF p* = 0.62*, rbs p* = 0.24, *spoT p* = 0.07, *topA p*= 0.99; accessory genome phylogeny: *pykF p* = 0.99*, rbs p* = 0.99, *spoT p* = 0.99, *topA p*= 0.38). In the context of our previous finding of significant mutation-by-strain interaction effects a low phylogenetic signal is consistent with rapid evolution of the genetic determinants that control these interactions. One way in which this could happen is if interactions are mediated indirectly through the effect of genotype on the physiology of recipient strains.

Of course, that we cannot reject the null hypothesis of no phylogenetic signal in determining the effect of introduced mutations does not mean that some kind of signal does not exist. For example, it could be that mutation effects tend to be similar in strains closely related to the strain in which they were selected, but fall off to be random in less closely related strains. To test this possibility we examined the relationship between the genetic distance of each recipient strain to the donor strain and the fitness effect of each mutation. In no case did we find any linear relationship between these variables (Fig. 2.3 & A2.3). Quadratic and exponential models that can accommodate a sharp drop and then a plateau of mutation effects in increasingly distantly related strains did not provide any substantial improvement in fit (Fig. A2.3). We conclude that the fitness effect of mutations selected in one strain and then transferred to the recipient strains considered here cannot be predicted by the genetic similarity of donor and recipient strains.

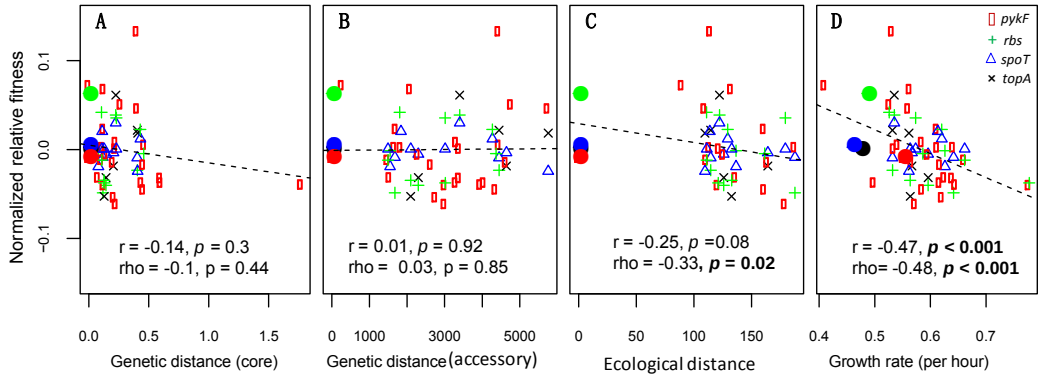

Figure 2.3 Relationship between fitness effect of added mutations and recipient strain attributes. Fitness effect of added mutations is compared against: core and accessory genome genetic distance of recipient strains relative to the lab strain in which mutations were originally selected (A and B, respectively), Biolog distance of recipient strains against the lab strain (C), and growth rate of recipient and original strains (D). Solid points indicate strains in which mutations initially were selected. Fitness effects are presented as the log ratio of fitness effect of a mutation in a given strain relative to its effect in the original strain normalized so that all mutations have an overall mean log ratio of zero. Significant *p*-value is in bold.

Mutation effects are not explained by ecological similarity

Genotypes adapted to a similar ecological niche may be more likely to have similar underlying genetic architectures and therefore respond similarly to a new mutation, even though they are not genetically closely related. We tested this possibility in two ways. First, we examined the relationship between ecological similarity — based on Biolog resource utilization profiles — of the different recipient strains, relative to the donor strain, and the fitness effect of the introduced mutations. Although there was a general trend for more ecologically distant strains to benefit less from *topA*, *spoT*, and *pykF* mutations, no linear or tested non-linear relationship was significant (Fig. 2.3 & A2.4). To account for the fact that recipient strains can have the same overall similarity relative to the progenitor, but be different from one another, we also tested for a signal of

mutation effect in the context of a phylogeny based on strain performance in each of the 95 Biolog resources. In no case was any significant phylogenetic signal observed. Together, these results indicate that current ecological similarity, as assessed by Biolog profiles, does not explain differences in the fitness effect of any of the four introduced mutations.

Fitter recipient strains benefit less from introduced mutations

Recent studies have found that epistatic interactions between beneficial mutations and their genetic backgrounds tend to become increasingly negative as the fitness of the genetic background increases (Chou et al. 2011, Khan et al. 2011, Kryazhimskiy et al. 2011, Rokyta et al. 2011, Wang et al. 2012). If this relationship also holds across the diverse recipient strains that we examine, we predict a negative relationship between the absolute fitness of a strain (measured as growth rate in the mutation assay environment) and the benefit conferred by addition of a mutation. We found that the fitness effect of two introduced mutations — in *spoT* and *pykF* — tended to decrease as the growth rate of recipients increased (*spoT*: r = -0.55, $p = 0.05$; ρ = -0.58, $p = 0.04$; *pykF* r = -0.51, $p = 0.01$; ρ = -0.35, $p = 0.11$). Negative, but statistically non-significant relationships were found for the *rbs* and *topA* mutations (Figs. 2.3 & A2.5). As judged by comparison of AIC scores, no tested non-linear relationship gave a substantially improved fit (Fig. A2.5).To increase our power to detect a relationship between growth rate and mutation effect we also considered all mutation effects together by normalizing individual effects to have the same mean. When we do this there is strong overall signal of a dependence of the fitness conferred by a mutation on the growth rate of the strain it is added to (r = -

0.429, p <0.001; ρ = -0.386, p = 0.005). This relationship remains significant when an outlying strain with a very high growth rate is omitted from the analysis (r = -0.307, p = 0.006; ρ = -0.316, p = 0.020) (Fig. A2.5).

Interactions between genotype, niche breadth and growth rate in explaining fitness effects

Even when a factor does not individually explain a significant proportion of the variation in a response, it may still contribute in combination with some other factor. To test for interactions between initial fitness, genetic and ecological similarity, and the effect of each introduced mutation, we performed a partial least square regression. This approach is robust to having relatively few observations relative to the number of predictor variables, as was the case especially for the *rbs* and *topA* mutation datasets considered here. For the *pykF, spoT*, and *topA* mutations, growth rate explained the largest proportion of variance in fitness effects (Fig. 2.4).

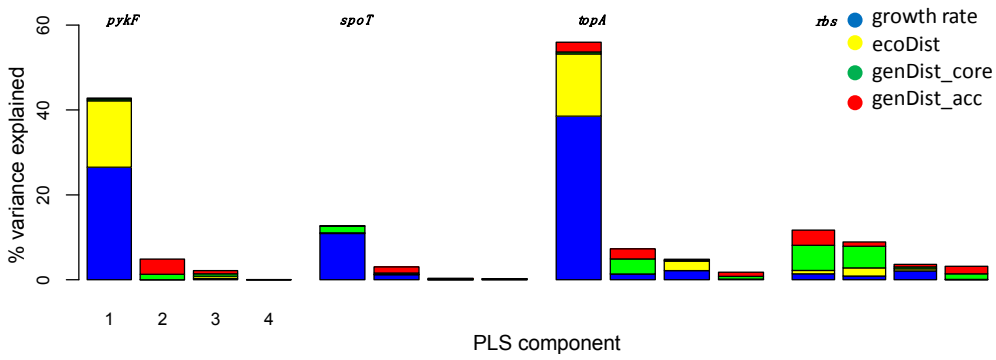

Figure 2.4 Partial least square (PLS) regression to determine contribution of genetic distance, Biolog distance, and growth rate, to mutation fitness effects. The original strains are excluded from this analysis because they have a disproportionate influence on variation of genetic and Biolog distance measures. Color of bars indicates the contribution of each strain attribute to that component. Only the first four components are shown.

Discussion

Studies that have examined the effect of mutation-by-background interactions on fitness have found that their influence is common and often significant (reviewed in (Chandler et al. 2013)). The advent of genome sequencing techniques has enabled the interactions to be assessed focusing on beneficial mutations and so the role of epistasis in adaptation to be examined. Those studies have, however, typically examined interactions that occur between a small set of focal mutations in the context of the same broader genetic background (Chou et al. 2011, Khan et al. 2011, Rokyta et al. 2011) but there are exceptions, (e.g., (Pearson et al. 2012, Wang et al. 2012)). We have extended this approach by estimating the effect of interactions between the broader genetic background and four mutations that conferred a benefit in a focal genetic background that was selected as part of a long-term evolution experiment. We found that epistasis had a major

influence on mutation effects. These effects did not depend strongly on the genetic or diet breadth relationships between recipient strains, but they did depend on each strain's initial fitness. As in previous studies considering interactions between focal mutations, the benefit of introducing a mutation declined with the fitness of the strain to which it was added (MacLean et al. 2010, Chou et al. 2011, Khan et al. 2011).

It is perhaps surprising that we did not find a strong phylogenetic signal in the effect of the mutations we considered, indicating that the genetic basis of mutation-background interactions was not predicted by the overall genetic relationships used to construct our core or accessory genome phylogenies. One explanation for this is that there may be a small number of genes and perhaps just sites in those genes, that interact with the transferred mutations and the identity of these genes/sites is not captured by genome-level phylogenies. A similar explanation has been proposed for the absence of a strong phylogenetic signal in the effect of a gene transferred into *Methylobacterium* strains (Michener et al. 2014). Discordance between the interacting genes and the overall phylogeny might be exacerbated if the interacting genes are under selection or if their identity changes between strains.

What could account for a strong relationship between genotype growth rate and benefit conferred by a transferred mutation? Several studies have identified and examined mutation interactions mediated through a specific biochemical pathway or within one enzyme (Lunzer et al. 2010, MacLean 2010, Chou et al. 2011, Chou et al. 2014). The adaptive target of the four mutations we consider are not known, so we cannot offer any

specific explanation that takes into account the particular mutations and backgrounds used in our study. Nevertheless, it is worthwhile to consider a general explanation that follows from the principles of metabolic control theory; that a key target of selection is a central physiological process has a saturating form such that further improvements confer diminishing benefits. This process should be common to all backgrounds and, because the mutations we consider also interact directly with each other following a pattern of diminishing returns (Khan et al. 2011), it should be a single process that is affected by many general beneficial mutations. The process of translation meets these requirements. Translational capacity is selected to follow a balance between being high enough to ensure sufficient expression of necessary genes but not being so high as to represent a wasteful investment in unused capacity (Maitra and Dill 2015). The ideal balance will depend on the distribution of translation activity across all genes in a cell. Translation provides a means to mediate interactions between seemingly disparate mutations, for example, considering two of the mutations examined in this work: the deletion of *rbs* genes, which might provide an advantage by reducing energy expenditure and ribosome allocation to unnecessary gene expression and *topA*, which impacts gene expression through altering DNA supercoiling. The pattern of gene expression change caused by the *spoT* mutation studied here is consistent with translation being affected by it (Cooper et al. 2003). Some support for this possibility comes from the finding that fitness effects of defects in translational and transcriptional capacity interact antagonistically with each other (MacLean 2010).

Mutation interactions play a key role in adaptation. Specific interactions clearly affect the specific genetic and fitness trajectory a population can follow and recent findings of general patterns of interactions suggests that they may cause general aspects of adaptive trajectories to be predictable (Kryazhimskiy et al. 2009, Wiser et al. 2013, Kryazhimskiy et al. 2014). Several studies have found a trend toward diminishing returns epistasis when assessing interactions between specific focal mutations (Chou et al. 2011, Hall and MacLean 2011, Khan et al. 2011). That is, the benefit of a mutation tends to decline as it is added to more fit backgrounds. This kind of diminishing returns epistasis can explain patterns of fitness increase and mutation accumulation in a well-studied lab-evolved population (Kryazhimskiy et al. 2009, Wiser et al. 2013, Kryazhimskiy et al. 2014). Our findings support the idea of a general, though not necessarily universal, relationship between strain fitness and the potential benefit of a specific mutation.

Although experimental studies have typically found a pattern of diminishing returns epistasis, most interactions considered did not lie along an adaptive path that was actually followed by an evolving population. For example, in a network representing combinations of the first five beneficial mutations to fix in a population, only one of the 120 possible mutation combinations can actually have been followed. Analysis of diverse evolutionary models has found that mutation interactions along an adaptive pathway tend to be positive, even in the midst of a tendency toward diminishing returns epistasis considering all possible interactions (Draghi and Plotkin 2013, Greene and Crona 2014). This discrepancy follows from the action of selection in favoring mutational steps of high benefit, thereby biasing toward new mutational steps that interact positively with the

previous genotype background. Our work does not follow mutation interactions along a specific trajectory, and so does not represent a direct test of the prediction the pattern of mutation interactions relevant to adaptation are not represented by the overall pattern of interactions. It does however broaden the basic design of experiment types consistent with diminishing returns epistasis from those focusing on interactions between single mutations to interactions between mutations and genetic backgrounds. Moreover, the pattern we see suggests the importance of physiology in mediating mutation interactions, something not considered in the models.

Mutation-background interactions also impact the ability of populations to adapt by influencing their ability to benefit from horizontally transferred mutations. Such transfer is common in many bacteria (Dagan and Martin 2007, Dagan et al. 2008), including *E. coli* (Touchon et al. 2009), and can be a major determinant of adaptation (Wiedenbeck and Cohan 2011). Horizontal transfer disconnects a beneficial mutation from the particular background it is initially selected in, so that its interaction with other possible mutations becomes relevant to determining both its own fate and that of recipient lineages. This process is clearly relevant to thinking about the nature of bacterial species. One bacterial species concept proposes that ecotypes — groups of genotypes that have similar fitness across a range of relevant environments — can be recognized as being subject to purifying selective sweeps in that beneficial mutations arising within one member of an ecotype will increase in frequency and so drive competing members of the ecotype extinct, as a consequence of competition within the same ecological niche (Wiedenbeck and Cohan 2011). In this way, selection within ecotypes is expected to

create a correlation between ecological and genetic similarity of strains. To the extent that horizontal transfer of new beneficial mutations occurs, genetic interactions between them and recipient genetic backgrounds can complicate this expectation. If beneficial mutations that arise in a focal strain tend to have equivalent effects in the ecologically similar strains that make up an ecotype, their transfer might retard selective sweeps within ecotypes. This would make each ecotype harder to recognize as a genotypic grouping. If a transferred beneficial mutation has beneficial effects in some members of multiple ecotypes, its increase in frequency could produce a pattern of population-level genetic change that is hard to interpret with reference to ecotypes. Our results are relevant to consequences of competition within, and perhaps between, ecotypes. Diminishing returns epistasis 'tilts the playing field' so that less fit lineages gain more than more fit lineages from transferred beneficial mutations. This process will act to reduce competitive differences between lineages within an ecotype, effectively promoting their co-existence and decreasing the genetic sweeps that bring genotypic cohesion to ecotypes, making them difficult to identify (Vos 2011).

Our results provide direct evidence that the effect of beneficial mutations varies dramatically over divergent strains of the same bacterial species. These effects were not predicted by the genetic or ecological similarity of species, though a significant portion of fitness variation was explained by the initial fitness of a recipient strain. This result supports an accumulating body of work consistent with an important role of physiology in mediating mutation interactions. While the current theoretical focus on predicting patterns of mutation interactions through abstract models is valuable in its production of

testable predictions, it will be important to also consider models that incorporate real and perhaps specific phenotypic interactions. Our results demonstrate that this addition may preserve the hope of identifying general patterns of mutation interactions that lead to some predictably of evolutionary processes.

Appendix 2

Table A2.1 Recipient strains used in this study

Strain ID*	Alternative ID	Obtained from or reference	Genome sequence
B175	TW15935	Broad Institute via MSU STEC center	Broad Institute†
B706	TW15943	Broad Institute via MSU STEC center	Broad Institute
B921	TW15945	Broad Institute via MSU STEC center	Broad Institute
BL21		ATCC	
E101	TW15947	Broad Institute via MSU STEC center	Broad Institute
E1118	TW15949	Broad Institute via MSU STEC center	Broad Institute
E560	TW15955	Broad Institute via MSU STEC center	Broad Institute
FBGM1	VS-151	Francisco Moore (University of Akron)	This study
FBGM2	VS-126	Francisco Moore (University of Akron)	This study
FBGM4	TA135	Francisco Moore (University of Akron)	This study
FBGM9	TA260	Francisco Moore (University of Akron)	This study
FBGM12	VS820	Francisco Moore (University of Akron)	This study
FBGM17	ECOR1	Francisco Moore (University of Akron)	This study
FBGM18	ECOR11	Francisco Moore (University of Akron)	This study
MG1655		CGSC	
H260	TW15964	Broad Institute via MSU STEC center	This study
H442	TW15976	Broad Institute via MSU STEC center	Broad Institute
M114	TW15991	Broad Institute via MSU STEC center	Broad Institute
R424	TW15997	Broad Institute via MSU STEC center	Broad Institute
REL606		(Khan et al. 2011)	(Jeong et al. 2009)
TC720 #			
TC960#			
TC941#			
TA141	TW16010	Broad Institute via MSU STEC center	Broad Institute
TA144	TW16012	Broad Institute via MSU STEC center	Broad Institute
TA271	TW16017	Broad Institute via MSU STEC center	Broad Institute
TA280	TW16018	Broad Institute via MSU STEC center	Broad Institute
TW10509	TW16023	Broad Institute via MSU STEC center	Broad Institute

* ID from (Moore and Woods 2006) or MSU STEC center documentation.

† Broad genome sequences downloaded from: http://www.broadinstitute.org/annotation/genome/escherichia_antibiotic_resistance/downloads.html

#Strains are derived from REL606 and differ by the addition of combinations of beneficial mutations as described in Materials and Methods.

Table A2.2 List of all strains used in full *E. coli* phylogeny.

Strain ID					
B008	E1118	FBGM9	H386	M056	TA008
B093	E1167	H001	H397	M114	TA014
B108	E1492	H120	H413	M605	TA054
B175	E1520	H185	H420	M646	TA103
B185	E267	H218	H442	M718	TA141
B354	E482	H220	H454	M863	TA143
B367	E560	H223	H461	M919	TA144
B574	E704	H252	H489	MG1655	TA206
B671	FBGM1	H260	H504	R424	TA249
B706	FBGM12	H263	H588	R527	TA255
B799	FBGM17	H288	H593	R529	TA271
B921	FBGM18	H296	H605	REL606	TA280
BL21	FBGM2	H299	H617	T408	TA435
E1002	FBGM20	H305	H660	T426	TA447
E101	FBGM3	H378	H730	TA004	TA464
E1114	FBGM4	H383	H736	TA007	TW10509

Table A2.3 Fitness effects of mutations of all recipient strains.

Strain ID	**Fitness effect of added mutation*:**			
	rbs	***topA***	***spoT***	***pykF***
REL606	1.012	NA	NA	NA
TC720	NA	1.089	NA	NA
TC960	NA	NA	1.093	NA
TC941	NA	NA	NA	1.093
B175	0.979	NA	0.845	1.124
B706	NA	NA	NA	1.13
B921	0.956	NA	0.942	1.039
BL21	NA	NA	NA	1.319
E101	NA	NA	NA	1.255
E1118	NA	NA	NA	1.02
E560	NA	NA	NA	1.178
E704	NA	NA	1.041	NA
FBGM1	0.978	1.142	0.896	1.353
FBGM12	0.99	1.041	NA	1.08
FBGM17	0.987	1.01	0.862	1.102
FBGM18	NA	NA	0.867	1.018
FBGM2	0.944	1.133	NA	1.242
FBGM20	1.026	NA	0.995	NA
FBGM4	1.07	1.25	1.033	1.12
FBGM9	NA	NA	0.935	1.007
H260	NA	NA	NA	1.139
H442	NA	NA	NA	1.025
M114	NA	NA	NA	1.074
MG1655	1.047	NA	NA	1.123
R424	1	0.962	0.873	1.306
TA141	NA	NA	NA	0.987
TA144	1	NA	0.92	1.087
TA271	1	NA	1.025	0.97
TA280	NA	NA	NA	1.025
TW10509	NA	NA	NA	1.037

* Fitness of natural isolate strain with added mutation relative to its progenitor strain.

Table A2.4a Descriptive characteristics of recipient strains obtained from the Broad Institute.

Strain ID	Core genome		Accessory genome		Ecological distance§	Growth rate (h^{-1})
	Panseq†	Gene‡	Panseq	Gene		
B175	0.117	0.332	1675.4	922.52	187.73	0.642
B706	0.467	0.77	4453.1	1852.8	NA	0.542
B921	0.08	0.4	1547.1	739.13	136.53	0.626
E101	0.27	0.619	4779.9	1472.8	NA	0.527
E1118	1.774	2.971	3963.2	1658.9	NA	0.646
E560	0.126	0.334	1714.1	938.12	161.103	0.612
E704	0.106	NA	1803.7	NA	111.074	0.526
H260	0.227	1.531	3321.7	1378.3	NA	0.585
H442	0.603	1.095	4052.4	1643.7	NA	0.617
M114	0.459	0.863	2648.5	1852.9	NA	0.618
R424	0.128	0.386	2097.1	1129.9	132.685	0.564
TA141	0.213	0.545	2767.0	1154.1	NA	0.622
TA144	0.108	0.323	1492.8	810.58	111.150	0.662
TA271	0.23	0.554	3011.5	1454.4	179.350	0.573
TA280	0.444	0.975	3307.6	1887.4	NA	0.499
TW10509	0.6	1.103	3395.9	1860.9	NA	0.640

† Panseq core and accessory genomes determine as described in SI. Units are patristic distances.

‡ Gene level core and accessory genomes determined as described in M&M and SI. Units are patristic distances.

§ Calculated as Euclidean distance to the donor REL606 strain over 94 resources measured in a Biolog PM1 plate (glucose measurement was excluded).

Table A2.4b Descriptive characteristics of recipient strains, including lab strains and the natural isolate strains obtained from Moore lab.

Strain ID	Core genome		Accessory genome		Ecological distance§	Growth rate (h^{-1})
	Panseq†	Gene‡	Panseq	Gene		
REL606	0	0	0	0	0	0.46
TC720	0	0	0	0	0	0.475
TC960	0	0	0	0	0	0.487
TC941	0	0	0	0	0	0.551
FBGM1	0.404	0.901	4447.9	1857.4	114.616	0.532
FBGM12	0.205	0.893	4643.9	1674.6	164.112	0.567
FBGM17	0.144	1.643	2301.2	2058.7	125.923	0.596
FBGM18	0.129	0.493	3016.4	1096.1	121.104	0.778
FBGM2	0.407	0.907	5755.6	1929.9	109.596	0.561
FBGM20	0.43	1.27	4269.8	1806.3	129.336	0.621
FBGM4	0.225	0.608	3400.3	1395.2	122.198	0.535
FBGM9	0.458	1.126	4367.3	1866.8	161.449	0.586
MG1655	0.112	0.321	1840.1	379	113.821	0.613
BL21	0.002	0.015	280	535.99	89.892	0.410

† Panseq core and accessory genomes determine as described in SI. Units are patristic distances.

‡ Gene level core and accessory genomes determined as described in M&M and SI. Units are patristic distances.

§ Calculated as Euclidean distance to the donor REL606 strain over 94 resources measured in a Biolog PM1 plate (glucose measurement was excluded).

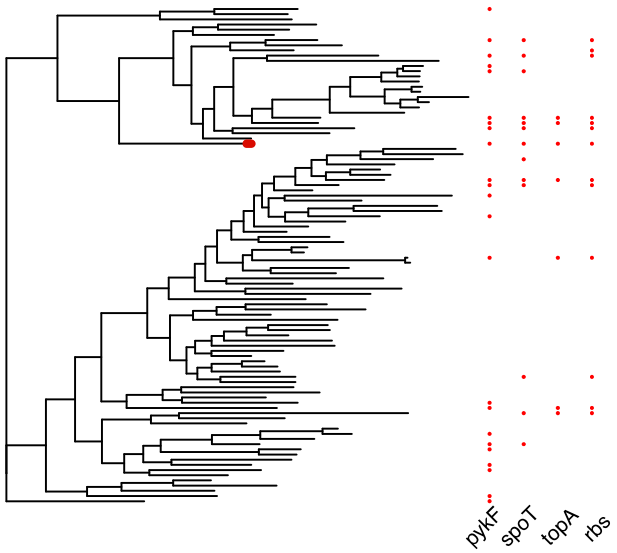

Figure A2.1 Phylogeny of 96 *E. coli* strains based on their shared accessory genome (see M&M for details) indicating strains to which the four beneficial mutations were transferred. Arrow indicates strains REL606, the strain in which mutations were initially selected as part of a lab evolution experiment.

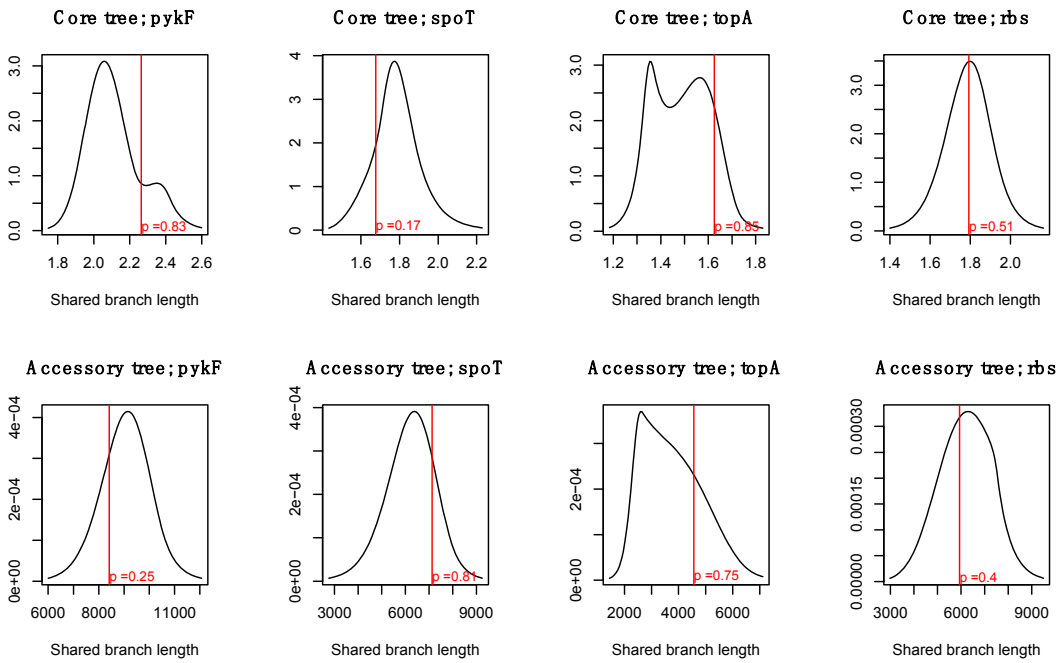

Figure A2.2 Recipient strains are randomly drawn from a larger phylogeny. The strains used as recipients for each gene were tested for being representative draws from core and accessory genome phylogenies. The solid black line indicates the distribution of shared branch lengths of 1000 draws of the same number of strains as used as recipients for the relevant gene transfer from the comprehensive phylogenies presented in Figures 2.1 and A2.1 The red line indicates the shared branch length of the actual recipient strains. P-values are calculated as the fraction of bootstrapped samples with a lower branch length than the actual sample. Mutation-phylogeny combinations are indicated in the title of each panel.

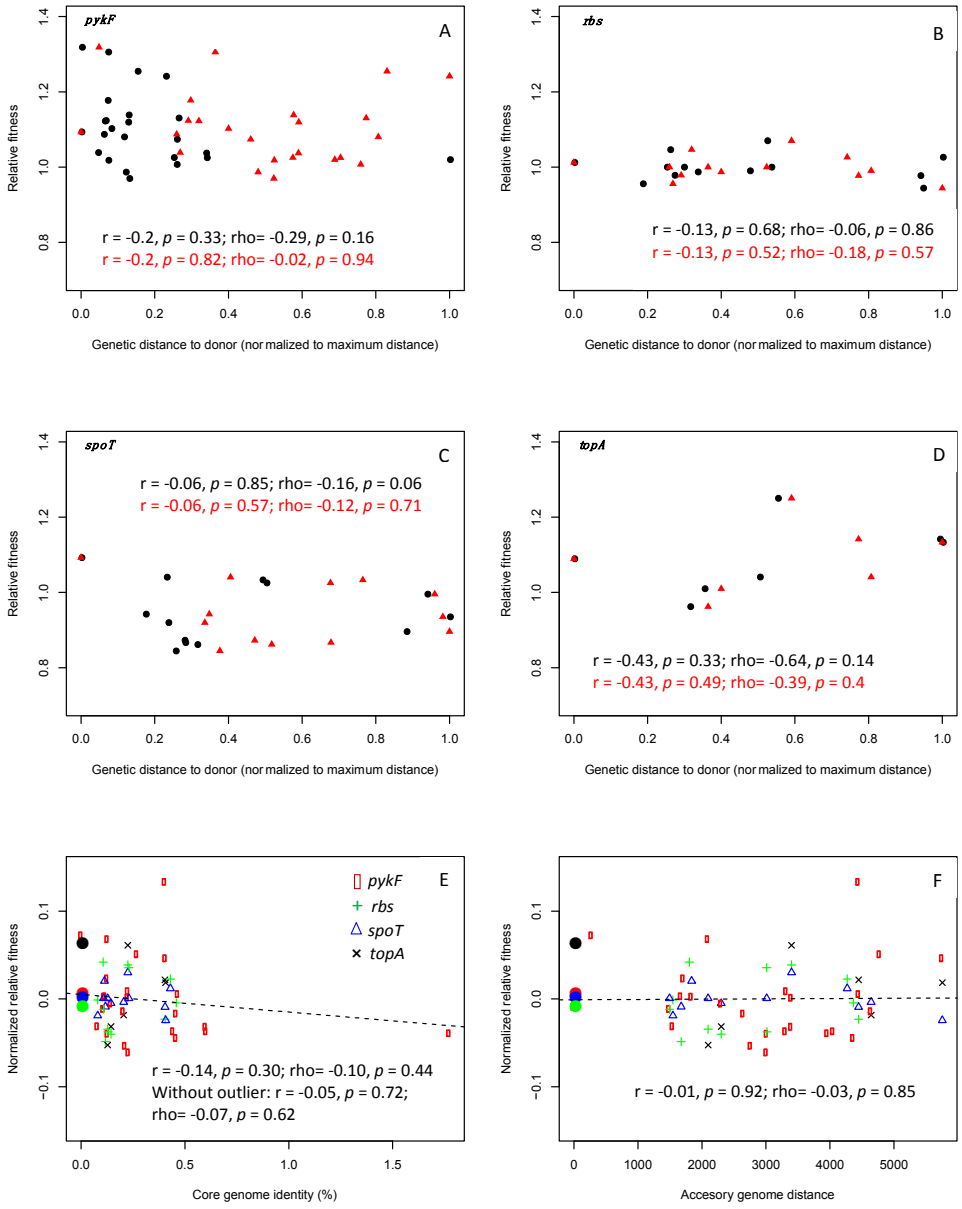

Figure A2.3 Pearson correlation and Spearman's rank correlation analysis between the fitness effect of individual beneficial mutation in *rbs*(A), *topA* (B), *spoT*(C) and *pykF*(D) and the genetic distance based on core genes (black) and accessory genes (red) to recipient strains. The correlation between the normalized relative fitness of all four mutations and genetic distance based on core genes and accessory genes (F).

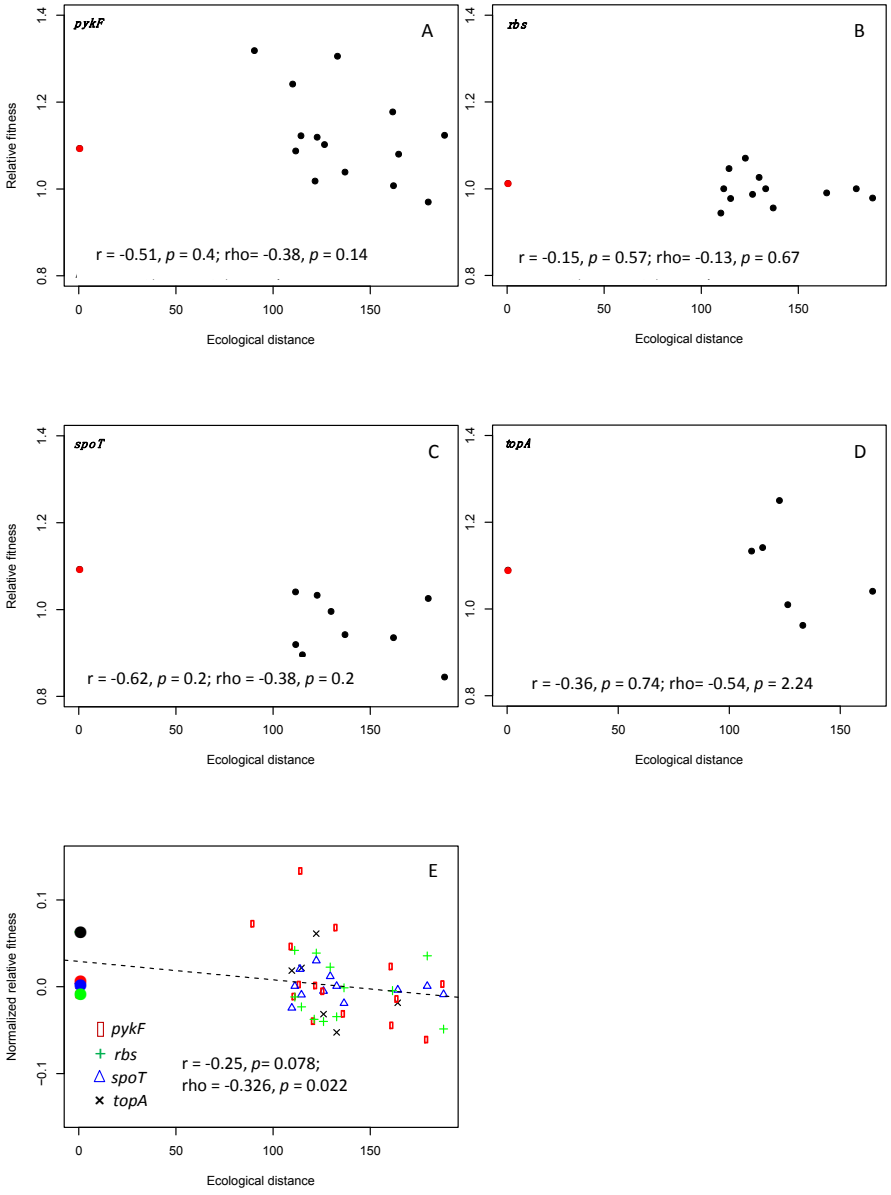

Figure A2.4 The correlation analysis between the fitness effect of individual beneficial mutation in *rbs*(A), *topA* (B), *spoT*(C) and *pykF*(D) and the ecological distance to recipient strains. The linear regression between the normalized relative fitness of all four mutations and ecologial distance to recipient strains (E). Red dots in figure A-D and the solid dots in figure E indicate the original genetic background in which the mutations arise.

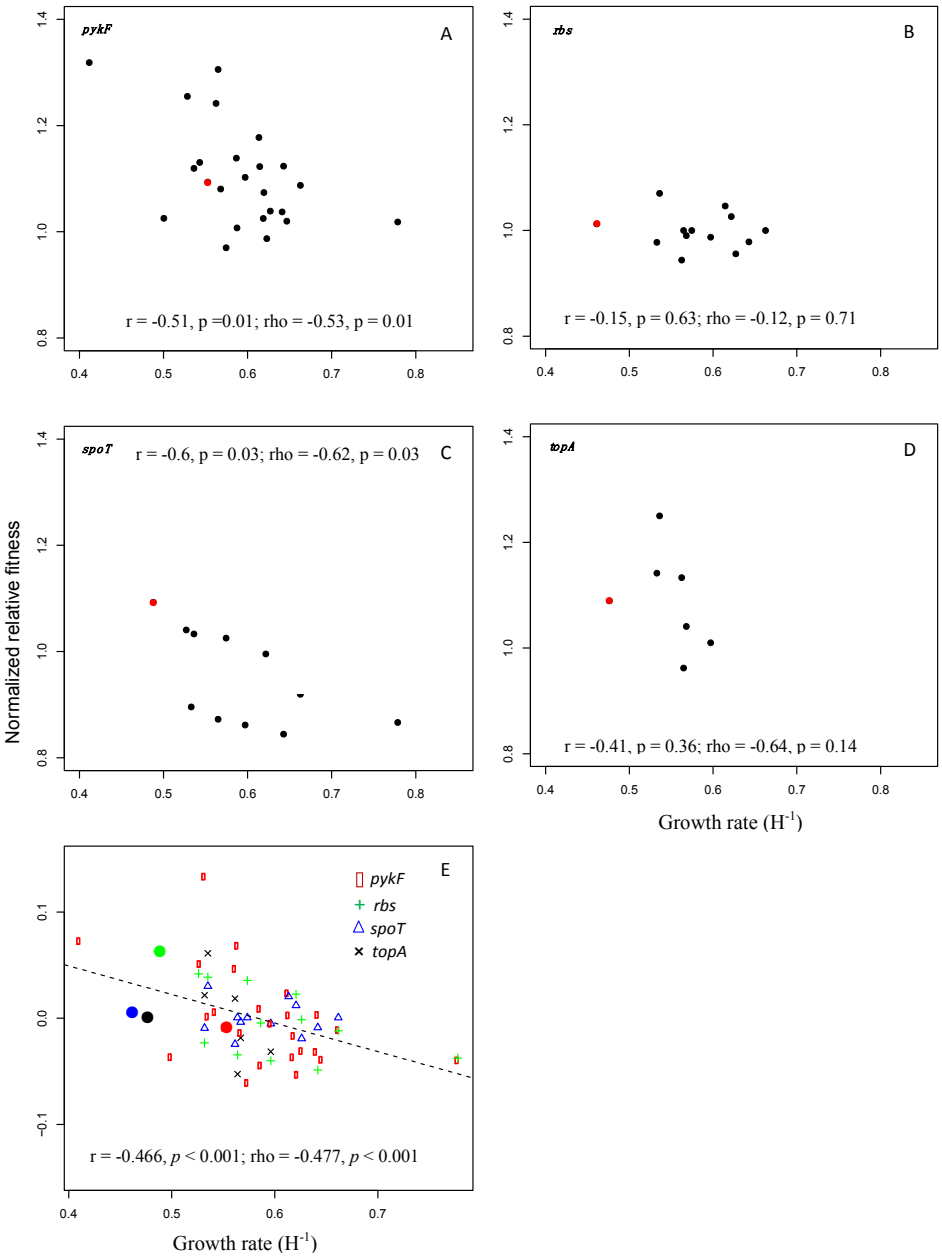

Figure A2.5 The correlation analysis between the fitness effect of individual beneficial mutation in *rbs*(A), *topA* (B), *spoT*(C) and *pykF*(D) and the growth rate of recipient strains. The linear regression between the normalized relative fitness of all four mutations and growth rate of recipient strains (E). Red dots in figure A-D and the solid dots in figure E indicate the original genetic background in which the mutations arise.

Acknowledgements

I thank Carolina Aremas, Daniel Stoebel for constructing strains. I thank Ethan Knapp and Kenny Flynn for sequencing analysis and biolog assay. I thank Fen Peng for help in doing fitness assay and growth rate assay and to Dr. Tim Cooper for phylogeny analysis. This chapter is a manuscript prepared for submission to a journal. Both my advisor Dr. Tim Cooper and I contributed to this manuscript significantly.

Chapter 3: Genetic background affects epistatic interactions between a pair of beneficial mutations*

Summary

The phenotypic effect of mutations can depend on their genetic background, a phenomenon known as epistasis. Many experimental studies have found that epistasis is pervasive, and some indicate that it may follow a general pattern dependent on the fitness effect of the interacting mutations. These studies have, however, typically examined the effect of interactions between a small number of focal mutations in a single genetic background. Here, we extend this approach by considering how the interaction between two focal beneficial mutations that were isolated from a population of laboratory-evolved *Escherichia coli*, changes when they are added to divergent natural isolate strains of *E. coli*. We find that interactions between the focal mutations and the different genetic backgrounds are common. Moreover, the pair-wise interaction between the focal mutations also depended on their genetic background, being more negative in backgrounds with higher absolute fitness. Together, our results indicate the presence of interactions between focal mutations, but also caution that these interactions depend quantitatively on the wider genetic background.

* This chapter has been published in Biology Letters (Wang, Y., C. Díaz Arenas, D. M. Stoebel, and T. F. Cooper. 2012. Genetic background affects epistatic interactions between two beneficial mutations. Biology Letters **9**:20120328.).

Introduction

The phenotypic effect of a mutation can depend on its genetic background. This dependence, known as epistasis, is important to many areas of developmental and evolutionary biology, including speciation (Orr and Turelli 2001, Brideau et al. 2006, Dettman et al. 2007, Anderson et al. 2010), the maintenance of sex (Azevedo et al. 2006, Kouyos et al. 2007), adaptation (Chou et al. 2011, Khan et al. 2011, Woods et al. 2011), the evolution of ploidy (Omholt et al. 2000), and evolutionary contingency (Weinreich et al. 2005, Weinreich et al. 2006, Salverda et al. 2011, Schaper et al. 2011). As the technology to identify and manipulate specific mutations becomes increasingly available, the influence of epistasis can be studied directly. For example, several studies have examined specific examples of epistasis causing evolutionary contingency (Weinreich et al. 2005, Salverda et al. 2011). Other studies have demonstrated a trend for beneficial mutations to interact antagonistically, causing the rate of adaptation to slow as beneficial mutations accumulate (Chou et al. 2011, Khan et al. 2011), as predicted by theory (Kryazhimskiy et al. 2009).

Of note, however, most experimental studies of epistasis focus on interactions between mutations arising within a single population or replicate populations evolved from a common ancestor. In these cases, relatively few mutational differences separate different genotypes, and consequently, there are relatively few opportunities for differential epistatic interactions to alter the effect of newly occurring mutations. It seems likely that the effects of a single mutation when added to relatively divergent genotypes

might reveal a much wider variety of phenotypes reflecting the many different mutation interactions. Indeed, background-dependent interaction has been proposed as a general explanation for the common failure to find genetic determinants that explain a large fraction of phenotypic variation in natural populations (Gibson 2012). Here, we test for higher-order interactions revealed when not just the main effect, but also the interaction between focal mutations, depends on the wider genetic background (Da Silva et al. 2010).

We examine the effect of the interaction between two mutations — in the genes *topA* and *pykF* — on fitness in a series of distinct genetic backgrounds. These mutations were isolated from a laboratory-evolved population of *Escherichia coli* and were beneficial in the genetic contexts in which they arose, although the magnitude of this benefit depended on the presence of other evolved mutations. By comparing the effects of these mutations individually and in combination in seven natural isolate strains, we isolate the effects of background dependence on the individual and epistatic interaction between these mutations.

Materials and Methods

Bacterial strains

Strains of *E. coli* were obtained from the Broad Institute (http://www.broadinstitute.org/annotation/genome/escherichia_antibiotic_resistance/MultiHome.html) (E267 and R424) and from Francisco Moore (University of Ohio, Akron; ECOR1, VS-126, VS-820, TA135 and TA105) (Table A3.1). Both collections were isolated with the aim of being representative samples of the genetic diversity of *E. coli*.

Strain REL606 was used as the ancestor to the evolution experiment in which the mutations we study were isolated. Preliminary sequencing and phylogenetic analysis indicate that the strains we chose represent the major clades of the species. The sequencing of *topA* in all recipient strains indicated no amino acid differences from REL606. The effect of transferring the evolved allele is therefore limited to intergenic interactions, at least at the protein level. Because the *pykF*-evolved allele represents a large deletion, introduction of this allele can only affect intergenic interaction.

Mutations and genetic manipulations

Mutations in *topA* (Woods et al. 2011) and *pykF* (Schneider et al. 2000) were identified in a population evolved in a minimal glucose environment as part of a long-term evolution experiment (Schneider et al. 2000). We moved the *topA* mutation (H33Y) and a deletion allele of *pykF* that was equivalent to the evolved *IS*150 insertion mutation (Khan et al. 2011) individually and in combination into each of our seven natural isolate strains using a suicide vector approach described previously (Khan et al. 2011). To address the possibility that secondary mutations occurred during the construction process, for each allele replacement strain we obtained a paired clone that went through the same construction process but that retained the original allele. The allele replacement strain was only kept if the control clone had fitness indistinguishable from the corresponding progenitor strain. We also introduced a mutation in *araA* into all natural isolate strains to allow for discrimination of constructed strains from their progenitor on tetrazolium arabinose (TA) indicator plates (Lenski et al. 1991).

Competitions and growth rate

The fitness of each constructed strain was measured relative to an *araA*-derivative of its progenitor using direct competition experiments carried out in Davis minimal (DM) medium supplemented with 25 μg ml^{-1} glucose (Lenski et al. 1991). In some strains, the *araA*-competition marker was not neutral. In these cases, we normalize relative fitness estimates to account for the marker effect. Maximum growth rates were estimated by growing strains in 96 well plates containing DM medium supplemented with 500 μg ml^{-1} glucose and measuring changes in OD450 in a VersaMax plate reader. All strains were pre-conditioned in the same media. An R script was used to estimate the maximum growth rate of each strain (Table 3.1).

Statistical analysis

We used a multiplicative model to test for epistasis because fitness effects are expected to combine exponentially. We note, however, that multiplicative and additive null models will be very similar for the fitness values we consider. We calculate epistasis as a relative magnitude: $E_m = \log_{10}\left(w_M / \prod_{i \in M} w_i \right)$, where w_M is the fitness of a mutant with a set of M mutations and w_i is the relative fitness of a mutant containing a single mutation from set M (Da Silva et al. 2010). This measure indicates the relative change in fitness due to epistasis. Our results are qualitatively unaffected if we instead use an absolute measure of epistasis (i.e., the difference between observed and expected genotype fitness).

One-way ANOVAs were used to test for differences in the effect of the *topA* and *pykF* mutations and their epistatic interaction over different genetic backgrounds. Genetic background was included in these models as a random factor. All analyses carried out in R 2.14.2 (http://cran.r-project.org/).

Results and discussion

The fitness effects of *top*A and *pykF* mutations depend on their genetic background

The *topA* and *pykF* mutations were beneficial in the contexts of the genetic backgrounds in which they first occurred (Khan et al. 2011). Individually the *topA* mutation was beneficial in five and neutral in three of the genetic backgrounds we consider, and the *pykF* mutation was beneficial in six and neutral in two (Fig.s3. 1, A3.1). One-way ANOVA found that the fitness effect of both mutations varied across the different genetic backgrounds, indicating the presence of different gene-by-genotype epistatic interactions that effect fitness (*topA*: $F_{7,45} = 26.6554$, $P < 0.0001$; *pykF:* $F_{7,45} = 38.518$, $P < 0.0001$).

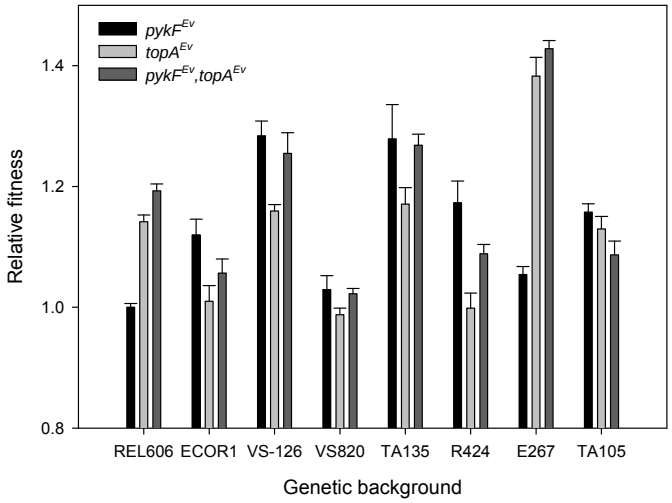

Figure 3.1 Relative fitness effect of *topA* and *pykF* mutations individually and in combination in each genetic background. Fitness was measured relative to the corresponding natural isolate strains. REL606 is the ancestor used to found the population in which the mutations were originally isolated. Other genotypes are natural isolates. Mean and s.e.m. are shown (natural isolates: n = 4; REL606: n = 25).

Epistasis depends on their genetic background

To test for epistasis between the *topA* and *pykF* mutations, we added them together in each genetic background. Epistasis was generally negative (Table 3.1). In six genetic backgrounds, the fitness of the double mutant was lower than that of the mutant having only the *pykF* mutation, although this difference was only significant in three cases (VS-126, TA105 and E267). The pair-wise relative magnitude of epistasis between the *topA* and *pykF* differed significantly between different backgrounds ($F_{7,45} = 20.041$, $P < 0.0001$). Of note, in TA105 the double mutant was less fit than either single mutant — an example of reciprocal sign epistasis (Weinreich 2005, Weinreich et al. 2006). Sign

epistasis has been shown to play an important role in constraining adaptation by causing adaptive landscapes to become rugged (Weinreich et al. 2005, Salverda et al. 2011) .

Table 3.1 Estimates of relative fitness and epistatic deviations of different genotypes.

Progenitor strain	**Mutations added**	**Fitness relative to progenitor (95% CI)**	**Absolute epistasis(95% CI)**	**Maximum growth rate of progenitor (H^{-1}) (s.e.m)**
ECOR1	*topA*	1.010 (0.082)		
	pykF	1.120 (0.082)		
	topA, pykF	1.060 (0.074)	-0.076 (0.193)	0.654 (0.016)
VS-126	*topA*	1.160 (0.033)		
	pykF	1.284 (0.078)		
	topA, pykF	1.255 (0.109)	-0.233 (0.085)	0.663 (0.021)
VS-820	*topA*	0.988 (0.035)		
	pykF	1.029 (0.074)		
	topA, pykF	1.022 (0.028)	0.006 (0.035)	0.612 (0.065)
TA135	*topA*	1.171 (0.087)		
	pykF	1.279 (0.181)		
	topA, pykF	1.268 (0.058)	-0.226 (0.111)	0.700 (0.069)
R424	*topA*	0.998 (0.080)		
	pykF	1.173 (0.114)		
	topA, pykF	1.089 (0.048)	-0.081 (0.131)	0.587 (0.005)
E267	*topA*	1.383 (0.042)		
	pykF	1.054 (0.098)		
	topA, pykF	1.428 (0.043)	-0.031 (0.116)	0.618 (0.031)
TA105	*topA*	1.130 (0.065)		
	pykF	1.157 (0.244)		
	topA, pykF	1.087 (0.112)	-0.225 (0.219)	0.726 (0.037)
REL606	*topA*	1.142 (0.023)		
	pykF	1.000 (0.013)		
	topA, pykF	1.193 (0.023)	0.051 (0.051)	0.591 (0.014)

topA and *pykF* mutations interact more negatively in strains with faster maximum growth rates

Recent studies have found that epistatic interactions between beneficial mutations tend to become increasingly negative as the fitness of the genetic background increases (Chou et al. 2011, Khan et al. 2011). This trend can explain the general tendency for the rate of adaptation to decline in experimentally evolving populations that are selected in constant environments (Chou et al. 2011, Khan et al. 2011). However, those studies considered genetic backgrounds that differed by only a few mutations, and it is not clear if the same pattern will be seen between mutations in divergent backgrounds. To test this, we estimated the correlation between each strain's maximum growth rate and its deviation in fitness owing to the interaction between *topA* and *pykF*. We found that these interactions tended to be more negative in strains with higher growth rates (Pearson: $r = -0.810$, $p = 0.015$; Spearman: $\rho = -0.666$, $p = 0.083$; Fig. 3.2). Notwithstanding our limited sample size, this result is consistent with the growth rate of a strain determining some portion of the overall direction and magnitude of epistasis. We note, however, that we do not rule out that specific interactions between the genetic background and the focal mutations could also play a role in determining epistatic effects.

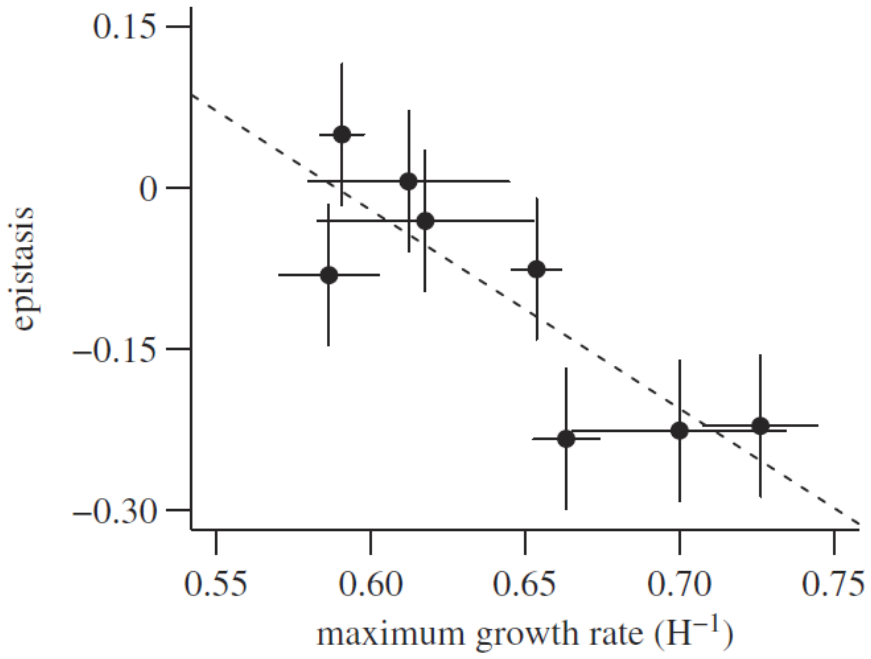

Figure 3.2 Relationship between maximum growth rate and the magnitude of epistasis. Points indicate the mean of estimates of maximum growth rate (n = 9) and magnitude of epistasis between *topA* and *pykF* mutations for the ancestor (n = 25) and seven natural isolate strains (n = 4). Error bars indicate 95% CI.

Our results demonstrate that both the direct fitness effect of two mutations that were beneficial in a laboratory-evolved population, and the epistatic interaction between these mutations, depend on genetic background. In addition, we found that the effect of epistasis between the *topA* and *pykF* mutations was negatively correlated with the growth rate of the strain containing the mutations, which is consistent with previous studies (Chou et al. 2011, Khan et al. 2011, Rokyta et al. 2011). This result suggests that our focal mutations may affect a common saturating physiological process, which is related to growth rate. Targeting this process in already fast-growing genotypes would lead to little additional growth rate improvement and, therefore, negative epistasis.

Appendix 3

Table A3.1 Detailed information about the natural isolate of *E. coli*.

Isolates of *E. coli*	**Source of strain**	**Species or environmental sites**
VS-126	Bird	*Atlapetes brunneinucha*
TA135	Mammal	*Petrogale lateralis*
VS-820	Mammal	*Perognathus penicillatus*
ECOR1	Human	*Homo sapiens* North America
TA105	Mammal	*Mus musculus*
E267	Environmental	Lake Burley Griffin, ACT, Australia
R424	Reptile	*Lampropholis quichenoti*

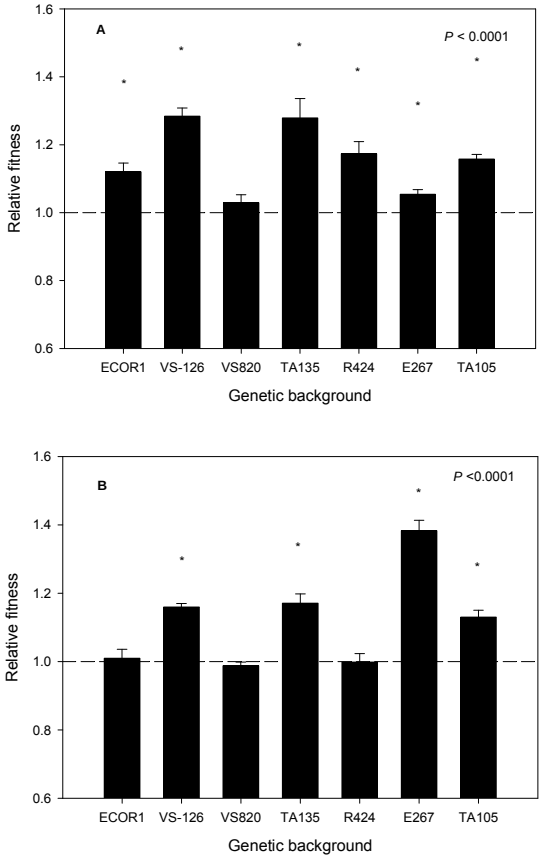

Figure A3.1 Relative fitness of different strains carrying beneficial mutation. Error bars represent standard error. * indicates significant difference from 1.0 (significance level of 0.05). (A) is the fitness effect of beneficial mutation in *pykF* and (B) is the fitness effect of beneficial mutation in *topA*.

Acknowledgements

This work was supported by NSF DEB-0844355.Thanks to Carolina Díaz, Daniel Stoebel for constructing strains, to V. Cooper and F. Moore for helpful discussion, and to Rolf Lohaus for development of the growth rate analysis script. I also thank my adviser, Dr. Tim Cooper, for revising my manuscript.

Chapter 4: Influence of adaptation in driving reproductive isolation

Summary

Speciation is one of the most important topics in evolutionary biology. Although much attention has been paid to the role of adaptation in driving reproductive isolation, the understanding of the genetic mechanisms by which reproductive isolation arises remains limited. In this study, I examined the role of adaptation in driving reproductive isolation at the initial stage of speciation, by studying the extent and underlying patterns of genetic interactions arising in crosses between pairs of 12 *Escherichia coli* populations evolved in two different environments. I found that, regardless of whether parents were selected in the same or different environments, the average fitness of recombinants was significantly lower than expected, consistent with there being a prevailing influence of incompatibility between the independently accumulated mutations. Such incompatibility suggests that both parallel and divergent adaptation can lead to differentiation, and eventually reproductive isolation, between independently evolving populations. Considering recombinants of parents evolved in different environments, I found that incompatibilities can be asymmetric, with fitness being lower than expected in the selective environment of one parent, but higher than expected in the selective environment of the other parent. This result provides a clear demonstration of the environment can play a role in producing reproductive isolation. Overall, my results shed

light on how parallel and divergent adaptation can lead to intrinsic genetic barriers at the initial stages of speciation.

Introduction

Speciation, defined as a single population dividing into two or more populations that can no longer interbreed with each other (Myar 1942), is a central topic in evolutionary biology. This process involves evolution of intrinsic or extrinsic barriers of genetic exchange between different populations (i.e. reproductive isolation). Reproductive isolation is the most commonly used criterion to define species and plays a critical role in the maintenance of different species. Many theoretical (Kirkpatrick and Ravigne 2002, Guerrero et al. 2012) and experimental (Schluter and Conte 2009, Moyle et al. 2012, Powell et al. 2014) studies have shown that natural selection can lead to the origination of new species. However, exactly how this occurs is still a matter of debate.

A general mechanism of reproductive isolation is postzygotic reproductive isolation caused by hybrid dysfunction. Adaptive evolution of geographically isolated populations is suggested as a major cause of postzygotic reproductive isolation, conferring hybrid inviability or sterility, based on patterns of molecular variation in genes contributing to hybrid dysfunction (Presgraves et al. 2003, Barbash et al. 2004). For example, Michalak et al. (2001) examined the genetic differentiation between two populations of *Drosophila* living on the north and south side of "Evolution canyon" in Israel. They analyzed microsatellites and *hsp70Ba* which is a gene-encoding inducible heat shock protein, markers in both populations. They found that there was a limited

genetic exchange between populations by analyzing microsatellites. In addition, they observed a great variation in Hsp70 between populations. The level of Hsp70 can be the target of thermometry selection, therefore they conclude that local adaption to the contrasting environments present on the two sides is responsible for the genetic and phylogenetic divergence between these two populations, providing strong evidence of adaptation-driven incipient speciation in nature. Dettman *et al.* (2007) studied experimental populations of yeast that were evolved in two different environments and found that divergent adaptation of populations evolved in different environments resulted in hybrids with lower fitness in either environment than the parental populations. Although much attention has been paid to the role of adaptation in reproductive isolation (Dettman et al. 2007, Dettman et al. 2008, Schluter 2009), our understanding of the origin of reproductive isolation is limited (Schluter 2001).

There are two main views on how new species can arise through natural selection. One is ecological speciation, defined as the development of reproductive isolation between populations as the consequence of differentiation on traits driven by natural selection under different environments. This mechanism is widely considered as the primary mechanism for the origin of species (Rundle et al. 2005, Schluter and Conte 2009, Tarroso et al. 2014), as ecological differences usually evolve at the early stages of speciation (Weissing et al. 2011, Arnegard et al. 2014). In this process, natural selection drives the divergent adaptation of populations in contrasting environmental conditions to different adaptive peaks by fixing different sets of mutations (Presgraves 2010). If the benefit of the adaptive mutations is specific to one or the other environment, hybrids

must be less well adapted to either environment than the parent selected in that environment.

A second speciation process, known as mutation-order speciation, involves evolution of reproductive isolation by accumulation of different mutations in different populations (Schluter and Conte 2009, Butlin et al. 2014). In this process, it is recognized that even populations adapted in similar selective environments may accumulate different mutations. In fact, even if populations ultimately arrive at the same adaptive peak, they are likely to transiently diverge from one another because the order in which mutations accumulate can differ. Mutations fixed in different populations might interact antagonistically with each other when brought together in hybrids of two populations, a process that underlies the Dobzhansky-Muller model of reproductive isolation (Dobzhansky, 1937, Muller, 1942). This model has been suggested to provide an explanation of how populations become reproductively isolated under adaptive evolution (Greig 2009) and been widely accepted as the most common route to the evolution of hybrid dysfunction (Barton 2001, Rundle 2002, Brideau et al. 2006, Tang and Presgraves 2009). Genetic incompatibilities represent an intrinsic barrier to genetic exchange between isolated populations that is likely to be irreversible once it is completed, which means that the separated populations will not be able to interbreed, even when the external barrier is removed (Weissing et al. 2011).

Experimental tests of the Dobzhansky-Muller model lag far behind the theory. As speciation is a long-term historical phenomenon it is extremely hard to examine how

adaptation leads to the process of speciation through direct observation in nature. Therefore, most relevant studies have tested for conflicts among interacting genes in closely related, but already established species (Axenovich et al. 1998, Fishman and Willis 2001, Brideau et al. 2006, Masly et al. 2006, Bikard et al. 2009, Cattani and Presgraves 2009, Mihola et al. 2009). For example, Brideau et al. (2006) found that one pair of genes - *Lhr* and *Hmr* - interacted negatively, and caused lethality in hybrids of two *Drosophila* species. However, studying incompatibility between well-established species does not lend itself to an understanding of how adaptation affects the initial development of reproductive isolation at the early stage of speciation, which is the most important point in the process of speciation. Thus it is hard to assess whether intrinsic genetic incompatibilities played a causal role in erecting a reproductive barrier, or arise as a by-product after speciation is complete. It is also difficult to estimate whether any incompatible genetic elements are adaptive, neutral, or even deleterious in the population in which they arise at the earliest stages of differentiation between populations.

Laboratory-evolution experiments offer an opportunity to examine the early events in the evolution of incompatibilities between independently evolving populations. The evolutionary history of lineages can be controlled, which allows us to directly examine evolution of genetic barriers between populations evolved from a common ancestor, as well as assess the role of adaptation in fixation of the very first alleles that contribute to these genetic barriers. Even so, the Dobzhansky-Muller incompatibility model has been rarely tested. I am aware of only a couple of studies performed using *Saccharomyces cerevisiae* (Dettman et al. 2007) and *Neurospora* (Dettman et al. 2008).

Both studies found that populations adapted to different environments had greater reproductive isolation than did populations independently adapted to the same environment. Both studies also observed antagonistic epistasis between alleles selected in different environments, suggesting an additional role of ecological speciation. Recently, the identity and function of a pair of genes involved in an incompatibility were examined. Anderson *et al*. (2010) reported that alleles of PMA1 and MKT1 that arose in two experimentally evolved yeast populations were incompatible with one another. However, we still lack a full understanding of the role of adaptation in reproductive isolation at the initial stage of speciation. Important unanswered questions include: what is the relative importance of parallel and divergent selection in driving reproductive isolation? Is reproductive isolation typically polygenic, or perhaps simply involving only one pair of incompatible genes?

In this study, we investigated the beginning of reproductive isolation by studying the extent and underlying patterns of genetic interactions in recombinant strains made by crossing strains isolated from experimental populations of *Escherichia coli* founded from a common ancestor and evolved in two different carbon-resource environments. We found that, in general, the average fitness of recombinants from two evolved populations was significantly lower than expected, consistent with their being a tendency toward negative epistatic interactions between independently accumulated mutations. This result held between populations evolved in the same environment and between populations evolved in different environments, suggesting that both parallel and divergent adaptation can lead to population differentiation and eventually reproductive isolation.

Recombinants of some parent combinations differed from this trend. We detected positive interactions in recombinants from two sets of parents evolved in different environments, suggesting the existence of positive interactions between independently evolved mutations. Finally, the effect of interactions was often asymmetric with the fitness of recombinants between parents evolved in different environments often being lower than expected in one parental environment, but higher than expected in the other parental environment. Overall, my results provide insight on the influence of both parallel and divergent adaptation in reproductive isolation at the initial stages of speciation.

Materials and Methods

Bacterial strains, growth conditions, and plasmids

Twelve replicate populations of *E. coli* were started from REL606 or REL607 ancestral strains and independently evolved in minimal medium supplemented with lactose (210mg/ml) or glucose (175mg/ml). REL606 and REL607 are isogenic except that the latter strain is able to grow on the sugar arabinose. Six populations were evolved in a minimal glucose environment (populations glu1-6) and six populations (populations lac1-6) in a minimal lactose environment. The metabolic pathways involved in utilization of these sugars do have substantial overlap, but previous studies have shown that beneficial mutations that arise during adaptation to one sugar are not necessarily beneficial, and can be costly during growth in the other environment (Quan et al. 2012). All 12 populations were propagated for 4,500 generations at 37 °C. One strain was isolated from each population and used in this study. The whole genome sequence of

strains isolated from populations glu2, glu5, lac5, and lac6 was obtained and used to identify evolved mutations.

In order to select the successful recombinants from conjugation, complementary pairs of auxtrophic or catabolic mutations — in the genes *trpA*, *cysE*, *lysA*, and *araA* (*araA* is required for growth on minimal arabinose medium) — were introduced into evolved strains as described below. Conjugation between two parents with complementary auxotropic/catabolic markers allowed selection of recombinants as only prototrophs are able to grow on arabinose. An F' plasmid was obtained from the Coli Genetic Stock Center (Yale University, New Haven, Connecticut, United States) (CGSC# 4824). This F plasmid, F8-3 (F'gal(*ts*)), has a temperature-sensitive origin of replication, thus is lost from strains following overnight incubation at 42°C. Another plasmid pBSL182, encoding a mini-Tn*10* encoded gentamicin resistance (Gm^r) gene, was used to construct a derived plasmid, F'gal (*ts*, Gm^r), which was used to mediate recombination between different evolved strains. The Gm^r gene facilitates screening for the loss of the F plasmid from recombinant strains.

Construction of strains with different selectable mutation markers -*araA*, *trpA*, *lysA*, and *cysE*.

The mutation markers -*araA, trpA, lysA*, and *cysE* were used to select recombinants from two parental strains. I divided markers into two pairs: *araA*- (position ~1min on the *E. coli* chromosome) and *lysA*- (~64 min), and *trp*A- (~24 min) and *cysE*- (~81 min). The separation of each mutation pair by approximately half the chromosome

means that prototrophic recombinants can only be produced if at least half a genome of DNA sequence is transferred from donor to recipient strain (Fig. 4.1). In fact, it is likely that a greater amount of DNA is transferred because the start site of transfer is determined essentially randomly by integration of the recombination mediating plasmid.

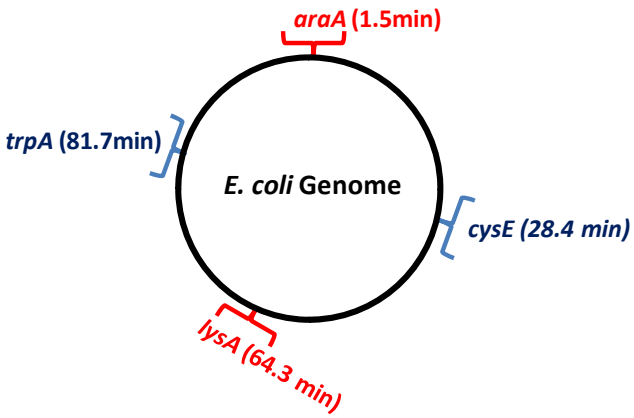

Figure 4.1 Schematic representation of the positions of the four auxotrophic or catabolic marker mutations created in the *E. coli* genome.

Populations started from ancestor REL607 encode a functioning *araA* gene and can grow on arabinose (Ara+). To replace *araA*+ with *araA*-, so that strains isolated from these populations could be used in the recombinant selection scheme outlined above, an 800 bp fragment of *araA* amplified from the Ara- strain, REL606, was amplified and cloned into the intermediate vector pCR2.1 using a TA cloning kit (Invitrogen). This fragment was cut from pCR2.1 and ligated into a suicide plasmid, pDS132, which is temperature and sucrose sensitive. To do this, pDS132 was digested with XbaI and SacI and purified using a PCR purification kit (Qiagen) to create sticky ends that match the insert fragment cut from pCR2.1 using the same enzymes. The insert fragment was

cloned into pDS132 by ligation using T4 ligase with overnight incubation at 14°C. The ligation reaction was purified and transformed into electro-competent cells (SM10) by electroporation. Cells were plated on lysogeny broth (LB) plates supplemented with 20 µg/ml of chloramphenicol (LB+Cm) and incubated overnight at 37°C. SM10 is sensitive to Cm, whereas pDS132 confers Cm resistance. Therefore only SM10 cells successfully transformed with pDS132 can grow on LB+Cm plates. Several colonies from LB+ Cm plates were restreaked to the same plates and incubated overnight at 37°C. To confirm that pDS132 had correct insert fragment, plasmid from cultured cells was purified, digested by XbaI and SacI, and run in an agarose gel, allowing a check that the insert fragment was the expected size. SM10 competent cells containing plasmid pDS132::*araA*- were identified and saved for use as a donor strain to replace *araA*+ with *araA*- through conjugation. Conjugations were carried out by mixing 10 µl of donor and 10 µl of recipient cells as a spot on LB plates, incubating for 2-4 hours, and then transferred to LB+Cm+Sm plates for up to 36 hours incubation at 37°C. Recipient cells are sensitive to Cm, whereas donor cells are sensitive to Sm. Thus, only recipient cells which received pDS132::*araA*- can grow on LB+Cm+Sm plates. Also, pDS132 is only stable when it is integrated into the chromosome of recipient cells. Therefore, only recipient cells that have received pDS132::*araA*- and integrated it into their chromosome are able to grow on these plates. Several colonies grown on LB+Cm+Sm plates were restreaked to the same plates to remove false positive colonies. A single colony from a restreaked LB+Cm+Sm plate was resuspended in DM0 and plated on LB+Sucrose plates (without NaCl) to select for excision of suicide plasmid pDS132, which encodes *sacB,* a

gene that confers sensitivity to sucrose. Colonies that grew on these plates were re-streaked to LB plates. The genotype of *araA*- was confirmed by streaking on TA (tetrazolium arabinose) indicator plates, on which *araA*- colonies appear red and *araA*+ colonies appear white or pink. In all constructed strains, the introduced Ara- marker was shown to be neutral (Fig. A4.1).

The populations evolved from REL606 are arabinose negative (*araA-).* To construct derivatives having only the *trpA-/cysE*- mutation markers, I first needed to replace the *araA*- allele with *araA*+. To do this, populations evolved from REL606 carrying *araA*- were taken out from -80°C freezer and inoculated in 3 ml LB broth. After overnight incubation, 1ml LB culture was centrifuged at 6000rpm for 5 minutes. The supernatant was discarded and the cell pellet was resuspended in 200 μl DM0 (Davis minimal medium without glucose), plated on a MA (Minimal arabinose) plate, and incubated at 37°C overnight. The colonies which are able to grow on MA plates are *araA*+, as the progenitor is not capable of growing on minimal media in which arabinose is the sole carbon resource. *araA*+ colonies were restreaked on MA plates again to get an isolated single colony. In order to confirm the genotype of colonies growing on MA plates- I then streaked them to TA indicator plates. Red Ara- colonies were grown in LB liquid media and saved at -80 ℃.

Each of three auxtrophic markers were separately introduced into focal evolved strains using red recombineering method (Datsenko 2000). *lysA*- was transferred into an *araA*- strain and the other two markers, *cysE*- and *trpA-,* were transferred into an *araA*+

strain. To transfer a mutation marker into an evolved strain, the *cat* (chloramphenicol resistance, Cm^r) gene was amplified from the pKD3 plasmid with primers pairs that contain tags homologous to the three target genes. Purified PCR product was electroporated into strains carrying the λ red expressing helper plasmid pKD46 (ampicillin resistant). Transformed cells were spread on LB agar plates supplemented with chloramphenicol and incubated at 37°C overnight. Transformants were restreaked onto LB and incubated at 43°C to facilitate loss of pKD46. Colonies forming after overnight incubation were tooth picked to LB, LB +Ap, and minimal glucose plates to test for loss of the helper plasmid and auxotrophy, which indicates successful incorporation of the auxotrophic marker. The helper plasmid pCP20, which expresses the FLP recombinase, was used to make derivative Cm^s strains by removing the Cm^r gene. FLP acts on target FRT sites that flank the resistance gene resulting in its excision from the genome. All gene deletion events were confirmed by PCR using primers outside of the target ORF and primers (Table A4.1) inside the deletion sequences to make sure that the deletion was made at the right position in the chromosome.

Once the evolved strains carrying each of the three markers were constructed, P_1*vir* transduction was used to transfer the auxtrophic mutation markers into other target strains. The *lysA*- and *trpA*- markers were introduced into *araA*- and *araA*+ strains, respectively. The *cysE*- marker was added to the same strains that received the *trpA*- marker. The *araA-/lysA*- and *trpA-/cysE*- mutation marker pairs were introduced into six and 12 evolved strains, respectively (Table 4.1).

Table 4.1 Constructed strains containing pairs of mutation markers. F' indicates the additional presence of a temperature-sensitive F plasmid to mediate conjugation between parental strains.

Mutation markers	Populations in which mutation markers were introduced
lysA-/araA-	glu3(F') glu5(F') glu6(F') lac3(F') lac5(F') lac6(F')
cysE-/trpA-	All 12 populations: glu1-6 and lac1-6

Conjugation experiment

A temperature-sensitive plasmid, F'gal (*ts*, Gm^r), was transferred into six strains carrying the *lysA-/araA-* marker pair (table 4.1). These F'-containing strains were used as donors in conjugation with recipient evolved strains carrying the alternative *cysE-/trpA-* marker pair, and recombinants were identified as prototrophic cells able to grow on arabinose as sole carbon source. Briefly, strains having F' (*ts*,Gm^r) were grown overnight in LB+Gm at 42 °C. The plasmid cannot independently replicate at temperatures above 32 °C, so selection for the plasmid encoded Gm^r at this high temperature selects those cells that chromosomally integrated the plasmid, thereby creating a so called 'high frequency of recombination' (Hfr) donor strain. Recipient strains carrying the complementary mutation pair were concurrently incubated in LB media overnight. Because of the low efficiency of the conjugation, concentrated overnight donor and recipient cultures were used in the conjugations (500 µl was concentrated to 20µl). 20µl of recipient cells and 20 µl of donor cells were mixed as a spot on LB plate and incubated at 37°C for 2-4 hours. The mixed cells were resuspended in 200 µl of minimal media, plated on minimal arabinose (MA) plates and incubated at 37°C for 36-48 hours. Only recombinants that inherit a marker combination that omits both mutation pairs present in

the parent strains can grow on this medium. Fifteen to 20 colonies appearing on MA plates were re-streaked to the same medium and grown at 37°C overnight to remove any false-positive colonies.

I carried out three different types of conjugation schemes (Fig. 4.2): 1) evolved strains carrying one pair of mutation markers crossed with the same strain carrying the complementary pair of mutation markers (Type I), 2) two strains independently evolved in the same environment crossed with each other (Type II), and 3) two strains evolved in different environments crossed with each other (Type III). These different combinations allowed me to test different aspects of the process of genetic incompatibility. Type I recombinants represent a control to test whether the process of strain construction and conjugation has some unanticipated effect on fitness. Type II and Type III recombinants allowed me to test and compare reproductive isolation originating during parallel and divergent selection, respectively.

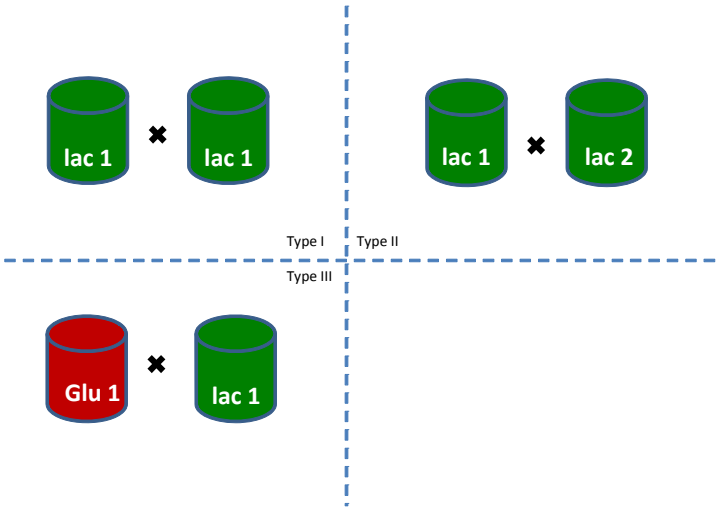

Figure 4.2 A schematic figure outlining the four conjugation schemes I carried out. Strains from three evolved populations - lac1, lac2, and glu1, and the ancestor, are taken as examples in the figure. The different types of conjugations were used to test for the fitness cost of conjugation between the same strains (Type I), and reproductive isolation developed from parallel (Type II) and divergent adaptation (Type III).

Fitness assays

The fitness of all evolved populations, strains, and recombinants was measured by direct head-to-head competition assays (Lenski et al. 1991). Due to the large number of strains whose fitness was estimated, a flow cytometer was used to distinguish competitors (Fig.4.3). A flow cytometer can identify cells with different fluorescent intensity and counts them separately. In my competition assays, relative fitness of each competitor was determined with respect to a derivative of the ancestral strain that expressed the green fluorescent protein (GFP) (Zhang et al. 2012). This reference strain, REL606 *rhaA*::*gfp*, expresses a fast maturing GFP fluorescent protein from a strong promoter, P_{A1} (Gallet et al. 2012) allowing the two competitors, ancestral and evolved cells, to be distinguished (all the evolved strains used were non-GFP). Flow cytometry was performed with an

Accuri C6 Flow Cytometer (BD Biosciences). Each population included a total of 5,000 cell events captured at a rate of 1,000-2,000 events/s.

Competitions were started with two competitor strains inoculated from -80°C stocks into separate tubes containing 3 ml of LB broth. After overnight incubation, populations were transferred to minimal glucose/lactose media for two days by daily transfer with 10^{-2} dilution factor. This step allowed both competitors to become physiologically pre-conditioned to the competition environment. Following preconditioning, competing strains (100 µl of each) were mixed and allowed to compete with each other for one day by incubating at 37 °C overnight. The number of cells of each competitor on the initial day and final day were counted in the flow cytometer by diluting mixture 1:200 into a mix of HPLC water, Davis minimal medium and a red fluorescent nucleic acid stain, SYTO17 (used at a final concentration 200 nM), which was used to distinguish bacterial cells from background noise. The fitness (W) of strains relative to the ancestral strain with the GFP marker was calculated as the ratio of the competitors Malthusian parameters (MA):

$$W = \frac{\text{MA (evolved strain or recombinants)}}{\text{MA (Ancestral strain ::}gfp\text{)}}$$

Malthusian parameters (MA) were estimated using the initial and final density of each competitor as:

$$\text{MA} = \ln\left(\frac{\text{Counts (final day)}}{\text{Counts (initial day)}}\right)$$

The GFP marker was not neutral. Therefore, I normalized all relative fitness estimates to account for the effect of the marker.

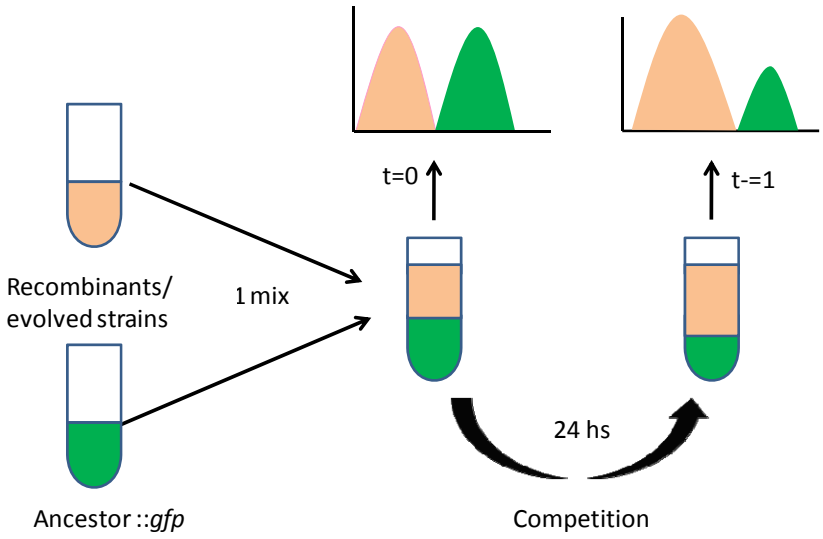

Figure 4.3 Schematic diagram of competition fitness assays. Both competitors were adapted physiologically to the competition environment before mixing. At the beginning of the competition, the competitors, GFP- strains (pink) and GFP+ strains (green), were mixed at 1:1 ratio and flow cytometry was used to count each type. After 24 hours competition, the mixture was again analyzed to count the number of each competitor. The pink and green peaks in the histogram indicate the number of events of GFP- and GFP+ strains counted in channel, respectively.

Mutations screened in recombinants

The whole genomes of strains from four evolved strains and eight recombinants were sequenced. I prepared genomic DNA and used illumina TruSeq DNA PCR-Free Library Preparation Kit to prepare DNA libraries for strains from two evolved populations, lac5 and lac6. The other 10 libraries were processed by one of my lab-mates. The DNA library was sent out for genome sequencing. A computational pipeline, Breseq, was used to identify mutations, including single-nucleotide mutations, point insertions

and deletions, and large deletions, relative to the sequence of the ancestral strain REL606. I also determined the combination of parental mutations present in each of 22 recombinants derived from a cross of lac5 and lac6 parents by using a combination of PCR, RFLP, and Sanger sequencing.

Data analysis

I used a multiplicative model to estimate the deviation from expected fitness of recombinant strains (Fig.4.4). The two parent strains are expected to make approximately equal contributions to the genotype, and therefore fitness, of each recombinant. We derive an expected recombinant fitness using the formula: $\varepsilon_{1,2} = W_{1,2} - (W_1^{1/2} \times W_2^{1/2})$, where W_1 and W_2 are the fitness of evolved populations 1 and 2, respectively. The terms $W_{1,2}$ and $W_1^{1/2} \times W_2^{1/2}$ (the geometric mean of the fitness of the two parents) represent the observed relative fitness and expected relative fitness of recombinants from evolved population 1 crossed with evolved population 2, respectively. Positive and negative values of $\varepsilon_{1,2}$ indicate a net positive or negative fitness of recombinant strains relative to their expected fitness, respectively. A schematic illustration of this approach is presented in Figure 4.4. Deviations of individual recombinant strains from their expected fitness can be due either to interactions between the variable alleles inherited from their parent strains or because they inherited more or fewer parental mutations than expected by chance. We cannot distinguish between these possibilities by comparing the fitness if a single recombinant strains to its parents. By considering the deviation from expected fitness averaged over a large number of recombinants, however, we reduce the influence

of random fluctuation in the number of mutations present in individual recombinants, and thereby increase the signal due to allele interactions.

One-way ANOVA was used to test for variation in fitness among recombinants obtained from independent conjugation of the same parent strains. Two-way ANOVA was used to test for the effect of assay environment, the selective environment and the interaction between them, to the deviation from expected fitness. *t*-tests were used to test if the fitness of evolved populations was different from 1 and if all the recombinants are different from expected fitness. All analyses were carried out in R v.3.1.3 (http://cran.r-project.org/).

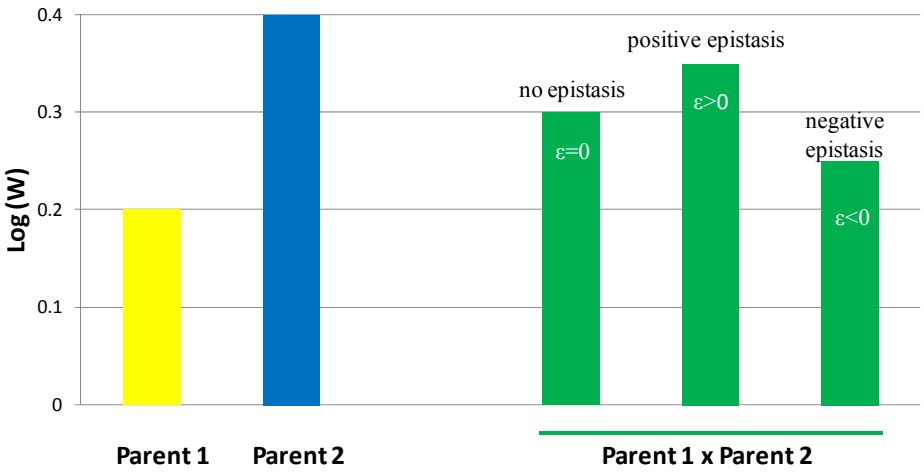

Figure 4.4 Schematic illustration of different types of deviations from expected fitness. W indicates the relative fitness of strains. Deviations are measured as $\varepsilon = W_{1,2} - (W_1^{1/2} \times W_2^{1/2})$, where W_1, W_2 and $W_{1,2}$ represent the relative fitness of Parent 1, Parent 2 and recombinants of Parent 1 × Parent 2, respectively.

Results

Response to selection

I measured the fitness of six glucose-evolved and the six lactose-evolved populations in both environments. I found that all 12 evolved populations displayed a significant increase in fitness relative to their ancestor in their selective environment (Fig. 4.5), indicating adaptation to that environment. On average, the glu and lac populations had a fitness advantage, relative to their ancestor, of 28% and 41.2%, respectively. I also found that all populations had increased in fitness in their non-selective environment, except for the glu2 population. However, the magnitude of fitness increase was significantly smaller in non-selective, than in selective, environments (i.e. all populations tended to increase their fitness most in their own selective environment), which indicates adaptation was, at least in part, environmentally specific.

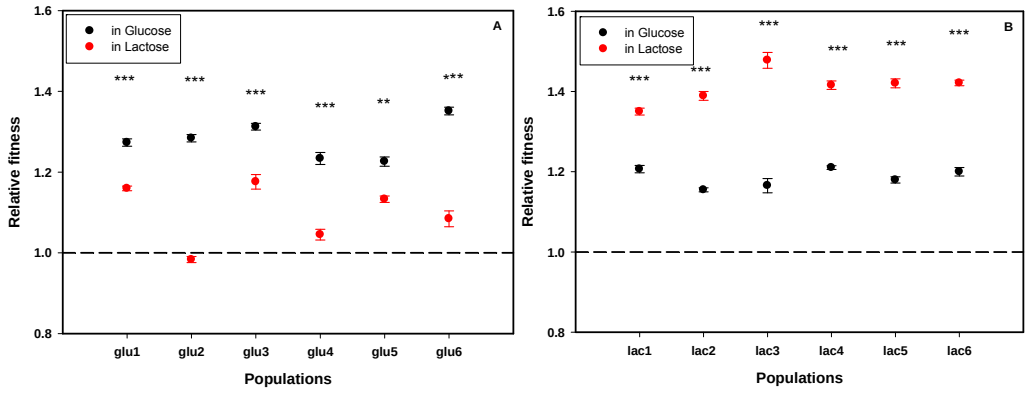

Figure 4.5 Fitness of 12 evolved populations, glu1-glu6, which were evolved in glucose (A), and lac1-lac6, which were evolved in lactose (B), relative to the common ancestor. Fitness estimates were made in both the minimal glucose and minimal lactose environments. Error bars indicate 95% confidence intervals (n=4). *** $p<0.001$, ** $p<0.01$ according to *t*-test.

Fitness cost of conjugation

To test if the process of plasmid-mediated recombination had any intrinsic fitness cost, I conjugated F' donors with derivatives of themselves carrying the alternative *trpA-/cysE-* marker pair (Fig. 4.2, type I). Prototrophic recombinants should be genetically identical to one another and to the original progenitor strain from which auxotrophic parents were made. In fact, I found that there was some variation between independent recombinants obtained from the same pair of strains (Fig. A4.2). For example, four out of 20 recombinants of glu5 × glu5 had significantly lower fitness than expected based on t-test. On average, however, there was no overall deviation from neutrality between the recombinants from all six pairs of conjugations (Fig. 4.6; significance assessed by 2-tailed t-test at $p < 0.05$). The average fitness of glu5xglu5 recombinants measured in glucose was significantly lower than their progenitor (Fig. 4.6), however, the effect was very small (mean fitness = 0.992 ± 0.005). Overall, I conclude that the process of conjugation does not introduce any meaningful bias in the fitness of recombinant strains relative to their parents when using our scheme of transfer and selection.

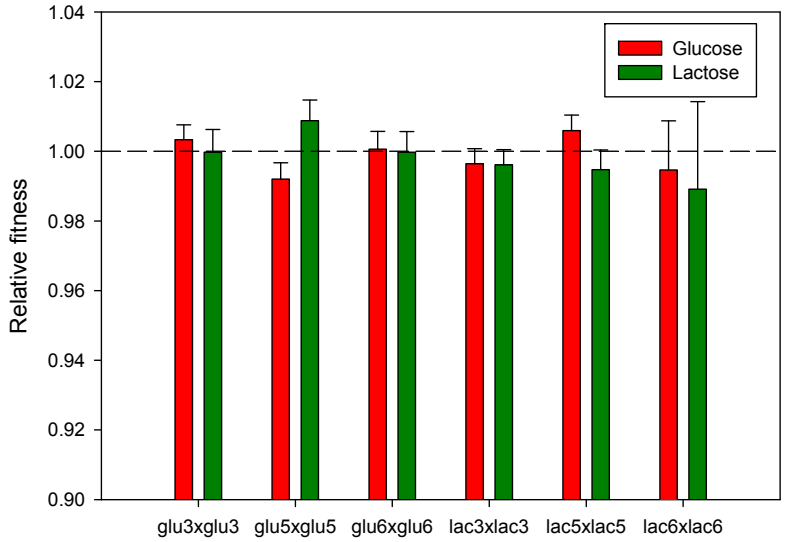

Figure 4.6 The fitness of recombinants generated from indicated parental strains measured in minimal glucose (red) and minimal lactose media (green). Fitness was measured relative to their progenitor strain. Error bars indicate the 95% confidence interval (recombinants assayed: lac3:n=20, lac5:n=20,lac6:n=2, glu3:n=20, glu5:n=20, glu6:n=19).

Fitness of recombinants from different evolved populations

To examine the form of genetic interactions between mutations present in strains isolated from different evolved populations I conducted eight crosses; four with strains evolved in the same environment and four with strains evolved in different environments. Considering first crosses from parents that evolved independently in the same selective environment, I found high variation in fitness between recombinants obtained, and one-way ANOVA showed that independent recombinants from the same cross had significantly different fitness (Fig. 4.7, Table 4.2). This variation indicates that, within a cross, independent recombinants have different sets of parental mutations, providing the potential for different interactions to occur. The mean fitness of recombinants from different parent strains is quite different. For example, most of the recombinants (18 out

of 24) from the glu6 × glu4 cross have lower fitness than expected when assayed in minimal glucose medium ($p < 0.01$ for each recombinant); however, the fitness of most recombinants from the lac5 × lac6 cross are not significantly different from expected. In all crosses, there was a minority of recombinants that had extremely low fitness.

Table 4.2 Variation in fitness between recombinants produced from the same parents.

Parent selection environment	**Parent**	**Assay environment**	***df***	**F**	***p*-value**
Same	lac3 × lac2	Glucose	17,89	44.47	<0.0001
		Lactose	17,90	11.08	<0.0001
	lac5 × lac6	Glucose	23,120	52.06	<0.0001
		Lactose	23,120	18.73	<0.0001
	glu5 × glu2	Glucose	23,119	22.74	<0.0001
		Lactose	23,119	5.86	<0.0001
	glu6 × glu4	Glucose	21,95	75.64	<0.0001
		Lactose	21,98	8.05	<0.0001
Different	lac5 × glu5	Glucose	31,129	38.36	<0.0001
		Lactose	31,131	78.78	<0.0001
	glu3 × lac1	Glucose	25,123	6.05	<0.0001
		Lactose	25,130	11.28	<0.0001
	lac3 × glu2	Glucose	21,110	18.20	<0.0001
		Lactose	21,110	106.02	<0.0001
	glu6 × glu4	Glucose	29,150	16.91	<0.0001
		Lactose	29,150	17.88	<0.0001

Table 4.3 Number of recombinants and mean deviation from expected fitness.

Parent selection environment	Parents	Number of recombinants	Mean deviation (± SEM) from expected fitness in assay environment:	
			Glucose	**Lactose**
Same	lac3 × lac2	18	**-0.037 ±0.023***	**-0.194 ±0.044**
	lac5 × lac6	22	**-0.030 ±0.014**	**-0.050 ± 0.022**
	glu2 × glu5	24	**-0.033 ± 0.014**	0.001 ± 0.019
	glu4 × glu6	24	0.001 ±0.011	**-0.063 ±0.011**
Different	lac1 × glu3	26	**-0.082 ±0.007**	**0.057 ±0.012**
	lac3 × glu2	22	**-0.046 ±0.019**	**-0.140 ±0.048**
	glu6 × lac4	30	**-0.027 ±0.010**	**0.220 ±0.017**
	lac5 × glu5	32	**-0.070 ±0.020**	**-0.040 ±0.022**

*Deviations that are significantly different from zero (based on *t*-test) indicated in bold.

To assess the overall effect of interactions between mutations recombined from the different sets of parents, I averaged the fitness of recombinants from each set of parents (Table 4.3). This averaging will reduce the influence of fitness deviations due to individual recombinants by chance receiving a smaller or greater than expected number of mutations from the parents, increasing the influence of interactions between mutations as the cause of fitness deviations. Three sets of recombinants had significantly lower fitness than expected when measured in the environment the parents were selected (Fig. 4.7), consistent with the presence of pervasive negative interactions between mutations accumulated in different populations. I did two-way ANOVA analysis to test if the interaction between mutations has general or environmental specific effect. The result showed that the assay environment did not have a significant effect on the mean deviation from expected fitness of recombinants, suggesting the interaction has general effect (Table 4.4).

Table 4.4 Two-way ANOVA examining basis of deviation from expected fitness of recombinants.

The environments and source of variation	***df***	**Sum of Squares**	**F**	***p***
Same environment where the two parents evolve from				
Cross	3	0.212	8.816	<0.0001
Assay environment	1	0.015	1.794	0.1822
Cross × Assay environment	3	0.272	10.940	<0.0001
Different environment where the two parents evolved from				
Cross	3	0.966	23.115	<0.0001
Assay environment	1	0.231	16.605	<0.0001
Cross × Assay environment	3	0.998	23.864	<0.0001

To examine the influence of divergent adaptation to reproductive isolation, I crossed four pairs of clones isolated from populations evolved in different environments and collected recombinant offspring (Fig. 4.7). Considering crosses of parents selected in different environments, I again found significant variation in the deviation from expected fitness between different recombinants collected from the same parental cross in both assay environments (Fig. 4.7-B, Table 4.3). In glucose, the mean deviation from expected fitness was significantly lower than expected in all crosses (Fig. 4.8). For two out of four crosses, the average deviation from expected fitness was negative when assayed in lactose environments (Fig. 4.8). Two-way ANOVA showed that there was a significant effect of assay environment on the deviation from expected fitness, indicating that the fitness reduction of recombinants from the parents depended on the assay environment (Table 4.4).

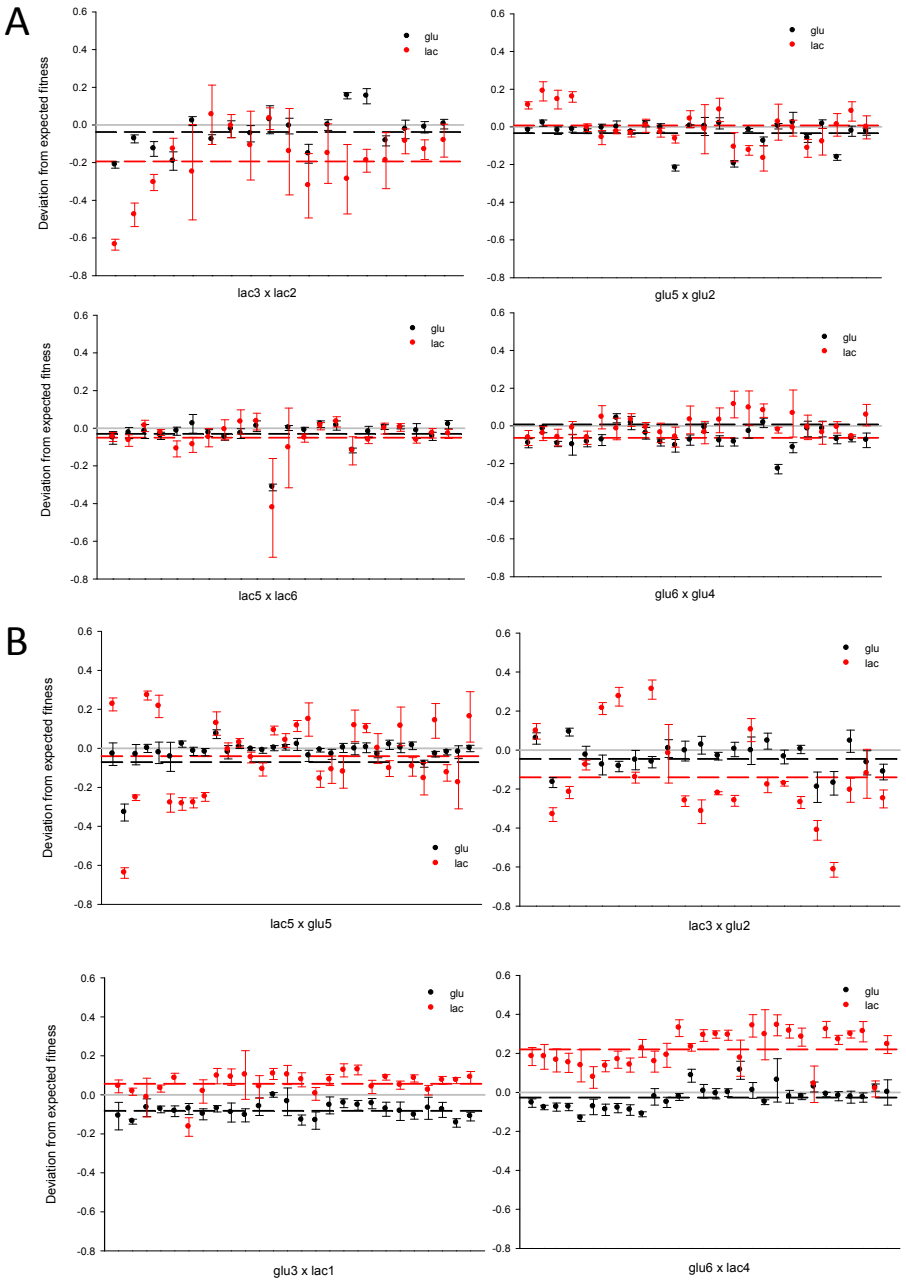

Figure 4.7 The fitness of recombinants generated from clones isolated from indicated populations evolved in the same (A) and different (B) environments. Fitness estimates were made in glucose (black) and lactose (red) environments. Solid gray, dashed black and dashed red lines represent the expected and grand mean of deviations in fitness in glucose and lactose, respectively. Error bars indicate 95% confidence intervals (recombinant fitness was estimated at least 4 replicates).

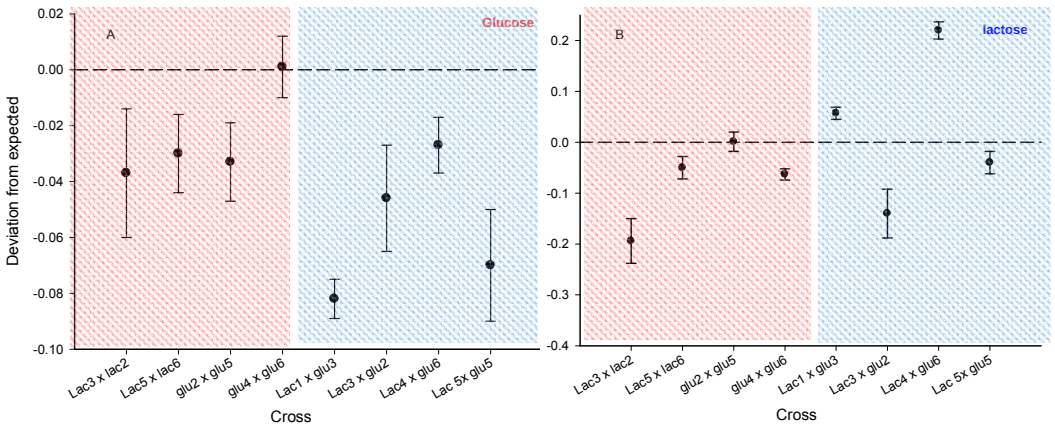

Figure 4.8 The grand mean of deviation from expected fitness of recombinants generated in eight crosses and measured in glucose (A) and lactose (B). The left and right halves of each panel indicate crosses from parents selected in the same and different environments, respectively. Error bars indicate 95% confidence intervals (recombinants in each cross: lac3× lac2, n=18; lac5 × lac6, n=22; glu2 × glu5, n=24; glu4 × glu6, n=24; lac1× glu3, n=26; lac3 × glu2, n=22; lac4 × glu6, n=30; lac5 × glu5, n=32).

To test for the effects of parallel and divergent adaptation on reproductive isolation, I compared the deviation from expected fitness of recombinants from parents selected in the same environment versus those selected in different environment. In the glucose assay environment I found that recombinants from parents evolved in different environments had larger deviations from expected fitness than did recombinants from parents had evolved in the same environment (p = 0.039). In the lactose assay environment, however, there was no significant difference between the fitness of recombinants from parents evolved in the same and different environments (p = 0.36). This remained true even when I excluded those crosses with positive deviations from expected fitness from the analysis (p=0.93) (Fig. 4.8). These results indicated that the assay environment is an important factor to consider. Moreover, I only found overall

positive deviations from expected fitness in recombinants from parents evolved in different environments assayed in lactose environment (Fig. 4.8).

The genetic basis of lower fitness found in recombinants

To better understand the genetic basis of the fitness lower than expected fitness among recombinants, I first sequenced the genome of four parent strains (glu2, glu5, lac5, and lac6). I found 17, 14, 14, and 15, nonsynonymous and indel mutations accumulated in these four populations, respectively (Table 4.5). There are many unique mutations comparing strains isolated from populations evolved in both the same and in different environments.

Table 4.5 Mutations found in clones from four evolved population.

Clone isolated from:			
lac5	**lac6**	**glu2**	**glu5**
argR	*lacI*	*ybbN*	*proY*
proX/ygaXY	*yccE*	*infA*	*argT/ubiX*
ppk	*mreD*	*mokB*	*fabB/mnmC*
ygfT	*fis*	*pykF*	*vacJ*
*malT(+50bp)**	*malT(25bp x 2)*	*yeaJ*	*malT(+25bp)*
nadR(insertion)	*nadR(R123C)*	*nadR(S178L)*	*nadR(+AAAG)*
spoT(M330I)	*spoT(G207D)*	*spoT(G207D)*	*spoT(R701Q)*
Δ rbsD	*Δ rbsD*	*Δ rbsD*	*kgtP/rrfG*
yijC	*fabB*	*yijC*	*dcuR/yjdI*
holC	*fabF*	*ECB02816*	*cysB*
yhiO/uspA	ECB_01998	*hemX*	*arcB*
hsdS	*nanK*	*nupC/yfeA*	*yehM*
ECB_01992	ECB_00726	*ptxA*	*IclR/metH*
trkD / insJ-5	*polA*	*ytfT*	
	recD	*yijZ*	
		yfbG	
		cysB	

*The allele information is given if different mutations arise in the same gene in different evolved strains

Next, I focused on recombinants with low fitness, to try to determine the nature of any genetic incompatibilities. To do this, I sequenced genomes of two low fitness recombinants from each of three crosses (Table 4.6). I also screened for the presence of all parental mutations in 22 recombinants from the lac5 × lac6 cross. In all cases, recombinants had mutations from both parent strains (Table 4.6 and Fig. 4.10). Of note, all recombinants had more mutations from the recipient than the donor parent, perhaps indicating some bias of conjugation or reflecting differences in the proximity of mutations to the different selective markers used in the recombinant selection procedure. The averaged proportion of mutations in recombinants from parents lac5 and lac6 were 28.7% and 66.7%, respectively. A Chi-squared test showed that there was a significant difference between the proportion of mutation number in recombinants from parental strains lac5 and lac6 ($p = 0.04$), but the averaged total number of mutations in recombinants was not different from the mean of the total number of parental mutations (t-test: $p = 0.25$).

Table 4.6 The mutations found in six recombinants from three pairs of evolved populations evolved for 4500 generations in either the same or different selective environments.

Selected recombinants					
glu5 × glu2-11*	**glu5 × glu2-15**	**lac5 × glu5-2**	**lac5 × glu5-19**	**lac5 × lac6-3**	**lac5 × lac6-11**
ybbN‡	*ybbN*	*argT/ubiX*	*argT/ubiX*	*lacI*	*lacI*
infA	*infA*	*fabB/mnmC*	*fabB/mnmC*	*yccE*	*yccE*
mokB	*mokB*	*vacJ*	*vacJ*	*nadR*	*nadR*
nupc/yfeA	*nupc/yfeA*	*kgtP/rrfG*	*kgtP/rrfG*	*fis*	*fis*
yijC	*yijC*	*arcB*	*arcB*	*fabF*	*mreD*
ytfT	*ytfT*	*malT* (+25bp)	*malT* (+50bp)	*nanK*	*ECB00726*
yijZ	*yiJZ*	*dcuR/yjdI*	*dcuR/yjdI*	*recD*	*malT*
nadR	*nadR*	*nadR*	*nadR*	*ECB01998*	*spoT* (M330I)
cysB	*cysB*	*yehM*	*yehM*	*rbsD*	*ΔrbsD*
yfbG	*pykF*	*iclR/metH*	*iclR/metH*	*polA*	*yhiO/uspA*
spoT (G207D)	*ECB02816*	*cysB*	*cysB*	*ECB00726*	*trkD/insJ-5*
malT	*hemX*	*spoT* (R701Q)		*proX/ygaXY*	*jijC*
arcB	*yeaJ*	*ΔrbsD*	*ΔrbsD*	*yhiO/uspA*	*yhiO*
recD	*spoT*	***Hns/tdk***		*holc*	*argR*
Δ(26 gene)#	**Δ(26 gene)**	**Δ(26 gene)**		*malT*	***hns***
				spoT	***gale/modF***

*Number indicates recombinant clone number that was sequenced.

‡Text color of parent strains used to indicate origin of mutations in recombinants.

#Mutations in bold are spontaneous mutations not found in either parent.

From genome sequencing data, I found that some of the recombinants had mutations that were not present in the parental strains. Among six genome sequenced recombinants, four recombinants had either insertion or deletion mutations not present in either parent (Table 4.6). Three recombinants, two from the same parents and one from another different pair of parents, had a large deletion mutation and two other

recombinants had *IS*1-mediated insertion mutations. These non-parental spontaneous mutations could be involved in the reduced fitness of some recombinants. This finding indicated that the lower fitness of the recombinants may caused by negative interactions between mutations or the spontaneous mutations from conjugation or both.

If the lower fitness was caused by the negative interaction between the mutations from different parental strains, we may expect that the fitness decrease is negatively correlated with the potential number of interacting mutations. To test this, I calculated the potential number of new pair-wise mutation interactions in each recombinant and tested for a relationship with that recombinant strains deviation from expected fitness. I found that as the number of potential interactions increase, the deviation from expected fitness becomes increasingly negative, but the relationship is not significant (Fig. 4.9). As well, there was no significant relationship between the deviation from expected fitness and the proportion of mutations either from either the lac5 or lac6 parent (Table 4.5 and Fig. 4.9). I noticed that there were only a couple of recombinants in each pair of conjugated populations whose fitness is substantially lower than expected. This may suggest that lower fitness of recombinants may be caused by the negative interactions between specific set of mutations. I compared the mutations present in recombinant with extremely low fitness, lac5 × lac6-11, with mutations present in the other 21 recombinants. I did not find any pair of parental mutations specific to the recombinant lac5 × lac6-11. Instead, I found that three mutations in genes *yijC*, *ECB01192*, and *yccE* only exist in this recombinant, thus these three mutations are candidates to be involved in

a negative interaction (Fig. 4.10). Among these three mutations, two mutations in *yijC* and *yccE* arise in lac5 and mutation in *ECB01192* arises in lac6.

Table 4.5 Genetic and fitness comparison of sequenced recombinants from cross between lac5 and lac6**.**

Recom-binant ID	Mutations from each parent (proportion of possible)		Potential pair-wise interactions		Deviation from expected fitness (glucose)	Deviation from expected fitness (lactose)
	lac5 (14 mutations)	lac6 (15 mutations)	Evolved - evolved	Evolved-ancestral		
1	8(0.27)	10(0.67)	80	98	-0.051	-0.047
2	8(0.57)	7(0.47)	56	104	-0.026	-0.066
3	5(0.36)	11(0.73)	55	89	-0.018	0.012
4	5(0.36	8(0.53)	40	101	-0.040	-0.023
5	4(0.29)	11(0.73)	44	84	-0.016	-0.110
6	0(0)	11(0.73)	0	44	0.024	-0.089
7	0(0)	14 (0.93)	0	14	-0.026	-0.050
8	5 (0.36)	8 (0.53)	40	101	-0.049	-0.008
9	4 (0.29)	8 (0.53)	32	96	-0.028	0.0331
10	6 (0.43)	6 (0.40)	36	102	-0.043	0.035
11	8 (0.57)	6 (0.40)	48	102	-0.314	-0.403
12	2 (0.14)	12 (0.80)	24	60	0.001	-0.104
13	9 (0.64)	8 (0.53)	72	101	-0.012	-0.052
14	2 (0.14)	11 (0.73)	22	68	0.021	0.011
15	2 (0.14)	14 (0.93)	28	38	-0.009	0.003
16	5 (0.36)	11 (0.73)	55	89	-0.118	-0.103
17	1 (0.07)	11 (0.73)	11	57	-0.019	-0.067
18	3 (0.21)	9 (0.60)	27	87	0.007	-0.002
19	5 (0.36)	9 (0.60)	45	99	0.004	0.004
20	4 (0.29)	11 (0.73)	44	84	-0.014	-0.064
21	1 (0.07)	13 (0.83)	13	39	-0.043	-0.027
22	2 (0.14)	11 (0.73)	22	68	0.019	-0.030
Average	4 (0.29)	10 (0.67)	36	78	-0.034	-0.051

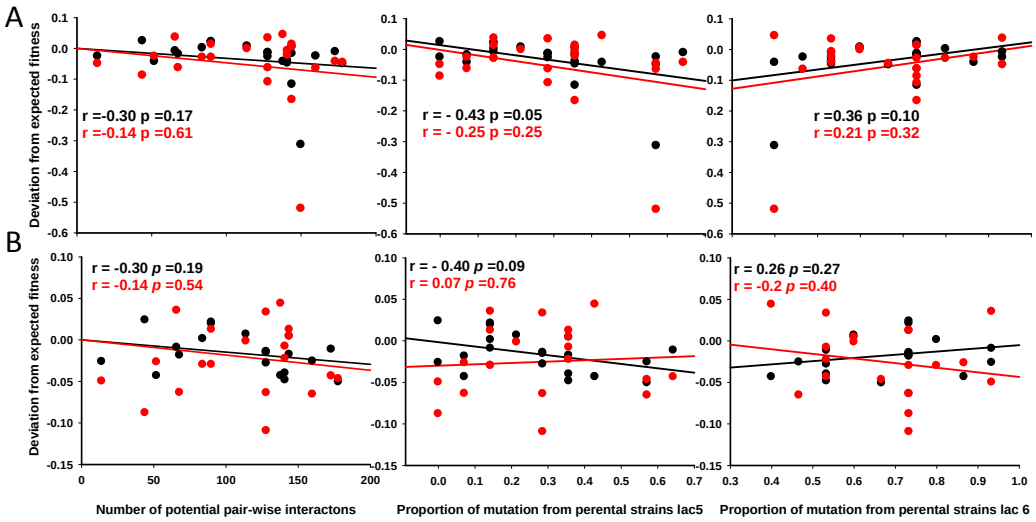

Figure 4.9 Relationship between fitness of recombinants (measured in glucose (red) and lactose (black)) and the deviation from expected in fitness of recombinants from lac5 × lac6 and the potential number of pair-wise interactions between mutations from the different parents, and the proportion of mutations in recombinants of from lac5and lac6 parents (A); The relationship between deviation from expected fitness and three independent variables was shown in (B) by omitting two recombinants with the very low fitness.

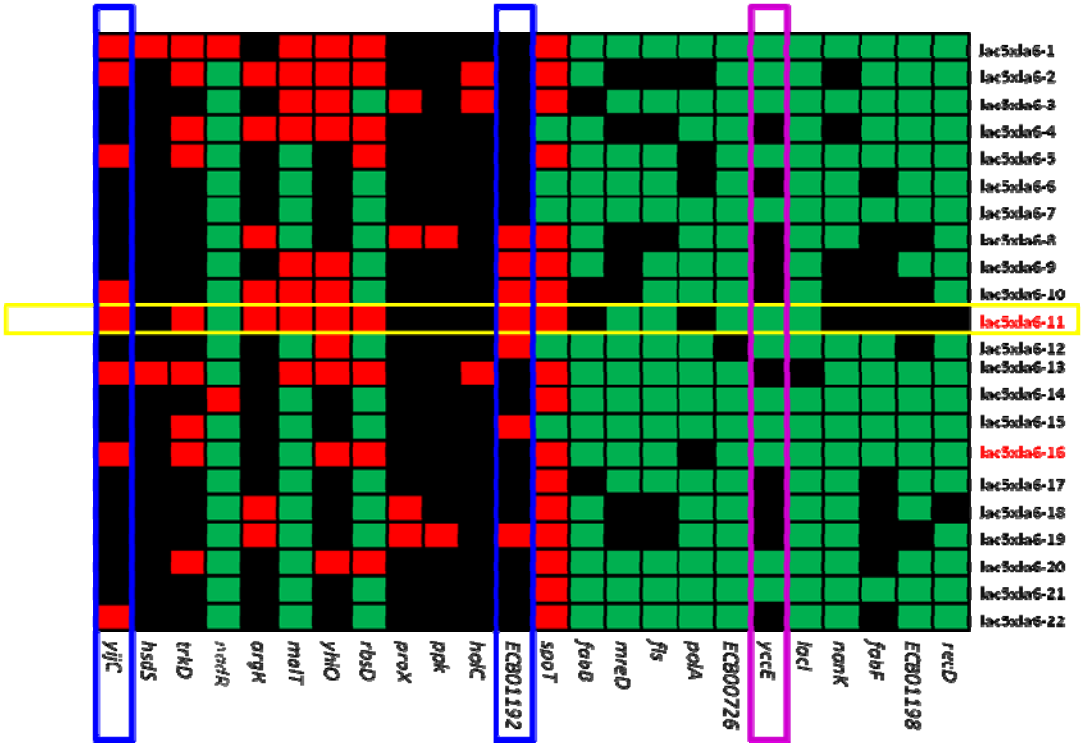

Figure 4.10 The mutations in 22 recombinants from lac5 × lac6. Red symbols indicate mutations derived from the lac5 parent and green symbols indicate mutation derived from the lac6 parent. Black symbols indicate wild type allele. The recombinant highlighted in yellow has extremely low fitness. The genes in blue and purple frame indicate candidate mutations which interact negatively.

Discussion

Despite the many theoretical and empirical studies that examine the role of natural selection in speciation, our understanding of the details of the process remains limited (Curie 2012). These limitations, particularly center on the role of adaptation in the initial development of reproductive isolation, the genetic basis of reproductive barriers at the initial stage of speciation, and the evolution and the relative importance of two theories of speciation, ecological and mutation-order speciation (Schluter 2009, Servedio et al. 2011). Here, I studied the extent of reproductive isolation arising following adaptation to defined

environmental conditions, by examining the extent, and aspects of the patterns of genetic interactions occurring in crosses of experimental populations of *Escherichia coli* evolved in two different carbon resource environments. I found two main results. First, in general, recombinants produced by crossing parents that evolved in both the same and different environments tended to be less fit than expected. This may be due to negative interactions between mutations arising in different populations, although we cannot rule out some contribution due to non-random inheritance of parental mutations. Second, analysis of recombinant genotypes indicates that negative interactions may be of higher than pair-wise order. Moreover, it is likely that indirect genetic incompatibilities that lead to the production of additional non-parental mutations in recombinants may also be involved in the lower than expected fitness of some recombinants. In general, the conclusion that adaptation promotes the development of reproductive isolation was supported. The pervasive incompatibility between multiple pairs of populations after 4500 generations evolution is consistent with the Dobzhansky-Muller model prediction of reproductive isolation through negative epistasis.

An important unanswered question about speciation is whether, and if so, to what extent, the process of population divergence and eventual speciation is a direct consequence, rather than an indirect by-product of adaptation (Curie 2012). Although natural selection has long been suggested to be involved in the process of speciation (reviewed in Coyne and Orr 2004), the role and importance of adaptive diversification in triggering reproductive isolation is still poorly understood. How adaptation affects reproductive isolation is essential to understand the relationship between natural selection

and speciation. There are two main views about how natural selection lead to new species, ecological speciation due to divergent adaptation, and mutation-order speciation due to parallel adaptation (Rundle et al. 2005, Schluter and Conte 2009, Weissing et al. 2011, Tarroso et al. 2014). Only a few studies have examined the role of adaptation in speciation, as it is hard to directly observe the long-term adaptation of organisms in nature. Thus the connection between adaptation to different environments and the evolution of postzygotic hybrid incompatibility remains poorly understood, and very little is known about the environmental dependency of early-acting isolating barriers. I am aware of only a few other experimental evolution studies in which the effects of divergent and parallel adaptation on the origination of reproductive isolation between populations at the initial stage of speciation were examined. The results of those studies are not consistent. Studies of experimentally evolved yeast populations clearly demonstrated that divergent adaptation resulted in significantly greater reproductive isolation than did parallel adaptation (Dettman et al. 2007, Anderson et al. 2010). Similarly, Moyle *et al.* (2012) examined the F1 hybrid sterility of *Collinsia sparsiflora*, a Californian endemic plant, which was adapted to two distinct soil types. They found that hybrids generated from parents adapted to different soil types were less fertile than hybrids from parents adapted to the same soil type, indicating that populations under divergent selective pressure had stronger reproductive isolation than the populations under similar selective pressure. Wright (2013) found that the wild flower *Mimulus guttatus,* adapted to copper mine soil, failed to produce viable hybrids following crosses with other populations. High-resolution genome mapping led them to conclude that selection on copper tolerance

only caused hybrid incompatibility. Studying *Drosophila*, neither Rundle et al. (2003) nor Kwan (2010) found evidence to support the hypothesis that divergent adaptation can produce reproductive isolation. In my study, I found, in general, that the average fitness of recombinants produced by crossing parents evolved either in the same or different environments was significantly lower than the expected fitness, indicating a prevailing influence of incompatibility between the independently accumulated mutations, which suggested that both parallel and divergent adaptation can lead to population differentiation and eventually reproductive isolation between adapted populations over time. However, I did not find any significant difference in the extent of incompatibility between parallel adaptation and divergent adaptation.

I found that reproductive isolation was environmentally dependent. In several cases the deviation of recombinants from their expected fitness was negative in only one of the assay environments, indicating a weak barrier to genetic exchange between strains isolated from different evolved populations. This asymmetry of reproductive isolation was also observed in a study examining *Drosophila* populations, evolved in both desiccation and starvation environments (Kwan and Rundle 2010). Recombinants had reduced fitness in the desiccation but not in the starvation environment. Two general reasons could explain this observation. (i) That population may not be adapted to the selective environment, i.e., the evolution time might not be long enough to produce distinct genetic variations between different populations, which was considered as a likely explanation for not detecting any reproductive isolation in two studies (Mooers et al. 1999, Rundle 2003). However, this is unlikely to be the case in our experiment since

all of our populations showed a dramatic fitness increase in their selective environment after 4500 generation evolution, and the sequenced strains of several evolve populations provide evidence that there are many unique mutations arise in different strains. (ii) Selection in two environments may not, in fact, be divergent, i.e., adaptation to one environment may not come at a cost in the alternative environment. This explanation is relevant to my experiment. Previous studies have shown that at least some beneficial mutations that arise during adaptation to lactose are costly during growth in glucose (Quan et al., 2012), but no glucose-selected mutation has yet been shown to be costly in lactose. In this study, we found that the populations evolved in its selective environment also confer a fitness benefit when assayed in the other environment, even not as high as in the original selective environment (Fig. 4.5). Therefore, absence of divergent selection might be one of the reasons why the deviation from expected fitness between populations evolved in different environment is asymmetric. This result also suggests that contrasting environments may be a key element to produce the evolution of reproductive isolation.

A theoretical genetic model of reproductive isolation, the Dobzhansky-Muller incompatibility model (DMI), has been widely accepted and considered as the most common genetic mechanism to promote speciation (Presgraves 2010, Agrawal et al. 2011, Silva et al. 2012). The generally lower than expected fitness of recombinants is consistent with the widespread presence of recombinant incompatibility. I only found a very small fraction of recombinants showing really low fitness, for instance, one out of 22 recombinant (lower than 5%) having really lower fitness for cross lac5 ×lac6. Thus this minority of recombinants from each cross, involving both parents from the same and

from different environments, showed very low fitness, suggesting that specific sets of mutations are involved in the potential negative interactions between evolved mutations.

Negatively interacting genes, which cause recombinant incompatibility, are termed DM genes. Pairs of such incompatible genes have been identified in evolved yeast populations (Anderson et al. 2010) and in *Drosphila* (Brideau et al. 2006). However, Kao *et al*. (2010) demonstrated that multiple genes could be involved in DM incompatibility between different organisms and epistasis among more than two genes can play a fundamental role in determining patterns and rates of evolution of isolation between diverging species. Moyle and Nakazato (2009) found complex genetic interaction between pairs of short chromosomal regions in *Solanum habrochaites* by introducing them together into another species *Solanum lycopersicum*. In my study, I screened all mutations in 22 recombinants from the same pair of parents and I compared the mutations present in the recombinants with extremely low fitness to those in recombinants from the same parents that had much higher fitness. I did not find any specific pair of mutations, independently evolved from different parent, which present in the recombinants with really low fitness but not in all the other recombinants, whereas I actually found a specific set of three mutations that were exclusively present in lower fitness recombinants, which are mutations in genes *yijC*, *yccE*, and *ECB01192*. This result suggests the possibility that interactions between more than two genes may be involved in the genetic barrier between divergent populations. Though I do not yet know exactly whether these three mutations are interacting negatively to decrease the fitness of that recombinant, this finding at least reveals the complexity of negatively interacting genes in the development

of reproductive isolation in incipient species. Overall, my results demonstrated that negative epistasis between mutations is likely a major contributor to low recombinant fitness, but the interaction is more likely to be polygenetic.

In my study, I also found that four out of six genome sequenced recombinants from different pairs of conjugations had mutations that were not present in the parental strains. These represent spontaneous mutations that may have occurred independently of the recombination process, or, more likely, given the low mutation rate in the ancestral strain, through some recombination-dependent process. Because control recombination experiments crossing strains with themselves did not result in any recombinants with low fitness, the new mutations may reflect some kind of physical incompatibility between parental genomes or a genetic incompatibility that has physical consequences. For example, a mutation combination could upregulate mobile elements, which are known to mediate chromosomal rearrangements and deletions in our strain (Schneider et al. 2000, Cooper et al. 2001). Supporting this, the same mutations, were found in multiple independent recombinant strains from the same parents, and in recombinants from different parents. Therefore these spontaneous mutations are a likely contributor to the very low fitness of these recombinants, suggesting other physical incompatibility which may cause chromosomal rearrangement between two parents was one of potential causes to overall decreased fitness from expected observed in this study.

Despite the prevailing overall negative deviation from expected fitness was observed in this study, I noticed that the averaged deviation from expected fitness of two

crosses, lac1 × glu3 and lac4 × glu6, were significantly positive when assayed in the lactose environment (Fig. 4.8). The positive deviation indicates either that some mutations evolved in different environment interact positively, or that mutations in one parent interact negatively in the non-selected environment. I rule out the possibility of our results reflecting an unconditionally deleterious mutation in one parent because if such a mutation was inherited by approximately half of all recombinants, it would have no mean effect on deviation from expected fitness. Of these possibilities, I favor that some mutations present in the glucose-selected strain increase fitness in lactose by positively interacting with the mutations accumulated in lactose. This possibility is consistent with the observation in the long-term evolution experiment running in our lab. It was found that pre-evolving an *E.coli* population in glucose for 1000 generations will accelerate subsequent adaptation in lactose, compared to the ancestor (unpublished). Moreover, I noted that the positive deviation was only found in recombinants from parents evolved in different environments, which may indicate that the mutations selected under same pressure tend not to positively interact with each other. This is consistent with findings that beneficial mutations selected under similar selective pressures tend to interact negatively (Chou et al. 2011, Khan et al. 2011, Rokyta et al. 2011, Tokuriki et al. 2012, Wang et al. 2012, Schenk et al. 2013).

In summary, I found that populations of *E. coli* evolved in either glucose and lactose environments are specialized to their selection environments. Limited genome sequencing is consistent with high genetic divergence between populations evolved both in the same and in different environments. On average, recombinants had lower than

expected fitness suggesting the presence of widespread, though environmentally dependent, negative interactions between mutations that accumulated independently in the parents. That negative interactions were common between parents evolved in the same and in different environments supporting that mutation interactions, rather than antagonistic pleiotropy, best explains our results.

Appendix 4

Table A4.1 The primers (inside and outside of the manipulated genes) used to confirm the successful introduction of auxtrophic mutation markers in evolved populations

Primers		**Mutation markers**		
		lysA-	*trpA*-	*cysE*-
Inside	left	5'-catctggaacaggtgtgtgg-3'	5'-cggcacaatgttttgaaatg-3'	5'-cactggcaatctttctgcaa-3'
	right	5'-catcaaccagcacaaagtgg-3'	5'-cctgctcgtgacagcaaata-3'	5'-cggtcaccaccagatttacc-3'
Outside	left	5'-agcaccgataccgatctcac-3'	5'-gaaagcgaggggaaatct-3'	5'-gccggtcattatctcatcgt-3'
	right	5'-tcatgcaaccagcgactaac-3'	5'-aggatttttccggcttcatt-3'	5'-cccatccccatactcaaatg-3'

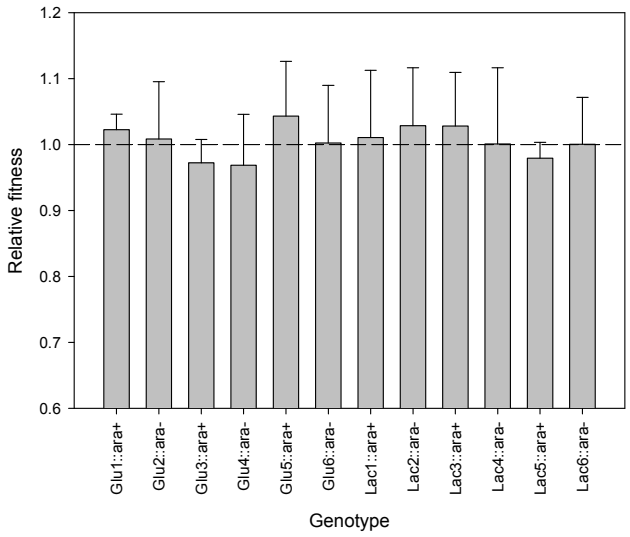

Figure A4.1 Fitness of *araA+/araA-* derivatives of 12 evolved populations relative to themselves assayed in their own selective environment. As expected, in all cases the Ara marker was neutral. Error bars indicate 95% confidence intervals (n=8/genotype).

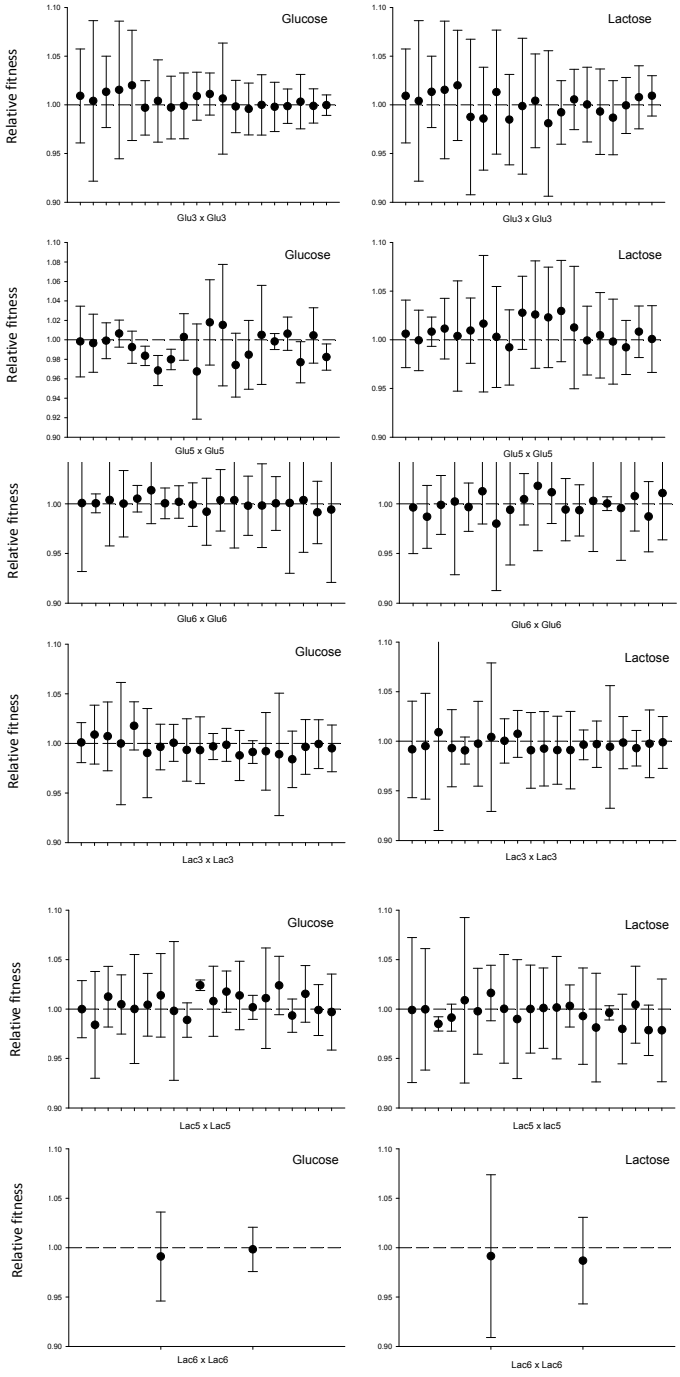

Figure A4.2 The fitness of recombinants from control conjugations designed to test for an effect of conjugation itself on fitness of recombinants. Fitness of recombinants was measured relative to the corresponding parent strain in minimal glucose (left column) and lactose (right column). Error bars indicate 95% confidence intervals based on n=4.

References

Acuna, R., B. E. Padilla, C. P. Florez-Ramos, J. D. Rubio, J. C. Herrera, P. Benavides, S.-J. Lee, T. H. Yeats, A. N. Egan, J. J. Doyle, and J. K. C. Rose. 2012. Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. Proceedings of the National Academy of Sciences of the United States of America **109**:4197-4202.

Agrawal, A. F., J. L. Feder, and P. Nosil. 2011. Ecological divergence and the origins of Intrinsic postmating isolation with gene flow. International Journal of Ecology **2011**:15.

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. Current Biology **20**:1383-1388.

Aragaki, C. C. and R. C. Barber. 2005. Higher order interaction estimation using epistasis among innate immunity SNP profile risk for severe sepsis. American Journal of Epidemiology **161**:S108-S108.

Arnegard, M. E., M. D. McGee, B. Matthews, K. B. Marchinko, G. L. Conte, S. Kabir, N. Bedford, S. Bergek, Y. F. Chan, F. C. Jones, D. M. Kingsley, C. L. Peichel, and D. Schluter. 2014. Genetics of ecological divergence during speciation. Nature **511**:307-311.

Axenovich, T. I., M. B. Rogatcheva, S. Oda, and P. M. Borodin. 1998. Inheritance of male hybrid sterility in the house musk shrew (*Suncus murinus*, Insectivora, Soricidae). Genome **41**:825-831.

Azevedo, R. B. R., R. Lohaus, S. Srinivasan, K. K. Dang, and C. L. Burch. 2006. Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. Nature **440**:87-90.

Barbash, D. A., P. Awadalla, and A. M. Tarone. 2004. Functional divergence caused by ancient positive selection of a *Drosophila* hybrid incompatibility locus. PLoS Biology **2**:839-848.

Barton, N. H. 2001. The role of hybridization in evolution. Molecular Ecology **10**:551-568.

Bateson, W. 1907. Facts limiting the theory of heredity. Science **26**:649-660.

Bikard, D., D. Patel, C. Le Metté, V. Giorgi, C. Camilleri, M. J. Bennett, and O. Loudet. 2009. Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. Science **323**:623-626.

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

Brown, J. R. 2003. Ancient horizontal gene transfer. Nature Review Genetics **4**:121-132.

Butlin, R. K., M. Saura, G. Charrier, B. Jackson, C. André, A. Caballero, J. A. Coyne, J. Galindo, J. W. Grahame, J. Hollander, P. Kemppainen, M. Martínez-Fernández, M. Panova, H. Quesada, K. Johannesson, and E. Rolán-Alvarez. 2014. Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. Evolution **68**:935-949.

Cattani, M. V. and D. C. Presgraves. 2009. Genetics and lineage-specific evolution of a lethal hybrid incompatibility between *Drosophila mauritiana* and its sibling species. Genetics **181**:1545-1555.

Chandler, C. H., S. Chari, and I. Dworkin. 2013. Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. Trends in Genetics : TIG **29**:358-366.

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.

Chou, H. H., N. F. Delaney, J. A. Draghi, and C. J. Marx. 2014. Mapping the fitness landscape of gene expression uncovers the cause of antagonism and sign epistasis between adaptive mutations. PLoS Genetics **10**:1-11.

Cooper, T. F., D. E. Rozen, and R. E. Lenski. 2003. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America **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. Journal of Bacteriology **183**:2834-2841.

Cooper, V. S., S. H. Vohr, S. C. Wrocklage, and P. J. Hatcher. 2010. Why genes evolve faster on secondary chromosomes in bacteria. Plos Computational Biology **6**:e1000732.

Coyne, J. A. 1992. Genetics and speciation. Nature **355**:511-515.

Curie, M. 2012. What do we need to know about speciation? Trends in Ecology & Evolution **27**:27-39.

Da Silva, J., M. Coetzer, R. Nedellec, C. Pastore, and D. E. Mosier. 2010. Fitness epistasis and constraints on adaptation in a human immunodeficiency virus type 1 protein region. Genetics **185**:293-303.

Dagan, T., Y. Artzy-Randrup, and W. Martin. 2008. Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. Proceedings of the National Academy of Sciences of the United States of America **105**:10039-10044.

Dagan, T. and W. Martin. 2007. Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. Proceedings of the National Academy of Sciences of the United States of America **104**:870-875.

Datsenko, K. A. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proceedings of the National Academy of Sciences of the United States of America **97**:6640-6645.

De Visser, J. and J. Krug. 2014. Empirical fitness landscapes and the predictability of evolution. Nature Reviews Genetics **15**:480-490.

De Visser, J. A. G. M., T. F. Cooper, and S. F. Elena. 2011. The causes of epistasis. Proceedings of the National Academy of Sciences of the United States of America **278**:3617-3624.

De Visser, J. A. G. M. and S. F. Elena. 2007. The evolution of sex: empirical insights into the roles of epistasis and drift. Nature Reviews Genetics **8**:139-149.

Dettman, J. R., J. B. Anderson, and L. M. Kohn. 2008. Divergent adaptation promotes reproductive isolation among experimental populations of the filamentous fungus *Neurospora*. Bmc Evolutionary Biology **8**:35.

Dettman, J. R., J. B. Anderson, and L. M. Kohn. 2010. Gnome-wide investigation of reproductive isolation in experimental lineages and natural species of neurospora: identifying candidate regions by microarray-based genotyping and mapping. Evolution **64**:694-709.

Dettman, J. R., C. Sirjusingh, L. M. Kohn, and J. B. Anderson. 2007. Incipient speciation by divergent adaptation and antagonistic epistasis in yeast. Nature **447**:585-588.

Dobzhansky, T. 1937. Genetics and the Origin of Species. Columbia University Press, New York.

Draghi, J. A. and J. B. Plotkin. 2013. Selection biases the prevalence and type of epistasis along adaptive trajectories. Evolution **67**:3120-3131.

Dworkin, I., E. Kennerly, D. Tack, J. Hutchinson, J. Brown, J. Mahaffey, and G. Gibson. 2009. Genomic consequences of background effects on scalloped mutant expressivity in the wing of *Drosophila melanogaster*. Genetics **181**:1065-1076.

Elena, S. F. and R. E. Lenski. 1997. Testofsynergistic interactions among deleterious mutations in bacteria. Nature **390**:395-398.

Elena, S. F. and R. E. Lenski. 2001. Epistasis between new mutations and genetic background and a test of genetic canalization. Evolution **55**:1746-1752.

Fisher, R. A. 1918. The correlations between relatives on the supposition of Mendelian inheritance. Transactions of the Royal Society of Edinburgh Earth Sciences **52**:399-433.

Fishman, L. and J. H. Willis. 2001. Evidence for Dobzhansky-Muller incompatibilites contributing to the sterility of hybrids between *Mimulus guttatus* and *M.nasutus*. Evolution **55**:1932-1942.

Fraïsse, C., J. A. D. Elderfield, and J. J. Welch. 2014. The genetics of speciation: are complex incompatibilities easier to evolve? Journal of Evolutionary Biology **27**:688-699.

Gallet, R., T. F. Cooper, S. F. Elena, and T. Lenormand. 2012. Measuring selection coefficients below 10(-3): method, questions, and prospects. Genetics **190**:175-U287.

Gerke, J., K. Lorenz, S. Ramnarine, and B. Cohen. 2010. Gene–environment interactions at nucleotide resolution. PLoS Genetics **6**:e1001144.

Gevers, D., F. M. Cohan, J. G. Lawrence, B. G. Spratt, T. Coenye, E. J. Feil, E. Stackebrandt, Y. Van de Peer, P. Vandamme, F. L. Thompson, and J. Swings. 2005. Re-evaluating prokaryotic species. Nature Reviews Microbiology **3**:733-739.

Gibson, G. 2012. Rare and common variants: twenty arguments. Nature Reviews Genetics **13**:135-145.

Gogarten, J. P., W. F. Doolittle, and J. G. Lawrence. 2002. Prokaryotic evolution in light of gene transfer. Molecular Biology and Evolution **19**:2226-2238.

Good, B. H. and M. M. Desai. 2015. The impact of macroscopic epistasis on long-term evolutionary dynamics. Genetics **199**:177-U639.

Greene, D. and K. Crona. 2014. The changing geometry of a fitness landscape along an adaptive walk. Plos Computational Biology **10**:e1003520.

Greig, D. 2009. Reproductive isolation in *Saccharomyces*. Heredity **102**:39-44.

Guerrero, R. F., F. Rousset, and M. Kirkpatrick. 2012. Coalescent patterns for chromosomal inversions in divergent populations. Philosophical Transactions of the Royal Society B: Biological Sciences **367**:430-438.

Hall, A. R. and R. C. MacLean. 2011. Epistasis buffers the fitness effects of rifampicin-resistance mutations in *Pseudomonas aeruginosa*. Evolution **65**:2370-2379.

Hayden, E. J., E. Ferrada, and A. Wagner. 2011. Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. Nature **474**:92-95.

Jeong, H., V. Barbe, C. H. Lee, D. Vallenet, D. S. Yu, S. H. Choi, A. Couloux, S. W. Lee, S. H. Yoon, L. Cattolico, C. G. Hur, H. S. Park, B. Segurens, S. C. Kim, T. K. Oh, R. E. Lenski, F. W. Studier, P. Daegelen, and J. F. Kim. 2009. Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). Journal of Molecular Biology **394**:644-652.

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.

Kirkpatrick, M. and V. Ravigne. 2002. Speciation by natural and sexual selection: models and experiments. American Naturalist **159**:S22-S35.

Kondrashov, A. S. 1988. Deleterious mutations and the evolution of sexual reproduction. Nature **336**:435-440.

Kondrashov, F. A. and A. S. Kondrashov. 2001. Multidimensional epistasis and the disadvantage of sex. Proceedings of the National Academy of Sciences of the United States of America **98**:12089-12092.

Kouyos, R. D., O. K. Silander, and S. Bonhoeffer. 2007. Epistasis between deleterious mutations and the evolution of recombination. Trends in Ecology & Evolution **22**:308-315.

Kreimer, A., E. Borenstein, U. Gophna, and E. Ruppin. 2008. The evolution of modularity in bacterial metabolic networks. Proceedings of the National Academy of Sciences of the United States of America **105**:6976-6981.

Kryazhimskiy, S., J. Dushoff, G. A. Bazykin, and J. B. Plotkin. 2011. Prevalence of epistasis in the evolution of influenza a surface proteins. PLoS Genetics **7**:e1001301.

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.

Kryazhimskiy, S., G. Tkacik, and J. B. Plotkin. 2009. The dynamics of adaptation on correlated fitness landscapes. Proceedings of the National Academy of Sciences of the United States of America **106**:18638-18643.

Kwan, L. and H. D. Rundle. 2010. Adaptation to desiccation fails to generate pre- and postmating isolation in replicate *Drosophila melanogaster* laboratory populations. Evolution **64**:710-723.

Laing, C., C. Buchanan, E. Taboada, Y. Zhang, A. Kropinski, A. Villegas, J. Thomas, and V. Gannon. 2010. Pan-genome sequence analysis using Panseq: an online tool for the rapid analysis of core and accessory genomic regions. Bmc Bioinformatics **11**:1-14.

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

Lehner, B. 2011. Molecular mechanisms of epistasis within and between genes. Trends in Genetics **27**:323-331.

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

Lunzer, M., G. B. Golding, and A. M. Dean. 2010. Pervasive cryptic epistasis in molecular evolution. PLoS Genetics **6**.

Macía, J., R. V. Solé, and S. F. Elena. 2011. The cause of epistasis in genetic networks. Evolution:no-no.

Mackenzie, S. M., M. R. Brooker, T. R. Gill, G. B. Cox, A. J. Howells, and G. D. Ewart. 1999. Mutations in the white gene of *Drosophila melanogaster* affecting ABC transporters that determine eye colouration. Biochim Biophys Acta **15**:173-185.

MacLean, R. C. 2010. Predicting epistasis: an experimental test of metabolic control theory with bacterial transcription and translation. Journal of Evolutionary Biology **23**:488-493.

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.

Maitra, A. and K. A. Dill. 2015. Bacterial growth laws reflect the evolutionary importance of energy efficiency. Proceedings of the National Academy of Sciences of the United States of America **112**:406-411.

Masly, J. P., C. D. Jones, M. A. F. Noor, J. Locke, and H. A. Orr. 2006. Gene transposition as a cause of hybrid sterility in *Drosophila*. Science **313**:1448-1450.

Michalak, P., I. Minkov, A. Helin, D. N. Lerman, B. R. Bettencourt, M. E. Feder, A. B. Korol, and E. Nevo. 2001. Genetic evidence for adaptation-driven incipient speciation of *Drosophila melanogaster* along a microclimatic contrast in "Evolution Canyon," Israel. Proccedings of National of Academy of Sciences of Unite States America **98**:13195-13200.

Michener, J. K., S. Vuilleumier, F. Bringel, and C. J. Marx. 2014. Phylogeny poorly predicts the utility of a challenging horizontally transferred gene in *Methylobacterium* strains. Journal of Bacteriology **196**:2101-2107.

Mihola, O., Z. Trachtulec, C. Vlcek, J. C. Schimenti, and J. Forejt. 2009. A mouse speciation gene encodes a meiotic histone H3 methyltransferase. Science **323**:373-375.

Mooers, A. Ø., H. D. Rundle, and M. C. Whitlock. 1999. The effects of selection and ottlenecks on male mating success in peripheral isolates. The American Naturalist **153**:437-444.

Moore, F. B. G. and R. Woods. 2006. Tempo and constraint of adaptive evolution in *Escherichia coli* (Enterobacteriaceae, Enterobacteriales). Biological Journal of the Linnean Society **88**:403-411.

Moyle, L. C., M. Levine, M. L. Stanton, and J. W. Wright. 2012. Hybrid sterility over tens of meters between ecotypes adapted to serpentine and non-serpentine soils. Evolutionary Biology **39**:207-218.

Moyle, L. C. and T. Nakazato. 2009. Complex epistasis for Dobzhansky-Muller hybridi incompatibility in *Solanum*. Genetics **181**:347-351.

Muller, H. J. 1942. Isolating mechanisms, evolution, and temperature. Biology Symposium **6**:71-125.

Myar, E. 1942. Systematics and the Origin of Species. Columbia University Press, New York.

Ochman, H., J. G. Lawrence, and E. A. Groisman. 2000. Lateral gene transfer and the nature of bacterial innovation. Nature **405**:299-304.

Omholt, S. W., E. Plahte, L. Oyehaug, and K. Xiang. 2000. Gene regulatory networks generating the phenomena of additivity, dominance and epistasis. Genetics **155**:969-980.

Orr, H. A. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. Genetics **139**:1805-1813.

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

Orr, H. A. and M. Turelli. 2001. The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. Evolution **55**:1085-1094.

Otto, S. P. and T. Lenormand. 2002. Evolution of sexresolving the paradox of sex and recombination. Nature Reviews Genetics **3**:252-261.

Pagel, M. 1999. Inferring the historical patterns of biological evolution. Nature **401**:877-884.

Pearson, V. M., C. R. Miller, and D. R. Rokyta. 2012. The consistency of beneficial fitness effects of mutations across diverse genetic backgrounds. Plos One **7**:e43864.

Pepin, K. M. and H. A. Wichman. 2007. Variable epistatic effects between mutations at host recognition sites In X174 bacteriophage. Evolution **61**:1710-1724.

Phillips, P. C. 2008. Epistasis - the essential role of gene interactions in the structure and evolution of genetic systems. Nature Reviews Genetics **9**:855-867.

Poelwijk, F. J., D. J. Kiviet, D. M. Weinreich, and S. J. Tans. 2007. Empirical fitness landscapes reveal accessible evolutionary paths. Nature **445**:383-386.

Popa, O. and T. Dagan. 2011. Trends and barriers to lateral gene transfer in prokaryotes. Current Opinion in Microbiology **14**:615-623.

Powell, T. H., A. A. Forbes, G. R. Hood, and J. L. Feder. 2014. Ecological adaptation and reproductive isolation in sympatry: genetic and phenotypic evidence for native host races of *Rhagoletis pomonella*. Molecular Ecology **23**:688-704.

Presgraves, D. C. 2010. The molecular evolutionary basis of species formation. Nature Reviews Genetics **11**:175-180.

Presgraves, D. C., L. Balagopalan, S. M. Abmayr, and H. A. Orr. 2003. Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. Nature **423**:715-719.

Quan, S., J. C. J. Ray, Z. Kwota, D. Trang, G. Balazsi, T. F. Cooper, and R. D. Monds. 2012. Adaptive evolution of the lactose utilization network in experimentally evolved populations of *Escherichia coli*. PLoS Genetics **8**:e1002444.

Remold, S. K. and R. E. Lenski. 2004. Pervasive joint influence of epistasis and plasticity on mutational effects in *Escherichia coli*. Nature Genetics **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 Genetics **7**: e1002075.

Ruane, L. G. and K. Donohue. 2007. Environmental effects on pollen-pistil compatibility between *Phlox cuspidata* and *P. drummondii* (Polemoniaceae): implications for hybridization dynamics. American Journal of Botany **94**:219-227.

Rundle, H. D. 2002. A test of ecologically dependent postmating isolation between sympatric sticklebacks. Evolution **56**:322-329.

Rundle, H. D. 2003. Divergent environments and population bottlenecks fail to generate premating isolation in *Drosophila pseudoobscura*. Evolution **57**:2557-2565.

Rundle, H. D., S. F. Chenoweth, P. Doughty, and M. W. Blows. 2005. Divergent selection and the evolution of signal traits and mating preferences. PLoS Biology **3**:e368.

Salverda, M. L., E. Dellus, F. A. Gorter, A. J. Debets, J. van der Oost, R. F. Hoekstra, D. S. Tawfik, and J. A. de Visser. 2011. Initial mutations direct alternative pathways of protein evolution. PLoS Genetics **7**:e1001321.

Sanders, L. H. and K. E. Whitlock. 2003. Phenotype of the zebrafish masterblind (mbl) mutant is dependent on genetic background. Developmental Dynamics **227**:291-300.

Sanjuan, R. 2006. Quantifying antagonistic epistasis in a multifunctional RNA secondary structure of the *Rous sarcoma* virus. Journal of General Virology **87**:1595-1602.

Sanjuan, R. and F. Elena. 2006. Epistasis correlates to genomic complexity. Proceedings of the National Academy of Sciences **103**:14402-14405.

Schaper, S., I. G. Johnston, and A. A. Louis. 2011. Epistasis can lead to fragmented neutral spaces and contingency in evolution. Proceedings of the Royal Society B: Biological Sciences **279**:1777-1783.

Schenk, M. F., I. G. Szendro, M. L. M. Salverda, J. Krug, and J. A. G. M. de Visser. 2013. Patterns of epistasis between beneficial mutations in an antibiotic resistance gene. Molecular Biology and Evolution **30**:1779-1787.

Schluter, D. 2001. Ecology and the origin of species. Trends in Ecology & Evolution **16**:372-380.

Schluter, D. 2009. Evidence for ecological speciation and its alternative. Science **323**:737-741.

Schluter, D. and G. L. Conte. 2009. Genetics and ecological speciation. Proceedings of the National Academy of Sciences of the United States of America **106**:9955-9962.

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.

Servedio, M. R., G. S. V. Doorn, M. Kopp, A. M. Frame, and P. Nosil. 2011. Magic traits in speciation: 'magic' but not rare? Trends in Ecology & Evolution **26**:389-397.

Silva, J. C. E., B. Potts, and P. Tilyard. 2012. Epistasis causes outbreeding depression in eucalypt hybrids. Tree Genetics & Genomes **8**:249-265.

Sorek, R., Y. Zhu, C. J. Creevey, M. P. Francino, P. Bork, and E. M. Rubin. 2007. Genome-wide experimental determination of barriers to horizontal gene transfer. Science **318**:1449-1452.

Spor, A., T. Nidelet, J. Simon, A. Bourgais, D. de Vienne, and D. Sicard. 2009. Niche-driven evolution of metabolic and life-history strategies in natural and domesticated populations of *Saccharomyces cerevisiae*. Bmc Evolutionary Biology **9**:1-12.

Tang, S. and D. C. Presgraves. 2009. Evolution of the Drosophila nuclear pore complex results in multiple hybrid incompatibilities. Science **323**:779-782.

Tarroso, P., R. J. Pereira, F. Martinez-Freiria, R. Godinho, and J. C. Brito. 2014. Hybridization at an ecotone: ecological and genetic barriers between three Iberian vipers. Molecular Ecology **23**:1108-1123.

Tokuriki, N., C. J. Jackson, L. Afriat-Jurnou, K. T. Wyganowski, R. Tang, and D. S. Tawfik. 2012. Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. Nature Communications **3**:1257.

Touchon, M., C. Hoede, O. Tenaillon, V. Barbe, S. Baeriswyl, P. Bidet, E. Bingen, S. Bonacorsi, C. Bouchier, O. Bouvet, A. Calteau, H. Chiapello, O. Clermont, S. Cruveiller, A. Danchin, M. Diard, C. Dossat, M. E. Karoui, E. Frapy, L. Garry, J. M. Ghigo, A. M. Gilles, J. Johnson, C. Le Bouguénec, M. Lescat, S. Mangenot, V. Martinez-Jéhanne, I. Matic, X. Nassif, S. Oztas, M. A. Petit, C. Pichon, Z. Rouy, C. S. Ruf, D. Schneider, J. Tourret, B. Vacherie, D. Vallenet, C. Médigue, E. P. C. Rocha, and E. Denamur. 2009. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. PLoS Genet **5**:e1000344.

Treangen, T. J. and E. P. C. Rocha. 2011. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. PLoS Genetics **7**:e1001284.

Vos, M. 2011. A species concept for bacteria based on adaptive divergence. Trends in Microbiology **19**:1-7.

Wang, Y., C. Díaz Arenas, D. M. Stoebel, and T. F. Cooper. 2012. Genetic background affects epistatic interactions between two beneficial mutations. Biology Letters **9**:20120328.

Weinreich, D. M. 2005. The rank ordering of genotypic fitness values predicts genetic constraint on natural selection on landscapes lacking sign epistasis. Genetics **171**:1397-1405.

Weinreich, D. M., N. F. Delaney, M. A. DePristo, and D. L. Hartl. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science **312**:111-114.

Weinreich, D. M., Y. H. Lan, C. S. Wylie, and R. B. Heckendorn. 2013. Should evolutionary geneticists worry about higher-order epistasis? Current Opinion in Genetics & Development **23**:700-707.

Weinreich, D. M., R. A. Watson, and L. Chao. 2005. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. Evolution **59**:1165-1174.

Weissing, F. J., P. Edelaar, and G. S. van Doorn. 2011. Adaptive speciation theory: a conceptual review. Behavioral Ecology and Sociobiology **65**:461-480.

Wiedenbeck, J. and F. M. Cohan. 2011. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. Fems Microbiology Reviews **35**:957-976.

Wielgoss, S., J. E. Barrick, O. Tenaillon, S. Cruveiller, B. Chane-Woon-Ming, C. Medigue, R. E. Lenski, and D. Schneider. 2011. Mutation rate inferred from synonymous substitutions in a long-term evolution experiment with Escherichia coli. G3-Genes Genomes Genetics **1**:183-186.

Wielgoss, S., J. E. Barrick, O. Tenaillon, M. J. Wiser, W. J. Dittmar, S. Cruveiller, B. Chane-Woon-Ming, C. Medigue, R. E. Lenski, and D. Schneider. 2013. Mutation rate dynamics in a bacterial population reflect tension between adaptation and

genetic load. Proceedings of the National Academy of Sciences of the United States of America **110**:222-227.

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

Wolf, J. 2000. Epistasis and the Evolutionary Process. Oxford University Press. Oxford.

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.

Wright, K. M., D. Lloyd, D. B. Lowry, M. R. Macnair, and J. H. Willis. 2013. Indirect evolution of hybrid lethality due to linkage with selected locus in *Mimulus guttatus*. PLoS Biology **11**:e1001497.

Zhang, W., V. Sehgal, D. M. Dinh, R. B. R. Azevedo, T. F. Cooper, and R. Azencott. 2012. Estimation of the rate and effect of new beneficial mutations in asexual populations. Theoretical Population Biology **81**:168-178.