The Genotype-Phenotype Map:

Origins, Properties, and Evolutionary Consequences

by

Wei-Chin Ho

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Ecology and Evolutionary Biology)
in the University of Michigan
2017

Doctoral Committee:

Professor George Zhang, Chair
Professor L. Lacey Knowles
Associate Professor Jun Li
Professor Patricia Wittkopp

Wei-Chin Ho

weichho@umich.edu

ORCID iD: 0000-0002-0848-2476

ACKNOWLEDGEMENTS

I am very fortunate to meet so many great people in my journey of earning a doctoral degree. I am usually a solo traveler, but this journey would be impossible without their assistance. First of all, I would like to thank my advisor George Zhang. His door is always open, and I am always welcome to discuss anything with him. He is very patient in listening to or reading my words and letting me know how to improve. His knowledge and experiences have substantially influenced both my research theme and my career development. The journey would be much bumpier if there were no his mentoring.

In addition, I would like to thank my three committee members, Jun Li, Lacey Knowles, and Trisha Wittkopp. They are always very encouraging. The discussion with them helps me improve the quality of works and expand the scope of my researches.

I would like to also thank all members I have met in Zhang lab. They are extremely supportive intellectually and personally. Especially I appreciate the countless helps and suggestions from people in a more computational side, including Bryan Moyers, Chungoo Park, Nagarjun Vijay, Xinzhu Wei, Jinrui Xu, Jianrong Yang, and Zhengting Zou. I also thank my cohort Chuan Li by all of scientific and nonscientific conversations. People in a more experimental side also give me great opportunities to learn what I did not know well, and the helps from Calum Maclean, Soochin Cho, and Wenfeng Qian are highly appreciated.

I also thank all of the people in the EEB department. My cohorts help me pass prelim, and I always have fun when talking with them. People in Wittkopp lab also always have a great conversation with me. I also need to thank all of the helps from EEB staffs, especially the current and previous program coordinator Cindy Carl and Jane Sullivan. Both of them have cleared up tons of my questions.

All of the supports from the university are appreciated, including the funding support from EEB department and Rackham graduate school, career development support from CRLT, and language support from ELI.

I also thank all of my friends I met in Michigan. They make my life much easier and more joyful here.

In the end, I want to thank all of the supports from my mother, my father and my younger brother. While they probably have no idea about what evolutionary biology is or what a graduate school is, they do not add any constrain on what I would like to do. That is highly remarkable in a Taiwanese family.

"A journey of a thousand miles begins with one step," said Laozi. Dissertation is not just the end of a journey but also rather the first step to start to a new journey. Thank you to all who have contributed to this first step, which is important for my life. I am looking forward to meeting everyone again in my following journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABSTRACT

Describing and understanding the relationship between genotypes and phenotypes, or the genotype-phenotype map, is of long-lasting interest in genetics and evolutionary biology. My dissertation focuses on understanding the origins, properties, and evolutionary consequences of genotype-phenotype maps. In Chapter 2, using yeast morphological traits, I showed that most traits are affected by a small proportion of genes, many of which have small effects while a few have large effects. To explain why many phenotypic effects are small, in the rest of Chapter 2 as well as in Chapter 3, I studied yeast morphological traits, yeast gene expression traits, and *E. coli* reaction flux traits and found evidence supporting the hypothesis of adaptive genetic robustness. In Chapter 4, by comparing the evolutionary rates of phenotypic traits of varying importance, I found evidence for that yeast morphological traits have evolved generally by adaptation while yeast gene expression traits have evolved largely neutrally. In Chapter 5, using yeast morphological traits, I found that increasing mutational correlation generally facilitates phenotypic evolution when the correlation is low, but constrains it when the correlation become very high. Thus, an intermediate level of mutation correlation is most conducive to phenotype evolution. In Chapter 6, using *E. coli* gene expression level traits and *E. coli* reaction flux traits, I found that genetic changes tend to reverse plastic changes when a population adapts to a new environment, suggesting that phenotypic plasticity does not generally serve as a steppingstone to genetic adaption. To sum up, this dissertation highlights the importance of incorporating genotype-phenotype

maps into the study of evolution, identifies influential factors in phenotypic evolution, and thus deepens our understanding of general principles of evolution.

Chapter 1

General Introduction

A genotype-phenotype map refers to the mapping from a set of genotypes to a set of phenotypes where the word "mapping" is used as in mathematics, describing the unidirectional relationship from a set of inputs toward a set of outputs. In other words, it describes what genotypes give rise to what phenotypes. The concept of genotype-phenotype map has been important since the field of genetics began. This concept is also central in the field of evolutionary biology given the importance of genetics in studying evolutionary biology. In order to improve our understanding of evolution, my dissertation focuses on several questions related to genotype-phenotype maps. Before presenting my research, in this chapter of general introduction, I will first discuss how the concept of genotype-phenotype map has evolved along with the development of biology. I will also discuss the importance of genotype-phenotype maps in specific subfields of biology. I will then describe important properties of genotype-phenotype maps. After that, the evolution of these properties and the effects of these properties on evolution will be discussed. In the end, I will introduce the methodology for constructing genotype-phenotype maps and briefly explain how my following research chapters relate to the concept of genotype-phenotype maps.

Synthesis of the concept of genotype-phenotype map

When Gregor Mendel studied the heredity of peas and published two fundamental laws of inheritance, he did not use the term genotype or phenotype (Mendel 1901). However, his work opened the door to the study of the relationship between genotypes and phenotypes. For example, noticing all of the hybridized offspring show the same state of a trait while the pure-line parents show different states of a trait, he proposed that one of the two states is dominant to the other state. This finding suggests that individuals with different genetic compositions can have the same appearance, so the mapping between genotypes and phenotypes must not be simply one-to-one.

The terms of genotypes and phenotypes were not coined until the work published by Wilhelm Johannsen (1911). While breeding beans, he noticed that some F1 individuals which come from two pure-line parents have different appearances such as different seed lengths. Therefore, there is a need to distinguish between genetic composition and organismal appearance. To address the need, he proposed the term "genotype" for the type of genetic composition. For example, the individuals of a pure line have the same genotype. He also proposed the term "phenotype" for the type of organismal appearances such as forms, structures, sizes, colors, or anything measurable. After genotypes and phenotypes had been distinctly defined, these two terms were widely used in genetics, and finding the links between genotypes and phenotypes became important in genetic studies.

With the progress in genetics, the challenge to Darwinian evolution was also brought up in the early 20th century. The tension between these two fields was later resolved by the works done by R.A. Fisher, J. B. S. Haldane, and Sewall Wright, known

as the modern synthesis (Huxley 1942). In modern synthesis, at least two important points about genotype-phenotype relationship were incorporated into evolutionary biology. First, in order to explain Darwin's gradualism theory in the context of Mendelian genetics, Fisher developed the theory which explains a continuous phenotypic distribution can result from multiple genes which affect the same trait (Fisher 1918; Fisher 1930). Therefore, the accumulation of small changes can lead to a large phenotypic change, and mutations with large phenotypic effects are not necessarily required in evolution. Second, Wright proposed the concept of fitness landscape and described the adaptation of a population as climbing the peak with high adaptive values (Wright 1932). Given that the adaptive value is considered as a trait, the concept of fitness landscape is an early example of genotype-phenotype maps because it illustrates the relationship between different genetic compositions and adaptive values.

Consistent with the concept of fitness landscape and population genetics theory developed during the modern synthesis, the evolution of a population is thought as the change of genotypic frequencies in the population. Lewontin (1974) further formulized that as a series of four transformation laws, which include (1) how the distribution of genotypes results in the distribution of phenotypes, (2) how natural selection alters the phenotypic distribution, (3) how the altered phenotypic distribution leads to the altered genotypic distribution, and (4) how the altered genotypic distribution determines the genotypic distribution in the next generation. The importance of genotype-phenotype relationship in understanding the evolution of a population is highlighted in the first and third laws of this formulation.

In addition to geneticists and evolutionary biologists, developmental biologists also have substantial contribution to the concept of genotype-phenotype relationship by studying how genetic materials determine organismal appearances. For example, C. H. Waddington proposed that trait development is mostly canalized (Waddington 1942). Therefore, multiple genotypes could give rise to the same phenotype. Based on the concept of canalization, Pere Alberch proposed that, in a genotype-phenotype map, the mapping function must have non-linearity and thresholds because the genotypic distribution is relatively continuous but phenotypic distribution is relatively discontinuous (Alberch 1980; Alberch 1991). This is an early example of quantitative statement on the mapping function between genotypes and phenotypes.

With the development of molecular technology, more genetic bases of complex traits were revealed, and their polygenic feature was supported. When considering the adaptive evolution of complex traits, Gunter Wagner and Lee Altenberg (1996) mentioned that considering their genotype-phenotype maps is fundamental. More importantly, they introduced the modular representation of genotype-phenotype maps. Its simplest version is a bipartite network which includes a set of traits, a set of genes, and the links between genes and traits for genic effect sizes on traits. Moreover, if some traits are mostly affected by the same set of genes, these genes and traits form a module in the map. Such bipartite representation is now widely used in this field (e.g. Wang et al. (2010) or Landry and Rifkin (2012)). Unless otherwise noted, the genotype-phenotype maps described in this dissertation are in this representation.

With the more recent progress of technology, more genotypes and phenotypes have been collected in a high-throughtput manner. Rather than organismal appearances,

the macromolecular traits such as transcript abundances, protein abundances, protein-protein interactions, metabolic fluxes, metabolites concentrations are more focused. These macromolecular traits constitute a hierarchical structure, in which each layer is a collection of traits in the same kind (Civelek and Lusis 2014; Ritchie et al. 2015). Focusing on the elements and their interaction in a hierarchical structure, the study of systems biology aims to synthesize holistic understanding of organisms. Such systems view also impact the study of evolutionary biology, stimulating the emergence of evolutionary systems biology. Beyond how the genotypic variations lead to phenotypic variations in a hierarchical structure of macromolecular traits, evolutionary systems biologists emphasize the accompanying fitness variations and evolutionary outcomes (Loewe 2009; Papp et al. 2011; Soyer and O'Malley 2013). By such approach, evolutionary mechanisms and principles are more comprehensively studied, evolutionary theories becomes more testable, and evolutionary outcomes could are more predictable.

Importance of genotype-phenotype maps in medicine

While the concept of genotype-phenotype map is important in genetics, developmental biology, and evolutionary biology, it is also of importance in other biological fields such as medicine. For example, if one focuses on the traits related to human diseases, studying such genotype-phenotype maps will help us understand the genetic basis and possible molecular mechanisms of the human diseases. In fact, the genic effects of a large number of Mendelian diseases have been mapped (Welter et al. 2014; Bush et al. 2016).

Moreover, if one considers the growth rate of cancer cells as a trait, studying such genotype-phenotype map will help us understand cancer biology and develop cancer

treatment (Yi et al. 2017). For example, if the mutations inhibiting the growth are found, the genes related to these mutations should be the targets for drug design or gene therapy. Moreover, a comprehensive genotype-phenotype map of cancer growth rate is desired because multiple target genes may be needed to avoid the evolution of drug resistance (Hu and Zhang 2016).

Recently, scientists have made progress in the field of precision medicine. For this purpose, it is important to construct detailed genotype-phenotype maps in the scale of single nucleotide polymorphisms (Ashley 2016; Hall et al. 2016). For example, in order to provide the most suitable therapy for each patient, one should identify the different causal genotypes for the disease. Moreover, one should study whether individuals with different genotypes have different outcomes when receiving the same treatment. With more comprehensive genotype-phenotype maps of many diseases and treatment outcomes, the field of precision medicine will be largely improved.

Parameters and properties in genotype-phenotype maps

In the modular representation, three basic parameters are necessary and sufficient to describe a genotype-phenotype map: a set of genes, a set of traits, and links representing genic effects on traits. Nevertheless, emergent properties such as robustness, pleiotropy, and modularity could be also summarized from a map. In this section, for each basic parameter and emergent property, I will first introduce the detailed definition and then discuss the hypotheses related to the evolution and origination.

The first basic parameter of a genotype-phenotype map is a set of genes. The number of genes existed in a genome could evolve, and the determining factors could be mutation, drift, selection, gene duplication, horizontal gene transfer, or their

combinations (Lynch 2007). In practice, however, the number of genes in a genotype-phenotype map is largely affected by the methodology. For example, null mutations in essential genes are often missing in the map because the phenotypic effects are often not measurable in dead individuals (with exceptions such as embryonic lethal phenotypes). In addition, forward genetic methods can only study the variations existed in the population, and the variations in a low-frequency suffer a low statistical power. To increase the number of genes in a map, one needs to improve the statistical power such as by increasing the sample size or to use a different approach such as reverse genetics.

The second basic parameter of a genotype-phenotype map is a set of traits. Given that anything measurable in an organism could be defined as a trait, it is impossible to exhaust the trait space. Therefore, the choice of traits included in a genotype-phenotype map largely depends on researchers' interests. As a result, the conclusion drew from a meta-analysis of literature may be biased. For example, when studying the overall strength of adaptation compared with neutrality in evolution, one may conclude that adaptation has a stronger effect by finding more adaptive cases reported in the literature than neutral cases. However, this conclusion is potentially biased due to the overall higher interests in reporting cases with adaptive evolution rather than neutral evolution. Owing to the recent technical improvement of high-throughput phenotyping, such bias may be avoided with more subjectivity of trait choices. Moreover, as previously mentioned, traits are located in different layers of a hierarchical structure. Different layers could have different contributions in evolution. For example, the fitness by itself is a trait at the highest level with direct exposure to natural selection while other traits such as morphological traits or gene expression traits are at lower levels with possibly lower

importance to natural selection. Therefore, sampling the traits on a different layer could verify the generality of the findings based on traits on one layer.

After both a gene set and a trait set are determined, the left parameters of a genotype-phenotype map are the genic effect sizes on each trait. Effect sizes are certainly evolvable. For example, mutational severity is divergent between multiple closely related groups (Liao and Zhang 2008; Dowell et al. 2010; Kim et al. 2010; Vu et al. 2015). There are at least three categories of mechanisms for changes of effect sizes. The first category consists of environmental reasons. For example, the function performed by a gene in the environment of one group becomes not useful anymore in a different environment of another group. Therefore, the effect size of this gene becomes zero. This scenario is supported by the existence of environment-specific expressed genes because these genes probably only useful in some environments. The second category consists of functional changes of focal genes. For example, when a null mutation of a gene is fixed in a population, the effect size of this copy of gene becomes zero. Besides the changes of focal genes, the changes of other genes make up the third category. For example, some genes are assumed to be capable to buffer genic effects of other genes and called modifier genes. Compared with the genetic background without modifier genes, the genic effects in the genetic background with modifier genes are largely reduced.

One of the evolutionary explanations for the third category above is the emergence of genetic robustness, which is defined as the ability of organisms to buffer genetic variations. This concept was first proposed by Waddington (1942) and was called canalization. Analogous to the water from different upstream branches which still flow into the same mainstream, individuals with genotypic variations could still have the same

phenotype after development. This argument is based on the observations that cell fates of embryonic development are usually tightly regulated as well as the phenotypically wild-type individuals are pervasive. The emergence of genetic robustness, according to Waddington, is due to the natural selection for buffering effects. Therefore, if a trait is more important to fitness, one will expect that trait is buffered more. As shown in a theoretical study, this trend is generally true except for traits with extremely high importance (Wagner et al. 1997). This prediction is important in testing the generality of adaptive hypothesis of genetic robustness.

Other than the adaptive hypothesis proposed by Waddington, there are two alternative hypotheses (de Visser et al. 2003). First, genetic robustness could be an intrinsic property of a biological system. Second, genetic robustness may arise as a byproduct of other selections, such as selection for robustness to environmental perturbations. Because environmental perturbations tend to decrease organismal fitness, robustness to such variations would be selectively favored. This hypothesis is commonly referred to as the congruence hypothesis. The congruence hypothesis and the intrinsic hypothesis, together with the adaptive hypothesis, need to be tested in understanding the evolution of genetic robustness.

Two kinds of molecular mechanisms for genetic robustness have been proposed: duplication of genetic materials and heatshock proteins. However, the identification of these mechanisms does not directly support the adaptive hypothesis of genetic robustness. In the first case, it is apparent that the effect size of the focal gene could be reduced when one more copy exists in the genome. Empirical studies of gene expression also demonstrate the buffering effects from repeated *cis*-regulatory elements (Frankel et al.

2010; Crocker et al. 2015). However, two major problems arise when the fitness benefits are considered. First, if two copies of genes are exactly the same, and one copy is sufficient for the wild-type's function, one of the two copies is subject to degeneration and thus not able to long persist in the genome. Second, increasing the copy number will also increase the size of mutational targets and get fitness costs. Therefore, these duplications may not be fixed in population purely by the benefits of buffering ability. In the second case, while heatshock proteins are found to be able to buffer phenotypic variations (Rutherford and Lindquist 1998; Fares et al. 2002), it is also possible that the evolution of heatshock proteins largely depends on their abilities to buffer environmental perturbations instead of genetic perturbations. Therefore, more studies are required for examining whether the genetic robustness is largely favored by natural selection.

If we summarize the number of nonzero effect sizes per gene, this number will be one measurement of gene pleiotropy. Given that the effect sizes are evolvable, the pleiotropy must be also evolvable. For example, with the evolution of genic robustness, the effect size on one trait for one gene could disappear. As a result, the number of traits affected by that gene decreases, and the pleiotropy of that gene also decreases.

If there were no pleiotropy, one gene could only affect at most one trait, and therefore no complicated structure in a genotype-phenotype map would be expected. With pleiotropy, the various distributions of the effect sizes across different genes and across different traits make genotype-phenotype maps look different. If the effect sizes are clustered within modules but diminished between modules, we will say this map has high modularity. Several hypotheses about the evolution of high modularity have been proposed but not fully tested, including the neutral process of duplication-elimination, a

byproduct of genetic robustness removing inter-modules genic effects, a byproduct of periodic natural selection favoring various modules, or having direct benefits (Wagner et al. 2007).

Consequences of genotype-phenotype maps in adaptive evolution

One of the important goals in evolutionary biology is to understand and predict the occurrence of adaptive evolution. To achieve this goal, it is desired to identify the factors which constrain or facilitate adaptive evolution. Several emergent properties of a genotype-phenotype map have been proposed to impact adaptive evolution. These impacts will be discussed below.

First of all, whether genetic robustness constrains or facilitates adaptive evolution is under debates. On the one hand, genetic robustness reduces the phenotypic effects of mutations, so it seems to constrain adaptive evolution. On the other hand, it has been argued that genetic robustness facilitates adaptation because it allows more mutations to accumulate within a population because the fitness costs of mutations are neutralized. When the environment changes, the genotype-phenotype map could change, and some of standing genetic variations could have beneficial effects in the new environment. Therefore, the phenotypic space could be explored more quickly (Wagner 2008; Masel and Trotter 2010). While the facilitating role of genetic robustness was shown by simulation (Draghi et al. 2010), the empirical support is only found in a few systems such as RNA secondary structures and ribozymes (Wagner 2012).

In addition to the effect of genetic robustness and standing genetic variations, phenotypic plasticity, defined as the ability of individuals to show a different phenotype without the change of genotype, might also impact adaptive evolution. Note that

phenotypic plasticity could be considered as the contrary to environmental robustness, defined as the ability of individuals to buffer environmental variations and show the same phenotype. It has been argued that environmental robustness and genetic robustness are congruent by sharing the same molecular mechanisms (Meiklejohn and Hartl 2002). Therefore, phenotypic plasticity is also related to genotype-phenotype maps.

Several different reasons have been proposed for why phenotypic plasticity could facilitate adaptive evolution. First, at the initial stage of an environmental change, if there is no phenotypic plasticity, the founder population may not survive the harshness of new environment, and thus no adaptation could happen. This argument is known as Baldwin effect (Baldwin 1896). Second, plastic changes may serve as the steppingstones for the following steps of adaptation such as genetic assimilation (Waddington 1953). This thought has been used to argue the importance of phenotypic plasticity in adaptation by extended evolutionary synthesis (Laland et al. 2014; Laland et al. 2015), but the underlying mechanisms remain unclear. Third, even when plastic responses are nonadaptive, they can still increase the rate of following adaptation (Ghalambor et al. 2007; Ghalambor et al. 2015). This is likely to be true because nonadaptive plasticity make the starting point in the fitness landscape lower and thus make the selection gradient stronger. However, more studies are required in order to know how general these arguments are.

Pleiotropy could also affect the rate of adaptive evolution. Due to pleiotropy, genes with beneficial effects for some traits could have deleterious effects on some other traits. Therefore, the mutations with net beneficial effects may be limited, and thus the adaptive evolution could be largely constrained. This idea is known as “the cost of

complexity," which argues that a more complex organism could be less evolvable during environmental changes (Fisher 1930; Orr 2000).

Two possible features in the genotype-phenotype maps may alleviate the cost of complexity. First, if pleiotropy scales up the effect sizes, mutations with high pleiotropy could more likely to become overall beneficial and fixable. This possibility has been demonstrated in the case of yeast morphological traits (Wang et al. 2010). Second, considering the modularity in genotype-phenotype maps, if the effect sizes of traits are organized as how correlated adaptation among traits frequently happen, the efficiency of adaptation could be indeed improved (Wagner and Zhang 2011; McGee et al. 2016). However, the generality of these two scenarios remains largely unknown.

Methodology of constructing genotype-phenotype maps

In order to construct a genotype-phenotype map, one needs to quantify the effect size on a set of traits for a set of genes. Traditionally, many maps in model organisms are constructed by forward genetic methods such as yeast (Ehrenreich et al. 2010), fruit flies, mice and human (Flint and Mackay 2009). Forward genetic methods include both linkage mapping or association mapping (Mackay et al. 2009). Both mapping methods look for the statistical association between variations of genetic markers and variations of phenotypes across individuals, assuming that the association for farther genetic markers will be broken by recombination. In linkage mapping, F2 individuals from true-breeding parents are studied. The candidate regions are usually larger due to larger haplotype blocks. Instead, in association mapping, individuals with unknown mating history are studied. While it could end up with smaller candidate regions, the prevalence of rare alleles increases the demand of sampling a larger population.

Reverse genetic methods, contrast to forward genetic methods, can be also used to construct genotype-phenotype maps. Some problems in forward genetics such as missing heritability (Maher 2008; Manolio et al. 2009) are less severe in reverse genetic methods. However, the critical step for reverse genetic methods is to acquire collections of systematically generated mutants, which was only done in model organisms in the past. Nowadays with the development of CRISPR techniques, the speed of making deletion collections is much faster (Sander and Joung 2014). For example, essential genes in various human cell lines have been detected in this manner (Blomen et al. 2015; Hart et al. 2015; Wang et al. 2015).

Besides the experimental methods, with the development of systems biology, more models and methods are available for predicting the genic effects of traits. The most successful one is perhaps the constraint-based metabolic network analysis (Lewis et al. 2012; Bordbar et al. 2014). With the constructed metabolic network and proper choice of objective function, one can predict how the metabolic network behaves. For example, in flux balance analysis, the objective function is to maximize the biomass output, and therefore how the network behaves after the full adaptation could be predicted (Orth et al. 2010). In addition, in minimization of metabolic adjustment (abbreviated by MOMA), the objective function is to minimize the flux adjustment from the original fluxes, and how the network promptly responds to perturbations could be predicted (Segre et al. 2002). A large number of metabolic network models in different organisms have been available in public (King et al. 2016). However, the prediction is not completely free of errors. Because evaluating the accuracy of prediction requires the comparison to empirical data, many models have not been verified except for those in model organisms such as *E. coli*.

In the future, after more empirical are available, and after sufficient verification of prediction is performed, these model prediction methods will substantially impact the field because of their ability in quickly generating a gigantic number of genotype-phenotype maps.

In my dissertation work, there is a combination of using genotype-phenotype maps constructed by reverse genetic methods and model prediction methods. The former examples include yeast morphological traits and gene expression traits, while the latter example is *E. coli* metabolic flux traits.

Preview of research works

In my dissertation, five chapters of researches related to genotype-phenotype maps and evolutionary biology are included. The questions I addressed in these chapters fall into three different categories: (1) summarizing the properties in genotype-phenotype maps, (2) studying how these properties originated, and (3) clarifying the evolutionary consequence of these properties. Below are more details on each of these five chapters.

In chapter 2, because the general pattern of parameters in genotype-phenotype maps is largely unknown, I measured the number of genes affecting a trait as well as the distribution of effect sizes for each trait using yeast morphological traits and yeast gene expression traits. To further explore the role of natural selection in the evolution of effect sizes, I examined the adaptive hypothesis of the emergence of genetic robustness, which reduces effect sizes.

In chapter 3, the adaptive hypothesis of the emergence of genetic robustness was further tested using *E. coli* metabolic reaction fluxes as traits. Because the details in

metabolic networks are given, I also identified the candidate metabolic reactions which may contribute to genetic robustness.

In chapter 4, noticing the potential bias of the previous curation study, I varified whether the phenotypic evolution is generally adaptive using yeast morphological traits and yeast expression level traits. The methodology relies on the measurement of trait importance and effect sizes, where the information is from the genotype-phenotype maps.

In chapter 5, I tested whether the pleiotropy-caused mutation correlation tends to constrain or facilitate phenotypic evolution using yeast morphological traits. The measurement of mutation correlation also relies on the information on the genotype-phenotype map.

In chapter 6, because whether phenotypic plasticity generally serves as steppingstones for genetic adaptations is under debates, I examined whether the genetic changes of adaptation tend to reverse or reinforce the initial plastic changes using expression level traits from various species and metabolic flux traits from *E. coli*. The theoretical explanation for the generality of observation is also provided.

References

Alberch P. 1980. Ontogenesis and Morphological Diversification. *Am Zool* **20**: 653-667.

Alberch P. 1991. From Genes to Phenotype - Dynamic-Systems and Evolvability. *Genetica* **84**: 5-11.

Ashley EA. 2016. Towards precision medicine. *Nat Rev Genet* **17**: 507-522.

Baldwin JM. 1896. A new factor in evolution. *Am Nat* **30**: 441-451.

Blomen VA, Majek P, Jae LT, Bigenzahn JW, Nieuwenhuis J, Staring J, Sacco R, van Diemen FR, Olk N, Stukalov A et al. 2015. Gene essentiality and synthetic lethality in haploid human cells. *Science* **350**: 1092-1096.

Bordbar A, Monk JM, King ZA, Palsson BO. 2014. Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet* **15**: 107-120.

Bush WS, Oetjens MT, Crawford DC. 2016. Unravelling the human genome-phenome relationship using phenome-wide association studies. *Nat Rev Genet* **17**: 129-145.

Civelek M, Lusis AJ. 2014. Systems genetics approaches to understand complex traits. *Nat Rev Genet* **15**: 34-48.

Crocker J, Abe N, Rinaldi L, McGregor AP, Frankel N, Wang S, Alsawadi A, Valenti P, Plaza S, Payre F et al. 2015. Low Affinity Binding Site Clusters Confer Hox Specificity and Regulatory Robustness. *Cell* **160**: 191-203.

de Visser JAGM, Hermisson J, Wagner GP, Meyers LA, Bagheri HC, Blanchard JL, Chao L, Cheverud JM, Elena SF, Fontana W et al. 2003. Perspective: Evolution and detection of genetic robustness. *Evolution* **57**: 1959-1972.

Dowell RD, Ryan O, Jansen A, Cheung D, Agarwala S, Danford T, Bernstein DA, Rolfe PA, Heisler LE, Chin B et al. 2010. Genotype to Phenotype: A Complex Problem. *Science* **328**: 469-469.

Draghi JA, Parsons TL, Wagner GP, Plotkin JB. 2010. Mutational robustness can facilitate adaptation. *Nature* **463**: 353-355.

Ehrenreich IM, Torabi N, Jia Y, Kent J, Martis S, Shapiro JA, Gresham D, Caudy AA, Kruglyak L. 2010. Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* **464**: 1039-1042.

Fares MA, Ruiz-Gonzalez MX, Moya A, Elena SF, Barrio E. 2002. Endosymbiotic bacteria - GroEL buffers against deleterious mutations. *Nature* **417**: 398-398.

Fisher RA. 1918. The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Philosophical Transactions of the Royal Society of Edinburgh* **52**: 399-433.

Fisher RA. 1930. *The Genetical Theory of Natural Selection*. Oxford, Clarendon.

Flint J, Mackay TFC. 2009. Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Research* **19**: 723-733.

Frankel N, Davis GK, Vargas D, Wang S, Payre F, Stern DL. 2010. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466**: 490-U498.

Ghalambor CK, Hoke KL, Ruell EW, Fischer EK, Reznick DN, Hughes KA. 2015. Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature* **525**: 372-375.

Ghalambor CK, McKay JK, Carroll SP, Reznick DN. 2007. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Funct Ecol* **21**: 394-407.
Hall MA, Moore JH, Ritchie MD. 2016. Embracing Complex Associations in Common Traits: Critical Considerations for Precision Medicine. *Trends Genet* **32**: 470-484.
Hart T, Chandrashekhar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, Mis M, Zimmermann M, Fradet-Turcotte A, Sun S et al. 2015. High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* **163**.
Hu XD, Zhang ZM. 2016. Understanding the Genetic Mechanisms of Cancer Drug Resistance Using Genomic Approaches. *Trends Genet* **32**: 127-137.
Huxley J. 1942. *Evolution: The Modern Synthesis*. George Alien & Unwin Ltd., London.
Johannsen W. 1911. The genotype conception of heredity. *Am Nat* **45**: 129-159.
Kim DU, Hayles J, Kim D, Wood V, Park HO, Won M, Yoo HS, Duhig T, Nam M, Palmer G et al. 2010. Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe (vol 28, pg 617, 2010). *Nat Biotechnol* **28**: 1308-1308.
King ZA, Lu J, Drager A, Miller P, Federowicz S, Lerman JA, Ebrahim A, Palsson BO, Lewis NE. 2016. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res* **44**: D515-D522.
Laland K, Uller T, Feldman M, Sterelny K, Muller GB, Moczek A, Jablonka E, Odling-Smee J. 2014. Does evolutionary theory need a rethink? - POINT Yes, urgently. *Nature* **514**: 161-164.
Laland KN, Uller T, Feldman MW, Sterelny K, Muller GB, Moczek A, Jablonka E, Odling-Smee J. 2015. The extended evolutionary synthesis: its structure, assumptions and predictions. *Proc Biol Sci* **282**: 20151019.
Landry CR, Rifkin SA. 2012. The Genotype–Phenotype Maps of Systems Biology and Quantitative Genetics: Distinct and Complementary. In *Evolutionary Systems Biology*, (ed. OS Soyer). Springer.
Lewis NE, Nagarajan H, Palsson BO. 2012. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol* **10**: 291-305.
Lewontin RC. 1974. *The genetic basis of evolutionary change*. Columbia University Press, New York.
Liao BY, Zhang JZ. 2008. Null mutations in human and mouse orthologs frequently result in different phenotypes. *P Natl Acad Sci USA* **105**: 6987-6992.
Loewe L. 2009. A framework for evolutionary systems biology. *Bmc Syst Biol* **3**.
Lynch M. 2007. *The origins of genome architecture*. Sinauer, Sunderland, MA.
Mackay TF, Stone EA, Ayroles JF. 2009. The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* **10**: 565-577.
Maher B. 2008. Personal genomes: The case of the missing heritability. *Nature* **456**: 18-21.
Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A et al. 2009. Finding the missing heritability of complex diseases. *Nature* **461**: 747-753.
Masel J, Trotter MV. 2010. Robustness and Evolvability. *Trends Genet* **26**: 406-414.

McGee LW, Sackman AM, Morrison AJ, Pierce J, Anisman J, Rokyta DR. 2016. Synergistic Pleiotropy Overrides the Costs of Complexity in Viral Adaptation. *Genetics* **202**: 285-295.
Meiklejohn CD, Hartl DL. 2002. A single mode of canalization. *Trends Ecol Evol* **17**: 468-473.
Mendel G. 1901. Experiments in Plant Hybridization. *Journal of the Royal Horticultural Society* **26**: 1-32.
Orr HA. 2000. Adaptation and the cost of complexity. *Evolution* **54**: 13-20.
Orth JD, Thiele I, Palsson BO. 2010. What is flux balance analysis? *Nat Biotechnol* **28**: 245-248.
Papp B, Notebaart RA, Pal C. 2011. Systems-biology approaches for predicting genomic evolution. *Nat Rev Genet* **12**: 591-602.
Ritchie MD, Holzinger ER, Li RW, Pendergrass SA, Kim D. 2015. Methods of integrating data to uncover genotype-phenotype interactions. *Nat Rev Genet* **16**: 85-97.
Rutherford SL, Lindquist S. 1998. Hsp90 as a capacitor for morphological evolution. *Nature* **396**: 336-342.
Sander JD, Joung JK. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* **32**: 347-355.
Segre D, Vitkup D, Church GM. 2002. Analysis of optimality in natural and perturbed metabolic networks. *P Natl Acad Sci USA* **99**: 15112-15117.
Soyer OS, O'Malley MA. 2013. Evolutionary systems biology: What it is and why it matters. *Bioessays* **35**: 696-705.
Vu V, Verster AJ, Schertzberg M, Chuluunbaatar T, Spensley M, Pajkic D, Hart GT, Moffat J, Fraser AG. 2015. Natural Variation in Gene Expression Modulates the Severity of Mutant Phenotypes. *Cell* **162**: 391-402.
Waddington CH. 1942. Canalization of development and the inheritance of acquired characters. *Nature* **150**: 563-565.
Waddington CH. 1953. Genetic Assimilation of an Acquired Character. *Evolution* **7**: 118-126.
Wagner A. 2008. Robustness and evolvability: a paradox resolved. *P R Soc B* **275**: 91-100.
Wagner A. 2012. The role of robustness in phenotypic adaptation and innovation. *P R Soc B* **279**: 1249-1258.
Wagner GP, Altenberg L. 1996. Perspective: Complex adaptations and the evolution of evolvability. *Evolution* **50**: 967-976.
Wagner GP, Booth G, Bagheri HC. 1997. A population genetic theory of canalization. *Evolution* **51**: 329-347.
Wagner GP, Pavlicev M, Cheverud JM. 2007. The road to modularity. *Nat Rev Genet* **8**: 921-931.
Wagner GP, Zhang JZ. 2011. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. *Nat Rev Genet* **12**: 204-213.
Wang T, Birsoy K, Hughes NW, Krupczak KM, Post Y, Wei JJ, Lander ES, Sabatini DM. 2015. Identification and characterization of essential genes in the human genome. *Science* **350**: 1096-1101.

Wang Z, Liao BY, Zhang JZ. 2010. Genomic patterns of pleiotropy and the evolution of complexity. *P Natl Acad Sci USA* **107**: 18034-18039.

Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L et al. 2014. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* **42**: D1001-D1006.

Wright S. 1932. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In *Proceedings of the Sixth International Congress on Genetics* pp. 355-366.

Yi S, Lin S, Li Y, Zhao W, Mills GB, Sahni N. 2017. Functional variomics and network perturbation: connecting genotype to phenotype in cancer. *Nat Rev Genet* doi:10.1038/nrg.2017.8.

Chapter 2

The Genotype-Phenotype Map of Yeast Complex Traits: Basic Parameters and the Role of Natural Selection[1]

2.1 ABSTRACT

Most phenotypic traits are controlled by many genes, but a global picture of the genotype-phenotype map (GPM) is lacking. For example, in no species do we know generally how many genes affect a trait and how large these effects are. It is also unclear to what extent GPMs are shaped by natural selection. Here we address these fundamental questions using the reverse genetic data of 220 morphological traits in 4718 budding yeast strains, each of which lacks a nonessential gene. We show that (1) the proportion of genes affecting a trait varies from <1% to >30%, averaging 6%, (2) most traits are impacted by many more small-effect genes than large-effect genes, and (3) the mean effect of all nonessential genes on a trait decreases precipitously as the estimated importance of the trait to fitness increases. An analysis of 3116 yeast gene expression traits in 754 gene-deletion strains reveals a similar pattern. These findings illustrate the power of genome-wide reverse genetics in genotype-phenotype mapping, uncover an enormous range of genetic complexity of phenotypic traits, and suggest that the GPM of cellular organisms has been shaped by natural selection for mutational robustness.

[1]This chapter was published as Ho and Zhang (2014) *Mol. Biol. Evol.* 31: 1568-1580.

2.2 INTRODUCTION

Describing, understanding, and utilizing the relationship between genotypes and phenotypes, or the genotype-phenotype map (GPM), are major goals of genetics (Wagner and Zhang 2011). Because most phenotypic traits, including those relevant to human diseases, are controlled by multiple genes (Falconer and Mackay 1996) and because most genes affect more than one trait (Wang et al. 2010), the GPM is a dense bipartite network of genes and traits, where an edge between a gene and a trait indicates that the gene affects the trait, with the width of the edge representing the effect size (Wang et al. 2010). Traditionally, the GPM is constructed by forward genetics, which uses linkage or association studies to identify the genetic variants underlying particular phenotypic variations among individuals of the same species (Mackay et al. 2009). Due to the limited power and efficiency of such analyses, the GPMs of human and most model organisms remain highly incomplete and uninformative (Mackay et al. 2009; Manolio et al. 2009). For example, a recent large-scale linkage analysis estimated the number of genes affecting each of 18 yeast traits (Ehrenreich et al. 2010). However, because only two strains were compared in the study, only those genetic variants that cause the phenotypic differences between these two strains were revealed. Consequently, neither the distribution of the number of genes that could affect a trait nor the distribution of the effect sizes of these genes on a trait is known. Estimating these fundamental parameters of the GPM is of vital importance, because they impact how variable a particular trait is in a population, determine the best strategy to identify the underlying genetic variants of phenotypic variations, and predict how robust and adaptable a population is to environmental challenges.

In contrast to forward genetics, reverse genetics identifies phenotypic differences among individuals of known genetic differences. If empowered by high-throughput phenotyping of

systematically generated mutants, reverse genetics can be an effective approach to the GPM. For example, the fraction of genes that affect each of 12 physiological and behavioral traits in the mouse *Mus musculus* has been estimated using 250 gene-knockout lines (Flint and Mackay 2009) (**Fig. 2.1a**). Similar estimates have been made for eight morphological, physiological, and behavioral traits in the fruit fly *Drosophila melanogaster* based on P-element insertion mutagenesis (Mackay 2010) (**Fig. 2.1b**). These and a few other studies (Winzeler et al. 1999; Ramani et al. 2012) showed that the proportion of genes impacting a trait can reach 10-40% of all genes in a genome. But how general these results are is unclear because the numbers of traits and species examined are small. Regarding the size distribution of the genic effects on a trait, two competing hypotheses exist. Mather's infinitesimal model (Mather 1941; Mackay 2001) asserts that numerous loci have small and similar effects, while Robertson (1967) posits that the distribution is approximately exponential, with a few large-effect and many small-effect loci. The effect size distributions of P-element insertions on the abdominal and sternopleural bristle numbers in *Drosophila* support Robertson's model (Lyman et al. 1996), but the generality of this conclusion is unknown. Although forward genetic studies from several species also support Robertson's model, definitive conclusions are hindered by the inherent biases and limitations of the method (Mackay 2001).

A deeper question about the GPM is why it looks the way it does. In principle, the GPM can evolve under mutation, drift, and selection, but the relative contributions of these forces are elusive. Waddington and others proposed that the GPM has been shaped by natural selection for mutational robustness, resulting in genetic canalization (Waddington 1942; de Visser et al. 2003). Similarly, natural selection may have led to organismal robustness to environmental perturbations, or environmental canalization. These two forms of canalization, if true, would

explain the surprising tolerance of living organisms to genetic and environmental disturbances, which are quite common in nature (Scharloo 1991; Flatt 2005; Wagner 2005b; Alon 2007). They also impact how adaptable and evolvable a population is in the face of mutations and environmental changes (Gibson and Wagner 2000; de Visser et al. 2003; Wagner 2005b; Draghi et al. 2010). While selection for environmental robustness is commonly agreed upon (Gibson and Wagner 2000), direct selection for genetic robustness is controversial (Gibson and Wagner 2000; de Visser et al. 2003) except when the deleterious mutation rate is exceedingly high and/or population size is huge (e.g., in viruses) (Wilke et al. 2001; Ciliberti et al. 2007; Sanjuan et al. 2007), because for cellular organisms such selection is expected to be weak (Wagner et al. 1997; Gibson and Wagner 2000) and previous tests with relatively small data yielded ambiguous results (Stearns and Kawecki 1994; Stearns et al. 1995; Houle 1998; Gibson and Wagner 2000; de Visser et al. 2003; Proulx et al. 2007). Apparently, larger and better data are needed to evaluate it critically.

To address these fundamental questions on the basic parameters of the GPM and the role of natural selection in shaping the GPM, we use the budding yeast *Saccharomyces cerevisiae*, in which 220 morphological traits have been quantitatively measured by analyzing fluorescent microscopic images of triple-stained cells of the wild-type strain and 4718 mutant strains that each of which lacks a nonessential gene (Ohya et al. 2005). The generality of the findings from the morphological traits is then verified by analyzing 3116 gene expression traits in the wild-type and 754 gene-deletion strains of *S. cerevisiae*.

2.3 RESULTS

2.3.1 Fraction of genes affecting a morphological trait

In the yeast phenotyping experiment, 220 morphological traits were measured in multiple wild-type cells from each of 123 replicate populations (Ohya et al. 2005). In addition, for each of the 4718 mutant strains, multiple isogenic cells from one population were measured for the 220 traits (Ohya et al. 2005). To determine if deleting a gene affects a trait, we used the Mann-Whitney *U* test to compare the trait values of multiple cells of the gene-deletion strain and those of the wild-type from an arbitrary replicate population. A gene deletion is tentatively considered to affect the trait if the *p*-value is lower than 0.05. The distribution of the fraction of genes affecting a trait (f_{mt}) is shown in **Fig. A.1.1a**. The mean and median of f_{mt} are 0.37 and 0.38, respectively. To remove the confounding factor of potential environmental differences between the mutant and wild-type strains in the experiment and to control for multiple testing, for each trait, we estimated the fraction (f_{wt}) of the other 122 wild-type populations in which the trait value differs significantly from that of the arbitrary wild-type population used. We found f_{wt} to be substantial (**Fig. A.1.1b**). Subtracting f_{wt} from f_{mt}, we obtained f_{genes}, the true fraction of genes that, when deleted, significantly impact the trait. We found that f_{genes} varies greatly among traits (**Fig. 2.1c**) and that this variation significantly exceeds the random expectation under homogenous f_{genes} ($p < 0.01$; permutation test). Specifically, 37.7%, 39.6%, 20.0%, and 2.7% of traits are each affected by <1%, 1% to 10%, 10% to 30%, and >30% of all nonessential genes in the yeast genome, respectively. The mean and median of f_{genes} are 0.06 and 0.04, respectively (**Fig. 2.1c**). These results remain similar regardless of the *p*-value cutoff used (**Fig. A.1.1c-d**). Use of another arbitrary replicate population of the wild-type strain yielded similar results. For each trait, we also estimated f_{genes} by examining whether the mean trait value of a mutant would be an outlier in the distribution of the 123 means of the wild-type replicate populations, but the results were similar (**Fig. A.1.1e**). Because some of the 220 traits are highly correlated, we

removed those traits whose genetic correlation coefficients exceed 0.7, resulting in a dataset with 54 traits. But, the mean and median values of f_{genes} are virtually unchanged (**Fig. A.1.2**).

Whether a genic effect is detectable depends on the statistical power of the experiment, which is determined by the precision of the phenotypic measurement, the sample size, and the constancy of the environment in which different strains are phenotyped. In the present case, the number of cells measured varied among traits and strains. On average, 91 and 95 cells were measured in the wild-type (per replicate population) and deletion strains, respectively. The effect of environmental variation is clearly seen in the 123 wild-type populations, because the standard deviation of the mean phenotypic value among the 123 populations is on average 2.48 times the mean strand error calculated from individual populations (**Fig. A.1.3**). This observation suggests that environmental fluctuation rather than sample size or measurement error is the dominant factor limiting the detection of genic effects in the present study.

2.3.2 Mean effect size of gene deletion on a morphological trait

We define the raw effect size (*ES*) of deleting a gene on a trait as the difference between the mean trait value of the deletion strain and the average of the mean trait values of the 123 replicate populations of the wild-type strain, divided by the average of the mean trait values of the 123 populations of the wild-type. The cumulative probability distribution of |*ES*|, or the absolute value of *ES*, of 4718 genes on each of the 220 traits is depicted by a curve in **Fig. 2.2a**. This distribution shows that, in most cases, a trait is affected by more genes of small effects than those of large effects, as proposed by Robertson (1967). Considering only statistically significant genic effects does not alter this conclusion. The mean |*ES*| of all nonessential genes on a trait varies substantially among traits, with an average value of 0.098 (**Fig. 2.2b**).

Due to inevitable environmental fluctuations among populations of cells that were phenotyped, we computed net |*ES*| by subtracting from raw |*ES*| a term called pseudo |*ES*|, which is the absolute effect size expected from environmental variation and sampling error arising from a limited sample size (see Materials and Methods). The cumulative probability distribution of net |*ES*| of 4718 genes on each of the 220 traits is depicted by a curve in **Fig. 2.2c**. Again, this distribution supports Robertson's model (1967). The mean net |*ES*| of all nonessential genes on a trait also varies substantially among traits, with an average value of 0.035 (**Fig. 2.2d**).

For each trait, the phenotypic variation among isogenic cells includes variations originating from stochastic noise of the trait, random measurement error, and environmental variation. We quantified the phenotypic variation among isogenic wild-type cells by the coefficient of variation (*CV* = standard deviation / mean), including both the variation among cells in a population and the variation among replicate populations (see Materials and Methods). The mean *CV* of the 220 traits examined is 0.41 (**Fig. 2.2e**).

2.3.3 More important morphological traits are more robust to various perturbations

After describing the basic parameters of the GPM for yeast morphologies, we explore the potential role of natural selection in shaping the GPM. The hypothesis of natural selection for environmental robustness predicts that traits that are more important to organismal survival and reproduction have smaller *CV,* because natural selection for the environmental robustness of a trait intensifies with the importance of the trait. Similarly, the hypothesis of natural selection for genetic robustness predicts that, under certain conditions, traits that are more important to organismal survival and reproduction have smaller net |*ES*|, because such a GPM minimizes the deleterious effects of random mutations (Wagner et al. 1997) (see Materials and Methods). To

test these hypotheses, we define trait importance (*TI*) by 100 times the reduction in fitness caused by 1% change in the phenotypic value of the trait concerned, and estimated it using the net |*ES*| estimates and the fitness values of the gene deletion strains in the medium where the morphological data were collected (Qian et al. 2012) (see Materials and Methods). When estimating *TI*, we used 2779 gene deletion strains whose fitness values relative to the wild-type are smaller than 1 (see Materials and Methods). As a result, 210 traits (out of 220) have $TI > 0$, 197 of which significantly exceed 0 (nominal $p < 0.05$). These 210 traits were subject to further analysis.

We found the *CV* of a trait to decrease with the rise of *TI* ($\rho = -0.692$, $p < 10^{-300}$; **Fig. 2.3a**). Because the phenotypic measurements of the wild-type and mutants were used in estimating *TI*, the correlation between *CV* and *TI* could be artifactual. To exclude this possibility, we estimated 1000 sets of pseudo *TI* values by randomly shuffling the fitness values among the gene deletion strains. In each set, negative *TI* values are ignored because they are biologically meaningless. We calculated the 1000 rank correlations between *CV* and the 1000 sets of positive pseudo *TI*. Because these rank correlations are not directly comparable due to different sample sizes, we converted the correlations (ρ) to Fisher's *z* scores by $z = 0.5 \ln [(1 + \rho)/(1 - \rho)]$. We found the true *z* score (converted from $\rho = -0.692$) to be more negative than all 1000 pseudo *z* scores ($p < 0.001$; **Fig. 2.3b**), suggesting that the negative correlation between *CV* and *TI* is genuine. Because the same phenotypic data were used in calculating the true *z* and the pseudo *z* scores, their disparity cannot be caused by measurement errors in phenotyping. Rather, it reveals smaller stochastic noise and environmental variation for more important traits, consistent with the hypothesis that natural selection has increased the phenotypic robustness of organisms to

stochastic noise (Batada and Hurst 2007; Lehner 2008; Wang and Zhang 2011) and environmental perturbation (Gibson and Wagner 2000).

We also observed a negative correlation between mean net |*ES*| across all deletion lines and *TI* (ρ= -0.793, $p < 10^{-300}$; **Fig. 2.3c**), indicating that the mean effect size of all nonessential genes on a trait decreases as the trait becomes more important, supporting the hypothesis of natural selection for mutational robustness. This result was verified by comparing the observed *z* (converted from ρ = -0.793) with 1000 pseudo *z* scores converted from the correlations between mean net |*ES*| and the 1000 sets of positive pseudo *TI* ($p < 0.001$; **Fig. 2.3d**).

Interestingly, we found no significant correlation between the *TI* of a trait and the number of genes affecting the trait (f_{genes}) (simulated $p = 0.07$; **Fig. A.1.4**). Hence, the lower mean net |*ES*| of important traits is not because there are fewer genes impacting important traits but because the individual impacts are smaller.

Because genetic robustness may be a byproduct of natural selection for environmental/stochastic robustness or vice versa (the congruence hypothesis) (Gibson and Wagner 2000; de Visser et al. 2003), it is important to examine whether the two types of robustness have independent origins. We found that the partial correlation between *CV* and *TI* after the control of mean net |*ES*| is ρ = -0.532 ($p = 1.1\times10^{-16}$), while the partial correlation between mean net |*ES*| and *TI* after the control of *CV* is ρ = -0.700 ($p = 4.7\times10^{-32}$). Hence, the environmental/stochastic robustness and genetic robustness are not entirely attributable to each other and must have their separate origins. These results were further confirmed by comparing with the random expectations from the 1000 sets of pseudo *TI* (**Fig. 2.3e**, **f**). Due to the potential difference in the fitness effects of positive and negative genic effects on a trait, we also

reanalyzed the data using positive (or negative) effects only, but found the results to be qualitatively unaltered (**Table A.1.1**; see Materials and Methods).

To confirm that the significant correlations among *CV*, mean net |*ES*|, and *TI* are not due to high genetic correlations among some traits, we used two approaches to generate less correlated traits. First, we removed highly correlated traits as was done for **Fig. A.1.2**, but the negative correlation between *CV* and *TI* and that between mean net |*ES*| and *TI* still exist, so do the partial correlations (**Fig. A.1.5**). Second, we performed a principal component analysis using the net |*ES*| matrix (see Materials and Methods). Using the principal component traits, we confirmed the negative correlation between mean net |*ES*| and *TI* (**Fig. A.1.6**).

The negative correlation between mean net |*ES*| and *TI* could mean a decrease in net |*ES*| for important traits or an increase in net |*ES*| for unimportant traits; only the former supports natural selection for genetic robustness. To distinguish between these two possibilities, we analyzed the 4718 genes separately. For each gene, we estimated the rank correlation ($\rho_{TI\text{-}|ES|}$) between the importance of a trait and the net |*ES*| of the gene on the trait among the 210 traits with estimated *TI*. We found $\rho_{TI\text{-}|ES|}$ to vary greatly among genes, although most (64.8%) genes have negative $\rho_{TI\text{-}|ES|}$ values (**Fig. 2.4a**). We halved the 210 traits into a group of less important and a group of more important traits. We then respectively calculated the mean net |*ES*| of the 20% most robust genes (i.e., with the most negative $\rho_{TI\text{-}|ES|}$ values) and 20% least robust genes (i.e., with the smallest $|\rho_{TI\text{-}|ES|}|$ values) on each group of traits. Natural selection for mutational robustness should intensify at more important traits. Thus, for the group of less important traits, we expect net |*ES*| to be similar between the least robust and most robust genes; but for the group of more important traits, net |*ES*| should be smaller for the most robust genes than for the least robust genes. An opposite pattern would be inconsistent with selection for mutational robustness.

We found that, for the less important traits, the mean net $|ES|$ is similar between the most robust and least robust genes ($p = 0.21$, Wilcoxon signed-rank test; **Fig. 2.4b**). But, for the more important traits, the mean net $|ES|$ of the most robust genes becomes significantly smaller than that of the least robust genes ($p = 3.8 \times 10^{-22}$, Wilcoxon signed-rank test; **Fig. 2.4b**). These findings demonstrate that the negative correlation between *TI* and mean net $|ES|$ is caused by the reduction of mean net $|ES|$ of a large fraction of genes on important traits, supporting natural selection for genetic robustness.

2.3.4 Fitness advantage of genetic robustness

One primary reason why natural selection for genetic robustness is controversial for cellular organisms is that its selection coefficient is expected to be small (Gibson and Wagner 2000). Below we show that the selection coefficient in the present case is large enough for the effect of natural selection to surpass that of genetic drift.

If we consider only null mutations but not other deleterious mutations, the fitness advantage (g) of a robustness modifier equals $\sum(\mu_i \Delta s_i / s_i)$, where μ_i is the null mutation rate at gene i and is on average 2.15×10^{-6} per gene per generation in yeast (see Materials and Methods), s_i and $s_i - \Delta s_i$ are the selection coefficients against the null mutation of gene i in the absence and presence of the modifier, respectively, and $\sum$ indicates summation over all genes considered (see Materials and Methods). The modifier is strongly selected for when $S = 2N_e g = 2 \times 10^7 \times 2.15 \times 10^{-6} \times \sum(\Delta s_i / s_i) = 43 \sum(\Delta s_i / s_i)$ greatly exceeds 1, where N_e is the effective population size and equals $\sim 10^7$ in yeast (Wagner 2005a). For example, $g = 2.15 \times 10^{-5}$ and $S = 430$ if the modifier buffers the null mutations of 20 genes with a mean $\Delta s_i / s_i = 0.5$.

With the above consideration, let us estimate the fitness advantage to yeast conferred by the observed genetic robustness. This advantage can be partitioned into two parts: (i) a reduction of the averaged mean effect size of all genes on all traits and (ii) a greater reduction of mean effect sizes on more important traits. Ideally, we should compare the organismal fitness resulting from the real GPM in the presence of mutation and the fitness resulting from the ancestral GPM in which the effect sizes had not been reduced by selection. This comparison, however, is infeasible, because of the difficulty in inferring ancestral effect sizes. Instead, we estimated the fitness resulting from a hypothetical GPM in which the effect sizes of a gene on various traits are randomly sampled (without replacement) from the observed net effect sizes of the gene on these traits. This procedure yields a conservative estimate of the fitness advantage of genetic robustness, because only part (ii) is estimated. Employing this approach, we created 1000 hypothetical GPMs.

We built a multivariate linear model in which the fitness values of 2779 gene deletion strains that are less fit than the wild-type are explained by the phenotypes of the 220 traits (see Materials and Methods). This model explains 45% of fitness variance among the deletion strains used. Using this model and a GPM, we can predict the fitness upon the deletion of a gene. For example, the predicted expected fitness upon the deletion of one of the 2779 nonessential genes from the real GPM is 0.9443, which is essentially identical to the experimentally determined mean fitness (differs by 4×10^{-16}) of the 2779 nonessential gene deletion strains. For the 1000 hypothetical GPMs, the fitness is predicted to drop to 0.8956, with a standard deviation of 0.0091, when a randomly picked nonessential gene is deleted. Thus, the deleterious effect of deleting an average nonessential gene from the real GPM has been reduced by an impressive fraction of $\Delta s/s$ = (0.9443-0.8956)/(1-0.8956) = 46.6%, compared with the GPMs with

randomized effect sizes. Consequently, the yeast fitness upon a gene deletion has risen by (0.9443-0.8956)/0.8956 = 5.4%. The total fitness improvement conferred by robustness to null mutations is $G = \sum(\mu_i \Delta s_i / s_i) = 2779 \times 2.15\times10^{-6} \times 46.6\% = 2.8\times10^{-3}$. Here we use G instead of g to denote the combined effect of multiple modifiers. Because deleterious mutations that do not completely abolish the function of a gene were not considered in the above calculation, the total fitness gain from mutational robustness should be greater than 2.8×10^{-3}. Of course, as deleterious mutations become less severe after the canalization, their equilibrium frequencies in the population will increase. Consequently, the mean fitness of the population will return to the previous value (Wagner et al. 1997).

2.3.5 Genetic robustness of gene expression traits

To examine if the genetic robustness observed from the morphological traits can be generalized to other traits, we turn to another large set of traits where the expression level of each of 3116 yeast genes is considered a trait. The genetic robustness of yeast expression traits was previously assessed by calculating how much the expression of each gene varies among gene deletion lines and testing if the degree of variation is correlated with the importance of the gene, but the results were mixed (Proulx et al. 2007). We expanded the analysis from considering the effects of 276 gene deletions (Proulx et al. 2007) to 754 by combining several microarray experiments performed in rich media (see Materials and Methods). We define the raw effect size (*ES*) of deleting gene 1 on the expression level of gene 2 by the difference in the expression level of gene 2 between the deletion strain and the wild-type strain, divided by the expression level of gene 2 in the wild-type. Because the expression levels were measured in several different microarray experiments, it is difficult to assess whether an effect is statistically significant. But

if we use raw $|ES|$ of 0.2, 0.3, 0.4, and 0.5 as potential cutoffs, the average proportion of gene deletions that affect the expression level of a gene is 20%, 10%, 5.5%, and 3.3%, respectively. Thus, the expression traits are of comparable complexity as the 220 morphological traits.

In principle, we should estimate the importance of gene expression traits by the approach used for estimating the *TI* of morphological traits. However, this would require direct comparison of gene expression levels across different microarray data, which is unlikely to be reliable. Instead, we followed a previous study (Proulx et al. 2007) to use the fitness effect of deleting a gene as a proxy for the *TI* of the expression trait of the gene. That is, the expression of a gene is more important if the fitness effect of deleting the gene is larger. This proxy for *TI* is reasonable because the fitness effect caused by a small expression change of a gene is highly correlated with that caused by deleting the gene (Wang and Zhang 2011). We found that, regardless of whether *TI* is measured by gene essentiality (i.e., categorical) or the fitness effect of gene deletion (i.e., continuous), there is a significant negative correlation between the importance of a trait (*TI*) and the mean absolute effect size of gene deletion on the trait ($|ES_G|$, the subscript G indicates genetic perturbation) (**Table 2.1**). Here, ES_G is defined in the same way as *ES* for morphological traits. We also found a negative correlation between *TI* and the mean absolute effect size of environmental changes ($|ES_E|$, the subscript E indicates environmental perturbation; see Materials and Methods). Similar to the results for the morphological traits, the correlation between *TI* and $|ES_G|$ remains significant after the control of $|ES_E|$, and the correlation between *TI* and $|ES_E|$ remains significant after the control of $|ES_G|$ (**Table 2.1**), suggesting that neither the genetic nor environmental robustness of gene expression is entirely caused by the other. Similar results were obtained when the wild-type gene expression level is controlled (**Table A.1.2**). We also analyzed an expanded set of environmental perturbations, and the results were similar

(**Table A.1.3**). After removing highly correlated expression traits (see Materials and Methods), we still observed qualitatively similar results for the 54 remaining traits (**Table A.1.4**).

Because essential genes tend not to have a canonical TATA box in their promoters (Han et al. 2013) and because the expression levels of TATA-less genes are less noisy, less sensitive to environmental changes, and more conserved among species than those of TATA-containing genes (Newman et al. 2006; Tirosh et al. 2006), one wonders whether the above findings are artifacts caused by covariations of both gene essentiality and expression insensitivity with the absence of TATA boxes. In other words, we may observe a negative correlation between *TI* and $|ES_E|$ and/or that between *TI* and $|ES_G|$ if certain genes must use TATA-less promoters for reasons other than environmental/genetic robustness. To exclude this possibility, we analyzed TATA-containing and TATA-less genes (Rhee and Pugh 2012) separately. We found that the hypothesis of adaptive origin of genetic robustness is supported for both TATA-containing and TATA-less genes (**Table A.1.5**). The evidence for an independent adaptive origin of environmental robustness is weakened for TATA-containing genes, but remains strong for TATA-less genes (**Table A.1.5**). Taken together, these analyses support that the signals for adaptive genetic and environmental robustness of gene expression traits are genuine.

2.4 DISCUSSION

Using genome-wide reverse genetics, we estimated the fraction of nonessential genes affecting a trait for a large number of traits for the first time in any organism. We discovered that this fraction is on average 6% for the 220 yeast morphological traits examined. An analysis of 3116 yeast gene expression traits revealed a comparable degree of genetic complexity. It is interesting to note that the fraction of genes affecting a trait is similar among yeast, fly, and

mouse ($p > 0.05$ in all pairwise comparisons; Mann-Whitney *U* test; **Fig. 2.1**), despite the rather small data from the latter two species and multiple differences in phenotyping, sample size, and type of traits examined. It is tempting to suggest that our observation from yeast, a unicellular eukaryote, may be widely applicable to other organisms, including mammals. More studies, however, are needed to verify this observation.

While the fraction of genes affecting a yeast trait appears intermediate on average (6% or ~300 nonessential genes), the among-trait variation of this quantity is huge. Nearly two fifths of traits are relatively simple, each affected by <1% of genes (i.e., ~50 genes). Two fifths of traits are of medium complexity, each affected by 1% to 10% of genes (i.e., 50-500 genes). Over one fifth of traits are highly complex, each affected by >10% of genes (i.e., >500 genes), including those impacted by >30% of genes (i.e., >1500 genes). A systems approach (Mackay et al. 2009) is not only preferred but also necessary for understanding why and how so many genes affect each of these highly complex traits. Theoretical studies are needed to understand how the discovered distribution of genetic complexity of phenotypic traits impacts phenotypic variation and evolution.

Our findings partially explain why forward genetics is inefficient in genotype-phenotype mapping. Among the large number of genes that potentially affect a complex trait, typically only a few are variable in the mapping population of each linkage analysis. In association studies, although the number of variable causal genes may be high when the mapping population is large, the statistical power is generally low because most causal genes are not highly polymorphic. If one is interested in the actual mutations causing a particular trait variation in a population, using forward genetics seems necessary. But if one is interested in the molecular genetic network responsible for and potentially impacting a trait, reverse genetics offers a more complete and

unbiased view. Although only gene deletions are considered here, genome-wide reverse genetics is applicable to other types of mutants when they become available, including gain-of-function mutants.

A common approach to verifying a candidate causal gene identified by forward genetics is to examine the phenotypic effect of deleting the gene from a wild-type strain. But, the validation can only prove that the gene affects the trait but cannot vindicate that the gene causes the trait variation seen in the mapping population. This is especially so for highly complex traits, where a randomly picked gene has a >10% chance to affect the trait. Additional tests, such as allelic replacement, will be necessary to reduce the false positive rate.

We showed that the phenotypic variation (*CV*) of a trait among isogenic wild-type individuals decreases with the rise of trait importance, consistent with the hypothesis of natural selection for environmental/stochastic robustness. We also showed that the mean effect size of gene deletion decreases as the trait becomes more important, consistent with the hypothesis of natural selection for genetic robustness. We found that the environmental/stochastic robustness and the genetic robustness cannot fully explain each other, rejecting the congruence hypothesis (de Visser et al. 2003) and suggesting separate origins of the two types of robustness. One rationale of the congruence hypothesis is that some genes underlying environmental robustness are also used for genetic robustness (Lehner 2010). A often cited example is the heat shock protein Hsp90 in *Drosophila* (Meiklejohn and Hartl 2002). But more recent work found that Hsp90 buffering of genetic perturbation is independent of environmental/stochastic robustness (Milton et al. 2003). Furthermore, mapping data from mouse, *Arabidopsis*, and yeast suggested that genetic robustness and environmental robustness are often controlled by different loci (Fraser and Schadt 2010). While genetic robustness may also originate from some intrinsic

properties of the gene interaction networks without direct selection for robustness (Siegal and Bergman 2002; Hermisson and Wagner 2004), this hypothesis cannot explain why the observed genetic robustness is greater for more important traits. Taken together, our results provide strong evidence for the action of natural selection in shaping the GPM and in improving the mutational robustness of relatively important traits in yeast.

Three reasons may explain why several earlier studies did not find clear evidence of natural selection for genetic robustness. First, natural selection for genetic robustness is expected to be weak unless the population size is large and the deleterious mutation rate is high (Wagner et al. 1997). The previous ambiguous results in fly (Stearns and Kawecki 1994; Stearns et al. 1995; Houle 1998) may reflect weaker selection for genetic robustness in fly than in yeast due to their difference in effective population size. In the light of this comparison, it is interesting to ask if genetic robustness potentially exists in humans, which have an effective population size of 10^4 and a null mutation rate of 1.5×10^{-5} per gene per generation (see Materials and Methods). We calculated that $S = 2N_e\sum(\mu_i\Delta s_i/s_i) = 0.3\sum(\Delta s_i/s_i)$ for a modifier buffering deleterious mutations in humans. Assuming an average $\Delta s_i/s_i$ of 0.5, S will exceed 1 if a modifier simultaneously affects > 6 genes. Because the total number of genes in humans is about four times that in yeast, if f_{genes} in human is not lower than that in yeast, it is possible that a modifier affects much more than six genes. Nonetheless, it is clear from this calculation that selection for genetic robustness is ~100 fold weaker in humans than in yeast. Second, trait importance is quantitatively estimated in our analysis but not in many previous studies, rendering our analysis more powerful and objective than those earlier analyses. But, it should be noted that we estimated morphological trait importance by the slope in the correlation between trait values and fitness values among 2779 gene deletion strains. As such, our estimates may not accurately

reflect causal relationships between the variation of a trait and fitness. However, it is virtually impossible to establish causal relationships between traits and fitness, because no trait is independent of all other traits such that one can manipulate a trait without affecting all other traits. The fact that our use of the inaccurate trait importance estimates still yields significant evidence supporting natural selection for genetic robustness suggests that the true signal is even stronger. In other words, our results are likely to be conservative. Note that in some earlier studies, trait importance was named or defined differently. For example, some researchers used the term "sensitivity", defined by the percentage of fitness change associated with a 1% or 10% phenotypic change (Stearns and Kawecki 1994; Stearns et al. 1995; Houle 1998), while Proulx et al. (2007) measured the "importance" of a gene expression trait by the growth defect caused by deleting the gene. Finally and probably most importantly, our data are much larger than those used in all previous studies, allowing detecting selection for genetic robustness and excluding the congruence hypothesis.

Our analysis has three caveats. First, the morphological variations of the wild-type yeast cells were measured in the same gross environment, which may underestimate *CV*, which in turn may lead to an overestimation of genetic robustness unexplainable by environmental robustness. But this criticism does not apply to the gene expression data analyzed here, because they include gross environmental variations. Although these environments do not resemble the historical natural environments of yeast, they include important environmental variables that yeast faces in nature, such as temperature, osmotic pressure, and amino acid concentrations. The overall similar findings of genetic robustness between the morphological and expression traits suggest that the lack of gross environmental variation has a minimal impact on the result of morphological traits, but this conclusion requires further confirmation.

Second, our measurement of effect size is limited to null mutations while in nature there are also abundant mutations that impact the function of a gene only slightly or moderately; their effect size would be smaller. If the effects of a genic mutation on various traits are proportionally smaller when the mutation reduces but not abolishes the gene function, all of our empirical results should still hold. Our calculation of the fitness advantage of genetic robustness is conservative, because considering additional deleterious (but not null) mutations will increase the benefit of gene robustness. For obvious reasons, our analysis is limited to the deletions of ~80% of yeast genes that are nonessential. Although we do not expect essential genes to behavior qualitatively differently from nonessential genes, future studies are required to validate this expectation.

Third, our study focused on lab strains of yeast because the deletion lines were all constructed in the genetic background of a lab strain. Whether our results apply to natural strains of yeast requires future research. Recent studies have revealed substantial genomic (Bergstrom et al. 2014) and morphological (Yvert et al. 2013) variations among yeast strains. Our analysis can be applied to strains of different genetic backgrounds when gene deletion lines in these backgrounds as well as their morphological data become available.

It is unknown what molecular genetic mechanisms are responsible for the observed reductions in the effect sizes of environmental and genetic perturbations on important traits. Previous yeast studies identified so-called capacitor genes, which could buffer phenotypic variations upon environmental perturbations. For example, it was found that genes with larger fitness effects upon deletion are more likely to be expression capacitors (Bergman and Siegal 2003) and genes with more genetic interactions are likely to be morphology capacitors (Levy and Siegal 2008). However, these studies did not examine whether the buffering effects on a trait

varies depending on trait importance. Consequently, the roles of these capacitors in the adaptive genetic and environmental robustness revealed here is unclear.

The observed mutational robustness of the GPM is a double-edged sword. On the one hand, it reduces the deleterious effects of mutations on important traits such that the severity of the associated defects is lessened. On the other hand, because of the reduction in effect size, the defects are less harmful and hence tend to spread more widely in a population. The full ramifications of a GPM that is robust to mutation await further study, so do the molecular mechanisms conferring the robustness.

2.5 MATERIALS AND METHODS

2.5.1 Morphological and fitness data

The phenotypic data of 501 morphological traits measured in the wild-type (123 replicate populations) and 4718 nonessential gene deletion yeast strains (each with one population) in the rich medium YPD (yeast extract, peptone, and dextrose) were generated by Ohya and colleagues (Ohya et al. 2005). We focused on 220 of the 501 traits, because these 220 traits were measured in individual cells whereas the other traits were measured for populations of cells. Using single-cell measurements is necessary for our analysis. The YPD fitness values of the deletion strains, relative to the wild-type, were recently measured by Qian and colleagues (Qian et al. 2012).

2.5.2 Fraction of genes affecting a trait

The mouse results (**Fig. 2.1a**) were from a summary of gene knockout studies (2009). The fruit fly results (**Fig. 2.1b**) were based on previously published data of P-element insertion lines (2010). Because each line typically contains multiple P-element insertions, we calculated

the fraction of single P-element insertions that affect a trait using the mean number of insertions per line (Mackay et al. 1992) under the assumption of no epistasis.

In the statistical analysis of yeast data, a gene is said to affect a trait when the gene deletion strain and the wild-type strain have a significant difference in the median trait value. We first calculated the *p*-value by comparing multiple cells of each deletion line and those of an arbitrarily selected wild-type replicate population (04his3-1) by the Mann-Whitney *U* test. A *p*-value of <5% was used to establish statistical significance. We thus obtained the faction of mutants in which the trait is affected (f_{mt}). To control for false positives, we similarly performed the Mann-Whitney *U* test between 04his3-1 and each of the other 122 wild-type populations and calculated the fraction (f_{wt}) of the 122 wild-type populations in which the trait deviates significantly from 04his3-1. The estimate of the fraction of genes affecting a trait (f_{genes}) equals $f_{mt} - f_{wt}$ if $f_{mt} > f_{wt}$; otherwise, we set $f_{genes} = 0$.

2.5.3 Less correlated morphological traits

To examine if some of our results were generated by highly correlated traits, we attempted to remove genetically highly correlated traits. We measured the genetic correlation between a pair of traits by correlating their trait values across the 4718 gene deletion strains. We then removed traits one by one from those with the highest absolute correlations until no two traits have a Pearson correlation whose absolute value is greater than 0.7. The final dataset has 54 traits and the distribution of f_{genes} is shown in **Fig. A.1.2**. Because morphological traits are naturally correlated to some extent, it remains to be determined whether the original 220 traits or the 54 less correlated traits better represent randomly sampled traits. The 54 less correlated traits were also used in **Fig. A.1.5**.

2.5.4 Raw effect size and net effect size

The raw effect size (ES_{ij}) of deleting gene *i* on trait *j* is defined as $(x_{ij} - w_j)/w_j$, where x_{ij} is the mean phenotypic value of trait *j* in the deletion strain *i*, and w_j is the corresponding value in the wild-type (averaged across all replicate populations). Conventionally, ES_{ij} is defined by $(x_{ij} - w_j)/SD_j$, where SD_j is the standard deviation of the trait in the wild-type (Mackay et al. 2009). We avoided using the conventional definition because the expected value of SD_j is in a large part determined by the precision of the trait measurement, rendering the comparison of mean $|ES|$ among traits primarily a comparison of the measurement quality rather than the biology of the traits. By contrast, the expected value of w_j is not affected by the imprecision of the measurement.

To estimate the net $|ES|$ of gene deletion on a trait, we generated 1000 pseudo phenotypic datasets. To generate a pseudo dataset, we randomly chose one wild-type replicate population and pick (with replacement) from this population the same number of cells as in the actual gene-deletion data. We then calculated pseudo $|ES|$ for each of these pseudo datasets and computed its mean value. Because 1000 pseudo datasets were generated, effectively all 123 wild-type populations were used. Net $|ES|$ equals raw $|ES|$ minus mean pseudo $|ES|$ if raw $|ES|$ > mean pseudo $|ES|$; otherwise, net $|ES| = 0$.

2.5.5 Phenotypic variation

The phenotypic variation in the wild-type was measured by *CV*. We calculated the variance among cells within each replicate population and averaged it across the 123 populations (V_{within}). We then calculated the variance among the mean phenotypic values of the 123

populations ($V_{between}$). $CV=\sqrt{V_{within}+V_{between}}/m$, where m is the average of the mean phenotypic values of the 123 populations.

2.5.6 Relative trait importance

For trait j, we conducted a linear regression $F_i = a_j - b_j$ (net $|ES_{ij}|$) for all i, where net $|ES_{ij}|$ is the absolute value of the net effect size of deleting gene i on trait j and F_i is the YPD fitness of the strain lacking gene i relative to the wild-type (Qian et al. 2012). The intercept a_j is the expected fitness when net $|ES_{ij}| = 0$, whereas the slope $b_j > 0$ is 100 times the reduction in fitness caused by 1% change in the phenotypic value of trait j. Thus, b_j is a measure of the relative importance of trait j to fitness, or trait importance (TI). Because we focused on deleterious mutations in this model, we used only those genes that decrease fitness when deleted. We also tried the log model $\log F_i = a_j - b_j$ (net $|ES_{ij}|$) and found the results to be similar (**Fig. A.1.7**).

In the above estimation of trait importance, we assumed that, to a trait, a positive effect and a negative effect of the same size have the same fitness effect, which may not be true to all traits. Because positive and negative effects are arbitrarily defined, we also considered only positive (or only negative) raw effects in subsequent analysis (**Table A.1.1**).

2.5.7 Principal component analysis of the net |*ES*| matrix

To examine if the non-independence among traits affects our results, we followed a previous study (Wang et al. 2010) to perform a principal component analysis to transform the net $|ES|$ matrix **M** (4718 genes × 220 traits). The principal component analysis was done by the "princomp" function in MATLAB. After this function returned a coefficient matrix **C** (220 × 220), we calculated **M'=MC** (4718 transformed effects × 220 principal traits), which provides

the net effect size of each gene on each of the 220 orthogonal principal component traits. We then used **M'** to estimate trait importance.

2.5.8 Predicting the fitness of a mutant strain given the genotype-phenotype map

We built a multivariate linear model of yeast fitness that includes all 220 traits and 2779 gene deletion strains that are less fit than the wild-type: $F_i = \alpha - \sum_j \beta_j \left(\text{net} \left|ES_{ij}\right|\right)$, where α is a constant. We estimated α and β_j for all 220 traits using the "glmfit" function in MATLAB. Based on the above formula and the estimated α and β_j values, we predicted F_i upon the deletion of gene i when either the original or randomly shuffled net $|ES_{ij}|$ values were used.

2.5.9 Null mutation rate per gene per generation

Based on the genome sequences of mutation accumulation yeast strains (Lynch et al. 2008), the point mutation rate in yeast is 3.3×10^{-10} per site per generation; the small (1-3 bp) indel mutation rate is 2×10^{-11} per site per generation; and the gene loss mutation rate is 2.1×10^{-6} per gene per generation. Taken together, we estimated the null mutation rate per gene per generation to be approximately $[3.3\times10^{-10}\times(3/63) + 2\times10^{-11}\times0.83] \times 1419 + 2.1\times10^{-6} = 2.146\times10^{-6}$. Here, 3/63 is the average probability that a random point mutation in a coding region is nonsense, 0.83 is the fraction of small indels that are not multiples of 3 nucleotides (Zhang and Webb 2003), 1419 is the mean number of coding nucleotides per yeast protein-coding gene (Zhang 2000).

In humans, the point mutation rate is 1.25×10^{-9} per site per year and the indel mutation rate is 1×10^{-10} per site per year (Zhang and Webb 2003). Based on copy number variations in humans, it has been estimated that the gene loss rate is $\sim10^{-5}$ per gene per generation (Zhang et al.

2009). Thus, the total null mutation rate per gene per generation is $[1.25\times10^{-9}\times(3/63) + 1\times10^{-10}\times0.83] \times 1341 \times 25 + 1\times10^{-5} = 1.48\times10^{-5}$. Here, 1341 is the mean number of coding nucleotides per human protein-coding gene (Zhang 2000) and 25 is the approximate number of years per human generation.

2.5.10 Fitness advantages of robustness modifiers

In a diploid population, let *A* be the wild-type allele at gene *i* and *a* represent all null alleles, which are assumed to be completely recessive to *A*. The fitness values of *AA*, *Aa*, and *aa* are 1, 1, and 1- s_i, respectively, where $s_i > 0$. Let the mutation rate from *A* to *a* be μ_i and the back mutation rate be 0. Under the mutation-selection balance (Hartl and Clark 1997), the equilibrium frequency of *aa* individuals is μ_i/s_i and the expected fitness of a randomly picked individual in the population is $1\times(1-\mu_i/s_i) + (1-s_i)\times(\mu_i/s_i) = 1-\mu_i$. Let us consider a robustness modifier that masks the deleterious effect of *a* such that *aa* individuals now have a fitness of $1-s_i+\Delta s_i$ ($0 < \Delta s_i < s_i$). The expected fitness of an individual with the modifier is $1\times(1-\mu_i/s_i) + (1-s_i+\Delta s_i)\times(\mu_i/s_i) = 1-\mu_i+\mu_i\Delta s_i/s_i$. Thus, the mean fitness advantage of the robustness modifier is $g = \mu_i\Delta s_i/s_i$. If gene *i* affects multiple traits, $\Delta s_i/s_i$ is determined by the fractional change in its combined fitness effect on these traits. Because reducing the genic effect on a trait of large fitness contribution is expected to contribute more to Δs_i than reducing the same amount of effect on a trait of small fitness contribution, modifiers that preferentially reduce the mutational effects on important traits are more advantageous than those that have no such preference. As a result, natural selection is expected to preferentially enhance the mutational robustness of important traits. If the modifier buffers the null mutations of multiple genes, its fitness advantage is $\sum(\mu_i\Delta s_i/s_i)$, where $\sum$ indicates summation over the multiple buffered genes. The above formula also works for

deleterious mutations in general when u_i is the total deleterious mutation rate at gene *i*, as long as all the mutations considered are completely recessive to the wild-type allele.

When *a* is not completely recessive to *A*, the fitness is 1, 1-$h_i s_i$, and 1-s_i for *AA*, *Aa*, and *aa*, respectively, where $0 < h < 1$ is the dominance of *a*, relative to *A*. Under mutation-selection balance (Hartl and Clark 1997), the equilibrium frequency of the *a* allele is $\mu_i/(s_i h_i)$ and the expected fitness of a randomly picked individual in the population is approximately 1-$2\mu_i$. Let us consider a robustness modifier that masks the deleterious effect of *a* such that *Aa* individuals now have a fitness of 1-$h_i s_i$+$h_i \Delta s_i$ ($0 < \Delta s_i < s_i$) and *aa* individuals have a fitness of 1-s_i+Δs_i. The expected fitness of an individual with the modifier will be 1-$2\mu_i$+$2\mu_i \Delta s_i/s_i$. Thus, the mean fitness advantage of the robustness modifier equals $2\mu_i \Delta s_i/s_i$. If the modifier buffers the deleterious mutations of multiple genes, its fitness advantage is $2\sum(\mu_i \Delta s_i/s_i)$, where $\sum$ indicates summation over the multiple buffered genes.

2.5.11 Gene expression data and analysis

The genome-wide gene expression data of yeast single-gene deletion lines were compiled from four studies (Hughes et al. 2000; Hu et al. 2007; van Wageningen et al. 2010; Lenstra et al. 2011). In each study, the wild-type and gene deletion strains were grown in YPD or synthetic complete (SC) medium. The microarray expression level of gene *j* in the strain lacking gene *i* was compared with that in the wild-type under the same medium to measure the effect of deleting gene *i* on the expression level of gene *j*. Strains lacking more than one gene were not considered. If a deletion line was analyzed in multiple studies, we used the data from the most recent study. In the end, the data contained expression changes of 4399 genes in 754 gene deletion lines. We then limited our analysis to the expression levels of 3116 (out of 4399) genes

because these genes reduce fitness when deleted (i.e. fitness ≤ 1 regardless of statistical significance) (Qian et al. 2012).

The effect size of deleting gene *i* on the expression level of gene *j* was defined by (x_{ij} - w_j)/w_j = x_{ij}/w_j -1, where x_{ij} is the expression level of gene *j* in the strain lacking gene *i* and w_j is the expression level of gene *j* in the wild-type. The x_{ij}/w_j value used was available from each dataset. Because the expression data were obtained for populations of cells rather than individual cells, net |*ES*| cannot be estimated. The measurement error of the expression level of a gene in microarray is mainly determined by the expression level of the gene. Thus, controlling the expression level (**Tables A.1.2 - A.1.5**) largely controls the measurement error. For these analyses, the expression levels from the wild-type strain in YPD (Nagalakshmi et al. 2008) were used. We quantified the effect sizes of 35 highly different environmental changes on the wild-type gene expression levels (Proulx et al. 2007) using the same formula, where x_{ij}/w_j is the fold change in the expression of gene *j* induced by the *i*th environmental change. These 35 environments were previously chosen from a total of 162 environments to represent the least correlated environmental challenges (Proulx et al. 2007). We performed a similar analysis using all 162 environmental challenges under which the wild-type gene expression changes were previously measured (Gasch et al. 2000) (**Table A.1.3**).

When examining the potential impact of highly correlated traits, we measured the genetic correlation between a pair of expression traits by correlating their expression levels across mutant strains. We then removed expression traits one by one from those with the highest correlations until no two traits have a Pearson correlation greater than 0.7. The final dataset included 2223 expression traits (**Table A.1.4**).

2.6 ACKNOWLEDGMENTS

We thank Dr. Yoshikazu Ohya for sharing the yeast morphological data and Drs. Calum Maclean and Jian-Rong Yang for valuable comments. This work was supported in part by the research grant GM103232 from the U.S. National Institutes of Health to J.Z.

2.7 REFERENCES

Alon U. 2007. An Introduction to Systems Biology: Design Principles of Biological Circuits London: Chapman & Hall.

Batada NN, Hurst LD. 2007. Evolution of chromosome organization driven by selection for reduced gene expression noise. Nat Genet 39:945-949.

Bergman A, Siegal ML. 2003. Evolutionary capacitance as a general feature of complex gene networks. Nature 424:549-552.

Bergstrom A, Simpson JT, Salinas F, Barre B, Parts L, Zia A, Nguyen Ba AN, Moses AM, Louis EJ, Mustonen V, et al. 2014. A high-definition view of functional genetic variation from natural yeast genomes. Mol Biol Evol 31:872-888.

Ciliberti S, Martin OC, Wagner A. 2007. Innovation and robustness in complex regulatory gene networks. Proc Natl Acad Sci U S A 104:13591-13596.

de Visser JA, Hermisson J, Wagner GP, Ancel Meyers L, Bagheri-Chaichian H, Blanchard JL, Chao L, Cheverud JM, Elena SF, Fontana W, et al. 2003. Perspective: Evolution and detection of genetic robustness. Evolution 57:1959-1972.

Draghi JA, Parsons TL, Wagner GP, Plotkin JB. 2010. Mutational robustness can facilitate adaptation. Nature 463:353-355.

Ehrenreich IM, Torabi N, Jia Y, Kent J, Martis S, Shapiro JA, Gresham D, Caudy AA, Kruglyak L. 2010. Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature 464:1039-U1101.

Falconer DS, Mackay TFC. 1996. Introduction to Quantitative Genetics. London: Pearson.

Flatt T. 2005. The evolutionary genetics of canalization. Q Rev Biol 80:287-316.

Flint J, Mackay TF. 2009. Genetic architecture of quantitative traits in mice, flies, and humans. Genome Res 19:723-733.

Fraser HB, Schadt EE. 2010. The quantitative genetics of phenotypic robustness. PLOS One 5:e8635.

Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. 2000. Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell 11:4241-4257.

Gibson G, Wagner G. 2000. Canalization in evolutionary genetics: a stabilizing theory? BioEssays 22:372-380.

Han HW, Bae SH, Jung YH, Kim JH, Moon J. 2013. Genome-wide characterization of the relationship between essential and TATA-containing genes. FEBS Lett 587:444-451.

Hartl DL, Clark AG. 1997. Principles of Population Genetics. Sunderland, Mass.: Sinauer.

Hermisson J, Wagner GP. 2004. The population genetic theory of hidden variation and genetic robustness. Genetics 168:2271-2284.

Houle D. 1998. How should we explain variation in the genetic variance of traits? Genetica 102-3:241-253.

Hu Z, Killion PJ, Iyer VR. 2007. Genetic reconstruction of a functional transcriptional regulatory network. Nat Genet 39:683-687.

Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD, et al. 2000. Functional discovery via a compendium of expression profiles. Cell 102:109-126.

Lehner B. 2008. Selection to minimise noise in living systems and its implications for the evolution of gene expression. Mol Syst Biol 4:170.

Lehner B. 2010. Genes confer similar robustness to environmental, stochastic, and genetic perturbations in yeast. PLOS One 5:e9035.
Lenstra TL, Benschop JJ, Kim T, Schulze JM, Brabers NA, Margaritis T, van de Pasch LA, van Heesch SA, Brok MO, Groot Koerkamp MJ, et al. 2011. The specificity and topology of chromatin interaction pathways in yeast. Mol Cell 42:536-549.
Levy SF, Siegal ML. 2008. Network Hubs Buffer Environmental Variation in Saccharomyces cerevisiae. PLOS Biol 6:2588-2604.
Lyman RF, Lawrence F, Nuzhdin SV, Mackay TF. 1996. Effects of single P-element insertions on bristle number and viability in Drosophila melanogaster. Genetics 143:277-292.
Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ, Okamoto K, Kulkarni S, Hartl DL, et al. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. Proc Natl Acad Sci U S A 105:9272-9277.
Mackay TF. 2010. Mutations and quantitative genetic variation: lessons from *Drosophila*. Phil Trans R Soc B 365:1229-1239.
Mackay TF, Lyman RF, Jackson MS. 1992. Effects of P element insertions on quantitative traits in Drosophila melanogaster. Genetics 130:315-332.
Mackay TF, Stone EA, Ayroles JF. 2009. The genetics of quantitative traits: challenges and prospects. Nat Rev Genet 10:565-577.
Mackay TFC. 2001. The genetic architecture of quantitative traits. Annu Rev Genet 35:303-339.
Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, et al. 2009. Finding the missing heritability of complex diseases. Nature 461:747-753.
Mather K. 1941. Variation and selection of polygenic characters. J Genet 41:159-193.
Meiklejohn CD, Hartl DL. 2002. A single mode of canalization. Trends Ecol Evol 17:468-473.
Milton CC, Huynh B, Batterham P, Rutherford SL, Hoffmann AA. 2003. Quantitative trait symmetry independent of Hsp90 buffering: Distinct modes of genetic canalization and developmental stability. Proc Natl Acad Sci U S A 100:13396-13401.
Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320:1344-1349.
Newman JR, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS. 2006. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature 441:840-846.
Ohya Y, Sese J, Yukawa M, Sano F, Nakatani Y, Saito TL, Saka A, Fukuda T, Ishihara S, Oka S, et al. 2005. High-dimensional and large-scale phenotyping of yeast mutants. Proc Natl Acad Sci U S A 102:19015-19020.
Proulx SR, Nuzhdin S, Promislow DE. 2007. Direct selection on genetic robustness revealed in the yeast transcriptome. PLOS One 2:e911.
Qian W, Ma D, Xiao C, Wang Z, Zhang J. 2012. The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. Cell Rep 2:1399-1410.
Ramani AK, Chuluunbaatar T, Verster AJ, Na H, Vu V, Pelte N, Wannissorn N, Jiao A, Fraser AG. 2012. The majority of animal genes are required for wild-type fitness. Cell 148:792-802.
Rhee HS, Pugh BF. 2012. Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature 483:295-301.

Robertson A. 1967. The nature of quantitative genetic variation. In: A Brink, editor. Heritage From Mendel. Madison: The Univeristy of Wisconsin Press.
Sanjuan R, Cuevas JM, Furio V, Holmes EC, Moya A. 2007. Selection for robustness in mutagenized RNA viruses. PLOS Genet 3:e93.
Scharloo W. 1991. Canalization: genetic and developmental aspects. Annu Rev Ecol Syst 22:65-93.
Siegal ML, Bergman A. 2002. Waddington's canalization revisited: Developmental stability and evolution. Proc Natl Acad Sci U S A 99:10528-10532.
Stearns SC, Kaiser M, Kawecki TJ. 1995. The differential genetic and environmental canalization of fitness components in Drosophila melanogaster. J Evol Biol 8:539-557.
Stearns SC, Kawecki TJ. 1994. Fitness sensitivity and the canalization of life-history traits. Evolution 48:1438-1450.
Tirosh I, Weinberger A, Carmi M, Barkai N. 2006. A genetic signature of interspecies variations in gene expression. Nat Genet 38:830-834.
van Wageningen S, Kemmeren P, Lijnzaad P, Margaritis T, Benschop JJ, de Castro IJ, van Leenen D, Groot Koerkamp MJ, Ko CW, Miles AJ, et al. 2010. Functional overlap and regulatory links shape genetic interactions between signaling pathways. Cell 143:991-1004.
Waddington CH. 1942. Canalization of development and the inheritance of acquired characters. Nature 150:563–565.
Wagner A. 2005a. Energy constraints on the evolution of gene expression. Mol Biol Evol 22:1365-1374.
Wagner A. 2005b. Robustness and Evolvability in Living Systems. Princeton, NJ: Princeton University Press.
Wagner G, Booth G, Bagheri-Chaichian H. 1997. A population genetic theory of canalization. Evolution 51:329-347.
Wagner GP, Zhang J. 2011. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. Nat Rev Genet 12:204-213.
Wang Z, Liao BY, Zhang J. 2010. Genomic patterns of pleiotropy and the evolution of complexity. Proc Natl Acad Sci U S A 107:18034-18039.
Wang Z, Zhang J. 2011. Impact of gene expression noise on organismal fitness and the efficacy of natural selection. Proc Natl Acad Sci U S A 108:E67-76.
Wilke CO, Wang JL, Ofria C, Lenski RE, Adami C. 2001. Evolution of digital organisms at high mutation rates leads to survival of the flattest. Nature 412:331-333.
Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, et al. 1999. Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science 285:901-906.
Yvert G, Ohnuki S, Nogami S, Imanaga Y, Fehrmann S, Schacherer J, Ohya Y. 2013. Single-cell phenomics reveals intra-species variation of phenotypic noise in yeast. BMC Syst Biol 7:54.
Zhang F, Gu W, Hurles ME, Lupski JR. 2009. Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet 10:451-481.
Zhang J. 2000. Protein-length distributions for the three domains of life. Trends Genet 16:107-109.
Zhang J, Webb DM. 2003. Evolutionary deterioration of the vomeronasal pheromone transduction pathway in catarrhine primates. Proc Natl Acad Sci U S A 100:8337-8341.

Table 2.1 Spearman's rank correlation between the importance of a gene expression trait to fitness and the mean effect size of gene deletion ($|ES_G|$) or environmental perturbation ($|ES_E|$).

Variables correlated	Variables controlled	Spearman's ρ	p-value
Fitness effect[1], $\|ES_E\|$		-0.146	2.1e-16
Fitness effect[1], $\|ES_G\|$		-0.180	3.7e-24
Fitness effect[1], $\|ES_E\|$	$\|ES_G\|$	-0.094	1.3e-07
Fitness effect[1], $\|ES_G\|$	$\|ES_E\|$	-0.141	1.9e-15
Essentiality[2], $\|ES_E\|$		-0.088	8.1e-07
Essentiality[2], $\|ES_G\|$		-0.125	2.2e-12
Essentiality[2], $\|ES_E\|$	$\|ES_G\|$	-0.051	4.8e-03
Essentiality[2], $\|ES_G\|$	$\|ES_E\|$	-0.103	9.6e-09

[1] Trait importance is measured by the fitness defect caused by deleting the gene.
[2] Essentiality = 0 for nonessential traits and 1 for essential traits.

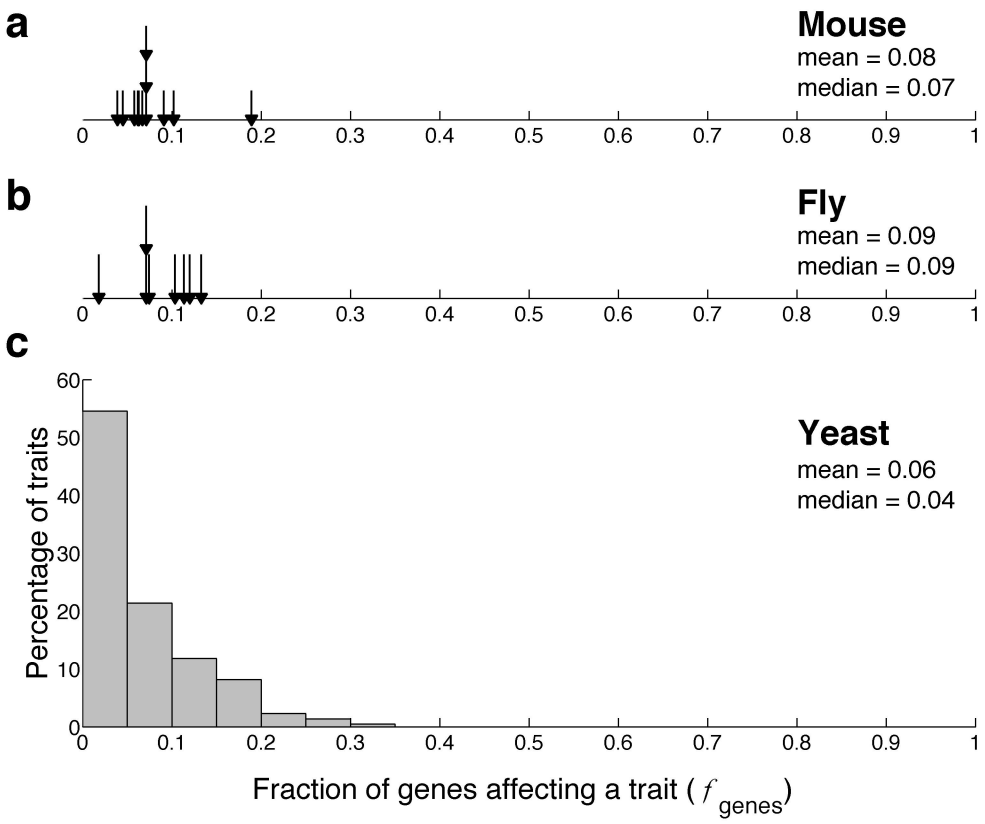

Figure 2.1 Fraction of genes affecting a trait, with the mean and median values indicated. (**a**) Patterns emerging from 12 traits examined in 250 lines of knockout mice. (**b**) Patterns emerging from eight traits examined in various *P*-element insertion lines of fruit flies. In (a) and (b), each arrow represents one trait. (**c**) Frequency distribution of the fraction of genes affecting a trait, derived from 220 morphological traits examined in 4718 nonessential gene deletion lines of yeast.

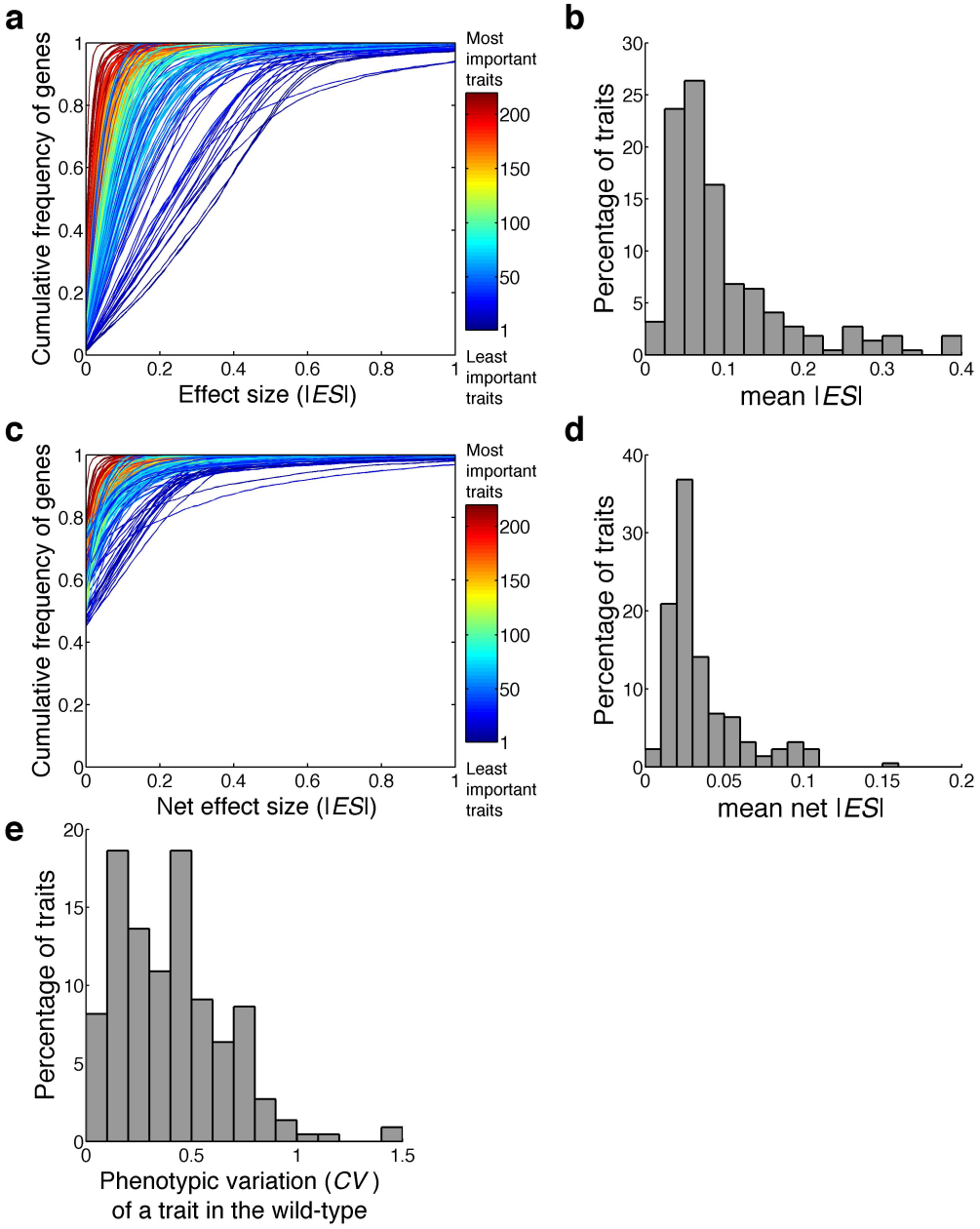

Figure 2.2 Distributions of the absolute values of the raw and net effect sizes (|*ES*|) of 4718 nonessential gene deletions on 220 morphological traits in yeast. (**a**) Cumulative probability distributions of raw |*ES*| of 4718 gene deletions on 220 traits. Each curve represents a trait and is colored according to trait importance. The distributions are shown only in the range of $0 < |ES| < 1$ to better distinguish among different curves. (**b**) Distribution of the mean raw |*ES*| among the 220 traits. (**c**) Cumulative probability distributions of net |*ES*| of 4718 gene deletions on 220 traits. (**d**) Distribution of the mean net |*ES*| among the 220 traits. (**e**) Distribution of the wild-type phenotypic variation (*CV*) among the 220 traits.

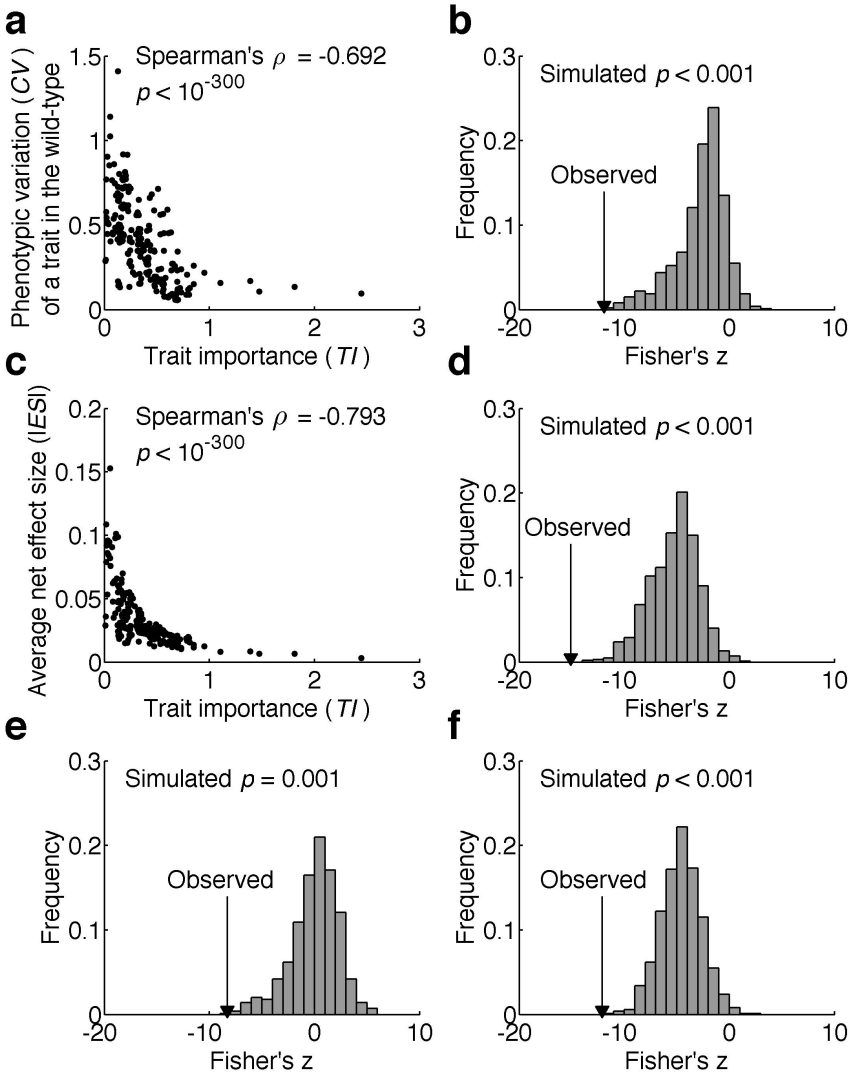

Figure 2.3 Environmental/stochastic robustness and genetic robustness of yeast morphological traits. (**a**) The phenotypic variation (*CV*) of a trait among isogenic wild-type cells decreases with the rise of the trait importance (*TI*). Each dot is a trait. (**b**) Distribution of Fisher's *z* derived from the rank correlation between *CV* and pseudo *TI*. (**c**) The mean net |*ES*| of gene deletion on a trait decreases with the rise of the trait importance (*TI*), demonstrating genetic robustness. Each dot is a trait. (**d**) Distribution of Fisher's *z* derived from the rank correlation between mean net |*ES*| and pseudo *TI*. (**e**) Distribution of Fisher's *z* derived from the partial rank correlation between *CV* and pseudo *TI*, after the control of mean net |*ES*|. (**f**) Distribution of Fisher's *z* derived from the partial rank correlation between mean net |*ES*| and pseudo *TI*, after the control of *CV*. In (b), (d), (e), and (f), the real *z* observed from the actual data is indicated by an arrowhead and the *p*-value is the probability that a randomly picked pseudo *z* is more negative than the real *z*.

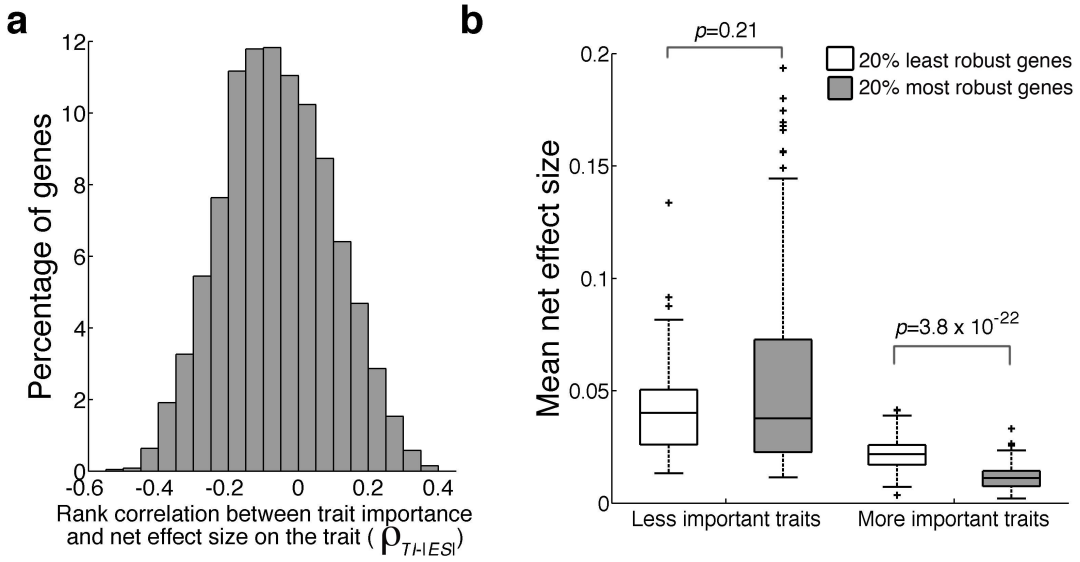

Figure 2.4 Among-gene variation in contribution to genetic robustness. (**a**) Frequency distribution of a gene's rank correlation (ρ) between its absolute net effect size (|*ES*|) on a trait and the trait importance (*TI*). Most genes show negative correlations. (**b**) Effect size differences between the 20% most robust (having the most negative ρ values in panel a) and 20% least robust (having the smallest |ρ| values) genes on traits of different importance. Traits are divided into two equal-size bins based on *TI*: less-important traits and more important traits. In the boxplot (see the scale marked on the left Y-axis), the lower edge and upper edge of a box represent the 25% quartile (q_1) and 75% quartile (q_3), respectively. The horizontal line inside a box indicates the median (*md*). The whiskers extend to the most extreme values inside inner fences, $md \pm 1.5(q_3 - q_1)$. The values outside the inner fences (outliers) are plotted by plus signs.

Chapter 3

Adaptive Genetic Robustness of *Escherichia coli* Metabolic Fluxes[1]

3.1 ABSTRACT

Genetic robustness refers to phenotypic invariance in the face of mutation and is a common characteristic of life, but its evolutionary origin is highly controversial. Genetic robustness could be an intrinsic property of biological systems, a result of direct natural selection, or a byproduct of selection for environmental robustness. To differentiate among these hypotheses, we analyze the metabolic network of *Escherichia coli* and comparable functional random networks. Treating the flux of each reaction as a trait and computationally predicting trait values upon mutations or environmental shifts, we discover that (1) genetic robustness is greater for the actual network than the random networks, (2) the genetic robustness of a trait increases with trait importance and this correlation is stronger in the actual network than in the random networks, and (3) the above result holds even after the control of environmental robustness. These findings demonstrate an adaptive origin of genetic robustness, consistent with the theoretical prediction that, under certain conditions, direct selection is sufficiently powerful to promote genetic robustness in cellular organisms.

[1]This chapter was published as Ho and Zhang (2016) *Mol. Biol. Evol.* 33: 1164-1176

3.2 INTRODUCTION

Genetic robustness, also known as genetic canalization, refers to the ability of a biological system to maintain phenotypic invariance upon mutation. Genetic robustness has been reported in many organisms (Rutherford, Lindquist 1998; Fares et al. 2002; Ciliberti, Martin, Wagner 2007; Ho, Zhang 2014; Yang, Ruan, Zhang 2014) at multiple levels of biological organization (Wagner 2005) and is an important characteristic of life. The evolutionary origin of genetic robustness, however, is controversial with three competing hypotheses (de Visser et al. 2003; Felix, Barkoulas 2015). The adaptation hypothesis states that genetic robustness results from direct positive selection, because mechanisms reducing the phenotypic effects of mutations are favored due to the overall deleterious nature of mutations (Waddington 1942). Although this hypothesis has been supported in digital organisms and viruses (Wilke et al. 2001; Sanjuan et al. 2007), its validity in cellular organisms is debated. The congruence hypothesis posits that genetic robustness is a byproduct of selection for environmental robustness, which is the ability to maintain phenotypic invariance upon environmental perturbations (Meiklejohn, Hartl 2002; de Visser et al. 2003). The intrinsic property hypothesis asserts that genetic robustness is an intrinsic property of biological systems (Siegal, Bergman 2002). For example, it was suggested that the genetic robustness of transcriptional networks originates from the functional constraint without selection for genetic robustness (Siegal, Bergman 2002).

When one mutation affects multiple traits of different levels of importance to fitness, robustness modifiers that preferentially buffer the mutational effects on relatively important traits have advantages over modifiers that equally buffer the mutational effects on all traits (Ho, Zhang 2014). Consequently, in the presence of selection for genetic robustness, genetic robustness is

expected to rise with trait importance. The population genetic theory of canalization also predicts that when trait importance is not too high (so that there is still sufficient genetic variation), genetic robustness should rise with trait importance (Wagner, Booth, Bagheri-Chaichian 1997). Nevertheless, this trend may also exist in the absence of selection for genetic robustness if there is selection for environmental robustness and if mechanisms for environmental robustness also confer genetic robustness (i.e., the congruence hypothesis). By contrast, the intrinsic property hypothesis does not predict an increase in genetic robustness with trait importance. Thus, examining the correlation between genetic robustness and trait importance allows distinguishing the intrinsic property hypothesis from the other two hypotheses, which can then be differentiated by computing the partial correlation after controlling environmental robustness. If the partial correlation remains significant, the adaptation hypothesis is supported; otherwise, the congruence hypothesis is supported.

Several previous studies tested the adaptation hypothesis using the above strategy, but obtained mixed results, possibly because the number of traits examined is small, the traits are not randomly sampled, the traits are not comparable with one another, and/or trait importance is not well defined (Stearns, Kawecki 1994; Stearns, Kaiser, Kawecki 1995; Houle 1998; Proulx, Nuzhdin, Promislow 2007). We recently applied the same strategy to 220 yeast morphological traits and found evidence for the adaptation hypothesis (Ho, Zhang 2014). While the traits were plentiful, unbiased, and comparable, the measure of trait importance was based on a correlation between the amount of phenotypic change in a trait and the fitness change upon gene deletion, and is hence indirect (Ho, Zhang 2014). Furthermore, although the intrinsic property hypothesis does not predict a positive correlation between genetic robustness and trait importance, it remains possible that some unknown intrinsic factors create this correlation in the absence of

selection for any robustness. For example, although some evidence for adaptive robustness of gene expression levels was presented (Ho, Zhang 2014), the intrinsic property hypothesis could not be completely excluded. Thus, it is desirable to test the adaptation hypothesis in a system where direct measures of trait importance are available and intrinsic properties can be explicitly scrutinized.

To this end, we take advantage of the well-developed mathematical analysis of metabolic networks of the model prokaryote *Escherichia coli*. We treat the flux of each reaction in a metabolic network as a trait and use flux balance analysis (FBA) (Orth, Thiele, Palsson 2010) to measure trait importance. We then use the method of minimization of metabolic adjustment (MOMA) (Segre, Vitkup, Church 2002) to quantify phenotypic alteration upon mutation or environmental perturbation. We compare our findings from the real network with those from comparable functional random networks. Our results provide unambiguous computational evidence for direct selection for the genetic robustness of metabolic fluxes in *E. coli*. We focus on reaction robustness rather than fitness robustness in this study, because the existence of multiple reactions allows analyzing the correlation between trait importance and genetic robustness, which is critical to our differentiation among the hypotheses of adaptation, congruence, and intrinsic property.

3.3 RESULTS

3.3.1 Traits and trait importance in metabolic networks

FBA is a powerful mathematical tool for metabolic network analysis. Under the steady state assumption, FBA maximizes the biomass production rate (i.e., cellular fitness) by simultaneously balancing all metabolic fluxes under a set of flux constraints and a given

nutritional environment, allowing the prediction of all fluxes as well as the fitness (Orth, Thiele, Palsson 2010). FBA predictions for model organisms such as the bacterium *E. coli* have been extensively validated experimentally (Ibarra, Edwards, Palsson 2002; Fong, Palsson 2004; Lewis et al. 2010) and have thus been widely used in the study of genotype-environment-phenotype relationships (He et al. 2010; Costenoble et al. 2011; Barve, Rodrigues, Wagner 2012; Harcombe et al. 2013; Bordbar et al. 2014).

We first applied FBA to the *E. coli* iAF1260 metabolic model (Feist et al. 2007) under the "glucose environment" where the only carbon source is glucose. Excluding immutable reactions such as simple diffusions, we obtained the wild-type fitness as well as the wild-type fluxes of 362 reactions that have nonzero fluxes. Unless otherwise noted, we focused on nonzero-flux reactions.

We then simulated mutations that caused loss of enzyme function by setting the flux of the corresponding reaction to 0. We used MOMA (Segre, Vitkup, Church 2002) to predict the new fluxes of all other reactions in the network and the fitness of the mutant relative to that of the wild-type (f). In addition to satisfying all the constraints in FBA, MOMA predicts each flux and the fitness by minimizing the sum of the squared change of each flux from its wild-type value over all reactions, based on the premise that the metabolic network undergoes minimal flux redistribution upon mutation. MOMA has been shown to outperform FBA in predicting fluxes and fitness upon mutation (Segre, Vitkup, Church 2002). After individually constraining the 362 fluxes to 0, we found that the frequency distribution of mutant fitness is U-shaped (**Fig. A.2.1A**), as was previously found (Wloch et al. 2001; Sanjuan, Moya, Elena 2004; Wang, Zhang 2009a). Note that some f values are extremely small (see **Fig. A.2.1B** where the distribution of f is plotted in $\log_{10}$ scale). We considered f values lower than 10^{-4} to be effectively 0 and regarded the

corresponding reaction as essential. By this definition, 257 reactions are essential, while the remaining 105 are nonessential. For each nonessential reaction, we quantified its importance by $s = 1\text{-}f$. In other words, the greater the drop in fitness upon the block of a reaction, the more important the reaction is.

3.3.2 Flux alteration upon mutation correlates with fitness reduction

The adaptation hypothesis of genetic robustness presumes that phenotypic changes from the wild-type are deleterious. Although this presumption appears reasonable and has been used in various simulation studies (Clark 1991; Rausher 2013), its validity has not been empirically confirmed for the flux traits examined here. To confirm this presumption, for each of the 105 nonessential reactions, we imposed a series of flux constraint and used MOMA to estimate the resultant fitness. That is, we reduced the maximal allowed flux of a focal reaction by a fraction ΔF, where $\Delta F = 0.05, 0.10, 0.15, \ldots$, and 1. For example, if the flux of the focal reaction equals v_0 in the wild-type, its new flux cannot exceed $0.95v_0$ when $\Delta F = 0.05$. Consistent with the expected behavior of constraint-based metabolic modeling, the fitness decreases as ΔF increases for each of the 105 nonessential reactions tested (blue lines in **Fig. 3.1A**). The same pattern was observed when essential reactions were examined (red lines in **Fig. 3.1A**). To test the presumption that flux alterations from the wild-type values are deleterious, we examined if the negative correlation between the flux alteration of a focal reaction and fitness exists even when mutations occur to other reactions (i.e., when non-focal reactions are constrained). Let $\gamma_k(i)$ be the fractional flux change for a nonessential reaction k in a mutant in which the flux of reaction i is constrained by ΔF relative to that in the wild-type (see Materials and Methods). For example, let k be the reaction PFK, which converts fructose 6-phosphate to fructose 1,6-bisphosphate in glycolysis. When individually constraining all 362 reactions by $\Delta F = 0.5$, we observed that the

resultant $\gamma_k(i)$ and fitness $f(i)$ are negatively correlated (Spearman's $\rho = -0.562$; $p < 10^{-300}$; **Fig. 3.1B**). That is, no matter which reaction is constrained, the more PFK changes in flux from the wild-type level, the lower the fitness. This result is not unique to PFK but is found for each nonessential reaction k examined (**Fig. 3.1C**).

Moreover, the above finding is not limited to $\Delta F = 0.5$. We calculated ρ between $\gamma_k(i)$ and $f(i)$ for each k at every level of $\Delta F < 0.6$, because constraining some essential reactions by 60% is lethal (**Fig. 3.1A**). For all ΔF values considered and all k, all ρ values except three are negative (**Fig. 3.1D**). Note that these three positive ρ values occurred under low ΔF values, where the flux alterations tend to be small and the correlation measures tend to be noisy. After we controlled for multiple testing by the conservative Bonferroni correction, 98.9% of the negative ρ values are significant at the 5% level while none of the positive ρ values are significant. Together, these results support the premise that flux changes from their wild-type levels are generally deleterious. In other words, flux invariance upon mutation is beneficial. For subsequent analyses, we used only $\Delta F = 0.5$ to reduce the computational demand.

3.3.3 Genetic robustness of *E. coli* reactions is significantly higher than those of random networks

Let γ_G be the fractional flux change of a focal reaction from its wild-type level upon a mutation, where the subscript G stands for genetic perturbation. We use $\bar{\gamma}_G$ to denote the mean of γ_G across all mutants in which a nonzero-flux reaction is constrained by $\Delta F = 0.5$. We then calculated $\bar{\bar{\gamma}}_G$, the mean $\bar{\gamma}_G$ across all nonzero-flux focal reactions. The smaller the $\bar{\bar{\gamma}}_G$, the greater the average genetic robustness of metabolic fluxes. We found $\bar{\bar{\gamma}}_G$ to be 0.302 for *E. coli*. To examine if *E. coli* metabolic fluxes are more robust to mutation than expected by chance, we generated 500 random metabolic networks that have not been subject to selection for robustness

and compared them with the *E. coli* metabolic network. These random networks were generated by the previously published method (Rodrigues, Wagner 2009; Barve, Wagner 2013), in which the reactions in the *E. coli* network were replaced with those randomly picked from the union of all known metabolic reactions of all organisms one at a time, under the condition that the network always has nonzero fitness under the glucose environment. All of these random networks have the same number of reactions as the *E. coli* network. On average, only 36.3% of *E. coli* reactions allowed for swapping appeared in a random network. This level of reaction overlap presumably reflects the constraint from the shared function between the *E. coli* network and the random networks. We calculated $\bar{\bar{\gamma}}_G$ for each of the 500 random networks, after excluding lethal mutations under $\Delta F = 0.5$. We found that only three of the 500 random networks have a $\bar{\bar{\gamma}}_G$ smaller than that of *E. coli* (**Fig. 3.2A**), indicating that the average genetic robustness of metabolic fluxes is significantly greater in *E. coli* than in comparable random networks ($P = 0.006$).

3.3.4 Environmental robustness of *E. coli* reactions is only marginally higher than those of random networks

We similarly studied environmental robustness. Let γ_E be the fractional flux change of a reaction upon an environmental shift from its level under the glucose environment, where the subscript E stands for environmental perturbation. We use $\bar{\gamma}_E$ to denote the mean of γ_E across many environmental changes. We then calculated $\bar{\bar{\gamma}}_E$, the mean $\bar{\gamma}_E$ across all nonzero-flux focal reactions. The smaller the $\bar{\bar{\gamma}}_E$, the greater the average environmental robustness of metabolic fluxes. We simulated 1000 nutritional environments by supplying a random combination of 258 different carbon sources (see Materials and Methods) and used MOMA to calculate the flux of each reaction. We found $\bar{\bar{\gamma}}_E$ to be 0.018 for *E. coli*. Interestingly, we found that 6.2% of the 500

random networks have $\bar{\bar{\gamma}}_E$ values smaller than that of *E. coli* (**Fig. 3.2B**), indicating that the average environmental robustness of *E. coli* metabolic fluxes is only marginally significantly greater than the chance expectation ($P = 0.062$).

3.3.5 The intrinsic property hypothesis for genetic robustness is refuted

As mentioned, selection for genetic robustness could result in higher genetic robustness of more important traits. To test this prediction of the adaptation hypothesis, we correlated the importance of a reaction (s) with its $\bar{\gamma}_G$. We considered only nonessential reactions as focal reactions in this analysis, because essential reactions all have $s = 1$ despite that different essential reaction may be of different importance. We found a strong, negative correlation between s and $\bar{\gamma}_G$ (Spearman's $\rho = -0.859$; $p = 1.5 \times 10^{-31}$; **Fig. 3.3A**), as predicted by the adaptation hypothesis.

To examine whether the above trend is truly adaptive or intrinsic, we analyzed each of the 500 random networks the same way we analyzed the *E. coli* network and computed for each random network Spearman's ρ between s and $\bar{\gamma}_G$. Because the number of nonessential reactions with nonzero fluxes varies among random networks, ρ values from different networks are not directly comparable. We therefore converted the ρ values of the 500 random networks and the *E. coli* network to standard z scores using Fisher's transformation of $z = \rho\sqrt{(n-3)/1.06}$, where n is the number of nonessential reactions with nonzero fluxes in the network (Fieller, Hartley, Pearson 1957). We found that the z score is more negative for the *E. coli* network than any of the 500 random networks ($P < 0.002$; **Fig. 3.3B**). We further confirmed that among 394 of the 500 random networks whose fitness under the glucose environment is higher than that of *E. coli*, none has a ρ that is more negative than that of *E. coli*. Hence, the negative correlation between s and $\bar{\gamma}_G$ observed in *E. coli* is beyond what the intrinsic property hypothesis can explain; natural selection for genetic or environmental robustness must be invoked.

3.3.6 The congruence hypothesis for genetic robustness is refuted

The congruence hypothesis makes three predictions. First, because the hypothesis asserts that environmental robustness results from direct selection, $\bar{\gamma}_{\mathrm{E}}$ and *s* should be negatively correlated in *E. coli* and the correlation should be stronger than what is exhibited in comparable random networks. Second, because the hypothesis posits that genetic robustness and environmental robustness are congruent, $\bar{\gamma}_{\mathrm{G}}$ and $\bar{\gamma}_{\mathrm{E}}$ should be positively correlated. Finally and most critically, because the hypothesis claims that genetic robustness is entirely a byproduct of selection for environmental robustness, $\bar{\gamma}_{\mathrm{G}}$ should no longer correlate with *s* after the control of $\bar{\gamma}_{\mathrm{E}}$.

We examined whether our data are consistent with the above three predictions of the congruence hypothesis. Consistent with the first prediction, $\bar{\gamma}_{\mathrm{E}}$ is negatively correlated with *s* ($\rho = -0.890$, $P = 5.5 \times 10^{-37}$; **Fig. 3.4A**) and the correlation is significantly stronger than those observed in the 500 comparable random networks ($P < 0.002$; **Fig. 3.4B**). Consistent with the second prediction, we found $\bar{\gamma}_{\mathrm{G}}$ to be highly and positively correlated with $\bar{\gamma}_{\mathrm{E}}$ (Spearman's $\rho = 0.911$, $P = 2.0 \times 10^{-41}$; **Fig. 3.4C**). But, contrary to the third prediction, the partial correlation between $\bar{\gamma}_{\mathrm{G}}$ and *s* after the control of $\bar{\gamma}_{\mathrm{E}}$ remains negative (Spearman's $\rho = -0.252$, $p = 9.6 \times 10^{-3}$) and is significantly stronger than the corresponding values from the 500 random networks ($P = 0.020$; **Fig. 3.4D**). Together, these results indicate that environmental robustness cannot fully explain genetic robustness. In other words, the congruence hypothesis for the origin of genetic robustness is rejected.

3.3.7 Results under single-carbon-source environments

Environmental robustness should be measured under the relevant environments of the species. Because of the paucity of such information for *E. coli*, in the above analyses, we applied 1000 random nutritional environments by using different combinations of carbon sources, which may have resulted in a less reliable estimate of environmental robustness. To examine the robustness of our results, we also used all possible single-carbon-source environments to quantify environmental robustness. Under the single-carbon-source environments, we found the new $\bar{\bar{\gamma}}_{\mathrm{E}}$ to be 0.53, much larger than the previous $\bar{\bar{\gamma}}_{\mathrm{E}}$ (**Fig. 3.2B**) and $\bar{\bar{\gamma}}_{\mathrm{G}}$ (Fig. **3.2A**), suggesting that on average the single-carbon-source environments represent severer challenges to *E. coli* than the 1000 random nutritional environments or flux constraints of $\Delta F = 0.5$ do. When applying the same analysis using single-carbon-source environments, we found all major results to remain qualitatively unchanged (**Fig. 3.5**). Because the single-carbon-source environments are the most extreme environments in terms of carbon source availability and represent all carbon-source challenges, this finding suggests that our conclusions are robust to the nutritional environments used and that the result in **Fig. 3.4D** is not an artifact of much smaller $\bar{\bar{\gamma}}_{\mathrm{E}}$ than $\bar{\bar{\gamma}}_{\mathrm{G}}$.

3.3.8 Environmental robustness as a side effect of genetic robustness

Although $\bar{\gamma}_{\mathrm{E}}$ is only marginally significantly smaller in *E. coli* than in the random networks (**Fig. 3.2B**), s and $\bar{\gamma}_{\mathrm{E}}$ are strongly correlated in *E. coli* (**Fig. 3.4A**) and this correlation is stronger than that in any of the 500 random networks (**Fig. 3.4B**). To examine whether the correlation between s and $\bar{\gamma}_{\mathrm{E}}$ is a byproduct of the adaptive genetic robustness, we examined the partial correlation between s and $\bar{\gamma}_{\mathrm{E}}$ after the control of $\bar{\gamma}_{\mathrm{G}}$. While this partial correlation is still significant ($\rho = -0.511$, $P = 2.8 \times 10^{-8}$), it is no longer significantly stronger in the *E. coli* network than in the random networks ($P = 0.060$; **Fig. A.2.2A**). This is also true under the single-carbon-

source environments (P = 0.39; **Fig. A.2.2B**). These results, coupled with those in **Fig. 3.4**, suggest that the environmental robustness of *E. coli* metabolic fluxes is likely a byproduct of the adaptive genetic robustness.

3.3.9 Capacitor reactions for genetic robustness of metabolic fluxes

What is the genetic mechanism of the genetic robustness of metabolic fluxes? One approach to this question is to identify reactions whose removal reduces the genetic robustness. Several previous studies used this approach to identify the so-called "capacitor" genes, which buffer genetic or environmental perturbations (Rutherford, Lindquist 1998; Levy, Siegal 2008; Takahashi 2013). We measured $\Delta\bar{\bar{\gamma}}_G$, the difference in $\bar{\bar{\gamma}}_G$ caused by the removal of an nonessential reaction. The more positive $\Delta\bar{\bar{\gamma}}_G$ is, the larger the contribution of the removed reaction to the genetic robustness of metabolic fluxes. To make the comparison fair, we only considered 96 reactions whose removal does not alter *E. coli* viability for each genetic perturbation considered. We found that $\Delta\bar{\bar{\gamma}}_G$ is positive in all 96 cases (**Fig. 3.6A**). The top 10% of the reactions in terms of the associated $\Delta\bar{\bar{\gamma}}_G$ values are marked in **Fig. 3.6A**, and these reactions contribute most to the genetic robustness of *E. coli* metabolic fluxes.

It is of particular interest to identify those reactions whose removal reduces the strength of the negative correlation between s and $\bar{\gamma}_G$ (i.e., renders ρ more positive or $\Delta\rho$ larger). **Fig. 3.6B** shows the frequency distribution of $\Delta\rho$ for the same 96 reaction removals, with those 10% of the reactions having the largest $\Delta\rho$ marked. Six genes are overlapped between the marked genes in **Fig. 3.6A and 6B** (marked blue). They are FUM (fumerate), SUCDi (succinate dehydrogenase), PSP_L (phosphoserine phosphatase), PSERT (phosphoserine transaminase), PGCD (phosphoglycerate dehydrogenase), and PPC (phosphoenolpyruvate carboxylase). FUM and SUCDi are part of the tricarboxylic acid (TCA) cycle, while PSP_L, PSERT, and PGCD

appear in serine anabolism and are closely related to some metabolites of the TCA cycle (**Fig. 3.6C**). In addition, PPC is an anaplerotic reaction replenishing metabolites in the TCA cycle that are largely consumed by anabolism (Nelson, Cox 2008). These observations suggest the biological importance of the TCA cycle for the genetic robustness of *E. coli*'s metabolic fluxes.

Because of the central role of the TCA cycle in aerobic metabolism, one wonders whether the six "capacitor" reactions simply reflect an intrinsic property of metabolic networks. We found that these reactions exist in 24% (PSP_L) to 39% (FUM) of the 500 random networks examined. We quantified the contributions of these reactions to the genetic robustness of the random networks by measuring $\Delta\bar{\bar{\gamma}}_G$ and $\Delta\rho$, as was conducted for the *E. coli* network. Under the same criteria used for examining the *E. coli* network, the number of usable random networks reduced to <5 for three reactions in serine metabolism (PSP_L, PSERT, and PGCD) and thus cannot be evaluated with any statistical meaning. For the other three reactions (FUM, PPC, and SUCDi), we found that $\Delta\bar{\bar{\gamma}}_G$ and $\Delta\rho$ are smaller in random networks than in *E. coli* (**Fig. 3.7**), suggesting that at least these three capacitor reactions indeed contribute to genetic robustness beyond the random expectation.

3.4 DISCUSSION

Several previous studies revealed the genetic robustness of metabolic networks by treating the viability as a trait (Edwards, Palsson 2000; Samal et al. 2010), but these studies did not and could not resolve the evolutionary origin of the observed genetic robustness. In this work, we first demonstrated that *E. coli* metabolic fluxes are robust to mutation. We then showed that neither the intrinsic property hypothesis nor the congruence hypothesis adequately explains the origin of the observed genetic robustness, supporting the hypothesis that direct

selection underlies genetic robustness. Our study has several caveats that are worth discussion. First, FBA and MOMA predictions of fitness and flux contain errors. In particular, MOMA minimizes the sum of squared flux changes upon mutation or environmental perturbation and thus may underestimate actual flux changes. However, because our conclusion is based on the comparison between real and random networks that are subject to the same analyses, these errors are not expected to bias our conclusion, although stochastic errors may have reduced the statistical power of our comparison and made our conclusion conservative. Second, under MOMA's criteria, large wild-type fluxes are expected to have bigger changes, which could bias our results because reaction importance and flux are positively correlated (Spearman's $\rho = 0.626$; $p = 8.7 \times 10^{-13}$). However, this bias is presumably removed by using fractional flux changes in calculating γ_G and γ_E. Further, the comparison between the real and random networks should remove the potential impact of any remaining bias on our conclusion. Third, we mimicked mutation by individually constraining metabolic fluxes by a certain fraction (ΔF). Real mutations of course occur in genes rather than reactions. Because one gene may affect multiple reactions and one reaction may be catalyzed by multiple isozymes or an enzyme made up of multiple peptides, the relationship between genes and reactions is not always one-to-one. Nevertheless, mutational effects on metabolism are usually manifested as different degrees of flux constraints. A mutation may occasionally lead to the gain of a new reaction or rewiring of the metabolic network. These possibilities are ignored in our estimate of genetic robustness because they are much less frequent than mutations that result in flux constraints. Mutations may also lead to a loosened flux constraint, which could increase rather than decrease fitness. These mutations are not considered because genetic robustness is about buffering deleterious

mutations and because advantageous mutations are orders of magnitude less common than deleterious mutations.

The intrinsic property hypothesis, positing that genetic robustness is an intrinsic property that arises without selection for robustness, was put forward based on observations made in simulations of regulatory network evolution (Siegal, Bergman 2002). Interestingly, we found that the correlation between *s* and $\bar{\gamma}_{\mathrm{G}}$ is significantly negative for 98.0% of the random metabolic networks (nominal $P < 0.05$), suggesting that intrinsic origins of a certain degree of genetic robustness may be quite common. Nonetheless, the observed genetic robustness of *E. coli* metabolic fluxes is well beyond the intrinsic level (**Fig. 3.3A**) and hence must involve selection. In this sense, genetic robustness may often have both intrinsic and adaptive origins.

Similar to genetic robustness, the correlation between *s* and $\bar{\gamma}_{\mathrm{E}}$ is significantly negative for 99.6% of the random metabolic networks, suggesting that environmental robustness could arise as an intrinsic property. Additional evidence (**Fig. A.2.2**), however, suggests that the environmental robustness of *E. coli* metabolic fluxes is likely a side effect of selection for genetic robustness, contrary to the congruence hypothesis that genetic robustness is a byproduct of selection for environmental robustness. The congruence hypothesis was proposed because environmental robustness was thought to be subject to stronger selection than was genetic robustness (Wagner, Booth, Bagheri-Chaichian 1997; Meiklejohn, Hartl 2002), based on the report that phenotypic variation caused by environmental perturbation is much larger than that caused by mutation (Lynch 1988). However, this observation may be a consequence of higher genetic robustness than environmental robustness, instead of a cause for stronger selection for environmental robustness than genetic robustness. To our knowledge, no experiment has been conducted to differentiate between these two scenarios. Consequently, whether environmental

robustness is subject to stronger or weaker selection than genetic robustness cannot be predicted theoretically.

While we identified capacitor reactions that make relatively large contributions to the adaptive genetic robustness of *E. coli*'s metabolic fluxes, it is worth discussing other possible mechanisms. First, chaperons such as GroEL in bacteria (Fares et al. 2002) and Hsp90 in *Drosophila* (Rutherford, Lindquist 1998) are known to provide genetic and environmental robustness by aiding protein folding. But the genetic robustness revealed here is presumably due to some systemic properties, because our analysis does not involve structure-function relations of individual proteins. Second, at the network function level, increasing the number of regulations was shown to enhance the genetic robustness of a simulated regulatory network (Siegal, Bergman 2002) and the power-law distribution of node connectivity was suggested to enhance the metabolic network robustness (Jeong et al. 2000). However, these results are not directly applicable to our study because we focus on individual fluxes rather than the entire network function. For instance, we found no significant correlation between the number of reactions that are directly connected to a focal reaction and $\bar{\gamma}_G$ of the focal reaction (ρ = -0.014, P = 0.79). Furthermore, the *E. coli* network and the random networks have no significant difference in the number of reactions that an average reaction is directly connected to, although the *E. coli* network has a lower $\bar{\bar{\gamma}}_G$ and a stronger correlation between s and $\bar{\bar{\gamma}}_G$ than those of the random networks (**Fig. A.2.3**). Third, in theory, functional redundancy can improve genetic robustness. Although reactions with various degrees of functional redundancy are common in the metabolic networks of *E. coli*, their potential backup role is not needed to explain their evolutionary maintenance (Wang, Zhang 2009a). In other words, functionally redundant reactions are unlikely to be the primary basis of the observed adaptive genetic robustness. Fourth, it is known

that the flux of a linear metabolic pathway is intrinsically robust to concentration changes of the enzymes catalyzing the reactions in the pathway and that the robustness increases with the pathway length (Kacser, Burns 1981; Wang, Zhang 2011). So, in principle natural selection could act on the length of a linear pathway to increase the genetic robustness of its constituent reactions. However, this mechanism is difficult to test empirically, because pathways are not easily discernable in a complex network (Wang, Zhang 2011).

The present finding of direct selection for genetic robustness of *E. coli* metabolic fluxes is consistent with our previous findings in 220 yeast morphological traits and over 3000 gene expression traits (Ho, Zhang 2014), suggesting that adaptive originations of genetic robustness may be widespread among different classes of phenotypes. These empirical results support the theoretical prediction that, under certain conditions, direct selection is sufficiently powerful to promote genetic robustness in cellular organisms (Wagner, Booth, Bagheri-Chaichian 1997; Ho, Zhang 2014). Our finding not only answers the long-stranding question on the origin of genetic robustness but also has other implications. For instance, one important question in the study of the genotype-phenotype relationship is why different traits are affected by mutations to different degrees. Our finding provides one explanation that, because of adaptive selection for genetic robustness, modifiers evolve to preferentially buffer the mutational effect on more important traits. In other words, the adaptive evolution of genetic robustness may result in rewiring of the genotype-phenotype map and impact the evolutionary trajectory.

3.5 MATERIALS AND METHODS

3.5.1 Metabolic models

We used *E.coli* metabolic network model iAF1260 (Feist et al. 2007), which includes 2381 reactions and 1039 metabolites. Among all reactions, 2082 are metabolic reactions and 299 are exchange reactions allowing the uptake of nutrients from the environment. The SMBL file of the iAF1260 model was downloaded from BiGG (Schellenberger et al. 2010) and parsed by COBRA (Becker et al. 2007). We chose iAF1260 because it outperforms other *E. coli* metabolic models in predicting gene essentiality and other properties (Feist et al. 2007).

3.5.2 Flux balance analysis (FBA)

Briefly, under the steady state assumption, FBA formulates a linear programming problem to determine the flux of each reaction to maximize the production of biomass (Orth, Thiele, Palsson 2010). Mathematically, the objective is to maximize the flux of the biomass reaction, which describes the relative contributions of various metabolites to the cellular biomass, under the constraints of $\boldsymbol{Sv} = \mathbf{0}$ and $\boldsymbol{\alpha} \leq \boldsymbol{v} \leq \boldsymbol{\beta}$. Here, $\boldsymbol{v}$ is a vector of reaction fluxes, $\boldsymbol{S}$ is a matrix describing the stoichiometric relationships among metabolites in each reaction, $\boldsymbol{\alpha}$ is a vector describing the lower bound of each flux, and $\boldsymbol{\beta}$ is a vector describing the upper bound of each flux.

We used the default $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in the metabolic model for all reactions to perform FBA. This default setting represents the parameters for wild-type cells in a minimal medium with limited glucose being the sole carbon source and some common inorganic compounds such as water, oxygen, carbon dioxide, and ammonium (Feist et al. 2007). Note that a reaction may be reversible or irreversible. For a reversible metabolic reaction i, the default $\alpha_i = -\infty$ and $\beta_i = \infty$, whereas for a reversible metabolic reaction i, the default $\alpha_i = 0$ and $\beta_i = \infty$. The optimized fluxes ($\boldsymbol{v}_0$) of the wild-type model under the glucose environment serve as the baseline for computing fractional flux changes upon mutation or environmental shift.

Under the glucose environment, the *E. coli* metabolic model has 387 nonzero-flux reactions. Note that some of them are simple diffusions of metabolites between different cellular compartments. Because these reactions do not have dedicated enzymes and are not "mutable", we excluded these reactions from our dataset of traits and used the remaining 362 reactions.

All linear programming problems in this study were solved by the cplexlp function of the IBM ILOG CPLEX Optimizer with MATLAB interface.

3.5.3 Minimization of metabolic adjustment (MOMA)

The objective of MOMA is to minimize $(\boldsymbol{v} - \boldsymbol{v}_0)^2$, where $\boldsymbol{v}$ is a vector of all reaction fluxes upon a genetic or environmental perturbation, whereas $\boldsymbol{v}_0$ is the corresponding vector for the wild-type in the glucose environment described in the previous section. The two constraints for FBA are also applied in MOMA. All quadratic programming problems in this study were solved by the cplexqp function of the IBM ILOG CPLEX Optimizer with MATLAB interface.

When introducing genetic perturbation to a reaction, we altered its lower bound flux (α) and upper bound flux (β) according to its wild-type flux in the glucose environment (v_0) and the fractional flux constraint (ΔF) imposed, but kept the α and β of other reactions unchanged. For example, constraining a flux by 10% means that the flux of the reaction cannot exceed 90% the wild-type value. Specifically, if this reaction is reversible, we use $-0.9|v_0| \leqslant v \leqslant 0.9|v_0|$; otherwise, we use $0 \leqslant v \leqslant 0.9v_0$.

When introducing environmental changes, we followed the approach previously published (Wang, Zhang 2009b) to simulate random environments. In iAF1260, there are 258 exchange reactions for 258 carbon sources (Feist et al. 2007). In each simulation, we randomly picked a number q for each carbon source following an exponential distribution with mean = 0.1. Here, q is the probability that the carbon source is available, and the actual availability is

determined stochastically based on q. Because in the default setting, the uptake rate of glucose was set as 10 mmol$g$DW^{-1}h^{-1} (Feist et al. 2007), we used the same rate for all organic chemicals when they are available. In addition, in each simulated random environment used, we required that *E. coli* and all 500 random networks are viable. In total, 1000 such environments were used. The first, second, and third quartiles of the number of carbon sources in a random environment are 24, 40, and 70.5, respectively. The first, second, and third quartiles of the number of random environments where a carbon source exists are 160, 180, and 200, respectively.

To examine the robustness of our results, we also used 257 single-carbon-source environments to mimic environmental changes from the glucose environment (**Fig. 3.5**). In this case, we did not require all 501 networks to be viable in an environment because none of the 257 single-carbon-source environments could support all 501 networks.

3.5.4 Fractional flux change

After calculating the flux of a reaction v from MOMA, we used the following formula to calculate fractional flux change (γ) from the flux of the reaction in the wild-type under the glucose environment (v_0). To make increasing and decreasing fluxes comparable, we normalize it in different ways depending on whether v is larger or smaller than v_0. A larger γ means larger difference.

$$\gamma = \begin{cases} 1 - v/v_0, & v < v_0 \\ 1 - v_0/v, & v \geq v_0 \end{cases}$$

3.5.5 Random networks

We used a previously published approach to generate 500 random networks (Rodrigues, Wagner 2009; Barve, Wagner 2013). We first compiled "the universe of reactions" by acquiring metabolic reactions listed in the REACTION section of the KEGG database (Kanehisa, Goto 2000) except when (i) a reaction appears in the *E. coli* metabolic network iAF1260; (ii) a

reaction involves polymer subunits with uncertain number of atoms; (iii) a reaction involves glycans; (iv) a reaction involves metabolites without information about their structure; (v) a reaction is unbalanced in mass or charges. Then we combined these reactions with the metabolic reactions in iAF1260 (excluding transport reactions). In total, there are 5001 reactions in the universe of reactions. When generating random networks, we performed a random walk in the space of networks by starting from the *E. coli* metabolic network iAF1260 and iteratively swapping between a randomly picked reaction from the current network and a randomly picked reaction from the universe of reactions that is different from any reaction in the current network, under the condition that the current network is viable in the glucose environment. Thus, the random networks generated have the same number of reactions as the *E. coli* network, but has been subject to neither selection for genetic robustness nor selection for environmental robustness. During the random walk, we sampled a random network after every 5000 swaps until we acquired 500 random networks. It is desirable for the random walk to effectively travel in the whole space with no bias, which can be shown by the saturation of the proportion of iAF1260 reactions that are absent from our sampled random networks in the time series (**Fig. A.2.4**). Among the random networks, the first, second, and third quartiles of network connectivity, defined by the number of reactions connected directly with a metabolite averaged across all metabolites, are 3.34, 3.36, and 3.38, respectively.

3.6 ACKNOWLEDGEMENTS

We thank Xinzhu Wei and Jian-Rong Yang for valuable comments. This work was supported in part by U.S. National Institutes of Health grant GM103232 and U.S. National Science Foundation grant MCB-1329578 to J.Z.

3.7 REFERENCES

Barve, A, JF Rodrigues, A Wagner. 2012. Superessential reactions in metabolic networks. Proc Natl Acad Sci U S A 109:E1121-1130.

Barve, A, A Wagner. 2013. A latent capacity for evolutionary innovation through exaptation in metabolic systems. Nature 500:203-206.

Becker, SA, AM Feist, ML Mo, G Hannum, BO Palsson, MJ Herrgard. 2007. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. Nat Protoc 2:727-738.

Bordbar, A, JM Monk, ZA King, BO Palsson. 2014. Constraint-based models predict metabolic and associated cellular functions. Nature reviews. Genetics 15:107-120.

Ciliberti, S, OC Martin, A Wagner. 2007. Innovation and robustness in complex regulatory gene networks. Proc Natl Acad Sci U S A 104:13591-13596.

Clark, AG. 1991. Mutation-selection balance and metabolic control-theory. Genetics 129:909-923.

Costenoble, R, P Picotti, L Reiter, R Stallmach, M Heinemann, U Sauer, R Aebersold. 2011. Comprehensive quantitative analysis of central carbon and amino-acid metabolism in *Saccharomyces cerevisiae* under multiple conditions by targeted proteomics. Molecular Systems Biology 7:464.

de Visser, JA, J Hermisson, GP Wagner, et al. 2003. Perspective: Evolution and detection of genetic robustness. Evolution 57:1959-1972.

Edwards, JS, BO Palsson. 2000. Robustness analysis of the *Escherichia coli* metabolic network. Biotechnology Progress 16:927-939.

Fares, MA, MX Ruiz-Gonzalez, A Moya, SF Elena, E Barrio. 2002. Endosymbiotic bacteria: groEL buffers against deleterious mutations. Nature 417:398.

Feist, AM, CS Henry, JL Reed, M Krummenacker, AR Joyce, PD Karp, LJ Broadbelt, V Hatzimanikatis, BO Palsson. 2007. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Molecular Systems Biology 3:121.

Felix, MA, M Barkoulas. 2015. Pervasive robustness in biological systems. Nature Reviews Genetics 16:483-496.

Fieller, EC, HO Hartley, ES Pearson. 1957. Tests for rank correlation coefficients .1. Biometrika 44:470-481.

Fong, SS, BO Palsson. 2004. Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. Nature Genetics 36:1056-1058.

Harcombe, WR, NF Delaney, N Leiby, N Klitgord, CJ Marx. 2013. The ability of flux balance analysis to predict evolution of central metabolism scales with the initial distance to the optimum. Plos Computational Biology 9:e1003091.

He, XL, WF Qian, Z Wang, Y Li, JZ Zhang. 2010. Prevalent positive epistasis in Escherichia coli and Saccharomyces cerevisiae metabolic networks. Nature Genetics 42:272-U120.

Ho, WC, JZ Zhang. 2014. The genotype-phenotype map of yeast complex traits: basic parameters and the role of natural selection. Molecular Biology and Evolution 31:1568-1580.

Houle, D. 1998. How should we explain variation in the genetic variance of traits? Genetica 102-3:241-253.

Ibarra, RU, JS Edwards, BO Palsson. 2002. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature 420:186-189.
Jeong, H, B Tombor, R Albert, ZN Oltvai, AL Barabasi. 2000. The large-scale organization of metabolic networks. Nature 407:651-654.
Kacser, H, JA Burns. 1981. The molecular basis of dominance. Genetics 97:639-666.
Kanehisa, M, S Goto. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research 28:27-30.
Levy, SF, ML Siegal. 2008. Network Hubs Buffer Environmental Variation in Saccharomyces cerevisiae. Plos Biology 6:2588-2604.
Lewis, NE, KK Hixson, TM Conrad, et al. 2010. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. Molecular Systems Biology 6:390.
Lynch, M. 1988. The rate of polygenic mutation. Genetical Research 51:137-148.
Meiklejohn, CD, DL Hartl. 2002. A single mode of canalization. Trends in Ecology & Evolution 17:468-473.
Nelson, D, M Cox. 2008. Lehniger Principles of Biochemistry. New York: W. H. Freeman and Company.
Orth, JD, I Thiele, BO Palsson. 2010. What is flux balance analysis? Nature Biotechnology 28:245-248.
Proulx, SR, S Nuzhdin, DE Promislow. 2007. Direct selection on genetic robustness revealed in the yeast transcriptome. PLoS One 2:e911.
Rausher, MD. 2013. The evolution of genes in branched metabolic pathways. Evolution 67:34-48.
Rodrigues, JFM, A Wagner. 2009. Evolutionary plasticity and innovations in complex metabolic reaction networks. Plos Computational Biology 5:e1000613.
Rutherford, SL, S Lindquist. 1998. Hsp90 as a capacitor for morphological evolution. Nature 396:336-342.
Samal, A, JFM Rodrigues, J Jost, OC Martin, A Wagner. 2010. Genotype networks in metabolic reaction spaces. Bmc Systems Biology 4:30.
Sanjuan, R, JM Cuevas, V Furio, EC Holmes, A Moya. 2007. Selection for robustness in mutagenized RNA viruses. Plos Genetics 3:939-946.
Sanjuan, R, A Moya, SF Elena. 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. Proc Natl Acad Sci U S A 101:8396-8401.
Schellenberger, J, JO Park, TM Conrad, BO Palsson. 2010. BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. Bmc Bioinformatics 11:213.
Segre, D, D Vitkup, GM Church. 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci U S A 99:15112-15117.
Siegal, ML, A Bergman. 2002. Waddington's canalization revisited: Developmental stability and evolution. Proc Natl Acad Sci U S A 99:10528-10532.
Stearns, SC, M Kaiser, TJ Kawecki. 1995. The differential genetic and environmental canalization of fitness components in *Drosophila melanogaster*. J Evol Biol 8:539-557.
Stearns, SC, TJ Kawecki. 1994. Fitness sensitivity and the canalization of life-history traits. Evolution 48:1438-1450.
Takahashi, KH. 2013. Multiple capacitors for natural genetic variation in Drosophila melanogaster. Molecular Ecology 22:1356-1365.

Waddington, CH. 1942. Canalization of development and the inheritance of acquired characters. Nature 150:563–565.
Wagner, A. 2005. Robustness and Evolvability in Living Systems. Princeton, NJ: Princeton University Press.
Wagner, G, G Booth, H Bagheri-Chaichian. 1997. A population genetic theory of canalization. Evolution 51:329-347.
Wang, Z, J Zhang. 2011. Impact of gene expression noise on organismal fitness and the efficacy of natural selection. Proc Natl Acad Sci U S A 108:E67-76.
Wang, Z, JZ Zhang. 2009a. Abundant indispensable redundancies in cellular metabolic networks. Genome Biology and Evolution 1:23-33.
Wang, Z, JZ Zhang. 2009b. Why is the correlation between gene importance and gene evolutionary rate so weak? Plos Genetics 5: e1000329.
Wilke, CO, JL Wang, C Ofria, RE Lenski, C Adami. 2001. Evolution of digital organisms at high mutation rates leads to survival of the flattest. Nature 412:331-333.
Wloch, DM, K Szafraniec, RH Borts, R Korona. 2001. Direct estimate of the mutation rate and the distribution of fitness effects in the yeast *Saccharomyces cerevisiae*. Genetics 159:441-452.
Yang, JR, S Ruan, J Zhang. 2014. Determinative developmental cell lineages are robust to cell deaths. Plos Genetics 10:e1004501.

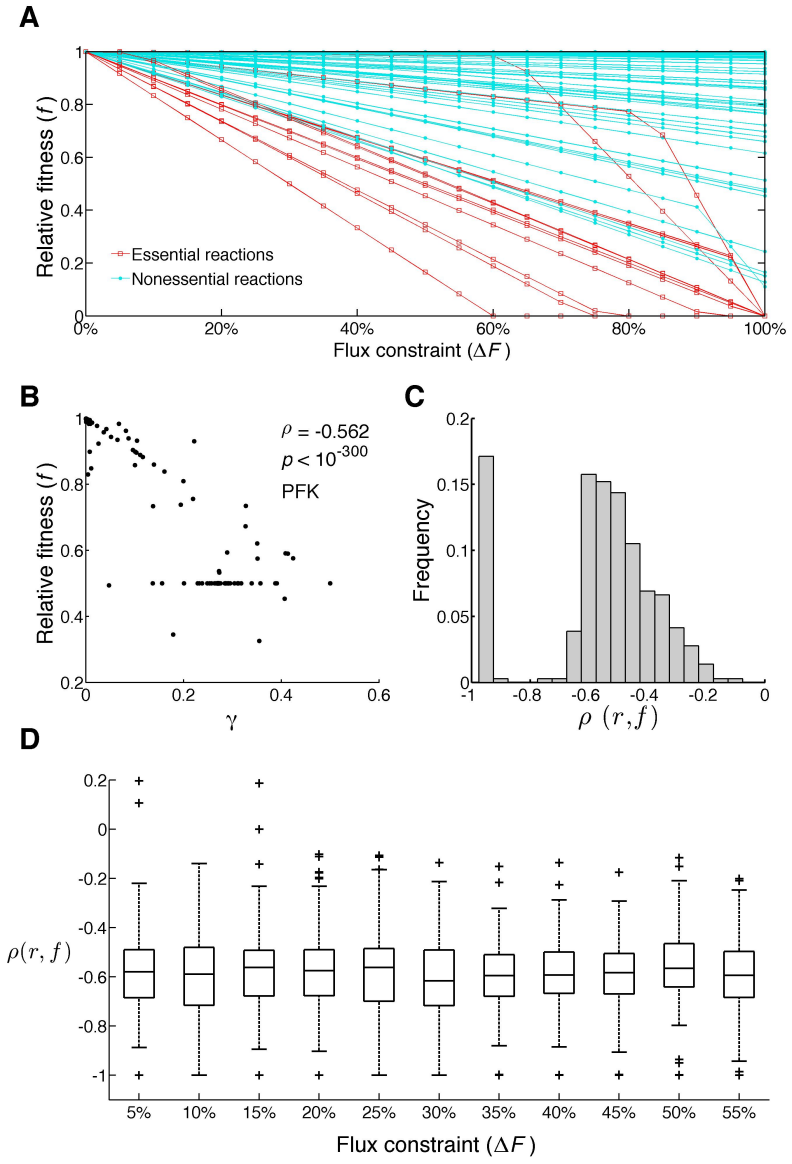

Figure 3.1 Changes of metabolic reaction fluxes from the wild-type levels predict fitness decreases. (A) FBA-predicted fitness decreases with the rise of the extent to which the flux of a reaction is constrained (ΔF). Each line represents a serial constraints imposed on the flux of one reaction. (B) Fitness (f) is negatively correlated with the fractional flux change (γ) for the focal reaction PGI upon the constraining of another reaction at $\Delta F = 50\%$. Each dot represents the result from constraining one reaction. (C) Frequency distribution among nonessential focal reactions of Spearman's correlation coefficient (ρ) between γ and f under $\Delta F = 50\%$. (D) Box plot of ρ between γ and f at various ΔF. In each box plot, the lower edge and upper edge of a box represent the first (q_1) and third quartiles (q_3), respectively. The horizontal line inside the box indicates the median (md). The whiskers extend to the most extreme values inside inner fences, $md \pm 1.5(q_3 - q_1)$. The values outside the inner fences (outliers) are plotted by plus signs.

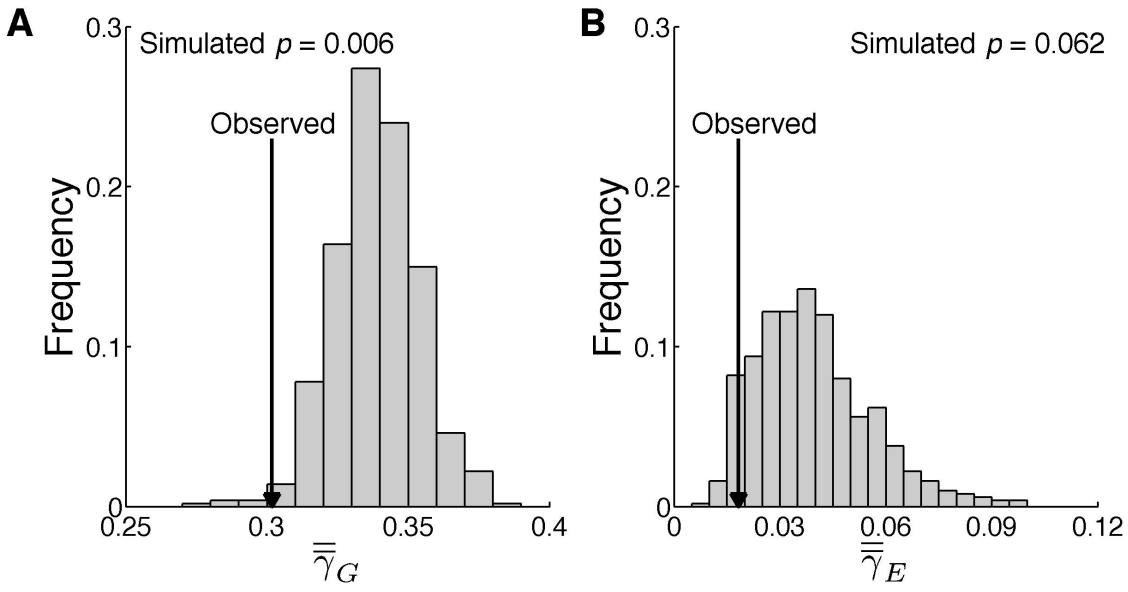

Figure 3.2 Comparison of genetic and environmental robustness between *E. coli* and random networks. (A) Distribution of the average fractional flux change across all focal reactions and mutants ($\bar{\bar{\gamma}}_G$) in 500 random networks. The arrow indicates the corresponding value observed in *E. coli*. (B) Distribution of the average fractional flux change across all focal reactions and 1000 random nutritional environments ($\bar{\bar{\gamma}}_E$) in 500 random networks. The arrow indicates the corresponding value observed in *E. coli*.

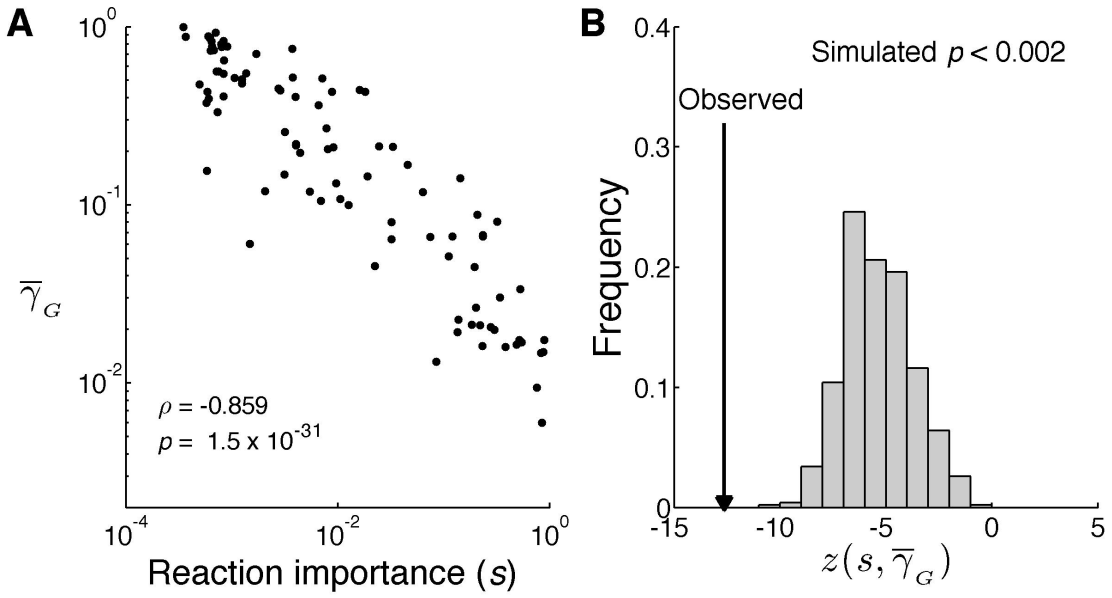

Figure 3.3 The intrinsic hypothesis of genetic robustness is rejected. (A) The average fractional flux change for a focal reaction among all mutants ($\bar{\gamma}_G$) decreases with the rise of reaction importance (s). Each dot represents one focal reaction. (B) Frequency distribution of the rank correlation (after conversion to z) between $\bar{\gamma}_G$ and s in 500 random networks. Arrow indicates the corresponding z observed in *E. coli*.

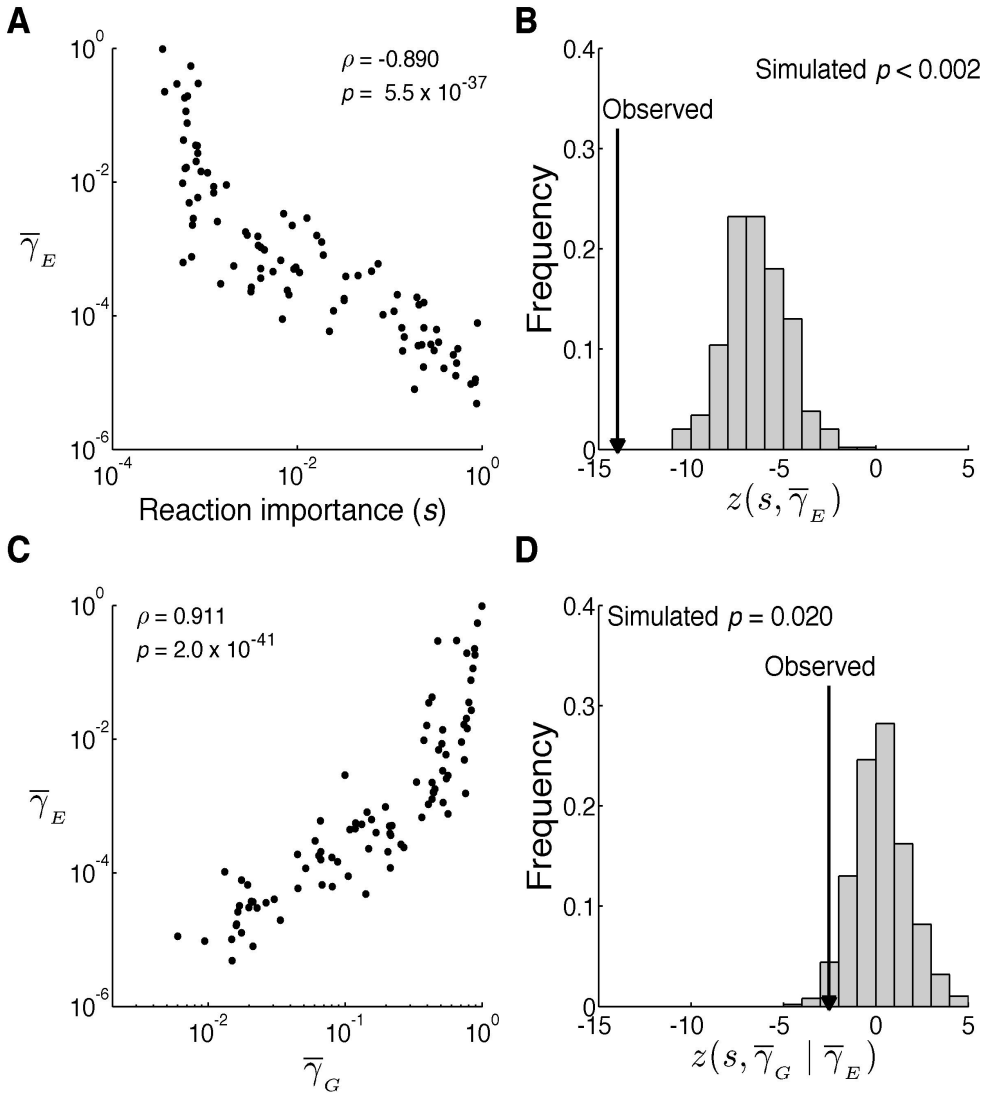

Figure 3.4 The congruence hypothesis of genetic robustness is rejected. (A) The average fractional flux change for a focal reaction among all 1000 environments examined ($\bar{\gamma}_E$) decreases with the rise of reaction importance (*s*). Each dot represents a focal reaction. (B) Frequency distribution of the rank correlation (after conversion to *z*) between $\bar{\gamma}_E$ and *s* in 500 random networks. Arrow indicates the corresponding *z* observed in *E. coli*. (C) Average fractional flux change for a focal reaction upon mutation and that upon environmental change are strongly correlated. Each dot represents a focal reaction. (D) Frequency distribution of the rank correlation (after conversion to *z*) between $\bar{\gamma}_G$ and *s* after the control of $\bar{\gamma}_E$ among the 500 random networks. Arrow indicates the corresponding *z* observed in *E. coli*.

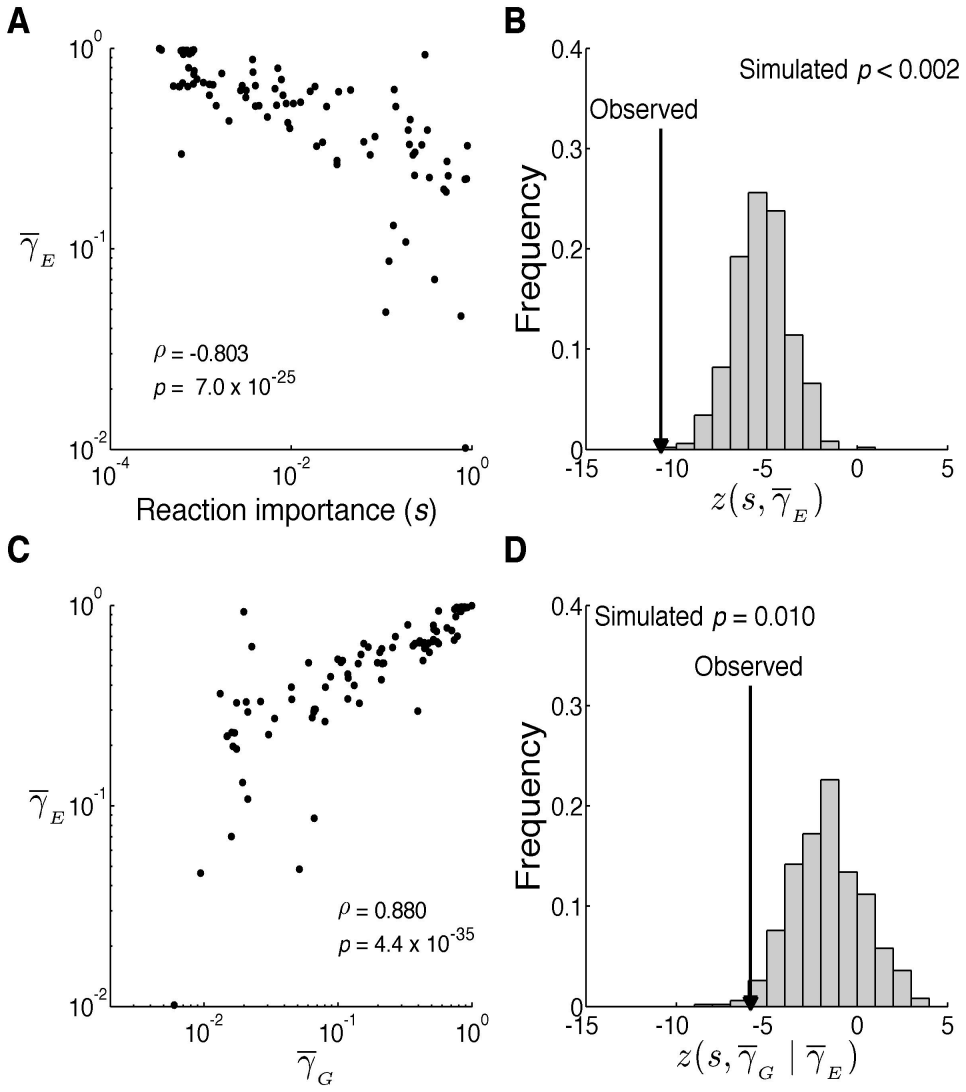

Figure 3.5 Confirmation of results using single-carbon-source environments. (A) The average fractional flux change for a focal reaction among all 1000 environments examined ($\bar{\gamma}_E$) decreases with the rise of reaction importance (*s*). Each dot represents a focal reaction. (B) Frequency distribution of the rank correlation (after conversion to *z*) between $\bar{\gamma}_E$ and *s* in 500 random networks. Arrow indicates the corresponding *z* observed in *E. coli*. (C) Average fractional flux change for a focal reaction upon mutation ($\bar{\gamma}_G$) and that upon environmental change ($\bar{\gamma}_E$) are strongly correlated. Each dot represents a focal reaction. (D) Frequency distribution of the rank correlation (after conversion to *z*) between $\bar{\gamma}_G$ and *s* after the control of $\bar{\gamma}_E$ among the 500 random networks. Arrow indicates the corresponding *z* observed in *E. coli*.

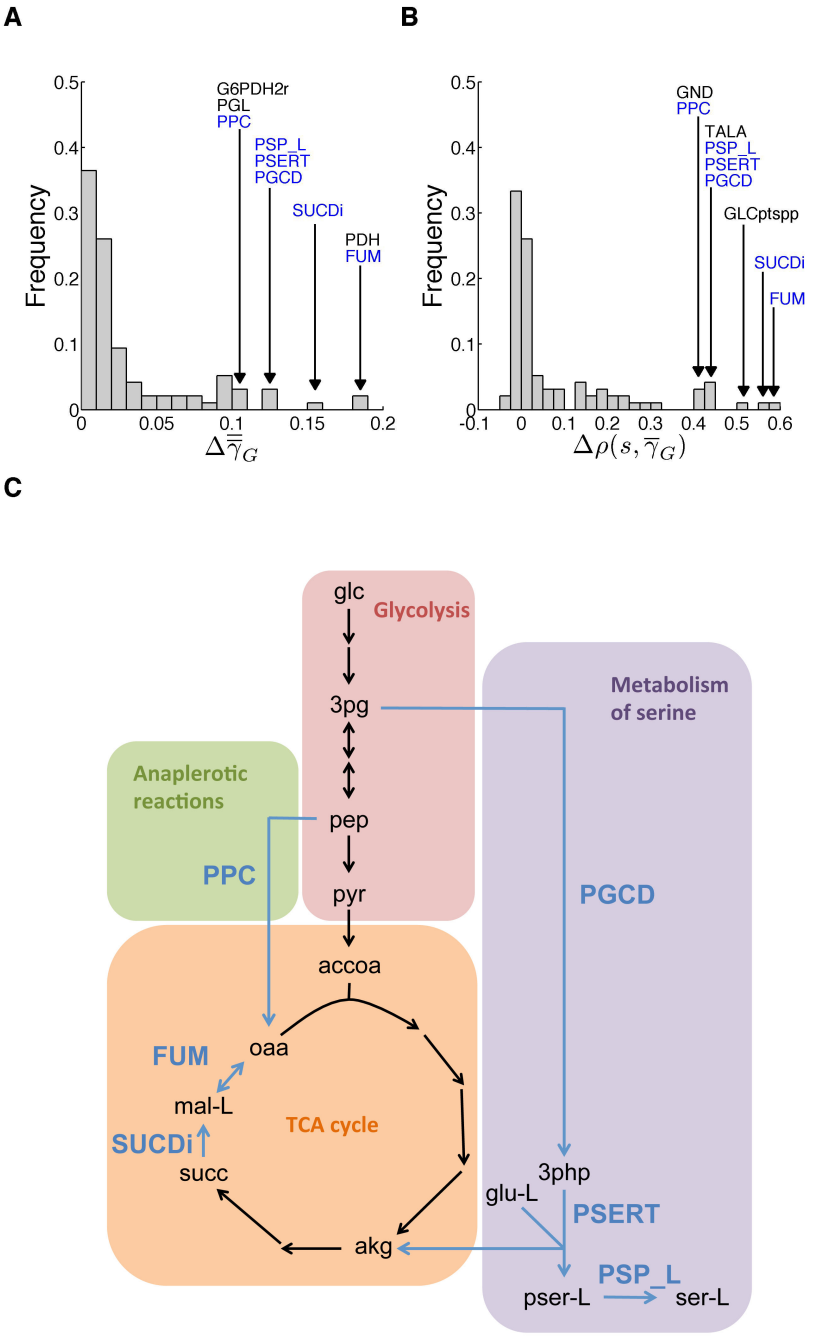

Figure 3.6 Removing reactions from the *E. coli* metabolic network affects how it behaves under genetic perturbations. (A) Frequency distribution of $\Delta\bar{\bar{\gamma}}_G$ caused by the removal of one of the 96 reactions examined. (B) Frequency distribution of $\Delta\rho(s, \bar{\gamma}_G)$ caused by the removal of one of the 96 reactions examined. In (A) and (B), reaction names are shown for those in the top 10% of the distributions, with the six common reactions in the top 10% sets of the two panels marked blue. (C) Part of the *E. coli* metabolic network containing the six capacitor reactions. The six capacitor reactions are marked blue. The abbreviations of reactions and metabolites follow (Feist et al. 2007).

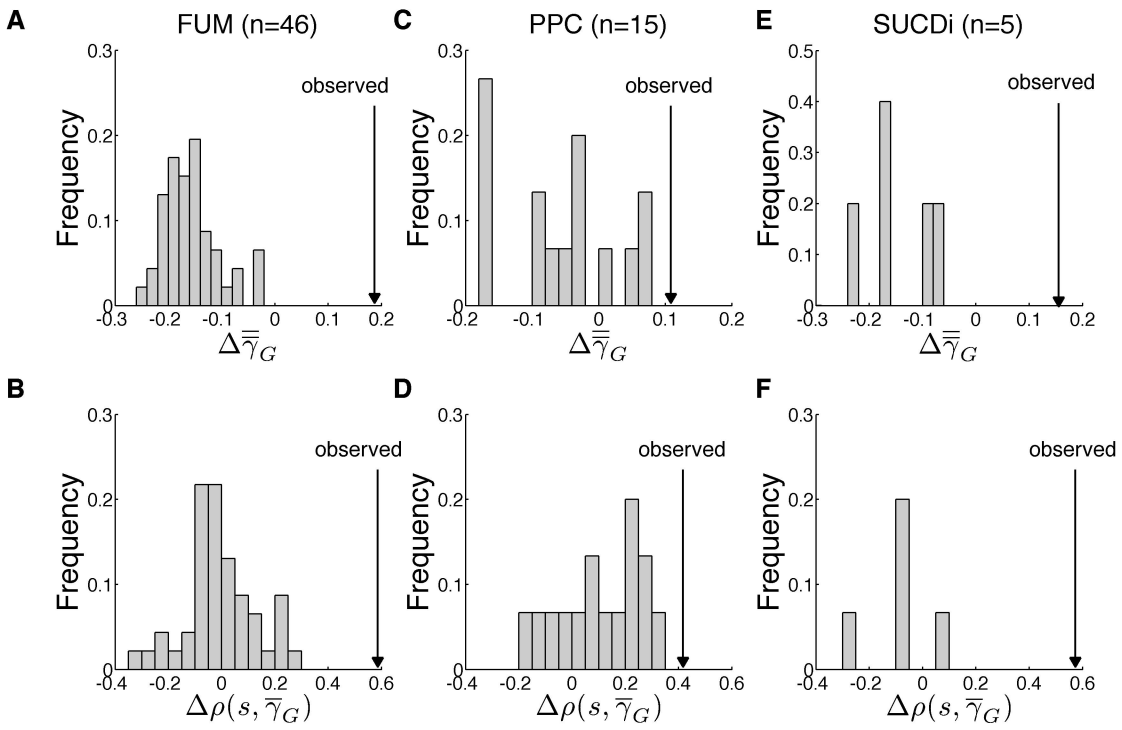

Figure 3.7 Removing capacitor reactions does not affect the genetic robustness in random networks as much as in *E. coli*. (A) Frequency distribution of $\Delta\overline{\overline{\gamma}}_G$ in 46 analyzed random networks upon the removal of FUM. (B) Frequency distribution of $\Delta\rho(s, \overline{\gamma}_G)$ in 46 analyzed random networks upon the removal of FUM. (C) Frequency distribution of $\Delta\overline{\overline{\gamma}}_G$ in 15 analyzed random networks upon the removal of PPC. (D) Frequency distribution of $\Delta\rho(s, \overline{\gamma}_G)$ in 15 analyzed random networks upon the removal of PPC. (E) Frequency distribution of $\Delta\overline{\overline{\gamma}}_G$ in 5 analyzed random networks upon the removal of SUCDi. (F) Frequency distribution of $\Delta\rho(s, \overline{\gamma}_G)$ in 5 analyzed random networks upon the removal of SUCDi.

Chapter 4

Testing the Neutral Hypothesis of Phenotypic Evolution

4.1 ABSTRACT

Although evolution by natural selection is widely regarded as the most important principle of biology, it is unknown whether most phenotypic variations within and between species are adaptive or neutral, due to the lack of relevant studies of large, unbiased samples of phenotypic traits. Here we examine 210 yeast morphological traits chosen purely on the basis of experimental feasibility irrespective of their potential adaptive values. After controlling for mutational size, we find faster evolution of more important morphological traits within and between species, rejecting the neutral hypothesis. By contrast, an analysis of 3466 gene expression traits that are similarly chosen fails to reject neutrality. Thus, yeast morphological evolution is largely adaptive, but the same may not apply to other classes of phenotypes.

4.2 INTRODUCTION

A large fraction of genomic sequence variations within and between species are neutral or nearly so (Kimura 1983). Whether the same is true for phenotypic variations is a central question in biology (Darwin 1859; Lande 1976; Lynch and Hill 1986; Mayr 2001; Nei 2007; Futuyma 2013). On the one hand, numerous phenotypic adaptations have been documented (Darwin 1859; Endler 1986; Kingsolver et al. 2001) and even Kimura, the champion of the neutral theory of molecular evolution, believed in widespread adaptive phenotypic evolution (Kimura 1983). On the other hand, phenotypic studies are strongly biased toward traits that are likely adaptive (Kingsolver et al. 2001), contrasting genomic studies that are typically unbiased. It is thus desirable to test the neutral hypothesis of phenotypic evolution using traits irrespective of their potential involvement in adaptation. Here we present such a test for 210 morphological traits measured in multiple strains and species of budding yeast. Our test is based on the premise that, under neutrality, the rate of phenotypic evolution declines as the trait becomes more important to fitness, analogous to the neutral paradigm that functional genes evolve more slowly than functionless pseudogenes (Li et al. 1981). Neutrality is rejected in favor of adaptation if important traits evolve faster than less important ones, parallel to the demonstration of molecular adaptation when a functional gene evolves faster than pseudogenes. After controlling for mutational size, we find faster evolution of more important morphological traits within and between species. By contrast, an analysis of 3466 yeast gene expression traits fails to reject neutrality. Thus, yeast morphological evolution is largely adaptive, but the same may not apply to other classes of phenotypes.

4.3 RESULTS

Analogous to the neutral hypothesis of molecular evolution (Kimura 1983), the neutral hypothesis of phenotypic evolution allows the presence of purifying selection; neutrality is rejected only when positive selection is invoked. Under the neutral hypothesis, compared with traits that are relatively unimportant to fitness, relatively important traits should be subject to stronger purifying selection and evolve more slowly given the same speed of mutational input (**Fig. 4.1A**). However, if relatively important traits evolve faster than relatively unimportant traits, the neutral hypothesis would no longer hold and the only reasonable explanation would be stronger positive selection acting on relatively important traits than relatively unimportant ones (**Fig. 4.1B**). This test of phenotypic neutrality differs from previous tests (Lande 1976; Lande 1977; Chakraborty and Nei 1982; Lynch and Hill 1986; Turelli et al. 1988; Lynch 1990; Spitze 1993), which consider only one trait at a time and effectively require the intensity of positive selection to surpass that of purifying selection to reject neutrality. Because this requirement is sufficient but not necessary for demonstrating positive selection, it is replaced with the criterion of a positive correlation between trait importance and evolutionary rate to improve the power of the test.

4.3.1 Adaptive intraspecific variations of yeast morphological traits

We first tested the neutral hypothesis of phenotypic evolution in a set of 210 morphological traits chosen purely on the basis of the feasibility of measurement (Ohya et al. 2005) rather than potential roles in adaptation. These traits were quantified by analyzing fluorescent microscopic images of triple-stained cells (Ohya et al. 2015) from

37 natural strains of *S. cerevisiae* (Yvert et al. 2013). For a given trait, we defined the phenotypic difference between two strains by the absolute difference in their trait value, relative to their average trait value. Phenotypic differences were corrected for potential environmental heterogeneity in the measurement and sampling error to allow among-trait comparison. We then estimated, for each trait, the mean evolutionary distance (*ED*) among all 666 pairs of the 37 strains by averaging their corrected pairwise phenotypic differences.

To test the neutral hypothesis, we used measures of trait importance (*TI*) for the 210 traits, where *TI* is 100 times the fitness effect caused by 1% change in trait value, estimated using the fitness and phenotype data of thousands of single gene deletion strains of *S. cerevisiae* (Ho and Zhang 2014). We found that mean *ED* decreases with *TI* (**Fig. 4.2A**), indicating that relatively important traits evolve more slowly than relatively unimportant ones. However, the rate of phenotypic evolution is determined by both the rate of mutational input and the direction and magnitude of natural selection. The rate of mutational input for a trait is the average effect size of a random mutation on the trait (i.e., mutational size or *MS*) multiplied by the mutation rate per genome per generation. Because the mutation rate is the same for all traits, we need only consider *MS*. We estimated the *MS* for a trait by the mean phenotypic effect of 4718 individual gene deletions on the trait (Ho and Zhang 2014). Although spontaneous mutations are expected to have smaller phenotypic effects than gene deletions, it is reasonable to assume that the average phenotypic effect of spontaneous mutations on a trait is approximately proportional to that of gene deletions on the trait. In other words, *MS* can be used as a proxy of spontaneous mutational size when different traits are compared. As

was previously discovered (Ho and Zhang 2014), *MS* decreases precipitously with *TI* (**Fig. 4.2A**). This observation indicates that relatively important traits are affected by mutations to a smaller degree than are relatively unimportant traits, which has likely resulted from stronger selection for mutational robustness of more important traits (Ho and Zhang 2014). To control the impact of mutational input on the rate of phenotypic evolution, we measured the evolutionary rate of a trait in the unit of its mutational size by dividing mean *ED* by *MS* for each trait. We found mean *ED*/*MS* to increase significantly with *TI* (**Fig. 4.2B**), suggesting that relatively important traits evolve faster than relatively unimportant traits in the unit of mutational size. This finding is inconsistent with the neutral hypothesis of phenotypic evolution (**Fig. 4.1A**), but supports the adaptive hypothesis (**Fig. 4.1B**).

To exclude the possibility that the above result is an artifact of our statistical analysis, we used two negative controls. First, we compared the wild-type BY strain with 666 randomly picked gene deletion strains of the BY background, under the premise that their phenotypic differences should not be adaptive. As expected, there is no significant correlation between mean *ED*/*MS* and *TI* (**Fig. 4.2B**), where mean *ED* for a trait is the average phenotypic difference between these deletion strains and the wild-type for the trait. Second, we compared between each of 89 mutational accumulation (MA) lines and their common ancestor (Geiler-Samerotte et al. 2016); the MA lines were produced in ~2000 mitotic generations with virtually no selection and therefore should not show adaptive signals when compared with their ancestor. Note that only 180 morphological traits are available in the data from MA lines for analysis. While mean *ED*/*MS* for the 37 natural strains still increases significantly with *TI* for these 180 traits (**Fig. A.3.1A**), mean

ED/*MS* between the MA lines and their ancestor shows no significant correlation with *TI* (**Fig. A.3.1B**).

Because the *MS* and *TI* used were estimated in haploid yeasts, while the 37 natural strains studied are diploid, we also estimated *MS* and *TI* using recently published morphological data of 130 diploid gene deletion strains (Yang et al. 2014). We confirmed that the significant positive correlation between *ED*/*MS* and *TI* holds (**Fig. A.3.2**).

To examine whether the detected adaptive signal among the 37 natural strains is attributable to a small number of strains or is a general phenomenon of the species, we estimated the rank correlation (ρ) between *ED*/*MS* and *TI* for each of the 666 strain pairs, using the *ED* value of the strain pair. We found ρ to be positive for the vast majority of the strain pairs (**Fig. 4.2C**), suggesting pervasive adaptive morphological evolution in *S. cerevisiae*.

Some of the 210 morphological traits are genetically highly correlated (Wang et al. 2010; Ho and Zhang 2014). To exclude the possibility that the adaptive signal is an artifact of the use of correlated traits, we estimated 210 principal component traits from the original traits (see Supplemental Experimental Procedures). Analysis of the principal component traits, which are independent from one another, shows even stronger signals of adaptive evolution (**Fig. A.3.3**).

If the prevalent positive ρ values among the 37 natural strains (**Fig. 4.2C**) truly arise from positive selection, we should expect that (i) the overall rate of morphological evolution is greater for strain pairs with higher ρ values and (ii) this rate disparity is primarily reflected in relatively important traits rather than relatively unimportant ones. We defined the overall rate of morphological evolution between two strains by their

morphological dissimilarity across all 210 traits divided by their fractional genomic sequence difference. Consistent with our expectation, the rate of morphological evolution increases with ρ ($p = 0.007$, one-tailed partial Mantel test with phylogenetic permutation to correct for both the nonindependence among strain pairs and phylogenetic relationships in the data; see Supplemental Experimental Procedures). Additionally, when we separated the 210 traits into two equal-sized bins based on *TI*, the correlation between the rate of morphological evolution and ρ remained significantly positive for the 105 traits with relatively high *TI* ($p < 0.001$) but not for the 105 traits with relatively low *TI* ($p = 0.805$). These observations support that, the greater the ρ value relative to 0, the stronger the positive selection on the morphological traits, especially relatively important ones.

What factors determine the ρ value of a strain pair? Because the 210 morphological traits were chosen purely based on experimental feasibility, we do not expect the detected adaptive signals to correlate with any obvious genetic or ecological factor. Indeed, using the partial Mantel test, we found no significant correlation between the ρ value of two strains and the strains' difference in genome sequence, ecological environment, population membership, or geographic location (**Table A.3.1**). These findings are consistent with a previous analysis showing that the morphological similarities among the 37 natural strains cannot be explained by the strains' similarities in population history or ecological environment (Yvert et al. 2013). Hence, the selective agents behind the detected morphological adaptations are unclear.

4.3.2 Adaptive interspecific variations of yeast morphological traits

To test the neutral hypothesis of morphological evolution beyond the species level, we collected comparable morphological data from two strains of *S. paradoxus*, the sister species of *S. cerevisiae*, and one strain of their outgroup species *S. mikatae* (see Supplemental Experimental Procedures). We first calculated the mean *ED*/*MS* between the 37 *S. cerevisiae* strains and *S. paradoxus* strain N17 for each trait. Across the 210 traits, we observed a significant, positive correlation between mean *ED*/*MS* and *TI* (**Fig. 4.3A)**. A similar pattern was observed between *S. cerevisiae* and *S. paradoxus* strain IFO1804 (**Fig. 4.3B**). Between *S. cerveisiae* and the more distantly related species of *S. mikatae*, however, the positive correlation is no longer significant (**Fig. 4.3C**), probably because of the lack of prevalent positive selection or a reduced statistical power as a result of using *MS* and *TI* values estimated from *S. cerevisiae* in tests involving distantly related species.

4.3.3 Neutral evolution of yeast gene expression levels

To investigate the generality of the above findings of adaptive phenotypic evolution, we turned to another class of traits that can be chosen regardless of their potential roles in adaptation: 3466 gene expression traits, each being the mRNA expression level of a yeast gene in a rich medium. Using microarray gene expression data, we quantified *ED* between two *S. cerevisiae* strains for each trait. We estimated *MS* from the microarray expression data of 1486 gene deletion strains, as in the case of morphological traits (Ho and Zhang 2014). *TI* of the expression level of a gene was measured by the fitness reduction caused by the deletion of the gene see (see Supplemental Experimental Procedures) (Ho and Zhang 2014). Similar to the

morphological data, gene expression data showed a negative correlation between *ED* and *TI* and a negative correlation between *MS* and *TI* (**Table 4.1**). However, contrary to the morphological traits, expression traits exhibited a significant, negative correlation between *ED*/*MS* and *TI* (**Table 4.1**). As a negative control, we estimated the standard deviation (SD_m) in relative expression level among four MA lines (Landry et al. 2007) and found no significant correlation between SD_m/*MS* and *TI* (**Table 4.1**). While adaptive evolution of the expression levels of some yeast genes have been suggested (Bullard et al. 2010; Fraser et al. 2010; Qian et al. 2012), overall we found relatively important genes to evolve more slowly in expression level than relatively unimportant ones upon the control of mutational size, consistent with the neutral hypothesis. Here we are restricted to intra-specific analysis, because the use of different probes in the microarrays of different species prohibits a reliable comparison of the rate of expression evolution among genes. Use of other types of expression data such as mRNA sequencing data is currently infeasible, because no comparable data are available for estimating *MS*.

4.4 DISCUSSION

In summary, our analysis of 210 yeast morphological traits with no *a priori* bias toward adaptive evolution reveals strong signals of adaptive evolution. To the best of our knowledge, this is the first demonstration of widespread adaptive evolution of a large set of random phenotypic traits in any organism. If these 210 traits are representative of yeast morphological traits, we must conclude that morphological evolution in yeast is frequently adaptive. Whether the same conclusion could be drawn in other species is unknown, but the methodology developed here is generally applicable, especially to

genetic model organisms, in which mutational size estimation and trait importance estimation are relatively straightforward. Although it is essential to use a random set of traits to evaluate the prevalence of adaptive phenotypic evolution, using such traits means that the selective agent would be difficult to discern if adaptation is detected, as in the present case. Future work should attempt to identify the selective agents acting on these morphological traits, which are necessary for a complete understanding of phenotypic adaptation.

It should be emphasized that, although the bar for rejecting neutrality is likely lowered in our test compared with that in most previous tests (Lande 1976; Chakraborty and Nei 1982; Lynch and Hill 1986; Turelli et al. 1988; Spitze 1993), it remains quite high. This is because important traits that could be subject to strong positive selection are also expected to be under strong purifying selection such that an adaptive signal becomes detectable only when the difference in the strength of positive selection among traits surpasses that in the strength of purifying selection. The high bar renders claims of adaptation conservative. But, it also means that failure to reject neutrality, as in the case of yeast gene expression evolution, neither proves neutrality nor refutes adaptation. Nonetheless, based on additional experiments and tests, we recently found unambiguous evidence that the vast majority of gene expression level variations within and between yeast species are not adaptive but neutral (Yang et al. 2016). Together, these analyses of yeast morphological and gene expression data raise the intriguing possibility that some classes of phenotypic traits evolve generally adaptively while others neutrally.

4.5 MATERIALS AND METHODS

4.5.1 A comparison between existing tests of the neutral hypothesis of phenotypic evolution and the newly proposed test

A number of neutrality tests of phenotypic evolution exist in the literature. They all test the neutrality of one trait and can be divided into two categories based on the rationale of the test. The first category compares the observed phenotypic distance in a trait between populations or species with the neutral expectation. Positive (or purifying) selection is inferred when the observed distance is significantly larger (or smaller) than the neutral expectation. The neutral expectation has been derived by considering the stochastic process of phenotypic evolution (Lande 1976) or genetic models of quantitative traits (Chakraborty and Nei 1982; Lynch and Hill 1986). Various test statistics have been proposed, including for example Lande's N_e (Lande 1976), Lande's F (Lande 1977), Chakraborty and Nei's B_t/V_t (Chakraborty and Nei 1982), Turelli et al.'s MDE test (Turelli et al. 1988), and Lynch's Δ (Lynch 1990). These tests usually require the information of effective population size and divergence time. They also require the information on the rate of mutational input such as narrow-sense heritability and mutational variance. If these parameters are unavailable, one may use an alternative approach by comparing Q_{ST} of a trait with F_{ST} of neutral loci, which are expected to be the same if the trait evolves neutrally (Spitze 1993). Positive selection is undetectable by any of the above tests unless the intensity of positive selection exceeds that of purifying selection. Because purifying selection is expected to be pervasive even for traits subject to positive selection, all tests in this category have a low power in detecting positive selection.

The second category is the QTL sign test, which relies on the information from QTL mapping of a trait. Typically, positive selection is inferred when the number of positive-effect QTLs differs significantly from that of negative-effect QTLs (Orr 1998). Although this category of test is model-free, it cannot distinguish between positive selection and relaxation from purifying selection without other information (Fraser et al. 2010).

Our test of the neutral hypothesis of phenotypic evolution compares the evolutionary rates of many traits with different levels of importance to fitness. Under the neutral hypothesis allowing purifying selection, relatively important traits evolve more slowly than relatively unimportant traits. If relatively important traits are found to evolve more rapidly than relatively unimportant ones, neutrality is rejected in favor of positive selection. Our test is expected to be more powerful than the above first category of tests, because detecting positive selection no longer requires the intensity of positive selection to exceed that of purifying selection. Rather, the criterion is that the difference in intensity between positive and negative selection is more positive for more important traits. Our test does not suffer from the problem in the above second category of tests, because it is extremely improbable for the relative importance of a large number of traits to be reversed in a second population, compared with that in the original population where trait importance is measured.

The yeast morphological data from natural strains and gene deletion strains, gene expression data, and the fitness data of gene deletion strains used for estimating trait importance were all collected under a laboratory rich medium. While this medium does not equal the many natural environments of various yeast strains, this discrepancy does

not affect our neutrality test, because all required by our test is that the same condition is used in phenotyping a trait and in estimating its trait importance. The main problem with mismatches between the experimental condition and the natural environments is that the selective agent becomes more difficult to discern.

4.5.2 Yeast morphological data

The morphological traits analyzed here were previously defined (Ohya et al. 2005). Briefly, yeast cultures were grown to 1×10^7 cells/ml in YPD or synthetic complete media. Cells were fixed with 3.7% formaldehyde and stained with fluorescein isothiocyanate-Con A, rhodaminephalloidin, and 4′,6-diamidino-2-phenylindole, which simultaneously mark cell wall, actin cytoskeleton, and nuclear DNA, respectively. After digital images were acquired, cell images were collected and processed by *CalMorph* (Ohya et al. 2015). In total, 501 morphological traits were measured. Among these morphological traits, we focused on 210 traits, which are defined for individual cells rather than cell populations, have positive trait importance values (Ho and Zhang 2014), and were measured in all the strains analyzed here. The morphological data of *S. paradoxus* strain N17, *S. paradoxus* strain IFO1804, and *S. mikatae* strain IFO1815 were generated with 15, 10, and five biological replicates, respectively. On average, ~60 cells were measured for each trait in each replicate. The morphological data from 37 natural strains of *S. cerevisiae* were previously collected, with five biological replicates per strain (Yvert et al. 2013). The morphological data of 4718 haploid and 130 diploid single gene deletion strains of *S. cerevisiae* were previously published (Ohya et al. 2005; Yang et al. 2014).

The morphological data of the derivatives of *S. cerevisiae* mutational accumulation (MA) lines and their common ancestor were generated previously (Geiler-Samerotte et al. 2016) and the data file "Raw_Data_Additional_Traits.Rfile" was downloaded for analysis. Only those measurements without the geldanamycin treatment were used. The ancestor and 89 MA derivatives with two replicates were used. A total of 180 traits that belong to the aforementioned 210 traits had relevant data for our analysis.

4.5.3 Evolutionary distance between two strains for a morphological trait

We estimated the mean phenotypic value for a trait in a strain by first calculating the mean trait value among all cells in a replicate population and then averaging this number across all replicate populations. Let x_i and x_j be the mean phenotypic values of a trait in strains *i* and *j,* respectively. We estimated the raw evolutionary distance for the trait between strains *i* and *j* by $ED_{ij} = |x_i - x_j| / [(x_i + x_j)/2]$. Different traits have different levels of environmental variation in phenotyping, different levels of random measurement error, and different levels of among-individual stochastic phenotypic variation. To allow a comparison among traits, we corrected the above estimated raw *ED* values for these factors (Ho and Zhang 2014). Let us assume that strain *i* has *m* replicate populations, with population sizes of $a_1, a_2, \ldots, a_m$, respectively, and strain *j* has *n* replicate populations, with population sizes of $b_1, b_2, \ldots, b_n$, respectively. We generated 100 sets of bootstrap samples for both strain *i* (i.e., each having *m* populations with sizes of $a_1, a_2, \ldots a_m$) and strain *j* (i.e., each having *n* populations with sizes of $b_1, b_2, \ldots, b_n$) using the data from the *n* populations of strain *j*. Note that the hierarchical structure of the data is retained in bootstrapping. We then estimated the average ED_{ij} across these 100

sets of bootstrapped samples, and denoted it by pseudo $ED_{j\rightarrow i}$. We similarly estimated pseudo $ED_{i\rightarrow j}$ using the data from the *m* populations of strain *i*. We then averaged $ED_{j\rightarrow i}$ and $ED_{i\rightarrow j}$ and subtracted this value from raw ED_{ij} to obtain the corrected ED_{ij}. If the corrected ED_{ij} is negative, we set it at 0. All *ED* values presented in the main text and figures are corrected ED_{ij}.

4.5.4 Mutational size

The mutational size for a trait was calculated by the average of the previously published net effect sizes of 4718 haploid single gene deletions on the trait (Ho and Zhang 2014). Briefly, we first calculated the raw effect size (ES_{ij}) of deleting gene *i* on trait *j* as $(x_{ij} - w_j)/w_j$, where x_{ij} is the mean phenotypic value of trait *j* in the deletion strain *i*, and w_j is the corresponding value in the wild-type (averaged across 123 replicate populations). Then we generated 1000 pseudo phenotypic datasets to estimate the net $|ES|$ of gene deletion on a trait. To generate a pseudo dataset, we randomly chose one of the 123 wild-type replicate populations and picked (with replacement) from this population the same number of cells as in the actual gene-deletion data. We then calculated mean pseudo $|ES|$ across all pseudo datasets; net $|ES|$ equals raw $|ES|$ minus mean pseudo $|ES|$ if raw $|ES|$ > mean pseudo $|ES|$ or zero if raw $|ES|$ < mean pseudo $|ES|$.

We also estimated mutational size by the same method but using the recently published morphological data of 130 diploid single gene deletion strains (Yang et al. 2014) along with the five replicate populations of the corresponding wild-type (BY4743) (Yvert et al. 2013). Due to the low number of wild-type replicates, 100 pseudo datasets were generated.

4.5.5 Morphological trait importance

The trait importance (*TI*) of each of the 210 morphological traits was previously estimated from the negative slope of the linear regression between the corrected phenotypic effect of a gene deletion on the trait and the fitness of the gene deletion strain across 2779 haploid single gene deletion strains (Ho and Zhang 2014). The deletion strains with fitness larger than 1 were not used. Briefly, for each trait *j*, we performed a linear regression $F_i = a_j - b_j$ (net $|ES_{ij}|$), where net $|ES_{ij}|$ is the absolute value of the net effect size of deleting gene *i* on trait *j*, and F_i is the fitness of the strain lacking gene *i* relative to the wild-type in YPD. In this regression, the estimated slope $b_j > 0$ is 100 times the reduction in fitness caused by 1% change in the phenotypic value of trait *j*, while the estimated intercept a_j is the expected fitness when net $|ES_{ij}| = 0$. Thus, b_j is a measure of the relative importance of trait *j* to fitness, or trait importance (*TI*).

Because *TI* is estimated by the correlation between phenotypic changes and fitness changes, it may not accurately reflect the causal relationship between the variation of a trait and fitness. The fact that our use of the inaccurate *TI* estimates still yields significant evidence for adaptive morphological evolution suggests that the true signal is even stronger. In other words, our results are likely to be conservative.

For diploid single gene deletion strains (Yang et al. 2014), we applied the same method to calculate *TI*, but in this case only 99 strains with fitness ≤ 1 could be used.

4.5.6 Principal component analysis

To examine if the non-independence among the 210 morphological traits affects

our results, we followed previous studies (Wang et al. 2010; Ho and Zhang 2014) to perform a principal component analysis to transform the net |*ES*| matrix **M** (4718 genes × 210 traits) described previously (Ho and Zhang 2014). Note that net |*ES*| is the corrected absolute effect size of a gene deletion on a trait. After this function returned a coefficient matrix **C** (210 × 210), we calculated $\boldsymbol{v'} = \boldsymbol{v}\,\mathbf{C}$ (1 × 210 principal traits) for each corrected *ED* vector $\boldsymbol{v}$ (1 × 210 traits) between two strains. The absolute values of $\boldsymbol{v'}$ provided the corrected *ED* for each of the 210 orthogonal principal component traits. We also used the transformed net |*ES*| matrix, $\mathbf{M'} = \mathbf{MC}$ (4718 genes × 210 principal traits), to estimate trait importance for the 210 principal component traits following the method previously used (Ho and Zhang 2014).

4.5.7 Mantel's test

We used Mantel's test (Mantel 1967; Sokal and Rohlf 1995) to evaluate the significance of the correlation between a biological distance (e.g., morphological dissimilarity between two strains) and the correlation (ρ) between *ED*/*MS* and *TI* across 666 pairs of the 37 natural strains of *S. cerevisiae*. In general, Mantel's test evaluates whether the correlation between two distance matrices is significant by comparing the test statistic (*z*) calculated by the observed matrices and the null distribution of *z* calculated by randomly shuffling one of the matrices. The test statistic could be an element-by-element product or a Pearson correlation coefficient between all elements. During each shuffling, two columns are randomly picked and swapped, and the corresponding rows are also swapped. In this study, we used the correlation coefficient between all elements of two matrices as the test statistic and generated the null

distribution of the test statistic by randomly shuffling one of the matrices 1000 times.

When the dataset has underlying phylogenetic relationships such as the present dataset in which all strains are connected by a phylogeny, the significance level may be inflated because of the shared evolutionary history among lineages (Felsenstein 1985). To control for the phylogenetic non-independence, we performed a partial Mantel test, in which the phylogenetic distance matrix is controlled for (Smouse et al. 1986). Among four different ways to implement the partial Mantel test, we used the so-called method 2, in which the residual matrix instead of the original matrix is shuffled, as suggested (Legendre 2000). In addition, we used phylogenetic permutation (PP) in partial Mantel test to obtain non-inflated type-I errors (Harmon and Glor 2010). Specifically, applying PP means that the chance to pick a strain pair to swap is proportional to their phylogenetic distance (Lapointe and Garland 2001). The phylogenetic tree of the 37 natural strains of *S. cerevisiae* was reconstructed using the single nucleotide polymorphism (SNP) data from Maclean et al. (Maclean et al. 2016) by the neighboring-joining method (Saitou and Nei 1987) with *p*-distance (Nei and Kumar 2000) implemented in MEGA 5.2.2 (Tamura et al. 2011), and the matrix of pairwise phylogenetic distances was obtained by APE in R (Paradis et al. 2004).

We used Spearman's correlation coefficient (ρ) between *ED/MS* and *TI* for a pair of strains as a measure of adaptation. By partial Mantel's test, we examined whether ρ is correlated with the following five parameters for the strain pair: morphological evolutionary rate, dissimilarity in genome sequence, dissimilarity in ecological environment, dissimilarity in population membership, and dissimilarity in geographic location. The morphological evolutionary rate between two strains was estimated as

follows. First, for each trait, we ranked the 37 strains by their mean phenotypic values for the trait. Second, we calculated Spearman's correlation between the 210 ranks of one strain with those of the other strain. The rate of morphological evolution was calculated by the negative of this Spearman's correlation divided by genomic sequence dissimilarity (see below). Genomic sequence dissimilarity was calculated by the SNP density between two strains (Maclean et al. 2016). Ecological environment dissimilarity was set as 0 if two strains were sampled from the same ecological environment and 1 if sampled from different environments (Maclean et al. 2016). Population membership dissimilarity was set as 0 if two strains belong to the same population based on the fastSTRUCTURE analysis of the SNP data of 190 yeast strains (Maclean et al. 2016); otherwise, it was set as 1. Strains designated as "mosaics" (Maclean et al. 2016) were excluded from this analysis because mosaics do not represent a population. Geographic location dissimilarity was set to 0 if two strains were sampled from the same continent; otherwise, it was set to 1. Note that CLIB382 was removed from this analysis because of the lack of SNP data in Maclean et al. (Maclean et al. 2016). Furthermore, any strain lacking information for any parameter was removed from the analysis for that particular parameter.

4.5.8 Gene expression data and analysis

The rich medium microarray gene expression ratio (r) between two *S. cerevisiae* strains, RM and BY, were previously measured for thousands of genes (Brem and Kruglyak 2005). We defined the intra-specific evolutionary distance (ED) of the expression level for a gene by $|x_{RM}\text{-}x_{BY}|/x_{BY}$, which equals $|r\text{-}1|$, because $r = x_{RM}/x_{BY}$,

where x_{RM} and x_{BY} are expression levels of the gene in RM and BY, respectively.

We estimated mutational size for each expression trait by the mean effect size of gene deletion on the expression trait across 1486 deletion lines, following a previously published method (Ho and Zhang 2014) but using a recently published large dataset (Kemmeren et al. 2014). The trait importance (*TI*) of a gene expression trait was defined by the fitness decrease caused by deleting the gene (Qian et al. 2012), and only those genes that cause a zero or positive fitness reduction were considered. After removing genes that miss any kind of data above, we obtained our final dataset with 3466 expression traits.

As a negative control in the neutraliy test for expression traits, we used the square root of variance in the expression level of an evolved line relative to that of the ancestral line (SD_m) among four mutation accumulation lines (Landry et al. 2007), and correlated between SD_m/MS and *TI*.

4.6 ACKNOWLEDGEMENTS

We thank Seiko Morinaga for technical assistance, Kerry Geiler-Samerotte for sharing the morphological data of mutation accumulation lines, and Xiaoshu Chen, Brian Metzger, and Jian-Rong Yang for valuable comments. This work was supported by U.S. National Institutes of Health research grant R01GM103232 to J.Z. and the Grants-in-Aid for Scientific Research grant 24370002 from the Ministry of Education, Culture, Sports, Science and Technology in Japan to Y.O.

4.7 REFERENCES

Brem RB, Kruglyak L. 2005. The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc Natl Acad Sci U S A* **102**: 1572-1577.

Bullard JH, Mostovoy Y, Dudoit S, Brem RB. 2010. Polygenic and directional regulatory evolution across pathways in Saccharomyces. *Proc Natl Acad Sci U S A* **107**: 5058-5063.

Chakraborty R, Nei M. 1982. Genetic differentiation of quantitative characters between populations or species .1. mutation and random genetic drift. *Genet Res* **39**: 303-314.

Darwin C. 1859. *On the Origin of Species by Means of Natural Selection*. J. Murray, London,.

Endler JA. 1986. *Natural Selection in the Wild*. Princeton University Press, Princeton, N.J.

Felsenstein J. 1985. Phylogenies and the Comparative Method. *American Naturalist* **125**: 1-15.

Fraser HB, Moses AM, Schadt EE. 2010. Evidence for widespread adaptive evolution of gene expression in budding yeast. *Proc Natl Acad Sci U S A* **107**: 2977-2982.

Futuyma DJ. 2013. *Evolution*. Sinauer Associates.

Geiler-Samerotte KA, Zhu YO, Goulet BE, Hall DW, Siegal ML. 2016. Selection Transforms the Landscape of Genetic Variation Interacting with Hsp90. *PLoS Biol* **14**: e2000465.

Harmon LJ, Glor RE. 2010. Poor statistical performance of the Mantel test in phylogenetic comparative analyses. *Evolution* **64**: 2173-2178.

Ho WC, Zhang J. 2014. The genotype-phenotype map of yeast complex traits: basic parameters and the role of natural selection. *Mol Biol Evol* **31**: 1568-1580.

Kemmeren P, Sameith K, van de Pasch LA, Benschop JJ, Lenstra TL, Margaritis T, O'Duibhir E, Apweiler E, van Wageningen S, Ko CW et al. 2014. Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. *Cell* **157**: 740-752.

Kimura M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

Kingsolver JG, Hoekstra HE, Hoekstra JM, Berrigan D, Vignieri SN, Hill CE, Hoang A, Gibert P, Beerli P. 2001. The strength of phenotypic selection in natural populations. *Am Nat* **157**: 245-261.

Lande R. 1976. Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30**: 314-334.

Lande R. 1977. Statistical tests for natural selection on quantitative characters. *Evolution* **31**: 442-444.

Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL. 2007. Genetic properties influencing the evolvability of gene expression. *Science* **317**: 118-121.

Lapointe FJ, Garland T. 2001. A generalized permutation model for the analysis of cross-species data. *J Classif* **18**: 109-127.

Legendre P. 2000. Comparison of permutation methods for the partial correlation and partial Mantel tests. *J Stat Comput Sim* **67**: 37-73.

Li WH, Gojobori T, Nei M. 1981. Pseudogenes as a paradigm of neutral evolution.

Nature **292**: 237-239.
Lynch M. 1990. The rate of morphological evolution in mammals from the standpoint of the neutral expectation. *American Naturalist* **136**: 727-741.
Lynch M, Hill WG. 1986. Phenotypic evolution by neutral mutation. *Evolution* **40**: 915-935.
Maclean CJ, Metzger BPH, Yang J-R, Ho W-C, Moyers B, Zhang J. 2016. Deciphering the genic basis of environmental adaptations by simultaneous forward and reverse genetics in Saccharomyces cerevisiae. *BioRxiv* **doi:** http://dx.doi.org/10.1101/087510.
Mantel N. 1967. Detection of disease clustering and a generalized regression approach. *Cancer Res* **27**: 209-&.
Mayr E. 2001. *What Evolution is*. Basic Books, New York.
Nei M. 2007. The new mutation theory of phenotypic evolution. *Proc Natl Acad Sci U S A* **104**: 12235-12242.
Nei M, Kumar S. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
Ohya Y, Kimori Y, Okada H, Ohnuki S. 2015. Single-cell phenomics in budding yeast. *Mol Biol Cell* **26**: 3920-3925.
Ohya Y, Sese J, Yukawa M, Sano F, Nakatani Y, Saito TL, Saka A, Fukuda T, Ishihara S, Oka S et al. 2005. High-dimensional and large-scale phenotyping of yeast mutants. *Proc Natl Acad Sci U S A* **102**: 19015-19020.
Orr HA. 1998. Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics* **149**: 2099-2104.
Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**: 289-290.
Qian W, Ma D, Xiao C, Wang Z, Zhang J. 2012. The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. *Cell Rep* **2**: 1399-1410.
Saitou N, Nei M. 1987. The neighbor-joining method - a new method for reconstructing phylogenetic trees. *Molecular biology and evolution* **4**: 406-425.
Smouse PE, Long JC, Sokal RR. 1986. Multiple-regression and correlation extensions of the Mantel test of matrix correspondence. *Syst Zool* **35**: 627-632.
Sokal RR, Rohlf FJ. 1995. *Biometry*. W. H. Freeman and Company, New York.
Spitze K. 1993. Population structure in *Daphnia obtusa* - quantitative genetic and allozymic variation. *Genetics* **135**: 367-374.
Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**: 2731-2739.
Turelli M, Gillespie JH, Lande R. 1988. Rate tests for selection on quantitative characters during macroevolution and microevolution. *Evolution* **42**: 1085-1089.
Wang Z, Liao BY, Zhang J. 2010. Genomic patterns of pleiotropy and the evolution of complexity. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 18034-18039.
Yang J-R, Maclean CJ, Park C, Zhao H, Zhang J. 2016. Intra- and inter-specific variations of gene expression levels in yeast are largely neutral. *BioRxiv doi:* http://dxdoiorg/101101/089995.
Yang M, Ohnuki S, Ohya Y. 2014. Unveiling nonessential gene deletions that confer

significant morphological phenotypes beyond natural yeast strains. *BMC Genomics* **15**: 932.

Yvert G, Ohnuki S, Nogami S, Imanaga Y, Fehrmann S, Schacherer J, Ohya Y. 2013. Single-cell phenomics reveals intra-species variation of phenotypic noise in yeast. *BMC Syst Biol* **7**: 54.

Table 4.1 Testing the neutral hypothesis of yeast gene expression evolution

Variables correlated	Spearman's ρ	p-value
ED, *TI*	-0.147	3.7×10^{-18}
MS, *TI*	-0.142	3.9×10^{-17}
ED/*MS*, *TI*	-0.087	3.2×10^{-7}
SD_m/*MS*, *TI*	-0.030	0.13

ED, evolutionary distance in gene expression level between *S. cerevisiae* strains.
TI, gene expression trait importance.
MS, mutational size measured by gene deletion.
SD_m, standard deviation in relative gene expression level among mutation accumulation lines.

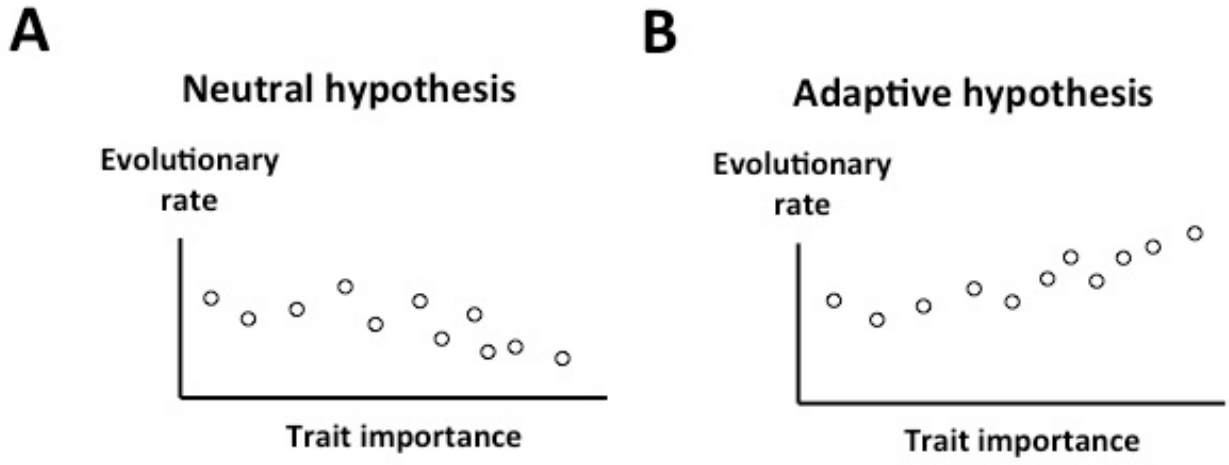

Figure 4.1 Schematic illustrating the test of the neutral hypothesis of phenotypic evolution by comparing evolutionary rates among traits of different levels of importance. (**A**) Under the neutral hypothesis, relatively important traits evolve more slowly than relatively unimportant traits. (**B**) Higher evolutionary rates of more important traits reject the neutral hypothesis and support the adaptive hypothesis. Each circle represents a trait.

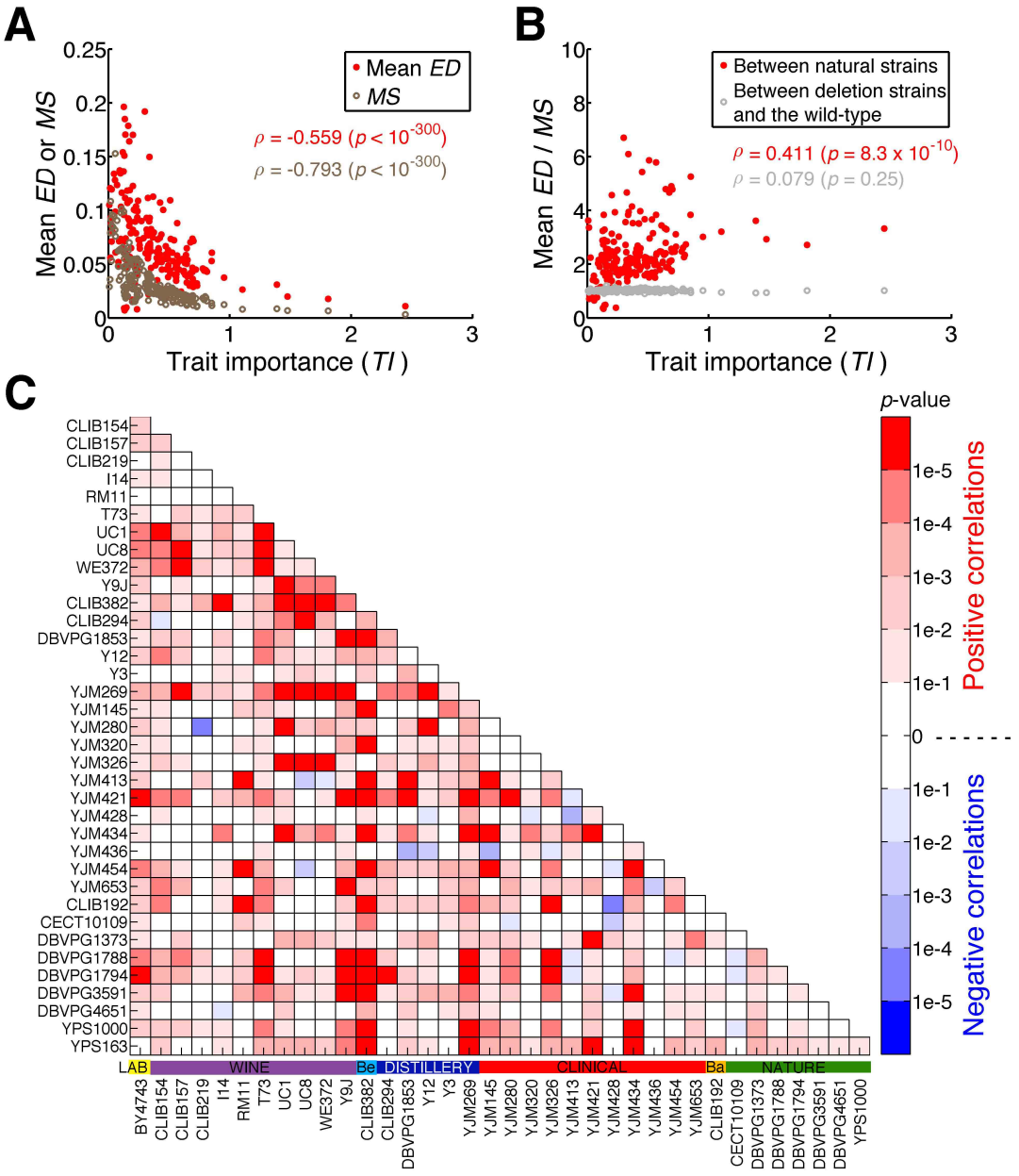

Figure 4.2 Prevalent adaptive evolution of morphological traits in the yeast *Saccharomyces cerevisiae*. (**A**) Mean evolution distance (*ED*) of 666 pairs of natural strains for a trait and the mutational size (*MS*) of the trait both decrease with trait importance (*TI*). Each dot represents a trait. ρ, Spearman's rank correlation coefficient. (**B**) Mean *ED* among 666 natural strain pairs for a trait relative to its *MS* increases significantly with *TI*, while the mean *ED* between 666 gene deletion strains and the wild-type relative to *MS* does not increase significantly with *TI*. (**C**) Nominal *p*-values for the Spearman's correlation between *ED*/*MS* and *TI* for all 666 pairs of natural strains. The horizontal colored bars above the strain names show the ecological environments of the strains. Be, beer; Ba, bakery.

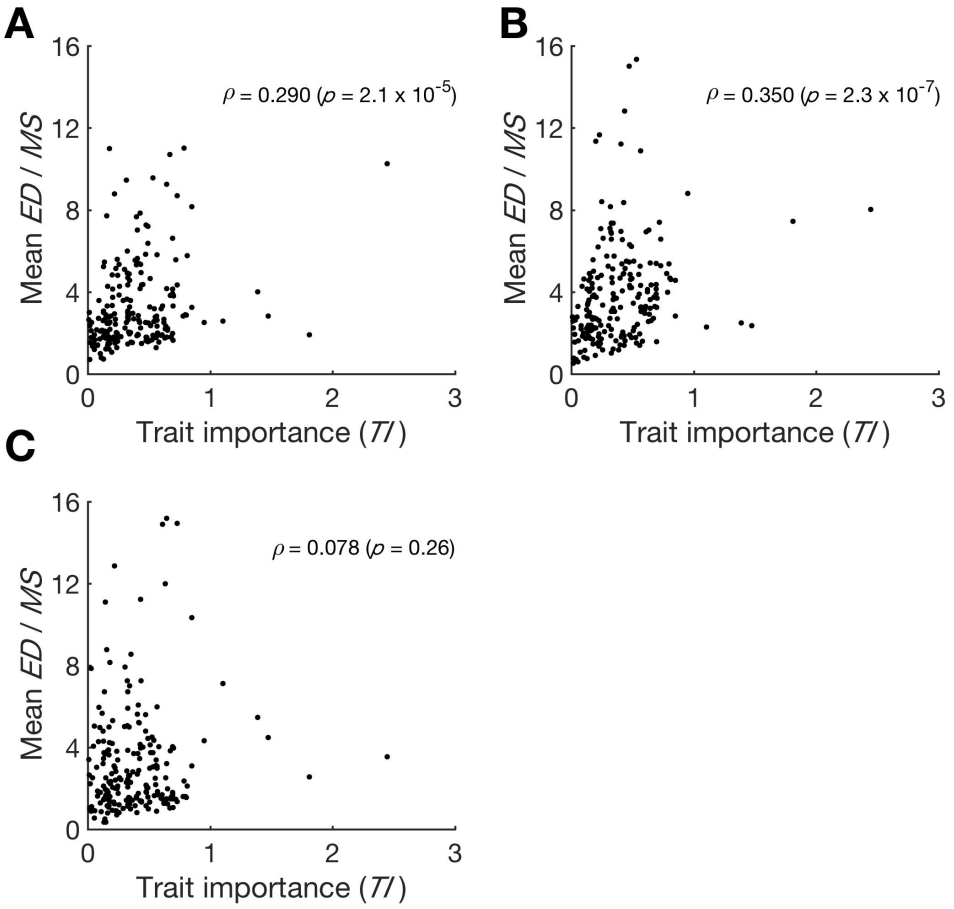

Figure 4.3 Adaptive morphological evolution between *Saccharomyces* species. (**A**) The mean *ED*/*MS* between 37 natural *S. cerevisiae* strains and *S. paradoxus* strain N17 for a trait increases significantly with trait importance (*TI*). (**B**) The mean *ED*/*MS* between 37 natural *S. cerevisiae* strains and the *S. paradoxus* strain IFO1804 increases significantly with *TI*. (**C**) The mean *ED*/*MS* between 37 natural *S. cerevisiae* strains and *S. mikatae* increases (but not significantly) with *TI*. Each dot represents a morphological trait.

Chapter 5

Does Mutational Correlation Constrain or Facilitate Phenotypic Evolution?

5.1 ABSTRACT

Phenotypic evolution of a trait is usually not independent from that of another trait. One source of the interdependence is mutational correlation due to pleiotropy. Whether mutational correlation constrains or facilitates phenotypic evolution is unresolved, and both scenarios are theoretically possible. Here we address this controversy using 210 yeast morphological traits measured in thousands of gene-deletion lines and dozens of divergent yeast strains. We found that, if two traits show a higher mutational correlation, their evolutionary rates also tend to be highly correlated. However, when focusing on individual traits, we did not observe a simple relationship between the average mutational correlation of the focal trait with all other traits and the evolutionary rate of the focal trait. Instead, a negative quadratic relationship was observed, suggesting that increasing mutation correlation speeds up evolution when the correlation is very low, but constrains evolution when it is very high. We discuss possible evolutionary scenarios consistent with these findings.

5.2 INTRODUCTION

Many traits do not evolve independently. One reason for the independence of phenotypic evolution is the genetic correlation caused by mutational effects, or mutational correlation. Because many mutations could simultaneously affect multiple traits due to gene pleiotropy, changes of one trait by mutations may inevitably change another trait (Wagner and Zhang 2011). If mutational effects on different traits are antagonistic, such mutational correlation could constrain phenotypic evolution with different severity levels. If the direct beneficial effect on one trait surpasses the indirect deleterious effects on other traits, the optimal value for that trait will remain reachable, but the path will be longer. However, if the indirect deleterious effects on other traits outcompete the direct beneficial effect, the optimal value for that trait will be unreachable because of the lack of mutations with net beneficial effects (**Fig. 5.1A**). In the contrary to antagonism, if mutational effects on different traits are accordant, such mutational correlation could facilitate phenotypic evolution, no matter if the direct benefit is larger or smaller than indirect benefits (**Fig. 5.1B**). Given these different possibilities, one interesting question is whether mutational correlation tends to constrain or facilitate phenotypic evolution. For example, if accordant effects are largely preserved, and antagonistic effects are largely eliminated by variational modularity, phenotypic evolution will be generally facilitated (Wagner and Altenberg 1996; Wagner et al. 2007; Melo et al. 2016). On the contrary, phenotypic evolution will be generally constrained if antagonistic effects are pervasive, and accordant effects are rare. Which one is more general remains largely unknown.

In the previous studies of correlated phenotypic evolution, the contribution of mutational correlation has not been exclusively studied. Most studies follow the quantitative genetic approach featuring the G-matrix, which summarizes additive genetic variances and covariances

among multiple traits (Lande 1979). Other than mutational inputs, the G-matrix could be also affected by genetic drift, selection or migration (Lande 1980; Turelli 1988; Steppan et al. 2002). In simulation studies of the G-matrix, it is easy to study these factors separately. For example, both mutational correlation and correlated selection could constrain the direction of phenotypic evolution in long-term evolution (Jones et al. 2003). In empirical studies of the G-matrix, however, it is difficult to tear apart these factors due to the lack of mutational data. For example, several mixed findings have been made such as the gradual decrease of constraining effect with evolutionary time (Schluter 1996), no general tendency of either constraining or facilitating the rate of fitness increase (Agrawal and Stinchcombe 2009), cases of strong constraining effect (Mitchell-Olds 1996; Hansen et al. 2003), or rapid evolution under strong genetic correlations (Conner et al. 2011). Nevertheless, it is unclear whether mutational correlation is sufficient to explain these results. In fact, mutation correlation was claimed to have smaller impacts in the G-matrix (Arnold et al. 2008), but this claim has not been fully tested.

Furthermore, the empirical studies of the G-matrix tend to focus on the traits with a clear pattern of directional selection in a relatively short-term evolution. Many are even less than ten generations (Conner 2012). Not only does the interest of researchers cause the biased focus, but the requirement of measured selection gradients in the G-matrix framework also does. Therefore, a good survey about the general role of mutational correlation on phenotypic evolution should avoid these biases. Using a large number of traits in a longer-term evolution is desired.

With the helps of systems genetics and high-throughput phenotyping, it is easier to estimate more mutational effects on more traits (Civelek and Lusis 2014). Therefore, mutational correlations for many trait pairs could be quantified. In this study, we will focus on 210 yeast morphological traits with phenotyping data in thousands of mutation lines and decades of

different strains and species. By testing the association between mutational correlations and evolutionary outcomes for a large amount of traits, a general understanding of the role of mutational correlation on phenotypic evolution will be reached.

5.3 RESULTS

5.3.1 Calculation of mutational correlations

We started the analysis from quantifying the mutational correlations among 210 yeast morphological traits. We took advantage of the net *ES* matrix (for simplicity "net" will be omitted below) previously measured in 4716 single-gene deletion lines for these yeast morphological traits (Ho and Zhang 2014). These traits were measured by triple-florescent staining of actin patches, cell walls, and nuclei in yeast cells and image analysis (Ohya et al. 2005), widely been used in many evolutionary studies (Ohya et al. 2015). Three steps are needed for calculating an *ES* for a gene deletion on a trait: (1) calculating the absolute difference between mean phenotypic values in the wild-type and mean phenotypic values in that gene, (2) normalizing the absolute difference by mean phenotypic values measured in the wild-type, and (3) subtracting the simulated pseudo effect size from the normalized raw effect size.

In the *ES* matrix, each trait has a vector of *ES* across all gene deletions. Using two such vectors from two traits, we defined their mutational correlation (COR_M) as the Pearson correlation coefficient between the two vectors. This measurement is suiTable 5.for the evolutionary scenario where traits are largely canalized with mean phenotypic value equal to the fittest value, and the adaptation is considered as shifts of fittest phenotypic values (**Fig. 5.1**). Such shifts, for example, may happen during the change of environmental factors at a location or the migration of organisms to a new location. This "shifting optimal means" scenario is more

consistent with the long term morphological data than other evolution scenarios are (Estes and Arnold 2007). In this scenario, a larger COR_M between two traits means that, when one trait moves away form its current mean, another trait is also more likely to move away from its current mean. Therefore, such mutational correlation has a fitness consequence because the second trait moves away from its fittest phenotypic value.

5.3.2 Prediction of evolutionary correlations by mutational correlations in trait pairs

After quantifying COR_M, we first tested whether COR_M was involved in phenotypic evolution. If two traits have high COR_M, one will expect their evolutionary outcomes should be also more correlated. To verify this, we took advantage of evolutionary distance (*ED*) previously quantified for the same set of 210 morphological traits (Yvert et al. 2013; Ho et al. 2016). Analogues to *ES*, three steps are needed to calculate *ED* between two taxa: (1) calculating the absolute difference between mean phenotypic values in two different taxa, (2) normalizing the absolute difference by mid-mean phenotypic values, and (3) subtracting the simulated pseudo evolutionary distance from the normalized raw evolutionary distance. In this way, potential measurement errors for *ED* could be corrected, and they are corrected by the same way for *ES*.

We first examined the prediction using inter-specific *ED*. Because we have 37 *S. cerevisiae* strains, in total we have 666 pairs of stains. Therefore, there is a vector of 666 *ED* for each trait. We then quantified the evolutionary correlation (COR_E) between two traits as the Pearson correlation coefficient between the two vectors of 666 *ED*. By plotting COR_E against COR_M for all trait pairs, we found a positive correlation (Spearman's $\rho = 0.37$, p-value $< 10^{-300}$; **Fig. 5.2A**). To account for autocorrelation, we randomly shuffled the *ES* matrix in two ways. In the first way, the vector of *ES* across different gene deletion lines for each trait gets permuted (randG). In the second way, the vector of *ES* across different traits for each gene deletion line

gets permuted (randP). The Spearman's ρ between COR_E and COR_M using each permuted *ES* matrix was also quantified by similar method. Comparing with the distribution of ρ of either set of permuted results, we found that the observed ρ between COR_E and COR_M is still significant higher (**Fig. A.4.1A & B**). These results show that the observed ρ between COR_E and COR_M is not purely a statistical artifact, suggesting the correlation in the level of mutations influence the correlation in the level of evolutionary differences.

Given the positive association between COR_M and COR_E, one interesting question is whether COR_M also predicts the Pearson's correlation coefficient between two vectors of phenotypic values across 37 *S. cerevisiae* strains, which is defined as phenotypic correlation (COR_P). After plotting COR_P against COR_M, we found that they also show a significantly positive correlation (Spearman's $\rho = 0.40$, p-value $< 10^{-300}$; **Fig. 5.2B**). In addition, this observed positive correlation is significantly larger than both randG and randP distribution of correlation coefficients (**Fig. A.4.1C & D**). Therefore, the correlation between phenotypic values across strains is substantially affected by the correlation existed in the level of mutations.

After finding the positive correlation between COR_M and COR_E within *S. cerevisiae*, we examined whether COR_M could be used in predicting inter-specific COR_E between *S. cerevisiae* and its sister species *S. paradoxus*. Note that because the *ES* matrix may evolve after the speciation event, the COR_M quantified by the *ES* matrix in *S. cerevisiae* may not have a good prediction power for inter-specific COR_E. With this concern, however, we are still able to demonstrate the prediction power for inter-specific COR_E in the following analysis. Using previous quantified *ED* between each of 37 *S. cerevisiae* strains and *S. paradoxus* strain N17 as well as *ED* between each of 37 *S. cerevisiae* strains and another *S. paradoxus* strain IFO1804, for each trait there is a vector of 74 *ED*. We then calculated COR_E for every trait pair by the

Pearson's correlation coefficient between two vectors of 74 *ED*. We found such COR_E and COR_M are also positively correlated (Spearman's $\rho = 0.21$, p-value $< 10^{-300}$; **Fig. 5.2C**), and this observed ρ is also significantly larger than both randG and randP distribution of ρ (**Fig. A.4.1E & F**). Therefore, mutational correlation also affects the correlation of evolutionary distances across different species.

5.3.3 No linear relationship between mutational correlations and evolutionary distances

In order to test whether mutational correlation tends to constrain or facilitate evolutionary distance for all traits, we quantified mean COR_M and mean *ED* for each trait. The mean COR_M was calculated by the mean across 209 COR_M between the focal trait and each of other traits, representing how much the focal trait is averagely correlated with other traits by mutations. We started the analysis using intra-specific *ED*, and the mean *ED* was calculated by the mean across 666 paired *ED* of 37 *S. cerevisiae* strains. After plotting mean *ED* against mean COR_M across all 210 traits, we did not find a significant Spearman correlation (Spearman's $\rho = -0.069$, p-value = 0.32; **Fig. 5.3A**). We also performed the analysis using inter-specific phenotypic distances. Similar to the intra-specific case, we quantified intra-specific mean *ED* for between 37 *S. cerevisiae* strains and either one of *S. paradoxus* strain N17 or another *S. paradoxus* strain IFO1804. When considering such mean *ED*, we again did not find a significant Spearman correlation between mean *ED* and mean COR_M (Spearman's $\rho = -0.011$, p-value = 0.87; **Fig. 5.3B**).

To make sure our methodology does not cause any biased contribution of ρ, we again performed similar analysis using randG and randP *ES* matrixes. We did not find any case showing the observed ρ significantly different from the simulated distributions of ρ (**Fig. A.4.2**). Therefore, the previous results of statistic tests are likely to be true.

5.3.4 Quadratic relationship between mutational correlations and evolutionary distances

By further examining the plots of mean *ED* against mean COR_M, we noticed that the traits showing high mean COR_M and low mean COR_M tend to have stronger reduced effects on mean *ED*. Therefore, it is suspected that mean COR_M and mean *ED* have a quadratic relationship. To validate this suspicion, we performed the quadratic regression for both intra-specific and inter-specific cases. For inter-specific case, compared with the linear model, the adjusted R^2 in the quadratic model is improved from 0.0099 to 0.071 (**Table 5.1**). In addition, the estimate of the quadratic coefficient in the quadratic model is -0.68, which is significantly different from zero (*p*-value = 1.8×10^{-4}, *t*-test). Therefore, a negative quadratic relationship between mean COR_M and intra-specific mean *ED* is evident.

In the inter-specific case, while we still found the same trend as shown in the intra-specific case, but the statistical support is much weaker. In the quadratic model, the adjusted R^2 is only 0.0095 (**Table 5.1**). While this value is small, it is already a big improvement from the adjusted R^2 in the linear model. Moreover, we still found a negative quadratic coefficient in the quadratic model ($b = -0.59$), but its difference from zero is only marginally significant (*p*-value = 0.057, *t*-test). Therefore, a negative quadratic relationship between mean COR_M and inter-specific mean *ED* is not strongly supported by this simple quadratic model.

5.3.5 Negative quadratic relationship between mutational correlations and mutational correlations in multivariate models

We noted that *ED* is a measurement affected by correlated factors such as the mutation input size or the selection strength for each trait. To make sure our results is not merely a byproduct of ignoring such factors, and potentially improve the inter-specific quadratic models, it is desired to perform a regression analysis with such factors controlled. Therefore, we

quantified two different potential correlated factors to incorporate into the quadratic model, mutational size (*MS*) and direct importance (*DI*). Theoretically, a trait with larger *MS* should tend to have larger *ED*, which may bias the true correlation between COR_M and *ED* if COR_M also correlates with *MS.* In practice, we calculated the mean *ES* across all gene deletions for each trait as a proxy for *MS* measurements.

In addition, *ED* may also be affected by how much each trait directly involved in natural selection, which may bias the true correlation between COR_M and *ED* if COR_M also correlates with the strength of natural selection. While the long-term evolution the history of natural selection is largely unknown, the importance of each trait contributed to fitness can be used as proxy because a more important trait is expected to create stronger selection gradient in natural selection. Previously the trait importance (*TI*) of each trait was quantified by the negative slope of the univariate regression model using *ES* of gene deletions to predict the fitness measurement of gene deletion lines (Ho and Zhang 2014). According to this definition, however, such *TI* not only directly caused by the focal trait but also indirectly caused by other traits correlated with it (Lande and Arnold 1983). To partition the direct effect from indirect effect of *TI*, we performed a multivariate regression model using matrix of *ES* to predict fitness of 2779 single-gene deletion lines with fitness smaller than the wild-type, and the absolute coefficient of each trait is defined as direct importance (*DI*). Note that the overall pattern presented below does not change in the analysis using traits only with either positive *DI* or negative *DI* (**Table A.4.1**). In addition, this multivariate regression model is able to explain ~50% of fitness variation across single deletion lines, which suggests the biological relevance of this model.

With both *MS* and *DI* available, we first calculated partial Spearman's ρ between different mean *ED* and mean COR_M with *MS* and *DI* controlled in both intra-specific and inter-

specific cases. As a result, we still found no significant correlations. The partial Spearman's ρ in the intra-specific case is -0.052 (p-value = 0.94) while the partial Spearman's ρ in the inter-specific case is 0.086 (p-value = 0.22). Therefore, our previous finding of no linear relationship is not a product of ignoring confounding factors such as *MS* and *DI*.

In addition, we attempted to build a multivariate quadratic model using mean COR_M, *MS,* and *DI* as explanatory variables and mean *ED* as the explaining variable in both intra-specific and inter-specific cases. We started from including all linear terms, quadratic terms and interaction terms in the regression model and found that only the linear and quadratic term of mean COR_M and *MS* have coefficients significantly different from zero (nominal p-value < 0.05 by t-test). Eliminating other terms, we performed the regression analysis with those four terms again (**Table 5.2**). As a result, we found both intra-specific and inter-specific cases show high adjusted R^2 values (0.51 and 0.28, respectively), suggesting *MS* has an important role in explaining mean *ED*. More importantly, the negative quadratic coefficients of mean COR_M in both cases are significantly different from zero. The coefficient is equal to -0.79 in the intra-specific cases (p-value = 7.2×10^{-9}, t-test) while the coefficient is equal to -0.73 in the intra-specific cases (p-value = 5.8×10^{-3}, t-test). Therefore, a negative quadratic relationship between mean COR_M and mean *ED* is evident for both intra-specific an inter-specific cases.

5.4 DISCUSSION

In this study, we present data showing the positive correlation between COR_M and COR_E across different pairs of traits. In addition, the association trend between COR_M and *ED* is not linear but negatively quadratic. Therefore, in general, COR_M facilitates *ED* for traits with low COR_M but constrains *ED* for traits with high COR_M. These two observations are not only found

in the inter-specific case but also in the intra-specific case. All of these results suggest a general pattern of multivariate phenotypic evolution below. Because we found the mixed evidence for constraining and facilitating, the frequency of scenarios of synergistic and antagonistic selection section is probably also mixed (**Fig. 5.1**). Therefore a certain level of synergistic selection exists, suggesting there is a certain level of modularity in the genotype-phenotype map. In the context of antagonistic selection, while the direction of phenotypic evolution is constrained by what kind of mutation is available, the indirect effects also rarely overpower the direct effects except for traits with strong indirect effects (bottom left panel in **Fig. 5.1A**). When direct effects can overcome the indirect effects, the mutations with net beneficial effects are still available, which makes new optimal values generally reachable (bottom right panel in **Fig. 5.1A**). Note that this is consistent with the finding that, in the multivariate regression model for fitness, roughly half of traits harbor negative coefficients while another half of traits have positive coefficients. Therefore, their combined effects for fitness could cancel out each other and ends up with a relatively small indirect effect.

We noted that COR_M and *DI* were inferred only using mutational data in *S. cerevisiae*, and we could not exclude the possibility that COR_M has been largely evolved between different *Saccharomyces* species. Such discrepancy may weaken the observed correlations if true correlations are existed. However, our correlation is not completely powerless given that we still found a significant correlation in the inter-specific case. In the future, when more mutational data in different species are available, it would be intriguing to perform these tests again and study whether the architecture of complex traits evolves differently and thus may have consequences on long-term phenotypic evolution.

It is noted that there is also a negative quadratic relationship between *MS* and *ED* (**Table 5.2**). In general, high *MS* predicts high *ED.* However, our results suggest when *MS* is too high, *ED* may be decreased. A possible explanation could be the extremely high *MS* increases the probability of overshooting the new optimal value and thus less likely to be not beneficial. This is consistent with the theoretical prediction that, in Fisher's geometric model, when the mutation size is too high, the rate of fitness increase could be reduced instead of increased (Orr 2000). However, whether this phenomenon is general in empirical data requires more studies.

In the context of correlated phenotypic evolution, pleiotropy is just one source of genetic correlation. There are at least two other kinds of sources which could lead to genetic correlations. First, genetic linkage makes the linked loci tend to be inherited together and thus increase genetic correlation (Futuyma 2013). The effects of genetic linkage, however, should be negligible in the long term because recombination could break linked loci with time being. Even in the short term, it has been shown that it is negligible when recombination rate is high and selection is weak (Lande 1980). Secondly, epistasis may increase genetic correlation between two traits if some combinations of mutations in two loci affecting two traits show non-addictive effects (Cheverud et al. 2004; Wolf et al. 2005). To assay the effect of epistasis, however, require the datasets of phenotypes measured in double mutations. Given the existence of epistasis in the architecture of complex traits (Wagner and Zhang 2011), it would be interesting to study its effects in the future when such dataset is available.

In this study, we demonstrated using a genotype-phenotype map helps the inference of the long-term history of selection. While it is often not easy to directly assay the long-term history of selection, it could be inferred by comparing the mutational inputs and evolutionary outcomes. Our analysis presented in this paper is generally applicable to datasets with both

mutational phenotypes and phenotypic differences measured. With more and more data of system genetics and high-throughput phenotyping coming out, one can study more on how a complex genetic architecture affects phenotypic evolution and infer a more long-term history of evolution.

5.5 MATERIALS AND METHODS

Effect size (*ES*) of a gene deletion for a trait was calculated following previous method (Ho and Zhang 2014). Briefly, raw *ES* was calculated by the absolute difference of mean phenotypic values in the wild-type versus in that gene deletion line and then normalized by mean phenotypic values measured in the wild-type. The pseudo *ES* was simulated by random samples of wild-type phenotypic measurements with the sample sizes unchanged. In the end, net *ES* was calculated by raw *ES* minus pseudo *ES* with a lower bound equal to zero. In the article, *ES* is for net *ES* unless otherwise notified.

Evolutionary distance (*ED*) for a trait between any two group of *Saccharomyces* yeasts was calculated following previous method (Ho et al. 2016). Briefly, raw *ED* was calculated by the absolute difference of mean phenotypic values between two groups and then normalized by mean phenotypic values between two groups. The pseudo *ES* was simulated by random samples of phenotypic measurements in each group separately with the sample sizes unchanged and averaged. In the end, net *ED* was calculated by raw *ED* minus pseudo *ED* with a lower bound equal to zero. In the article, *ED* is for net *ED* unless otherwise notified.

Direct importance (*DI*) for each trait was measured by building a multivariate linear regression model of yeast fitness. In this model, the fitness of 2,779 gene deletion strains that are less fit than the wild type were used as response variables, which was previously measured (Qian

et al. 2012). *ES* matrix of 210 traits for these 2779 gene deletion lines were used as explanatory variables. The coefficient in this multivariate linear regression model for each trait was calculated by using the "glmfit" function in MATLAB.

All of the regression analysis is also performed by the "glmfit" function in MATLAB.

5.6 ACKNOWLEDGEMENTS

We thank Zhang lab members for their valuable comments.

5.7 REFERENCES

Agrawal, A. F. and J. R. Stinchcombe. 2009. How much do genetic covariances alter the rate of adaptation? P Roy Soc B-Biol Sci 276:1183-1191.

Arnold, S. J., R. Burger, P. A. Hohenlohe, B. C. Ajie, and A. G. Jones. 2008. Understanding the Evolution and Stability of the G-Matrix. Evolution 62:2451-2461.

Cheverud, J. M., T. H. Ehrich, T. T. Vaughn, S. F. Koreishi, R. B. Linsey, and L. S. Pletscher. 2004. Pleiotropic effects on mandibular morphology II: Differential epistasis and genetic variation in morphological integration. J Exp Zool Part B 302b:424-435.

Civelek, M. and A. J. Lusis. 2014. Systems genetics approaches to understand complex traits. Nat Rev Genet 15:34-48.

Conner, J. K. 2012. Quantitative Genetic Approaches to Evolutionary Constraint: How Useful? Evolution 66:3313-3320.

Conner, J. K., K. Karoly, C. Stewart, V. A. Koelling, H. F. Sahli, and F. H. Shaw. 2011. Rapid Independent Trait Evolution despite a Strong Pleiotropic Genetic Correlation. Am Nat 178:429-441.

Estes, S. and S. J. Arnold. 2007. Resolving the paradox of stasis: Models with stabilizing selection explain evolutionary divergence on all timescales. Am Nat 169:227-244.

Futuyma, D. J. 2013. Evolution. Sinauer Associates.

Hansen, T. F., C. Pelabon, W. S. Armbruster, and M. L. Carlson. 2003. Evolvability and genetic constraint in Dalechampia blossoms: components of variance and measures of evolvability. J Evolution Biol 16:754-766.

Ho, W.-C., Y. Ohya, and J. Z. Zhang. 2016. Testing the neutral hypothesis of phenotypic evolution. BioRxiv:089987.

Ho, W.-C. and J. Z. Zhang. 2014. The Genotype-Phenotype Map of Yeast Complex Traits: Basic Parameters and the Role of Natural Selection. Mol Biol Evol 31:1568-1580.

Jones, A. G., S. J. Arnold, and R. Borger. 2003. Stability of the G-matrix in a population experiencing pleiotropic mutation, stabilizing selection, and genetic drift. Evolution 57:1747-1760.

Lande, R. 1979. Quantitative Genetic-Analysis of Multivariate Evolution, Applied to Brain - Body Size Allometry. Evolution 33:402-416.

Lande, R. 1980. The Genetic Covariance between Characters Maintained by Pleiotropic Mutations. Genetics 94:203-215.

Lande, R. and S. J. Arnold. 1983. The Measurement of Selection on Correlated Characters. Evolution 37:1210-1226.

Melo, D., A. Porto, J. M. Cheverud, and G. Marroig. 2016. Modularity: Genes, Development, and Evolution. Annu Rev Ecol Evol S 47:463-486.

Mitchell-Olds, T. 1996. Pleiotropy causes long-term genetic constraints on life-history evolution in Brassica rapa. Evolution 50:1849-1858.

Ohya, Y., Y. Kimori, H. Okada, and S. Ohnuki. 2015. Single-cell phenomics in budding yeast. Mol Biol Cell 26:3920-3925.

Ohya, Y., J. Sese, M. Yukawa, F. Sano, Y. Nakatani, T. L. Saito, A. Saka, T. Fukuda, S. Ishihara, S. Oka, G. Suzuki, M. Watanabe, A. Hirata, M. Ohtani, H. Sawai, N. Fraysse, J. P. Latge, J. M. Francois, M. Aebi, S. Tanaka, S. Muramatsu, H. Araki, K. Sonoike, S. Nogami, and S. Morishita. 2005. High-dimensional and large-scale phenotyping of yeast mutants. P Natl Acad Sci USA 102:19015-19020.

Orr, H. A. 2000. Adaptation and the cost of complexity. Evolution 54:13-20.
Qian, W., D. Ma, C. Xiao, Z. Wang, and J. Zhang. 2012. The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. Cell Rep 2:1399-1410.
Schluter, D. 1996. Adaptive radiation along genetic lines of least resistance. Evolution 50:1766-1774.
Steppan, S. J., P. C. Phillips, and D. Houle. 2002. Comparative quantitative genetics: evolution of the G matrix. Trends Ecol Evol 17:320-327.
Turelli, M. 1988. Phenotypic Evolution, Constant Covariances, and the Maintenance of Additive Variance. Evolution 42:1342-1347.
Wagner, G. P. and L. Altenberg. 1996. Perspective: Complex adaptations and the evolution of evolvability. Evolution 50:967-976.
Wagner, G. P., M. Pavlicev, and J. M. Cheverud. 2007. The road to modularity. Nat Rev Genet 8:921-931.
Wagner, G. P. and J. Z. Zhang. 2011. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. Nat Rev Genet 12:204-213.
Wolf, J. B., L. J. Leamy, E. J. Routman, and J. M. Cheverud. 2005. Epistatic pleiotropy and the genetic architecture of covariation within early and late-developing skull trait complexes in mice. Genetics 171:683-694.
Yvert, G., S. Ohnuki, S. Nogami, Y. Imanaga, S. Fehrmann, J. Schacherer, and Y. Ohya. 2013. Single-cell phenomics reveals intra-species variation of phenotypic noise in yeast. BMC systems biology 7:54.

Table 5.1 Statistics for linear and quadratic regression models using mean mutational correlations (COR_M) explaining intra-specific or inter-specific mean evolutionary differences (*ED*)

		Source of mean *ED*	
		intra-species	inter-species
Linear model	adjusted R^2	0.0099	-0.0032
	b_1	-0.033	-0.019
	p-value for b_1	0.081	0.56
Quadratic model	adjusted R^2	0.071	0.0095
	b_1	0.34	0.31
	p-value for b_1	7.5×10^{-4}	0.078
	b_{11}	-0.68	-0.59
	p-value for b_{11}	1.8×10^{-4}	0.057

Note- the linear model is $y = b_0 + b_1x$ while the quadratic model is $y = b_0 + b_1x + b_{11}x^2$ where y is mean *ED* and x is mean COR_M; each *p*-value is calculated by the t-tests against the null hypothesis where the coefficient equals zero.

Table 5.2 Statistics for multivariate quadratic regression models using mean mutational correlations (COR_M) and mutational sizes (MS) explaining intra-specific or inter-specific mean evolutionary differences (ED)

	Source of mean ED	
	intra-species	inter-species
adjusted R^2	0.51	0.28
b_1	0.42	0.40
p-value for b_1	4.4×10^{-8}	6.7×10^{-3}
b_2	2.6	3.5
p-value for b_2	1.2×10^{-21}	1.2×10^{-11}
b_{11}	-0.79	-0.73
p-value for b_{11}	7.2×10^{-9}	5.8×10^{-3}
b_{22}	-15	-21
p-value for b_{22}	5.6×10^{-12}	1.4×10^{-6}

Note- the multivariate quadratic model is $y = b_0 + b_1x_1 + b_2x_2 + b_{11}x_1^2 + b_{22}x_2^2$ where y is mean ED, x_1 is mean COR_M, and x_2 is MS; each p-value is calculated by the t-tests against the null hypothesis where the coefficient equals zero.

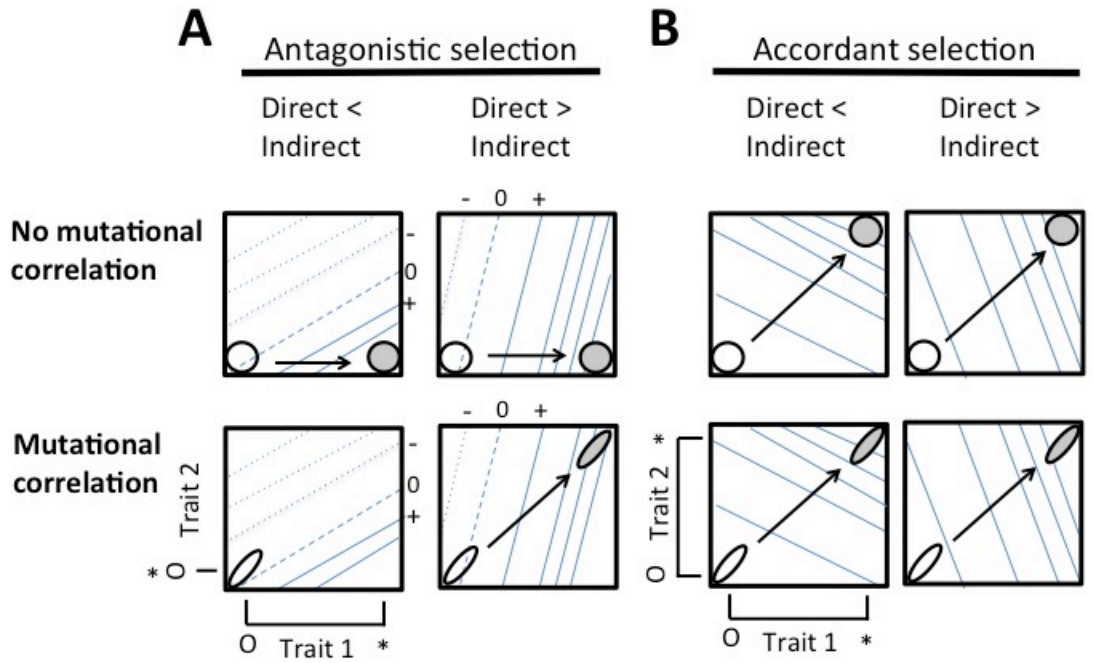

Figure 5.1 Mutational correlation constrains phenotypic evolution when indirect effects overpower direct effects under antagonistic selection. The hypothetical evolutionary scenario considers the phenotypic evolution of two traits with original optimal value at O and new optimal value at *. One trait is on x-axis while the other trait is on y-axis. Each panel represents a combination of three factors: with or without mutational correlation, with antagonistic selection (trait 1 shifted but trait 2 stabilized) or accordant selection (trait 1 and trait 2 both shifted), and direct effects (focusing on trait 1) stronger or weaker than indirect effects (due to trait 2). Within each panel, the open circle or ellipse represents the area accessible by mutations for the initial population before the shift of optimal values. With strong mutational correlation, only mutations around the diagonal are allowed, making an ellipse area. With no mutational correlation, every direction is accessible, making a circle area. The closed circle or ellipse represents the evolutionary outcome in the shift of optimal values. The arrow shows the moving direction form the origin to the outcome. The lines inside each panel represent isoclines for fitness. Compared with the original location, dotted lines represent the isoclines with reduced fitness (-), dashed lines represent the isoclines with no change of fitness (0), and solid lines represent the isoclines with increased fitness (+). Under mutational correlation, antagonistic selection, and direct effects weaker than indirect effects, no beneficial mutation is accessible in the beginning, shown by the ellipse not crossing the isocline 0, and therefore the phenotypic evolution is constrained.

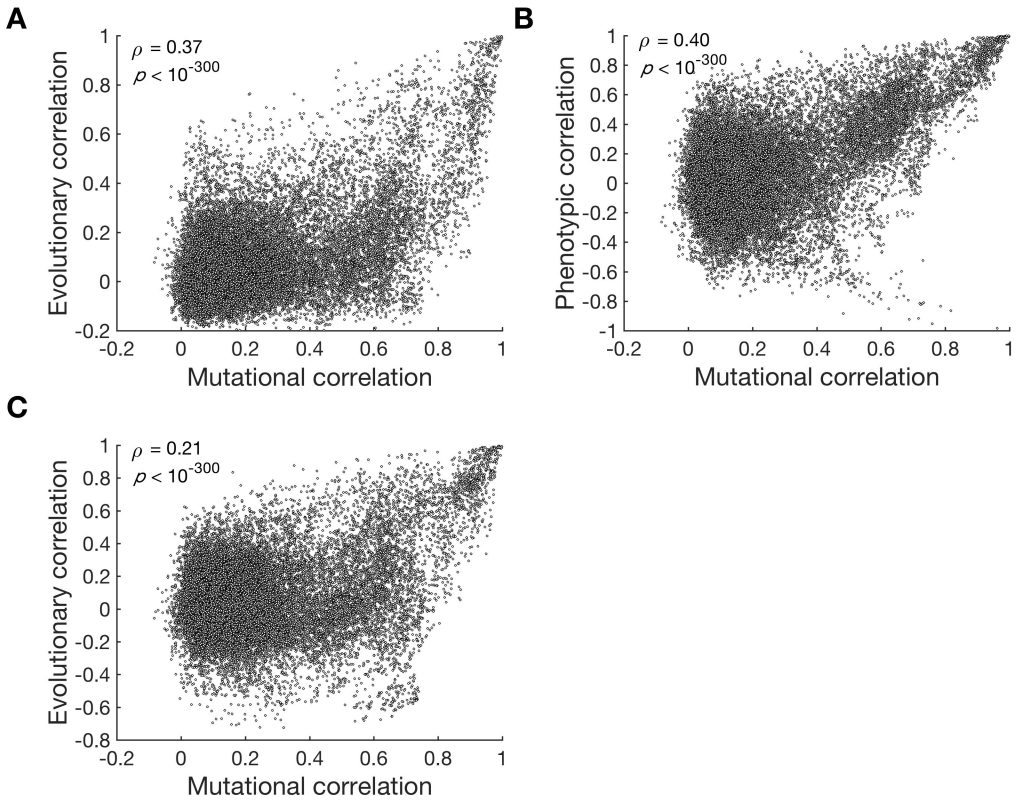

Figure 5.2 Mutational correlations (COR_M) predict evolutionary correlations (COR_E) or phenotypic correlations (COR_P). Each dot is a pair of traits. **(A)** COR_E calculated using evolutionary difference (*ED*) of 666 pairs of every two of 37 *Saccharomyces cerevisiae* strains. **(B)** COR_P calculated using mean phenotypic value in 37 *S. cerevisiae* strains. **(C)** COR_E calculated using *ED* between 37 *S. cerevisiae* strains and *S. paradoxus.*

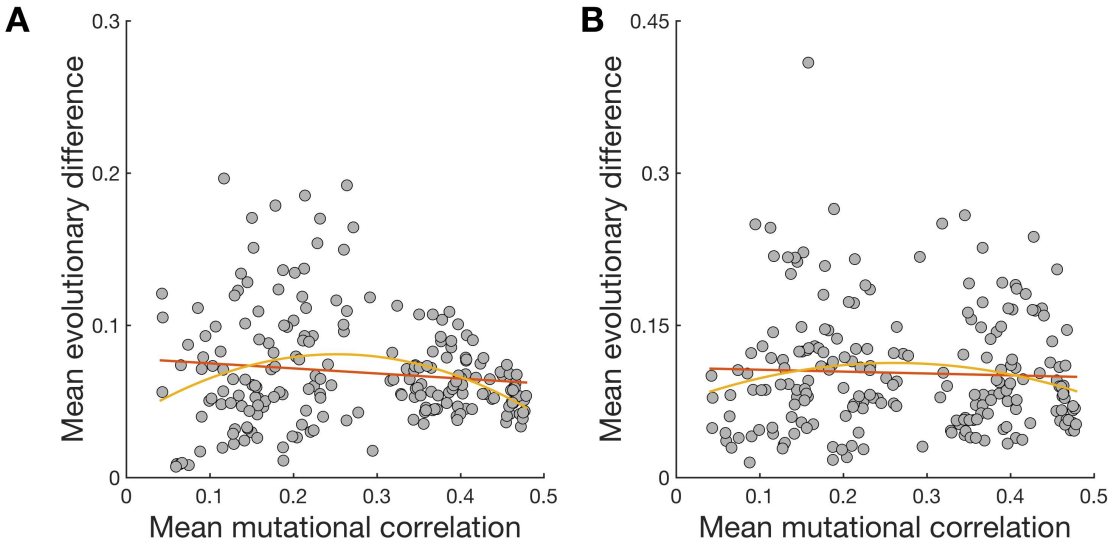

Figure 5.3 Negative quadratic relationship between mean evolutionary differences (*ED*) and mean mutational correlations (COR_M) across yeast morphological traits. (A) *ED* among 37 *Saccharomyces cerevisiae* strains. **(B)** *ED* between *S. cerevisiae* strains and *S. paradoxus* strains. Each dot is a trait. The linear brown line represents the fitting of a linear model while the curved orange line represents the fitting line of a quadratic model, where the fitting statistics are listed in **Table 5.1**.

Chapter 6

Evolutionary Adaptations to New Environments Generally Reverse Plastic Phenotypic Changes

6.1 ABSTRACT

Organismal adaptation to a new environment typically starts with plastic phenotypic changes followed by genetic changes, but whether the plastic changes are steppingstones to genetic adaptations is debated. Here we address this question by investigating gene expression and metabolic flux changes in the two-phase adaptation process using transcriptomic data from multiple experimental evolution studies and computational metabolic network analysis, respectively. We discover that genetic changes more frequently reverse than reinforce plastic phenotypic changes in virtually every adaptation. Metabolic network analysis reveals that, even in the presence of plasticity, organismal fitness drops precipitously after environmental shifts, but largely recovers through subsequent evolution. Such fitness trajectories during adaptation explain why plastic phenotypic changes are genetically compensated rather than strengthened. While phenotypic plasticity may serve as an emergent response to new environments that is necessary for organismal survival, we conclude that it does not generally facilitate genetic adaptation.

6.2 INTRODUCTION

Phenotypic adaptation to a new environment consists of two phases (**Fig. 6.1a**). In the first phase, the environmental shift induces phenotypic changes without genetic mutation; such changes are referred to as plastic changes irrespective of their fitness effects. In the second phase, phenotypes are altered by mutations that accumulate during the adaptive evolution. While most past evolutionary studies focused on the second phase, recent years have seen a growth in the argument for the importance of the first phase in adaptation (Price et al. 2003; West-Eberhard 2003; Pigliucci et al. 2006; Lande 2009; Pfennig et al. 2010; Moczek et al. 2011; Laland et al. 2014; Laland et al. 2015; Levis and Pfennig 2016). Specifically, it is suggested that plastic phenotypic changes are often necessary for organismal survival in a new environment (Baldwin 1896; Robinson and Dukas 1999), which is essential because no adaptive evolution is possible if the environmental shift kills all individuals. Furthermore, it is argued that plasticity moves the phenotypic value of an organism closer to the adapted state in the new environment and hence is a steppingstone to adaptation (Waddington 1953; Price et al. 2003; West-Eberhard 2003) (**Fig. 6.1b**). While some case studies appear to support this latter assertion (Suzuki and Nijhout 2006; Ledon-Rettig et al. 2008; Levis and Pfennig 2016), its general validity remains unclear (Ghalambor et al. 2007). Answering this question is of special significance, because the school of extended evolutionary synthesis believes that plasticity is critical to adaptation and hence is requesting a major revision of the modern synthesis of evolutionary biology, where the role of plasticity in adaptation is thought to be largely neglected (Laland et al. 2014; Laland et al. 2015).

For a trait, its plastic phenotypic change induced by an environmental shift and the subsequent genetic change during the adaption to the new environment could be in the same direction toward the optimal phenotypic value in the new environment. In this case, the plastic

change is reinforced by the adaptive genetic change and hence is considered adaptive (Ghalambor et al. 2007; Ghalambor et al. 2015) (**Fig. 6.1b**). The plastic change and the subsequent genetic change could also be in opposite directions. In this case, the plastic change is reversed by the adaptive genetic change and is thus commonly considered non-adaptive (Ghalambor et al. 2007; Ghalambor et al. 2015) (**Fig. 6.1c**). The hypothesis that plasticity facilitates adaptation is supported if reinforcement is more prevalent than reversion in a large sample of traits during a large number of adaptations; otherwise, the hypothesis is refuted. Thus, a test of the hypothesis can be performed by respectively phenotyping and comparing adapted organisms in the original and new environments as well as the organisms right after the environmental shift (i.e., after plastic changes but before genetic changes). Early tests used morphological, physiological, or behavioral traits, but the number of traits examined was small and the results varied among studies (Ghalambor et al. 2007). Recent tests with transcriptome data suggested that gene expression level reversion is more prevalent than reinforcement during experimental evolution (Fong et al. 2005; Sandberg et al. 2014; Ghalambor et al. 2015; Rodriguez-Verdugo et al. 2016). Although the number of traits is large in these recent studies, their analyses vary, rendering the interpretation and among-study comparison difficult. We thus reanalyze using a uniform method the transcriptome data from these studies as well as those from another study that did not address the role of plasticity in adaptation (Tamari et al. 2016).

More importantly, five considerations prompt us to expand the analysis from gene expression levels to metabolic fluxes. First, it is desirable to test the hypothesis across diverse environmental shifts, but experimental evolution studies with transcriptome data are currently limited in this aspect. By contrast, fluxes in well-annotated metabolic networks can be computationally predicted with reasonably high accuracy under a wide range of environments

(Segre et al. 2002; Orth et al. 2010). Second, it is necessary to examine if the finding from gene expression traits applies to other phenotypic traits. Third, organisms acquired at the end of experimental evolution are usually partially rather than fully adapted to the new environment, making the distinction between reinforcement and reversion less certain. Fourth, in experimental evolution, it is unknown whether an observed gene expression change is beneficial, neutral, or even deleterious. For example, an expression change accompanying adaptation could be responsible for, a result from, or even unrelated to the fitness gain. Some authors assume that expression changes observed in replicate experiments are beneficial (Ghalambor et al. 2015), but it is also possible that they are consequences of adaptation and have positive, zero, or negative fitness effects. Thus, not all expression changes observed in experimental evolution are relevant to the hypothesis that plasticity is a steppingstone to genetic adaptation. By contrast, in the metabolic network analysis, all flux changes observed in the maximization of fitness are required and therefore are by definition beneficial. It has been shown, for instance, that upon the maximization of fitness, alteration of any non-zero flux would be deleterious (Ho and Zhang 2016). Last and most importantly, because the regulatory and evolutionary mechanisms of gene expression changes are not well understood, it would be difficult to discern the mechanistic basis of expression level reinforcement or reversion. By contrast, patterns of computationally predicted flux changes can be understood mechanistically by the metabolic model used in the prediction. We thus test whether plasticity facilitates adaptation by computational metabolic flux analysis of the model bacterium *Escherichia coli*. Our analyses of transcriptome and fluxome changes in numerous adaptations consistently show that phenotypic reinforcement is not only no more but actually less prevalent than reversion, indicating that plasticity is not a steppingstone to

genetic adaptation. More importantly, we uncover the underlying cause of the preponderance of phenotypic reversion.

6.3 RESULTS

6.3.1 Preponderance of gene expression level reversion in experimental evolution

We identified five studies that conducted six different adaptation experiments and collected transcriptome data that suit our study. These six experiments included 10 replicates of *E. coli* adapting to a high-temperature environment (Sandberg et al. 2014), six replicates of another strain of *E. coli* adapting to a high-temperature environment (Rodriguez-Verdugo et al. 2016), seven replicates of *E. coli* adapting to a glycerol medium (Fong et al. 2005), seven replicates of *E. coli* adapting to a lactate medium (Fong et al. 2005), one replicate each of 12 different yeast (*Saccharomyces cerevisiae*) strains adapting to an xylulose medium (Tamari et al. 2016), and two replicates of guppies (*Poecilia reticulata*) adapting to a low-predation environment(Ghalambor et al. 2015). In total, we analyzed 44 cases of adaptation.

In each case, transcriptome data were respectively collected for the organisms in the original environment (*o* for the original stage), shortly after their exposure to the new environment (*p* for the plastic stage), and at the conclusion of the experimental evolution in the new environment (*a* for the adapted stage) (**Fig. 6.1a**). Let the expression levels of a gene at the *o*, *p*, and *a* stages be L_o, L_p, and L_a, respectively. In each experiment, we first identified genes with appreciable plastic changes (*PC*) in expression level by requiring $PC = |L_p - L_o|$ to be greater than a preset cutoff. We also identified genes with appreciable genetic changes (*GC*) in expression level by requiring $GC = |L_a - L_p|$ to be greater than the same preset cutoff. For those genes showing both appreciable plastic and appreciable genetic changes, we ask whether the two

changes are in the same direction (i.e., reinforcement) or opposite directions (i.e., reversion) (**Fig. 6.1b-c**). We used 20% of the original gene expression level (i.e., $0.2L_o$) as the cutoff in the above analysis. The fraction of genes exhibiting expression level reinforcement (C_{RI}) is smaller than the fraction of genes exhibiting reversion (C_{RV}) in 42 of the 44 adaptations, and the difference between C_{RI} and C_{RV} is significant in 40 of these 42 cases (nominal $P < 0.05$; two-tailed binomial test) (**Fig. 6.1d**). For the remaining two adaptations, $C_{RI} > C_{RV}$, but their difference is significant in only one of the two cases (**Fig. 6.1d**). The general preponderance of expression level reversion (i.e., 42 of 44 cases) in adaptation is statistically significant ($P = 1.1\times10^{-10}$, two-tailed binomial test). The same trend is evident when the cutoff is altered to $0.05L_o$ (**Fig. A.5.1a**) or $0.5L_o$ (**Fig. A.5.2a**), suggesting that the above finding is robust to the cutoff choice. Clearly, the transcriptomic data do not support the hypothesis that plasticity facilitates genetic adaptation.

6.3.2 Predominance of metabolic flux reversion in environmental adaptations

To study the generality of the above finding and understand its underlying cause, we expanded the comparison between phenotypic reinforcement and reversion to metabolic fluxes. Specifically, we computationally predicted plastic and genetic flux changes during environmental adaptations using *i*AF1260, the reconstructed *E. coli* metabolic network (Feist et al. 2007). We used flux balance analysis (FBA) to predict the optimized fluxes of adapted organisms in the original and new environments, respectively, under the assumption that the biomass production rate, a proxy for fitness, is maximized by natural selection (Orth et al. 2010). FBA predictions match experimental measures reasonably well for organisms adapted to their environments (Edwards et al. 2001; Ibarra et al. 2002; Fong and Palsson 2004; Papp et al. 2004;

Wang and Zhang 2009a; Lewis et al. 2010) and are commonly used in the study of genotype-environment-phenotype relationships (Papp et al. 2004; Segre et al. 2005; Pal et al. 2006; Wang and Zhang 2009a; He et al. 2010; Costenoble et al. 2011; Wang and Zhang 2011; Barve et al. 2012; Harcombe et al. 2013; Bordbar et al. 2014; Ho and Zhang 2016). We employed minimization of metabolic adjustment (MOMA) to predict plastic flux changes upon environmental shifts, because MOMA can recapitulate the immediate flux response to environmental or genetic perturbations (Segre et al. 2002). We treated the flux of each reaction in the metabolic network as a trait, and modeled environmental shifts by altering the carbon source available to the network. There are 258 distinct exchange reactions in *i*AF1260, each transporting a different carbon source. We therefore examined 258 different single-carbon source environments.

We started the analysis by using glucose as the carbon source in the original environment, because this environment was the benchmark in *i*AF1260 construction (Feist et al. 2007). We then considered the adaptations of *E. coli* to the 257 new environments each with a different single-carbon source. We found that these new environments are naturally separated into two groups in the MOMA-predicted biomass production rate, a proxy for the fitness at stage p (f_p) (**Fig. A.5.3**). One group shows $f_p < 10^{-4}$, suggesting that *E. coli* is unlikely to sustain in these new environments. We therefore focused on the remaining 50 new environments with $f_p > 10^{-4}$, to which *E. coli* can presumably adapt.

Defining flux reinforcement and reversion and using the cutoff of $0.2L_o$ as in the transcriptome analysis, we found C_{RV} to be significantly greater than C_{RI} (nominal $P < 10^{-10}$, two-tailed binomial test) in each adaptation. The chance probability that all 50 adaptations show $C_{RV} > C_{RI}$ is 1.8×10^{-15} (two-tailed binomial test; **Fig. 6.2a**), suggesting a general predominance of

flux reversion. The mean and median C_{RV} are 30.2% and 30.5%, respectively, while those for C_{RI} are only 1.0% and 0.8%, respectively. The above trend holds when we alter the cutoff to $0.05L_o$ (**Fig. A.5.1b**) or $0.5L_o$ (**Fig. A.5.2b**). Because an FBA or MOMA problem may have multiple solutions, the order of the reactions in the stoichiometric matrix could affect the specific solution provided by the solver. Nevertheless, when we randomly shuffled the reaction order in *i*AF1260, the general pattern of $C_{RV} > C_{RI}$ is unaltered (**Fig. A.5.4**). Because quadratic programming required by MOMA is harder to solve than linear programming used in FBA, C_{RV} could have been overestimated compared with C_{RI}. To rectify this potential problem, we designed a quadratic programming-based MOMA named "MOMA-b" and used it instead of FBA to predict fluxes at the stage *a* (see Methods), but found that C_{RV} still exceeds C_{RI} (**Fig. A.5.5**). Thus, this trend is not a technical artifact of the solver difference between MOMA and FBA.

6.3.3 Flux reversion largely restores the fluxes in the original environment

To examine whether the flux reversion during genetic adaptation restores the fluxes at stage *o*, we compared the total change $TC = |L_a - L_o|$ with $0.2L_o$ for each reaction with flux reversion, in each adaptation. If $TC < 0.2L_o$, the flux is considered restored (**Fig. 6.2b**). Otherwise, we further compare *PC* with *GC*. If $GC > PC$, the flux is over-restored; otherwise, it is under-restored (**Fig. 6.2b**). Across the 50 adaptations, the mean fractions of reactions showing "restored", "over-restored", and "under-restored" flux reversion are 26.4%, 3.1%, and 0.7%, respectively, and the medians are 30.2%, 0.3%, and 0.1%, respectively (**Fig. 6.2c**). Clearly, flux reversion largely restores the fluxes at the *o* stage.

6.3.4 Predominance of flux reversion irrespective of the original environment

To investigate the generality of our finding of the predominance of flux reversion, we also examined adaptations with a non-glucose original environment. For many original environments, however, only a few new environments are adaptable by the *E. coli* metabolic network. We thus focused on 41 original environments (including the previously used glucose environment) that each has more than 20 adaptable (i.e., $f_p > 10^{-4}$) new environments. For each of these original environments, we calculated the C_{RI}/C_{RV} ratio for each adaptable new environment, and found it to be typically lower than 0.1 (**Fig. 6.2d**). We then computed the median C_{RI}/C_{RV} across all adaptable new environments from each original environment. Across the 41 original environments, the largest median C_{RI}/C_{RV} is 0.11 and the median of median C_{RI}/C_{RV} is only 0.02. Hence, regardless of the original environment, flux reversion is much more prevalent than reinforcement during genetic adaptations to new environments.

6.3.5 Why phenotypic reversion is more frequent than reinforcement

Our finding that phenotypic reinforcement is not only no more but actually much less frequent than reversion is unexpected and hence demands an explanation. The observation of this trend in both transcriptomic and fluxomic analyses suggests a general underlying mechanism. Geometrically, it is obvious that when $PC > TC$, the genetic change must reverse the plastic change (left box in **Fig. 6.3a**). By contrast, when $PC < TC$, reversion and reinforcement are equally likely if no other bias exists (right box in **Fig. 6.3a**). Let the probability of $PC > TC$ be $q > 0$. C_{RI}/C_{RV} is expected to be $[0.5(1-q)]/[0.5(1-q)+q] = (1-q)/(1+q) < 1$. In other words, as long as $PC > TC$ for a few traits, reversion is expected to be more frequent than reinforcement (under no other bias).

To seek empirical evidence for the above explanation, for each of the 44 cases of experimental evolution, we calculated the fraction of genes whose expression changes satisfy *PC* > *TC* (**Fig. 6.3b**). The mean and median fractions are 0.51 and 0.48, respectively. Furthermore, after we remove all genes for which *PC* > *TC*, there is no longer a preponderance of reversion (**Fig. A.5.6a**), indicating the sufficiency of our explanation. Similarly, we computed the fraction of metabolic reactions showing *PC* > *TC* in the adaptation of the *E. coli* metabolic network from the glucose environment to each of the 50 new environments (**Fig. 6.3c**). The mean and median fractions are 0.85 and 0.93, respectively. Similarly, after the removal of reactions showing *PC* > *TC*, there is no general trend of more reversion than reinforcement across the 50 adaptations (**Fig. A.5.6b**). These findings from transcriptome and fluxome analyses support that the abundance of reversion relative to reinforcement is explainable by the occurrence of greater *PC* than *TC* for non-negligible fractions of traits.

Why does *PC* exceed *TC* for many traits? A likely reason is that plastic changes allow organisms to survive upon a sudden environmental shift but the fitness is much reduced compared with that in the original environment as well as that after the adaptation to the new environment. Thus, the overall physiological state of the organisms may be quite similar between the adapted stages in the original and new environments, but is much different in the low-fitness plastic stage right after the environmental shift. This may explain why *PC* exceeds *TC* for many traits, regardless of whether the trait values are causes or consequences of the organismal fitness and physiology.

We found strong evidence for the above model by metabolic network analysis. First, using the predicted biomass production rate as a proxy for fitness, we compared the *E. coli* fitness at the plastic stage (f_p) and that after adaptation to a new environment (f_a), relative to that

in the original glucose environment, for each of the adaptations to the 50 new environments. In all cases, $f_p < 1$ (**Fig. 6.3d**), confirming that environmental shifts cause fitness drops before genetic adaptation. We found that f_a is typically close to 1, although in a few new environments it is much greater than 1 (**Fig. 6.3d**). In a $\log_{10}$ scale, f_p is more different from 1 than is f_a in 43 of the 50 adaptations ($P = 1.0\times10^{-7}$; one-tailed binomial test). Second, our model assumes an association between flux changes and fitness changes (Ho and Zhang 2016). Across the 50 adaptions from the glucose environment, there is a strong negative correlation between f_p and mean *PC* (Spearman's $\rho = -0.98$, $P < 10^{-300}$; **Fig. 6.3e**). An opposite correlation exists between f_a and mean *TC* ($\rho = 0.57$, $P = 1.1\times10^{-5}$; **Fig. 6.3f**). Together, our analyses demonstrate that the primary reason for a higher frequency of phenotypic reversion than reinforcement during adaptation is that, in terms of fitness and associated phenotypes, organisms at stage *p* are more different than those at stage *a*, when compared with those at stage *o*.

6.3.6 Predominance of phenotypic reversion in random metabolic networks

The plastic and genetic changes in gene expression level and metabolic flux during adaptations depend respectively on the regulatory network and metabolic network of the species concerned. Because these networks result from billions of years of evolution, one wonders whether the predominance of phenotypic reversion is attributable to the evolutionary history of the species studied, especially the environments in which the species and its ancestors have been selected in the past, or an intrinsic property of any functional system. To address this question, we applied the same analysis to 500 functional random metabolic networks previously generated (Ho and Zhang 2016). These networks were constructed from *i*AF1260 by swapping its reactions with randomly picked reactions from the universe of all metabolic reactions in KEGG

(Kanehisa et al. 2016) as long as the network has a non-zero FBA-predicted fitness in the glucose environment upon each reaction swap (Barve and Wagner 2013).

Only 20 new environments that *i*AF1260 can adapt to (from the glucose environment) are adaptable by at least 20 of the 500 random networks. We thus analyzed the adaptations of random networks to each of these 20 new environments, with the glucose environment being the original environment. For each new environment, the median C_{RV} of all random networks that can adapt to this environment is generally around 0.1 (boxplots in **Fig. 6.4a**), with the median of median C_{RV} being 0.11. By contrast, median C_{RI} across random networks for a new environment is generally below 0.01 (boxplots in **Fig. 6.4b**), with the median of median C_{RI} being 0.0033. Median C_{RI}/C_{RV} ratio across random networks for a new environment is generally below 0.05 (boxplot in **Fig. 6.4c**), with the median of the median C_{RI}/C_{RV} being 0.0033. Clearly, the predominance of flux reversion is also evident in functional random networks, suggesting that this property is intrinsic to any functional metabolic network rather than a product of particular evolutionary histories. Indeed, the mechanistic explanation for this property in actual organisms (**Fig. 6.3**) holds in the random metabolic networks. Specifically, the fraction of reactions exhibiting $PC > TC$ is substantial (**Fig. 6.4d**) and f_p is mostly lower than 1 (**Fig. 6.4e**). Furthermore, f_p is generally more different from 1 than is f_a in a $\log_{10}$ scale, because $|\log_{10} f_p| - |\log_{10} f_a|$ is largely positive (**Fig. 6.4f**).

Intriguingly, however, for 19 of the 20 new environments, C_{RV} in the *E. coli* metabolic network exceeds the median C_{RV} in the random networks (**Fig. 6.4a**). A similar but less obvious trend holds for C_{RI} (**Fig. 6.4b**). For 16 of the 20 new environments, C_{RI}/C_{RV} from *E. coli* is smaller than the median C_{RI}/C_{RV} of the random networks ($P = 0.012$, two-tailed binomial test; **Fig. 6.4c**). Hence, although both the *E. coli* metabolic network and random networks show a

predominance of flux reversion, this phenomenon is more pronounced in the former than the latter. Mechanistically, this disparity is explainable at least qualitatively by our model in the previous section. Specifically, for 15 of the 20 new environments, the fraction of *E. coli* reactions with $PC > TC$ exceeds the corresponding median fraction in random networks (P = 0.021, one-tailed binomial test; **Fig. 6.4d**). For all 20 new environments, f_p of *E. coli* is lower than the median f_p of random networks ($P = 9.5\times10^{-7}$, one-tailed binomial test; **Fig. 6.4e**). For 19 of the 20 new environments, $|\log_{10} f_p| - |\log_{10} f_a|$ for *E. coli* is larger than the corresponding median value for the random networks ($P = 2.0\times10^{-5}$, one-tailed binomial test; **Fig. 6.4f**). But, why is f_p of *E. coli* lower than that of random networks? One potential explanation is that the composition and structure of the *E. coli* metabolic network have been evolutionarily optimized for growth in the glucose environment and/or related environments, while the same is not true for the random networks, which were only required to be viable in the glucose environment. As a result, when glucose is replaced with a new carbon source in a new environment, the fitness of *E. coli* drops substantially, but those of random networks may only drop mildly. Although the absolute fitness in the plastic stage may well be higher for *E. coli* than the random networks, the relative fitness, which f_p is, is expected to be lower for *E. coli* than the random networks. Thus, the higher prevalence of flux reversion relative to reinforcement in *E. coli* than random networks is likely a byproduct of stronger selection of *E. coli* compared with random networks in the original environment used in our adaptation analysis.

6.4 DISCUSSION

Using the transcriptome data collected in a total of 44 cases of six different experimental evolutionary adaptations of three species (*E. coli*, yeast, and guppy) and the computationally

predicted fluxomes of *E. coli* in hundreds of different environmental adaptations, we showed that genetic adaptations to new environments overwhelmingly reverse, rather than reinforce plastic phenotypic changes. Our fluxome analyses have several caveats worth discussion. First, because MOMA minimizes the total squared flux difference from the original flux, plastic changes could have been underestimated, but this bias would only make our conclusion more conservative. Second, a bias could exist owing to potentially different accuracies of MOMA and FBA that are respectively used to predict plastic and genetic flux changes. To tackle this problem, we designed a MOMA-based algorithm to infer both plastic and genetic changes, but found the results to be qualitatively unchanged (**Fig. A.5.5**). Third, we considered only single-carbon source environments in our analyses while the natural environments of *E. coli* can be much more complex. We thus simulated adaptations from the glucose environment to environments with mixed carbon sources (see Methods), but found our conclusion unaltered (**Fig. A.5.7**). Fourth, computational flux predictions by FBA and MOMA inevitably contain errors. But, the fact that our fluxome-based conclusion qualitatively match the transcriptome-based conclusion suggests that our fluxome analysis is reliable. Furthermore, some of our flux analyses are largely immune to flux prediction errors. For example, because the *E. coli* metabolic network and random metabolic networks were analyzed using the same method, their difference discovered is unlikely explainable by flux prediction errors. As mentioned, our transcriptome analysis also has a caveat. Because the organisms were not fully adapted to the new environments at the end of experimental evolution, it is possible that a trait currently not considered to show reversion or reinforcement due to insufficient genetic change would show one of these two patterns if allowed to adapt further. However, because our results are robust to

different cutoffs used ($0.05L_o$ to $0.5L_o$) in the definition of genetic changes (**Figs. S2-S3**), our finding of the preponderance of expression level reversion is minimally impacted by this caveat.

In our analyses, we compared the directions of plastic and genetic changes for all traits with plastic and genetic changes greater than a preset cutoff. One may argue that evolutionary biologists are interested only in those traits whose optimal values in two different environments are different. We thus focused on traits whose $TC > 0.2L_o$. Of the 44 cases of experimental evolution, 33 showed more expression level reversion than reinforcement ($P = 0.0013$, two-tailed binomial test; **Fig. A.5.8a**). Of the 50 environmental adaptations of the *E. coli* metabolic network, three cases had equal numbers of flux reversion and reinforcement. Of the remaining 47 cases, 22 showed more reversion than reinforcement, while 25 showed the opposite ($P = 0.77$, two-tailed binomial test; **Fig. A.5.8b**). Hence, even among traits with $TC > 0.2L_o$, there is no evidence for significantly more reinforcement than reversion either.

In all analyses, we regarded phenotypic reinforcement as evidence for the steppingstone role of plasticity in adaptation, and regarded phenotypic reversion as evidence against this hypothesis (Ghalambor et al. 2015). One could argue that although reinforcement supports the hypothesis, reversion is not necessarily against the hypothesis. Specifically, if a plastic change moves the organismal phenotype closer to the optimum in the new environment but overshoots, the genetic change required to bring the phenotype to the optimum may be smaller than that in the absence of plasticity. To investigate this scenario, we considered all traits with *PC* and *GC* both larger than a preset cutoff as was done in the definition of reinforcement and reversion. We then regarded the plastic change of a trait as constructive if $GC < TC$, or destructive if $GC > TC$. We computed the fraction of traits showing constructive changes (C_{CON}) and that showing destructive changes (C_{DES}) in each adaptation. In 32 of the 44 cases of experimental evolution,

C_{DES} exceeds C_{CON}, demonstrating an overall preponderance of destructive plasticity ($P = 3.7\times10^{-3}$, two-tailed binomial test; **Fig. A.5.9a**). Furthermore, C_{DES}/C_{CON} is likely underestimated in the above analysis, because the fact that adaptions to new environments had not ceased by the end of experimental evolution means that cases currently classified as constructive can become destructive. This is because *GC* will probably increase in further adaptations while *TC* will either increase by the same amount as *GC*−resulting in no change in *GC* relative to *TC*−or decrease, causing an increase in *GC* relative to *TC*. For the adaptations of the *E. coli* metabolic network from the glucose environment to the 50 new environments, the above underestimation does not exist, and C_{DES} is found to exceed C_{CON} in every adaptation ($P = 1.8\times10^{-15}$, two-tailed binomial test; **Fig. A.5.9b**). Thus, the comparison between constructive and destructive plasticity also refutes the hypothesis that plasticity is a steppingstone to adaptation.

We provided evidence that the cause for the preponderance of phenotypic reversion is that, even with plasticity, organismal fitness drops precipitously after environmental shifts, but more or less recovers through subsequent evolution; such fitness trajectories dictate that many fitness-associated traits are drastically altered at the plastic stage but are then restored via adaptive evolution. Our model is supported by the observation that stress response is frequently associated with growth cessation as well as reductions in the expression levels of growth-related genes and concentrations of central metabolites (Gasch et al. 2000; Lopez-Maury et al. 2008; Jozefczuk et al. 2010). It is also consistent with the notion that genetic adaptation tends to rebalance the energy allocation in growth that is broken in stress response and that the physiological state of organisms after the rebalance in the new environment is similar to that in the original environment (Fong et al. 2005; Grether 2005; Lopez-Maury et al. 2008; Carroll and

Marx 2013; Rodriguez-Verdugo et al. 2016). Together, these considerations suggest that plastic phenotypic changes in new environments represent emergent responses that may be important for organismal survival, but are otherwise not steppingstones for genetic adaptations to the new environments. The similar observation in functional random metabolic networks suggests that our conclusion is likely to be general to most functional systems regardless of the specific evolutionary histories of the systems.

It is important to stress that our study focuses exclusively on adaptations to new environments that the organisms have not experienced at least in the recent past. For those environments that have been experienced by the organisms in the recent past, it is possible that mutations conferring plastic phenotypic changes that are beneficial in these environments have been fixed and there is no controversy that adaptive plasticity can evolve under this scenario.

6.5 MATERIALS AND METHODS

6.5.1 Gene expression analysis

Transcriptome datasets from six experimental adaptations were acquired from five studies. For each replicate of each adaptation, the data included gene expression levels of ancestral organisms in the original environment (stage *o*), ancestral organisms in the new environment (stage *p*), and evolved organisms in the new environment (stage *a*). For each dataset, we removed genes with any missing expression levels and then normalized gene expression levels such that the mean expression level of all genes is the same across all datasets.

The first dataset came from the experimental evolution of *E. coli* K-12 MG1655 in a 42°C medium with ten replicates (Sandberg et al. 2014). The authors performed RNA sequencing (RNA-seq) using (i) the ancestral line at 37°C, (ii) ancestral line at 42°C, and (iii) ten

parallelly evolved lines at 42°C, and these data were respectively used to estimate the L_o, L_p, and L_a of 4341 genes. All expression levels measured in FPKM were available in their Dataset S3.

The second dataset came from the experimental evolution of *E. coli* B REL1206 in a 42°C medium (Rodriguez-Verdugo et al. 2016). The authors performed RNA-seq using (i) the ancestral line at 37°C, (ii) ancestral line at 42°C, and (iii) two evolved lines at 42°C, and (iv) four lines each carrying a distinct adaptive mutation at 42°C. We respectively used (i) to estimate L_o, (ii) to estimate L_p, and both (iii) and (iv) to estimate L_a of 4202 genes. All expression levels measured by DESeq were provided by the authors.

The third and fourth datasets came from the experimental evolution of *E. coli* K-12 MG1655 in glycerol and lactate medium, respectively (Fong et al. 2005). The authors used Affymetrix *E. coli* Antisense Genome Arrays to profile the transcriptome of (i) the ancestral line in glucose, (ii) ancestral line in glycerol, (iii) ancestral line in lactate, (iv) seven parallelly evolved lines in glycerol on day 21, (v) seven parallelly evolved lines in glycerol on day 44, (vi) seven parallelly evolved lines in lactate on day 20, and (vii) seven parallelly evolved lines in lactate on day 60. Each line has three replicates, except that the profile (iii) has six replicates. We averaged gene expression levels across replicates for each line. For the adaptation to the glycerol medium, we respectively used (i) to estimate L_o, (ii) to estimate L_p, and (v) to estimate L_a. For the adaptation to the lactate medium, we respectively used (i) to estimate L_o, (iii) to estimate L_p, and (vii) to estimate L_a. Transcriptomes of (ii)-(vii) were downloaded from GEO with the accession number GSE33147, whereas that of (i) was provided by the authors. In total, 3745 genes were considered.

The fifth dataset came from the experimental evolution of 12 different strains of yeast (*Saccharomyces cerevisiae*) in an xylulose medium (Tamari et al. 2016). The authors performed

RNA-seq using (i) 12 ancestral lines in a glucose medium, (ii) 12 ancestral lines in the xylulose medium, and (iii) 12 evolved lines in the xylulose medium. Each line has two replicates, and the averaged expression levels of the two replicates were used. We respectively used (i) to estimate L_o, (ii) to estimate L_p, and (iii) to estimate L_a of 2235 genes. All expression levels in terms of UMI scoring normalized counts were downloaded from GEO with the accession number GSE76077.

The sixth dataset came from the experimental evolution of Trinidadian guppies (*Poecilia reticulata*) originating from streams with *h*igh numbers of cichlid *p*redators (HP environment) in cichlid-free streams (LP environment) (Ghalambor et al. 2015). The authors performed RNA-seq of brain tissues from (i) guppies caught in HP, (ii) guppies caught in HP but reared in LP, and (iii) two populations of guppies in LP after experimental evolution. We respectively used (i) to estimate L_o, (ii) to estimate L_p, and (iii) to estimate L_a of 37,493 genes. All expression levels in terms of TMM-normalized counts measured by edgeR were provided by the authors.

6.5.2 Metabolic network analysis

The SMBL file of the *E. coli* metabolic network model *i*AF1260 (Feist et al. 2007) was downloaded from BiGG(Schellenberger et al. 2010) and parsed by COBRA(Schellenberger et al. 2011). All linear and quadratic programming problems in this study were solved by the barrier method using Gurobi optimizer with MATLAB (method = 2). Numerical differences smaller than 10^{-4} were ignored in the analysis. The codes are available upon request.

We used FBA to estimate the fluxes of the *E. coli* network when it is fully adapted to an environment. FBA assumes a metabolic steady state and maximizes the rate of biomass

production (Orth et al. 2010). Mathematically, FBA is a linear programming question in the following form

$$\text{maximize } \boldsymbol{c}^{\mathrm{T}}\boldsymbol{v}, \text{ subject to } \boldsymbol{S}\boldsymbol{v} = \boldsymbol{0} \text{ and } \boldsymbol{\alpha} \leq \boldsymbol{v} \leq \boldsymbol{\beta},$$

where $\boldsymbol{v}$ is a vector of reaction fluxes that need to be optimized, $\boldsymbol{c}^{\mathrm{T}}$ is a transposed vector describing the relative contributions of various metabolites to the cellular biomass, $\boldsymbol{S}$ is a matrix describing the stoichiometric relationships among metabolites in each reaction, $\boldsymbol{\alpha}$ is a vector describing the lower bound of each flux, and $\boldsymbol{\beta}$ is a vector describing the upper bound of each flux.

The model *i*AF1260 includes 258 exchange reactions each of which allows the uptake of one carbon source. In the estimation of the fully adapted flux distribution in one environment, the uptake rate of the focal carbon source was set at 10 mmol$g\mathrm{DW}^{-1}\mathrm{h}^{-1}$, which follows the setting in a previous study for a glucose-limited medium (Feist et al. 2007), while the uptake rates of other carbon sources were set at zero. The uptake rates of non-carbon sources such as water, oxygen, carbon dioxide, and ammonium were set as in the previous study (Feist et al. 2007). Note that some reactions are simple diffusions between different cellular compartments. Because these reactions do not have dedicated enzymes and are not "mutable", we excluded them from the list of phenotypic traits considered. In total, 1811 reactions were considered.

We used minimization of metabolic adjustment (MOMA) to simulate plastic flux changes when *E. coli* is shifted from one environment to another (Segre et al. 2002). The mathematical form of MOMA is

$$\text{minimize } (\boldsymbol{v} - \boldsymbol{v}_0)^2, \text{ subject to } \boldsymbol{S}\boldsymbol{v} = \boldsymbol{0} \text{ and } \boldsymbol{\alpha} \leq \boldsymbol{v} \leq \boldsymbol{\beta},$$

where $\boldsymbol{v}$ is the vector of all reaction fluxes upon the environmental shift and is the variable to optimize, $\boldsymbol{v}_0$ is the vector of all reaction fluxes in the original environment and are predetermined using FBA. $\boldsymbol{S}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$ are the same as described for FBA.

To ensure that our results are not artifacts of different optimization accuracies of FBA and MOMA, we designed MOMA-b and used it to predict the fluxes in organisms adapted to new environments. In addition to having the same objective function and constraints as in MOMA, MOMA-b has a biomass constraint. Specifically, we set the biomass production rate in MOMA-b to be the same as what FBA predicts for organisms adapted to the new environment. The mathematical form of this new optimization question is

$$\text{minimize } (\boldsymbol{v} - \boldsymbol{v}_0)^2, \text{ subject to } \boldsymbol{S}\boldsymbol{v} = \boldsymbol{0}, \boldsymbol{\alpha} \leq \boldsymbol{v} \leq \boldsymbol{\beta}, \text{ and } \boldsymbol{c}^{\mathrm{T}}\boldsymbol{v} = b,$$

where the variables $\boldsymbol{v}$ and parameters $\boldsymbol{v}_0$, $\boldsymbol{S}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$ are the same as described for MOMA, and b is the FBA-predicted biomass production rate in the new environment. This optimization problem is still a quadratic programming problem and its solution can differ from that of FBA.

In addition to using single-carbon source environments, we followed a previous study (Wang and Zhang 2009b) to generate 100 environments with multiple carbon sources. In each environment, we generated a random number g from an exponential distribution with a mean of 0.1 for each of the 258 carbon sources. Here, g is the probability that the carbon source is available. The actual presence or absence of the carbon source is then determined stochastically using g. These random environments have a mean of 28 and a median of 21 carbon sources per environment.

6.6 ACKNOWLEDGEMENTS

We thank Chuan Li, Wenfeng Qian, Xinzhu Wei, Jian-Rong Yang, and Zhengting Zou for valuable comments. This work was supported in part by the U.S. National Institutes of Health research grant GM103232 to J.Z.

6.7 REFERENCES

Baldwin JM. 1896. A new factor in evolution. *American Naturalist* **30**: 441-451.

Barve A, Rodrigues JF, Wagner A. 2012. Superessential reactions in metabolic networks. *Proc Natl Acad Sci U S A* **109**: E1121-1130.

Barve A, Wagner A. 2013. A latent capacity for evolutionary innovation through exaptation in metabolic systems. *Nature* **500**: 203-+.

Bordbar A, Monk JM, King ZA, Palsson BO. 2014. Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet* **15**: 107-120.

Carroll SM, Marx CJ. 2013. Evolution after introduction of a novel metabolic pathway consistently leads to restoration of wild-type physiology. *PLoS Genet* **9**: e1003427.

Costenoble R, Picotti P, Reiter L, Stallmach R, Heinemann M, Sauer U, Aebersold R. 2011. Comprehensive quantitative analysis of central carbon and amino-acid metabolism in *Saccharomyces cerevisiae* under multiple conditions by targeted proteomics. *Mol Syst Biol* **7**: 464.

Edwards JS, Ibarra RU, Palsson BO. 2001. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. *Nat Biotechnol* **19**: 125-130.

Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO. 2007. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology* **3**.

Fong SS, Joyce AR, Palsson BO. 2005. Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome Res* **15**: 1365-1372.

Fong SS, Palsson BO. 2004. Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nat Genet* **36**: 1056-1058.

Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. 2000. Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* **11**: 4241-4257.

Ghalambor CK, Hoke KL, Ruell EW, Fischer EK, Reznick DN, Hughes KA. 2015. Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature* **525**: 372-375.

Ghalambor CK, McKay JK, Carroll SP, Reznick DN. 2007. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Funct Ecol* **21**: 394-407.

Grether GF. 2005. Environmental change, phenotypic plasticity, and genetic compensation. *Am Nat* **166**: E115-123.

Harcombe WR, Delaney NF, Leiby N, Klitgord N, Marx CJ. 2013. The ability of flux balance analysis to predict evolution of central metabolism scales with the initial distance to the optimum. *PLoS Comput Biol* **9**: e1003091.

He XL, Qian WF, Wang Z, Li Y, Zhang JZ. 2010. Prevalent positive epistasis in *Escherichia coli* and *Saccharomyces cerevisiae* metabolic networks. *Nat Genet* **42**: 272-U120.

Ho W-C, Zhang J. 2016. Adaptive genetic robustness of *Escherichia coli* metabolic fluxes. *Mol Biol Evol* **33**: 1164-1176.

Ibarra RU, Edwards JS, Palsson BO. 2002. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* **420**: 186-189.

Jozefczuk S, Klie S, Catchpole G, Szymanski J, Cuadros-Inostroza A, Steinhauser D, Selbig J, Willmitzer L. 2010. Metabolomic and transcriptomic stress response of *Escherichia coli*. *Mol Syst Biol* **6**: 364.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **44**: D457-D462.

Laland K, Uller T, Feldman M, Sterelny K, Muller GB, Moczek A, Jablonka E, Odling-Smee J. 2014. Does evolutionary theory need a rethink? - POINT Yes, urgently. *Nature* **514**: 161-164.

Laland KN, Uller T, Feldman MW, Sterelny K, Muller GB, Moczek A, Jablonka E, Odling-Smee J. 2015. The extended evolutionary synthesis: its structure, assumptions and predictions. *Proc Biol Sci* **282**: 20151019.

Lande R. 2009. Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. *J Evolution Biol* **22**: 1435-1446.

Ledon-Rettig CC, Pfennig DW, Nascone-Yoder N. 2008. Ancestral variation and the potential for genetic accommodation in larval amphibians: implications for the evolution of novel feeding strategies. *Evol Dev* **10**: 316-325.

Levis NA, Pfennig DW. 2016. Evaluating 'Plasticity-First' Evolution in Nature: Key Criteria and Empirical Approaches. *Trends Ecol Evol* **31**: 563-574.

Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, Adkins JN, Schramm G, Purvine SO, Lopez-Ferrer D et al. 2010. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol Syst Biol* **6**: 390.

Lopez-Maury L, Marguerat S, Bahler J. 2008. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat Rev Genet* **9**: 583-593.

Moczek AP, Sultan S, Foster S, Ledon-Rettig C, Dworkin I, Nijhout HF, Abouheif E, Pfennig DW. 2011. The role of developmental plasticity in evolutionary innovation. *Proc Biol Sci* **278**: 2705-2713.

Orth JD, Thiele I, Palsson BO. 2010. What is flux balance analysis? *Nature Biotechnology* **28**: 245-248.

Pal C, Papp B, Lercher MJ, Csermely P, Oliver SG, Hurst LD. 2006. Chance and necessity in the evolution of minimal metabolic networks. *Nature* **440**: 667-670.

Papp B, Pal C, Hurst LD. 2004. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* **429**: 661-664.

Pfennig DW, Wund MA, Snell-Rood EC, Cruickshank T, Schlichting CD, Moczek AP. 2010. Phenotypic plasticity's impacts on diversification and speciation. *Trends Ecol Evol* **25**: 459-467.

Pigliucci M, Murren CJ, Schlichting CD. 2006. Phenotypic plasticity and evolution by genetic assimilation. *J Exp Biol* **209**: 2362-2367.

Price TD, Qvarnstrom A, Irwin DE. 2003. The role of phenotypic plasticity in driving genetic evolution. *Proc Biol Sci* **270**: 1433-1440.

Robinson BW, Dukas R. 1999. The influence of phenotypic modifications on evolution: the Baldwin effect and modern perspectives. *Oikos* **85**: 582-589.

Rodriguez-Verdugo A, Tenaillon O, Gaut BS. 2016. First-Step mutations during adaptation restore the expression of hundreds of genes. *Mol Biol Evol* **33**: 25-39.

Sandberg TE, Pedersen M, LaCroix RA, Ebrahim A, Bonde M, Herrgard MJ, Palsson BO, Sommer M, Feist AM. 2014. Evolution of *Escherichia coli* to 42 degrees C and subsequent genetic engineering reveals adaptive mechanisms and novel mutations. *Mol Biol Evol* **31**: 2647-2662.

Schellenberger J, Park JO, Conrad TM, Palsson BO. 2010. BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics* **11**: 213.

Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S et al. 2011. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature protocols* **6**: 1290-1307.

Segre D, Deluna A, Church GM, Kishony R. 2005. Modular epistasis in yeast metabolism. *Nat Genet* **37**: 77-83.

Segre D, Vitkup D, Church GM. 2002. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 15112-15117.

Suzuki Y, Nijhout HF. 2006. Evolution of a polyphenism by genetic accommodation. *Science* **311**: 650-652.

Tamari Z, Yona AH, Pilpel Y, Barkai N. 2016. Rapid evolutionary adaptation to growth on an 'unfamiliar' carbon source. *BMC Genomics* **17**: 674.

Waddington CH. 1953. Genetic assimilation of an acquired character. *Evolution* **7**: 118-126.

Wang Z, Zhang J. 2009a. Abundant indispensable redundancies in cellular metabolic networks. *Genome Biol Evol* **1**: 23-33.

Wang Z, Zhang J. 2011. Impact of gene expression noise on organismal fitness and the efficacy of natural selection. *Proc Natl Acad Sci U S A* **108**: E67-76.

Wang Z, Zhang JZ. 2009b. Why Is the Correlation between Gene Importance and Gene Evolutionary Rate So Weak? *Plos Genetics* **5**.

West-Eberhard MJ. 2003. *Developmental Plasticity and Evolution*. Oxford Univ. Press, New York, USA.

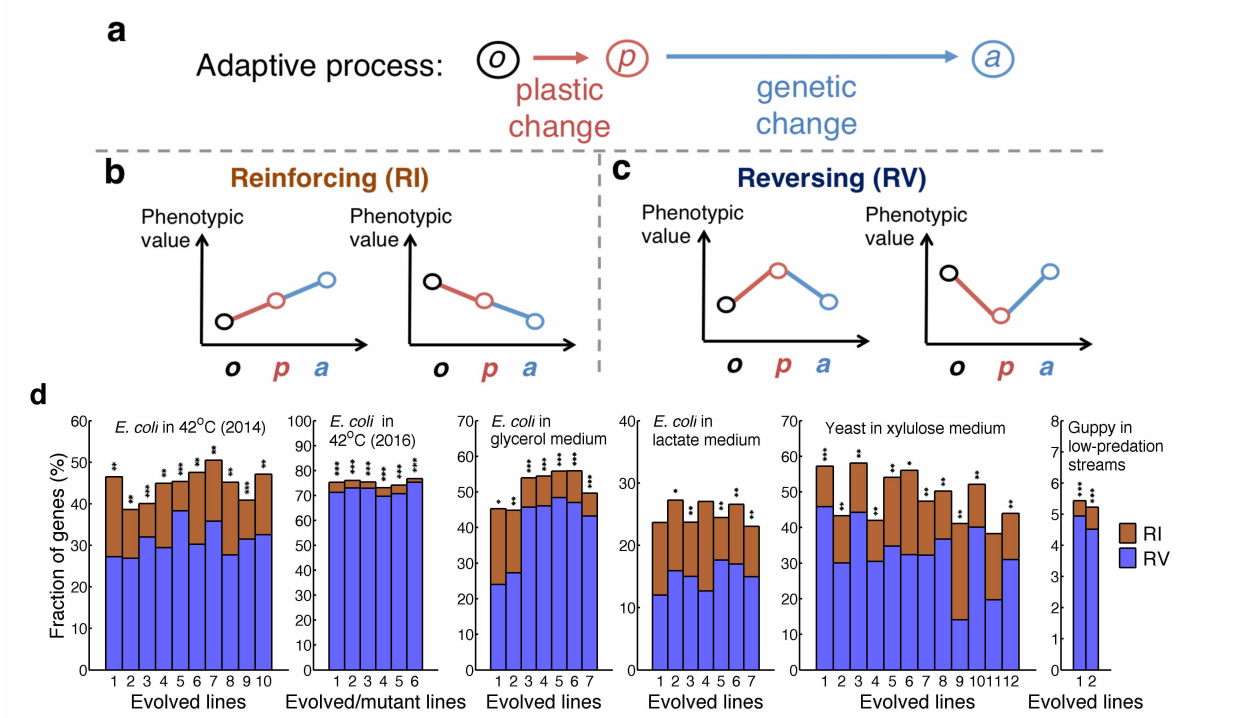

Figure 6.1 Genetic adaptations in experimental evolution more frequently reverse than reinforce plastic gene expression changes. (**a**) Phenotypic adaptation is studied by comparing the phenotypic values of a trait at three stages: ancestral organisms adapted to the original environment measured in the original environment (stage *o*), ancestral organisms measured in the new environment (stage *p*), and evolved organisms adapted to the new environment measured in the new environment (stage *a*). Plastic changes refer to changes from stage *o* to *p*, while genetic changes refer to changes from stage *p* to *a*. (**b**) A pair of plastic and genetic phenotypic changes of a trait are said to be reinforcing if both are larger than a preset cutoff and are in the same direction. (**c**) A pair of plastic and genetic phenotypic changes of a trait are said to be reversing if both are larger than a preset cutoff but are in opposite directions. (**d**) Fractions of genes with reinforcing and reversing expression changes, respectively, in experimental evolution. Organisms as well as the new environments to which the organisms were adapting to are indicated. Each bar represents an adaptation. The equality in the fraction of reinforcing and reversing genes in each adaptation is tested by a two-tailed chi-squared test, with *P*-values indicated as follows: *, $P < 0.05$; **, $P < 10^{-10}$; ***, $P < 10^{-100}$.

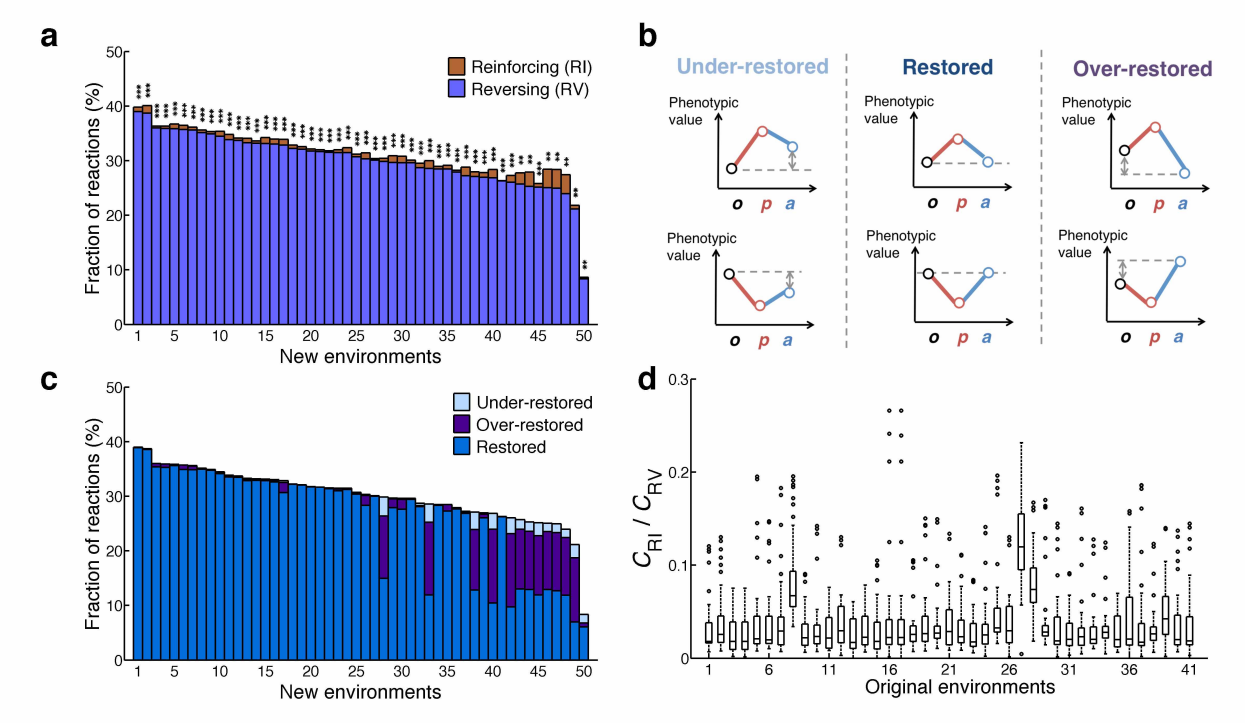

Figure 6.2 Predominance of flux reversion in the environmental adaptations of *E. coli*. **(a)** Fractions of reactions with reinforcing and reversing flux changes, respectively, in the adaptation from the glucose environment to each of 50 new environments. Each bar represents the adaptation to a new environment. The equality in the fraction of reinforcing and reversing reactions is tested by a two-tailed chi-squared test, with *P*-values indicated as follows: *, $P < 0.05$; **, $P < 10^{-10}$; ***, $P < 10^{-100}$. (**b**) Classification of reversion to three categories based on whether the phenotypic value in the original environment is under-restored, restored, or over-restored. (**c**) Fractions of the three categories of reversion in each of the 50 adaptations. (**d**) Fraction of reinforcing reactions relative to fracion of reversing reactions (C_{RI}/C_{RV}) in *E. coli* adaptations to at least 20 new environments from each of 41 original environments examined. The C_{RI}/C_{RV} ratios for all adaptations from each original environment are presented in a box plot, where the lower and upper edges of a box represent the first (qu_1) and third quartiles (qu_3), respectively, the horizontal line inside the box indicates the median (md), the whiskers extend to the most extreme values inside inner fences, $md \pm 1.5(qu_3-qu_1)$, and the circles represent values outside the inner fences (outliers).

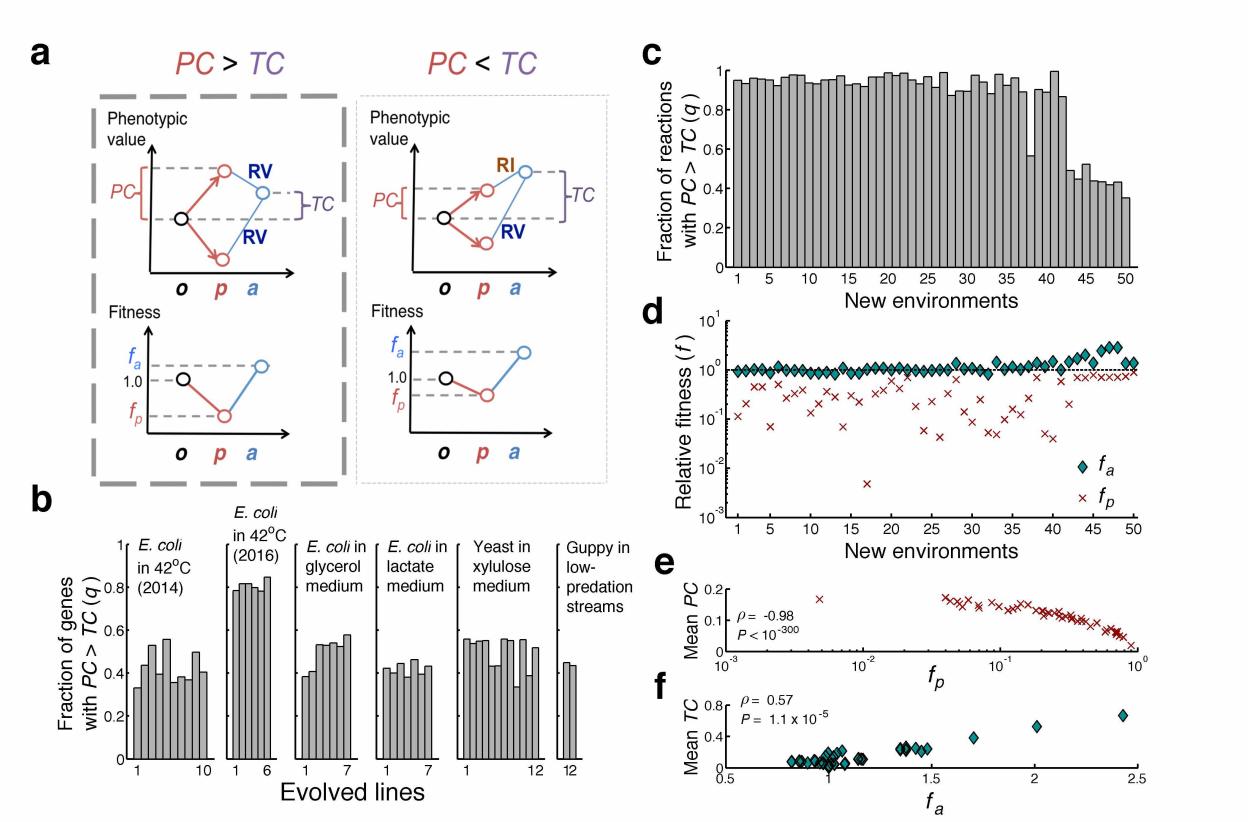

Figure 6.3 Cause of the preponderance of phenotypic reversion in adaptation. **(a)** Diagram illustrating the model. The left box shows that if the plastic change (*PC*) is bigger than the total change (*TC*), the genetic change must reverse the plastic change. One reason for *PC* > *TC* is that the fitness difference between organisms at stages *o* and *p* is larger than that between stages *o* and *a*. The right box shows that if *PC* < *TC*, the genetic change may reinforce or reverse the plastic change. This may occur if the fitness difference between organisms at stages *o* and *p* is smaller than that between *o* and *a* or if the phenotype is unassociated with fitness. (**b**) Fraction of genes showing expression *PC* > *TC* during each of 44 experimental evolutionary adaptations. **(c)** Fraction of reactions showing flux *PC* > *TC* during each of the *E. coli* metabolic adaptations from the glucose environment to the 50 new environments. (**d**) Fitness at stage *p* and that at stage *a*, relative to that at stage *o*, predicted by metabolic network analysis, for each of the 50 adaptations in panel c. The dotted line shows the fitness at stage *o*. **(e)** Mean *PC* across all fluxes negatively correlates with the relative fitness at stage *p* (f_p) among the 50 adaptations in panel c. **(f)** Mean *TC* across all fluxes positively correlates with the relative fitness at stage *a* (f_a) among the 50 adaptations in panel c.

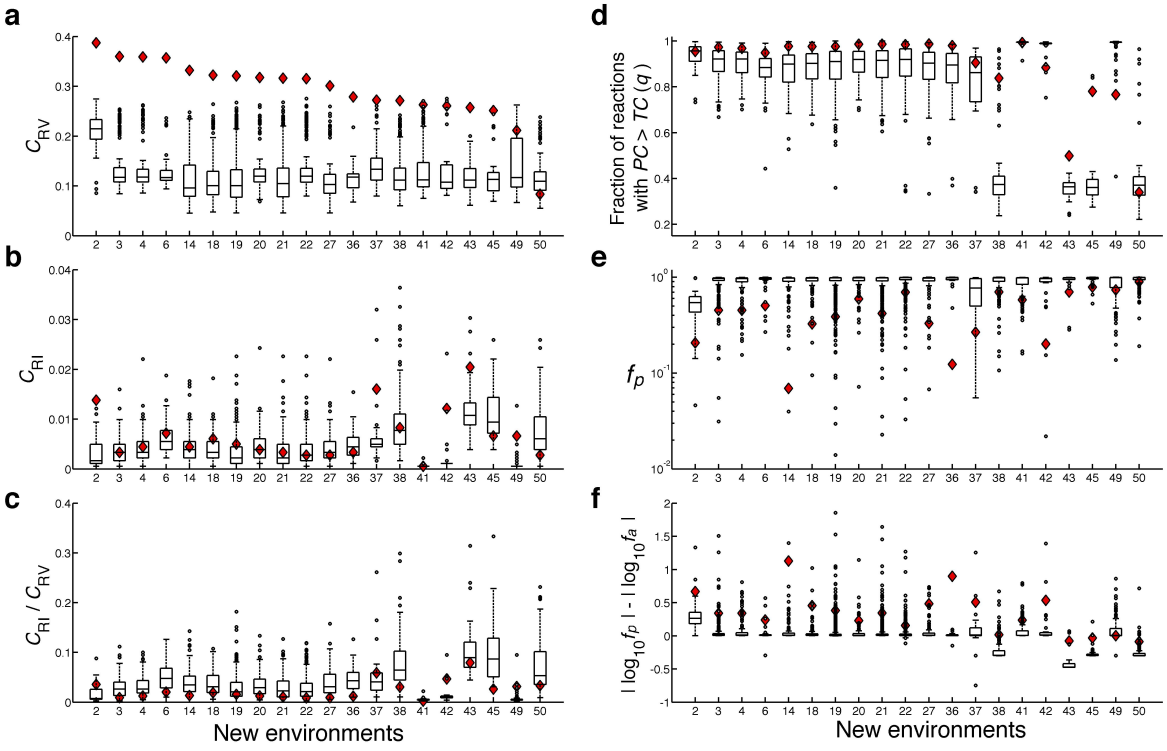

Figure 6.4 Predominance of flux reversion in random metabolic networks. Fractions of reactions showing flux reversion (C_{RV}) (**a**), fractions of reactions showing flux reinforcement (C_{RI}) (**b**), C_{RI}/C_{RV} ratios (**c**), fraction of reactions showing $PC > TC$ (**d**), relative fitness at stage p (f_p) (**e**), and $|\log_{10}f_p| - |\log_{10}f_a|$ (**f**) in the adaptations of random networks from the glucose environment to each of the 20 new environments examined. For each new environment, values estimated from different random networks are shown by a box plot, with symbols explained in the legend to Figure 6.2. The corresponding values for the *E. coli i*AF1260 network are shown by red diamonds.

Chapter 7

Conclusions

Before discussing the limitations and future directions, I will first summarize the findings presented in each chapter of the dissertation.

In chapter 2, I gave a statistical summary of parameters in genotype-phenotype maps. Specifically, I found the averaged proportion of genes affecting a trait is 6% - 9% with a range from <1% to >30%. In addition, consistent with Robertson's model, I found that most traits are affected by many small-effect genes as well as few large-effect genes.

In the following part of chapter 2 and the entire chapter 3, using yeast morphological traits, yeast gene expression traits, and *E. coli* metabolic flux traits, I found the evidence supporting the adaptive hypothesis of genetic robustness as well as refuting intrinsic hypothesis and congruent hypothesis. The main evidence is the observation that more important traits tend to be genetically more buffered, and this trend remains even after the control of the ability to buffer environmental variations. Therefore, genotype-phenotype maps evolve as the effect sizes are reduced by adaptive genetic robustness.

In chapter 4, in yeast morphological traits, I found that more important traits tend to evolve faster relative to their mutational sizes. Therefore, the phenotypic evolution of these traits is generally adaptive. However, I did not find such trend when using in yeast expression level traits, suggesting they are generally under neutral evolution.

In chapter 5, I found that pleiotropy-caused mutational correlations predict the correlation of evolutionary distances using yeast morphological traits. Moreover, high mutational correlation constrains the phenotypic distances while small amount of mutational correlation facilitates it.

Finally, in chapter 6, I found that the plastic responses tend to be on the contrary direction compared with the next step of genetic change in adaptation using expression level traits from multiple species and metabolic flux traits from *E. coli*. Therefore phenotypic plasticity does not generally serve as a steppingstone for adaptation.

Below I will point out some limitations and suggest some future directions which potentially deal with these limitations.

Traits choices in genotype-phenotype maps

Because different kinds of traits have different properties such as their importance to fitness, my findings may not be generally applicable to other kinds of traits. However, some of my findings are consistent across different classes of traits such as the adaptive hypothesis of genetic robustness and the role of phenotypic plasticity on adaptation. At the same time, not all of my findings are the same when different kinds of traits are focused. For example, when testing the adaptive hypothesis of phenotypic evolution, yeast morphological and yeast gene expression traits show opposite results. Given that the applicability is not general, it is of interests to explain when and why some results are applicable. For example, the discrepant results of testing phenotypic evolution are consistent with the view that different classes of traits have different important levels to fitness. In particular, morphological traits should be more important to fitness than expression level traits should be. Many kinds of trait have not been tested in my works

such as protein abundances, various ways of expression regulation, metabolite concentrations, and so on. Potentially, these classes of traits should show the importance level somewhere between transcript abundances and morphology. With the help of related omics data, the generality of my findings could be further tested. More importantly, when different classes of traits show different results, it will be interesting to examine the association between the results and their importance to fitness.

Focused organisms in genotype-phenotype maps

In my dissertation, due to data availability and quality, the microorganisms such as *E. coli* and yeast are tended to be studied. Because they have larger effective population size, their evolution is expected to be more affected by natural selection. In the future, it will be interesting to examine whether my findings are still true in other species. In particular, when nature selection is involved as an evolutionary force of the finding, it is expected to see the signal strength of that finding decreases in the species with a smaller effective population size. On the contrary, when natural selection is not involved, it is expected to see no association between the signal strength and the effect population size. Such approach will not only further test the generality of my findings but also synthesize a more universal theory in explaining and predicting evolution consequences.

Epistasis in genotype-phenotype maps

In the gene-trait bipartite expression of genotype-phenotype maps, epistasis, defined as the nonadditivity between mutational effects, has been ignored. Accordantly, epistasis is not considered in my presented works here. However, in other expressions of genotype-phenotype maps such as fitness landscapes, epistasis has been the central issue

because it could largely limit the number of possible adaptive trajectories and make the evolution more predictable (Phillips 2008; de Visser and Krug 2014). Several attempts have been made in systematically quantify epistasis, including forward genetic methods (Mackay 2014), reverse genetic methods (Tong et al. 2001; Li et al. 2016; Puchta et al. 2016), and model prediction methods (He et al. 2010). It will be interesting to study the maps of epistasis effects, their evolution and their impacts on evolution. The potential questions include whether epistasis is more likely to be positive or negative, whether epistasis is buffered by genetic robustness, whether epistasis evolves adaptively, and whether epistasis-caused genetic correlation impacts the rate of phenotypic evolution.

Using genotype-phenotype maps as a paradigm in studying evolution

Evolutionary outcomes result from both mutations and selections. Therefore, comparing evolutionary outcomes with mutational inputs could infer the historical action of selection. Such approach is widely used in molecular evolution. For example, in order to test positive or negative selection after speciation, one could compare nonsynonymous and synonymous sequence substitutions controlled by the mutation inputs form population data, which is known as McDonald–Kreitman test (Mcdonald and Kreitman 1991). In phenotypic evolution, however, such approach is rare due to the lacking of known mutational inputs. With the availability of genotype-phenotype maps, it is expected to see more incorporation of such approach in studying phenotypic evolution (e.g. Metzger et al. 2015).

In addition, while genotype-phenotype maps are claimed to be useful in predicting evolutionary outcomes, the verification of the prediction power is generally lack. Such verification can be done in the context of experimental evolution. For example, given

organisms with a certain genetic background, one can construct its genotype-phenotype map in a certain environment. After performing experimental evolution in that new environment, by comparing the spectrum of mutational inputs and evolutionary outcome, one can study how much mutational inputs predict evolutionary outcomes. For example, one can ask whether any kind of mutational property, such as effect size or pleiotropy, determines the chance of mutation to reach fixation in evolution. Note that, in most published studies of experimental evolution, the repeatability is studied by parallelism (Lenski and Travisano 1995; Cooper et al. 2003; Kryazhimskiy et al. 2014; Venkataram et al. 2016). However, without knowing the mutational inputs, it is unclear that the repeatability comes from selection or biased mutational inputs. In addition, it is interesting to ask whether any emergent property, such as robustness or modularity, determines the evolutionary repeatability. By understanding the variability of repeatability, the predictability in evolutionary biology will be largely improved.

Finally, given that the properties of genotype-phenotype maps could have consequences in evolution, it will be also interesting to improve the prediction of evolution of those properties. However, due to the demand of mapping highly complex traits or higher-order interactions, this topic is expected to be more difficult to study. In the future, if the methodological improvement makes constructing genotype-phenotype maps much easier, answering these difficult high-dimensional questions on genotype-phenotype maps should shed more lights in finding general principles of evolution.

References

Cooper TF, Rozen DE, Lenski RE. 2003. Parallel changes in gene expression after 20,000 generations of evolution in Escherichiacoli. *Proc Natl Acad Sci U S A* **100**: 1072-1077.

de Visser JAGM, Krug J. 2014. Empirical fitness landscapes and the predictability of evolution. *Nat Rev Genet* **15**: 480-490.

He XL, Qian WF, Wang Z, Li Y, Zhang JZ. 2010. Prevalent positive epistasis in Escherichia coli and Saccharomyces cerevisiae metabolic networks. *Nat Genet* **42**: 272-U120.

Kryazhimskiy S, Rice DP, Jerison ER, Desai MM. 2014. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* **344**: 1519-1522.

Lenski RE, Travisano M. 1995. Dynamics of adaptation and diversification: A 10,000 generation experiment with bacterial populations. *Tempo and Mode in Evolution*: 253-273.

Li C, Qian WF, Maclean CJ, Zhang JZ. 2016. The fitness landscape of a tRNA gene. *Science* **352**: 837-840.

Mackay TFC. 2014. Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat Rev Genet* **15**: 22-33.

Mcdonald JH, Kreitman M. 1991. Adaptive Protein Evolution at the Adh Locus in Drosophila. *Nature* **351**: 652-654.

Metzger BPH, Yuan DC, Gruber JD, Duveau F, Wittkopp PJ. 2015. Selection on noise constrains variation in a eukaryotic promoter. *Nature* **521**: 344-+.

Phillips PC. 2008. Epistasis - the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* **9**: 855-867.

Puchta O, Cseke B, Czaja H, Tollervey D, Sanguinetti G, Kudla G. 2016. Network of epistatic interactions within a yeast snoRNA. *Science* **352**: 840-844.

Tong AHY, Evangelista M, Parsons AB, Xu H, Bader GD, Page N, Robinson M, Raghibizadeh S, Hogue CWV, Bussey H et al. 2001. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**: 2364-2368.

Venkataram S, Dunn B, Li Y, Agarwala A, Chang J, Ebel ER, Geiler-Samerotte K, Herissant L, Blundell JR, Levy SF et al. 2016. Development of a Comprehensive Genotype-to-Fitness Map of Adaptation-Driving Mutations in Yeast. *Cell* **166**: 1585-1596 e1522.

APPENDICES

A.1 Supplementary tables and figures for chapter 2

Table A.1.1 Rank correlation between trait importance (*TI*) and mean net effect size (|*ES*|) or phenotypic variation (*CV*) of the wild-type when only positive effects or negative effects are considered.

Effects considered	Variables correlated	Variables controlled	Spearman's ρ	p-value
positive	*TI*, *CV*		-0.712	<1e-300
positive	*TI*, mean net |*ES*|		-0.604	<1e-300
positive	*TI*, *CV*	mean net |*ES*|	-0.610	7.1e-22
positive	*TI*, mean net |*ES*|	*CV*	-0.434	1.2e-10
negative	*TI*, *CV*		-0.569	<1e-300
negative	*TI*, mean net |*ES*|		-0.485	<1e-300
negative	*TI*, *CV*	mean net |*ES*|	-0.392	7.9e-09
negative	*TI*, mean net |*ES*|	*CV*	-0.209	2.9e-03

Table A.1.2 Spearman's rank correlation between the importance of a gene expression trait to fitness and the mean effect size of gene deletion ($|ES_G|$) or environmental perturbation ($|ES_E|$) after expression levels are controlled.

Variables correlated	Variables controlled	Spearman's ρ	p-value
Fitness effect[1], $\|ES_E\|$	Expression level[3]	-0.196	2.2e-16
Fitness effect[1], $\|ES_G\|$	Expression level[3]	-0.185	2.6e-25
Fitness effect[1], $\|ES_E\|$	Expression level[3] and $\|ES_G\|$	-0.145	3.5e-16
Fitness effect[1], $\|ES_G\|$	Expression level[3] and $\|ES_E\|$	-0.129	4.5e-13
Essentiality[2], $\|ES_E\|$	Expression level[3]	-0.117	5.7e-11
Essentiality[2], $\|ES_G\|$	Expression level[3]	-0.127	1.3e-12
Essentiality[2], $\|ES_E\|$	Expression level[3] and $\|ES_G\|$	-0.080	7.7e-06
Essentiality[2], $\|ES_G\|$	Expression level[3] and $\|ES_E\|$	-0.094	1.6e-07

1. Fitness defect caused by deleting the gene.
2. Essentiality = 0 for nonessential traits and 1 for essential traits.
3. Expression level of the gene in the wild-type.

Table A.1.3 Spearman's rank correlation between the importance of a gene expression trait to fitness and the mean effect size of gene deletion ($|ES_G|$) or environmental perturbation ($|ES_E|$) when $|ES_E|$ were measured in all environmental changes.

Variables correlated	Variables controlled	Spearman's ρ	p-value
Fitness effect[1], $|ES_E|$		-0.191	6.6e-27
Fitness effect[1], $|ES_G|$		-0.180	3.7e-24
Fitness effect[1], $|ES_E|$	$|ES_G|$	-0.138	9.4e-15
Fitness effect[1], $|ES_G|$	$|ES_E|$	-0.123	5.5e-12
Essentiality[2], $|ES_E|$		-0.117	6.4e-11
Essentiality[2], $|ES_G|$		-0.125	2.2e-12
Essentiality[2], $|ES_E|$	$|ES_G|$	-0.078	1.3e-05
Essentiality[2], $|ES_G|$	$|ES_E|$	-0.091	4.1e-07
Fitness effect[1], $|ES_E|$	Expression level[3]	-0.234	5.8e-40
Fitness effect[1], $|ES_G|$	Expression level[3]	-0.185	2.6e-25
Fitness effect[1], $|ES_E|$	Expression level[3] and $|ES_G|$	-0.183	7.8e-25
Fitness effect[1], $|ES_G|$	Expression level[3] and $|ES_E|$	-0.111	4.5e-10
Essentiality[2], $|ES_E|$	Expression level[3]	-0.141	2.4e-15
Essentiality[2], $|ES_G|$	Expression level[3]	-0.127	1.3e-12
Essentiality[2], $|ES_E|$	Expression level[3] and $|ES_G|$	-0.104	6.9e-09
Essentiality[2], $|ES_G|$	Expression level[3] and $|ES_E|$	-0.083	4.0e-06

1. Fitness defect caused by deleting the gene.
2. Essentiality = 0 for nonessential traits and 1 for essential traits.
3. Expression level of the gene in the wild-type.

Table A.1.4 Spearman's rank correlation between the importance of a gene expression trait to fitness and the mean effect size of gene deletion ($|ES_G|$) or environmental perturbation ($|ES_E|$) after the removal of highly correlated expressional traits.

Variables correlated	Variables controlled	Spearman's ρ	p-value
Fitness effect[1], $\|ES_E\|$		-0.156	6.2e-10
Fitness effect[1], $\|ES_G\|$		-0.181	5.1e-13
Fitness effect[1], $\|ES_E\|$	$\|ES_G\|$	-0.099	8.9e-05
Fitness effect[1], $\|ES_G\|$	$\|ES_E\|$	-0.136	6.3e-08
Essentiality[2], $\|ES_E\|$		-0.104	4.0e-05
Essentiality[2], $\|ES_G\|$		-0.131	2.1e-07
Essentiality[2], $\|ES_E\|$	$\|ES_G\|$	-0.062	1.5e-02
Essentiality[2], $\|ES_G\|$	$\|ES_E\|$	-0.101	6.5e-05
Fitness effect[1], $\|ES_E\|$	Expression level[3]	-0.196	4.7e-15
Fitness effect[1], $\|ES_G\|$	Expression level[3]	-0.185	1.7e-13
Fitness effect[1], $\|ES_E\|$	Expression level[3] and $\|ES_G\|$	-0.141	2.1e-08
Fitness effect[1], $\|ES_G\|$	Expression level[3] and $\|ES_E\|$	-0.124	8.2e-07
Essentiality[2], $\|ES_E\|$	Expression level[3]	-0.126	5.3e-07
Essentiality[2], $\|ES_G\|$	Expression level[3]	-0.132	1.7e-07
Essentiality[2], $\|ES_E\|$	Expression level[3] and $\|ES_G\|$	-0.085	7.7e-04
Essentiality[2], $\|ES_G\|$	Expression level[3] and $\|ES_E\|$	-0.093	2.3e-04

1. Fitness defect caused by deleting the gene.
2. Essentiality = 0 for nonessential traits and 1 for essential traits.
3. Expression level of the gene in the wild-type.

Table A.1.5 Spearman's rank correlation between the importance of a gene expression trait to fitness and the mean effect size of gene deletion ($|ES_G|$) or environmental perturbation ($|ES_E|$) for genes with or without canonical TATA boxes, respectively.

Variables correlated	Variables controlled	TATA (n=563)		Non-TATA (n=2726)	
		Spearman's ρ	p-value	Spearman's ρ	p-value
Fitness effect[1], $\|ES_E\|$		-0.185	2.4E-04	-0.178	6.7E-21
Fitness effect[1], $\|ES_G\|$		-0.267	8.5E-08	-0.155	4.0E-16
Fitness effect[1], $\|ES_E\|$	$\|ES_G\|$	**-0.062**	**2.2E-01**	-0.139	3.4E-13
Fitness effect[1], $\|ES_G\|$	$\|ES_E\|$	-0.206	4.4E-05	-0.107	2.3E-08
Essentiality[2], $\|ES_E\|$		-0.168	8.4E-04	-0.105	3.7E-08
Essentiality[2], $\|ES_G\|$		-0.201	6.2E-05	-0.109	1.3E-08
Essentiality[2], $\|ES_E\|$	$\|ES_G\|$	**-0.081**	**1.1E-01**	-0.076	7.4E-05
Essentiality[2], $\|ES_G\|$	$\|ES_E\|$	-0.138	6.6E-03	-0.080	2.6E-05
Fitness effect[1], $\|ES_E\|$	Expression level[3]	-0.232	3.6E-06	-0.214	1.2E-29
Fitness effect[1], $\|ES_G\|$	Expression level[3]	-0.294	3.4E-09	-0.147	1.5E-14
Fitness effect[1], $\|ES_E\|$	Expression level[3] and $\|ES_G\|$	-0.112	2.8E-02	-0.179	5.6E-21
Fitness effect[1], $\|ES_G\|$	Expression level[3] and $\|ES_E\|$	-0.215	1.9E-05	-0.085	9.3E-06
Essentiality[2], $\|ES_E\|$	Expression level[3]	-0.205	4.6E-05	-0.124	7.9E-11
Essentiality[2], $\|ES_G\|$	Expression level[3]	-0.221	1.1E-05	-0.102	9.0E-08
Essentiality[2], $\|ES_E\|$	Expression level[3] and $\|ES_G\|$	-0.118	2.0E-02	-0.097	3.8E-07
Essentiality[2], $\|ES_G\|$	Expression level[3] and $\|ES_E\|$	-0.144	4.5E-03	-0.067	4.9E-04

1. Fitness defect caused by deleting the gene.
2. Essentiality = 0 for nonessential traits and 1 for essential traits.
3. Expression level of the gene in the wild-type.
4. **bold** numbers: p>0.05.

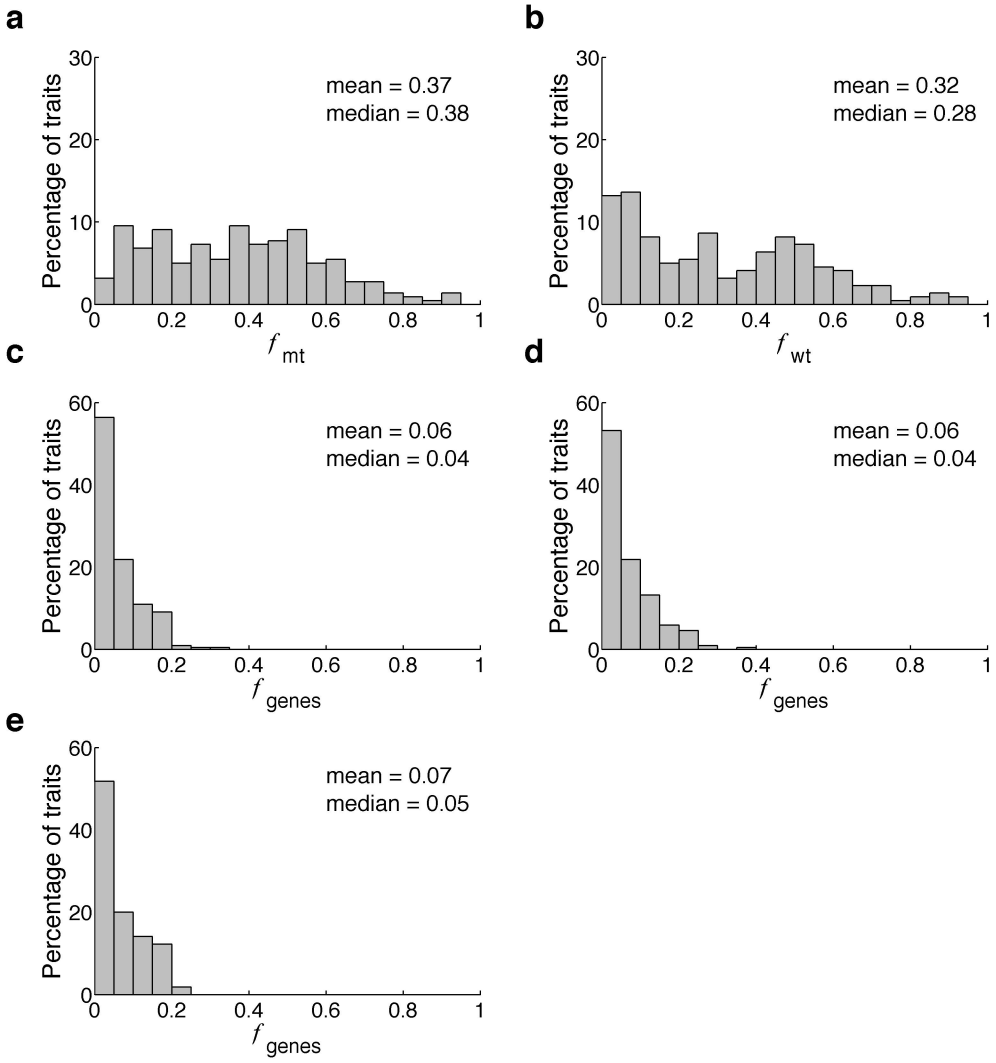

Figure A.1.1 Fraction of genes affecting a trait under different cutoffs, with or without corrections for environmental variation and multiple testing. (**a**) Fraction of genes (f_{mt}) affecting a trait under $p < 0.05$, without corrections. (**b**) Fraction (f_{wt}) of wild-type replicate populations that show significant deviations from an arbitrarily selected wild-type population. The corrected fraction of genes affecting a trait (f_{genes}) is shown in Fig. 2.1c. (**c**) Corrected fraction of genes (f_{genes}) affecting a trait under $p < 0.01$. (**d**) Corrected fraction of genes (f_{genes}) affecting a trait under $p < 0.1$. (**e**) Corrected fraction of genes (f_{genes}) affecting a trait estimated by examining whether the mean trait value of a mutant would be an outlier in the distribution of the 123 means of the wild-type replicate populations. A deletion is annotated to affect a trait if the mean trait value of the mutant is in either the left or right 2.5% tail of the distribution of the 123 mean trait values of the 123 wild-type replicate populations. However, given that 4718 deletions are tested per trait, there are 236 expected deletions located in these two tails simply by chance. Therefore, the total number of genes affecting a trait is either corrected by subtracting 236 or set as zero if the original number is less than 236.

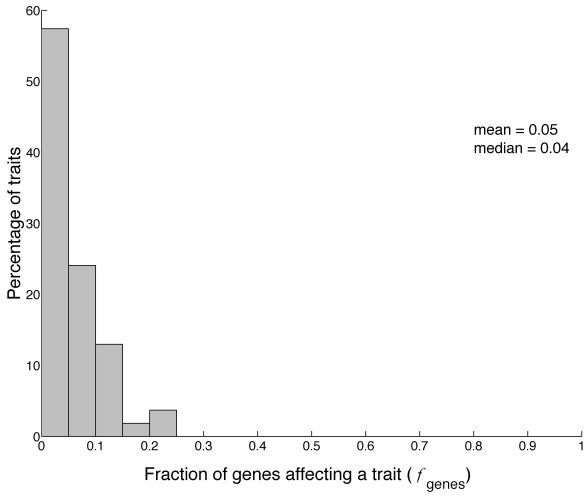

Figure A.1.2 Fraction of genes (f_{genes}) affecting a trait after the removal of highly correlated traits.

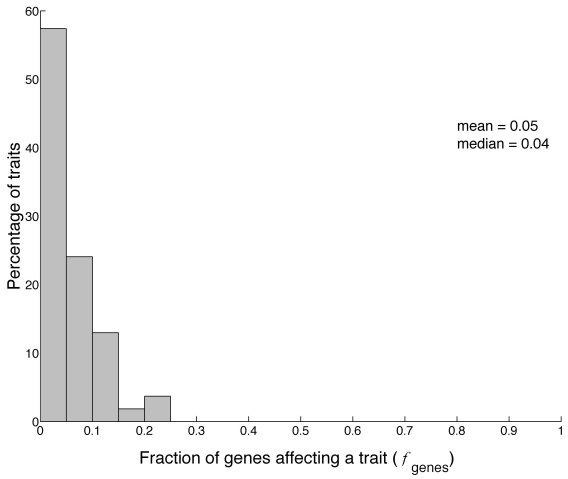

Figure A.1.3 Sources of phenotypic variations of wild-type cells. Standard deviation among population means generally exceeds the average standard error of individual replicate populations, indicating the existence of environmental variation among replicate populations. The dotted line shows the expectation when there is no environmental difference among replicate populations.

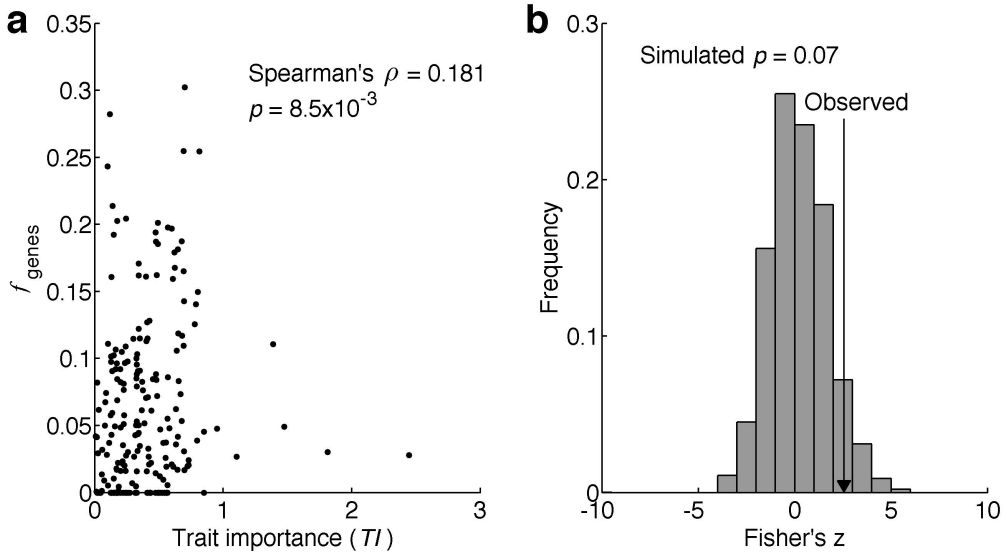

Figure A.1.4 No significant correlation between the corrected fraction of genes affecting a trait (f_{genes}) and trait importance (*TI*). (**a**) Weak positive correlation between the corrected fraction of genes affecting a trait (f_{genes}) and trait importance (*TI*). Each dot is a trait. (**b**) Distribution of Fisher's *z* derived from the rank correlation between the corrected fraction of genes affecting a trait (f_{genes}) and trait importance (*TI*). The real *z* observed from the actual data is indicated by an arrowhead and the *p*-value is the probability that a randomly picked pseudo *z* is more positive than the real *z*.

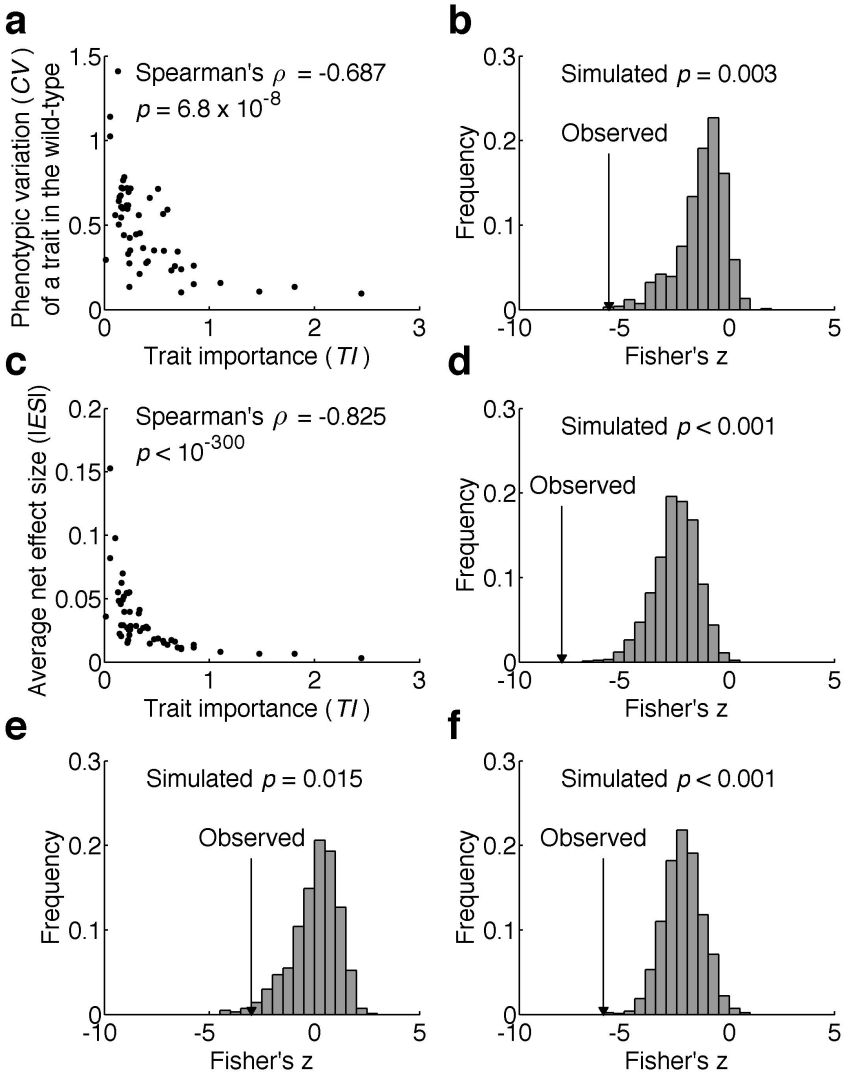

Figure A.1.5 Environmental/stochastic robustness and genetic robustness of yeast morphological traits using less correlated traits. (**a**) The phenotypic variation (*CV*) of a trait among isogenic wild-type cells decreases with the rise of trait importance (*TI*), demonstrating environmental/stochastic robustness. Each dot is a trait. (**b**) Distribution of Fisher's *z* derived from the rank correlation between *CV* and pseudo *TI*. (**c**) The mean net |*ES*| of gene deletion on a trait decreases with the rise of trait importance (*TI*), demonstrating genetic robustness. Each dot is a trait. (**d**) Distribution of Fisher's *z* derived from the rank correlation between mean net |*ES*| and pseudo *TI*. (**e**) Distribution of Fisher's *z* derived from the partial rank correlation between *CV* and pseudo *TI*, after the control of mean net |*ES*|. (**f**) Distribution of Fisher's *z* derived from the partial rank correlation between mean net |*ES*| and pseudo *TI*, after the control of *CV*. In (b), (d), (e), and (f), the real *z* observed from the actual data is indicated by an arrowhead and the *p*-value is the probability that a randomly picked pseudo *z* is more negative than the real *z*.

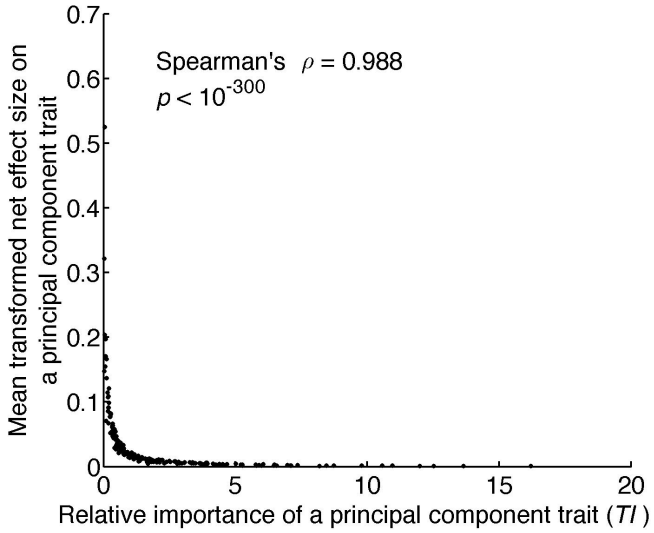

Figure A.1.6 Negative correlation between transformed mean net effect size (|*ES*|) and the importance of principal component traits. Each dot is a principal component trait.

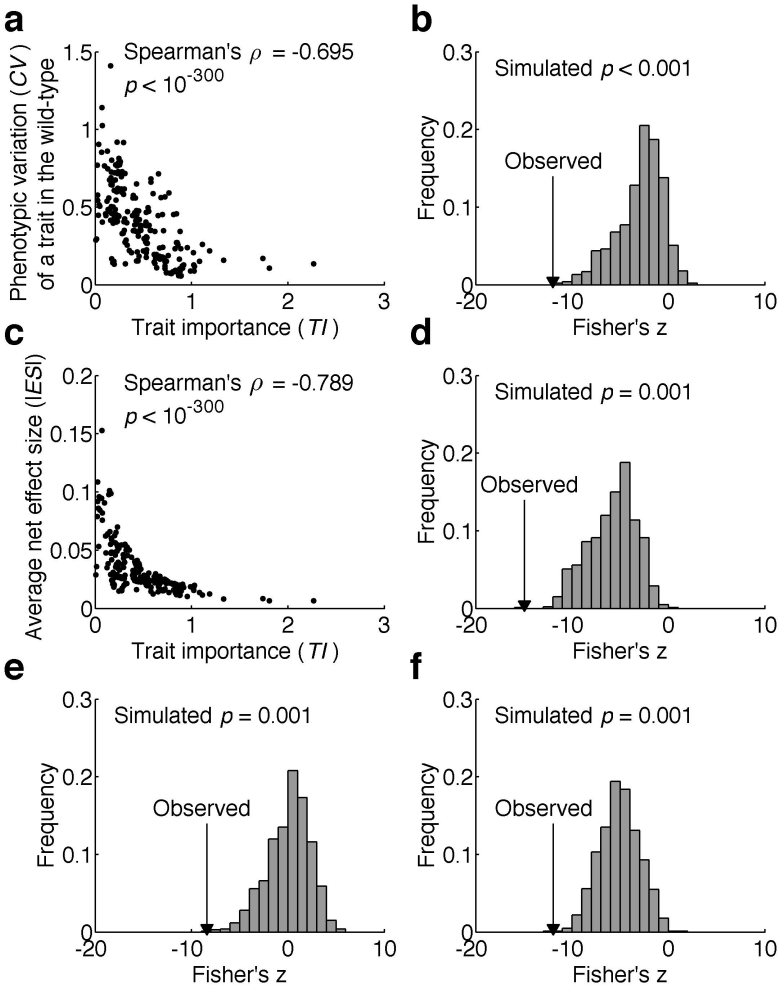

Figure A.1.7 Environmental/stochastic robustness and genetic robustness of yeast morphological traits when trait importance is estimated using a log model (see Materials and Methods). (**a**) The phenotypic variation (*CV*) of a trait among isogenic wild-type cells decreases with the rise of trait importance (*TI*), demonstrating environmental/stochastic robustness. Each dot is a trait. (**b**) Distribution of Fisher's *z* derived from the rank correlation between *CV* and pseudo *TI*. (**c**) The mean net |*ES*| of gene deletion on a trait decreases with the rise of trait importance (*TI*), demonstrating genetic robustness. Each dot is a trait. (**d**) Distribution of Fisher's *z* derived from the rank correlation between mean net |*ES*| and pseudo *TI*. (**e**) Distribution of Fisher's *z* derived from the partial rank correlation between *CV* and pseudo *TI*, after the control of mean net |*ES*|. (**f**) Distribution of Fisher's *z* derived from the partial rank correlation between mean net |*ES*| and pseudo *TI*, after the control of *CV*. In (b), (d), (e), and (f), the real *z* observed from the actual data is indicated by an arrowhead and the *p*-value is the probability that a randomly picked pseudo *z* is more negative than the real *z*.

A.2 Supplementary tables and figures for chapter 3

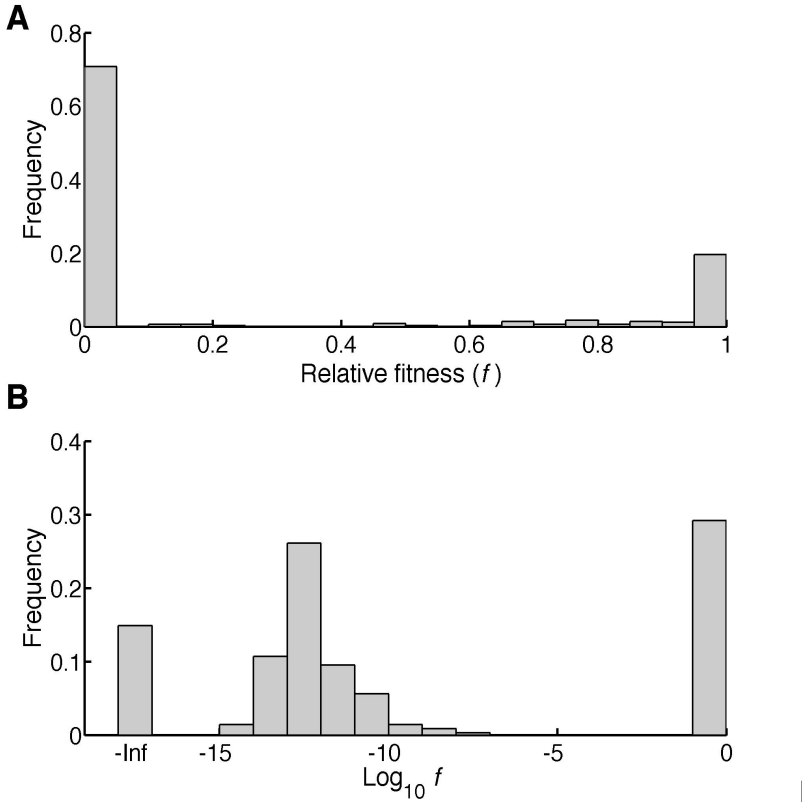

Figure A.2.1 Frequency distribution of FBA-predicted mutant fitness relative to the wild-type upon the removal of a reaction in *E. coli*. (A) The distribution shown with a linear scale on the X-axis. (B) The distribution shown with a log scale on the X-axis.

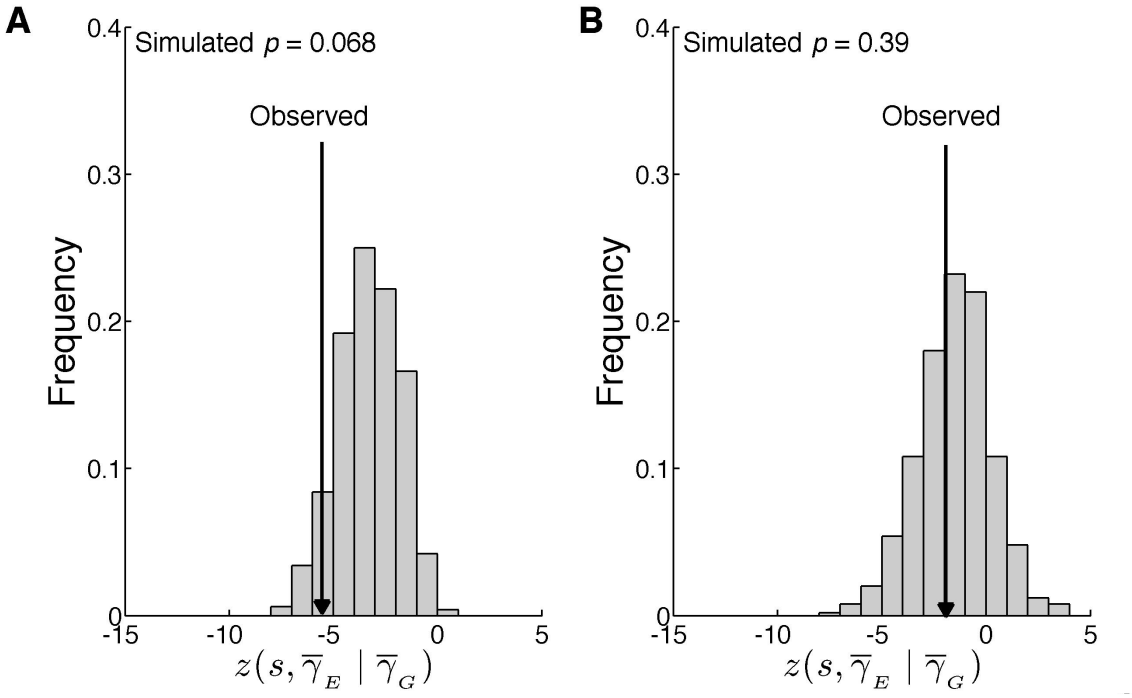

Figure A.2.2 Frequency distribution of the partial rank correlation (after conversion to *z*) between s and $\overline{\gamma}_E$ upon the control of $\overline{\gamma}_G$ among the 500 random networks. Arrow indicates the corresponding *z* observed in *E. coli* using 1000 random environments (A) or single-carbon-source environments (B).

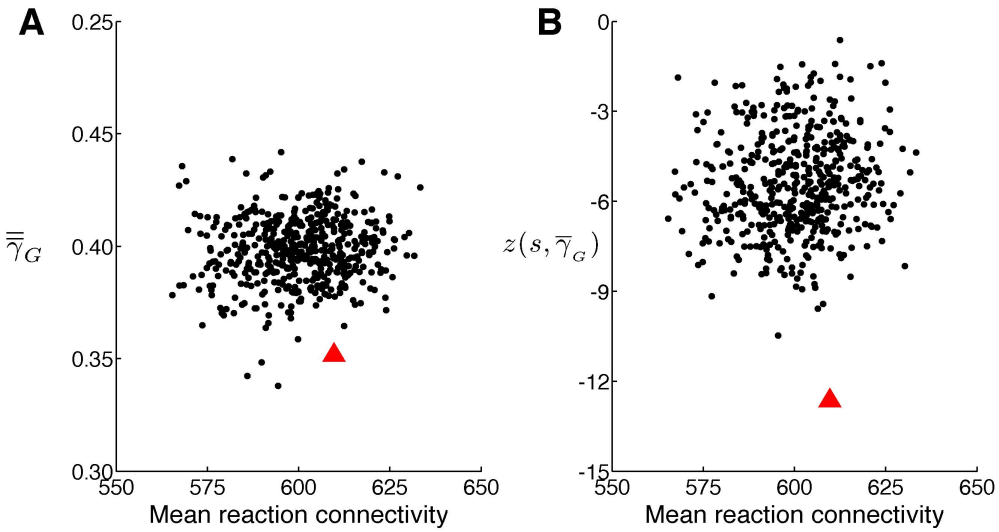

Figure A.2.3 Similar mean reaction connectivities between the *E. coli* network and the 500 random networks. The mean reaction connectivity of a network is the number of reactions directly connected to reaction averaged across all reactions in a network. (A) While the average fractional flux change across all focal reactions and mutants ($\bar{\bar{\gamma}}_G$, same as shown in Figure 3.2A) is lower in the *E. coli* network (red triangle) than in the random networks (black dots), the mean reaction connectivities are similar. (B) While the rank correlation (after conversion to *z,* same as shown in Figure 3.3B) between $\bar{\gamma}_G$ and *s* is more negative for the *E. coli* network (red triangle) than for the random networks (black dots), the mean reaction connectivities are similar.

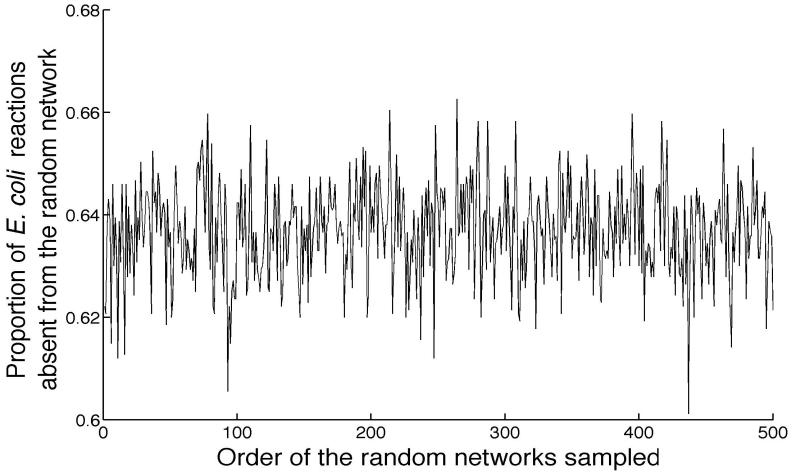

Figure A.2.4 Proportion of *E. coli* reactions that are absent from the 500 sampled random networks in the order of sampling.

A.3 Supplementary tables and figures for chapter 4

Table A.3.1 *P*-values of using partial mantel test for the predictors of morphological adaptation among wild strains in *Saccharomyces cerevisiae*

	Original traits	PCA traits
Genomic dissimilarity	0.067	0.743
Ecological environment dissimilarity	0.161	0.456
Population membership dissimilarity	0.395	0.725
Geographic difference	0.320	0.610

Note- All tests are one-tail test to the positive correlation.

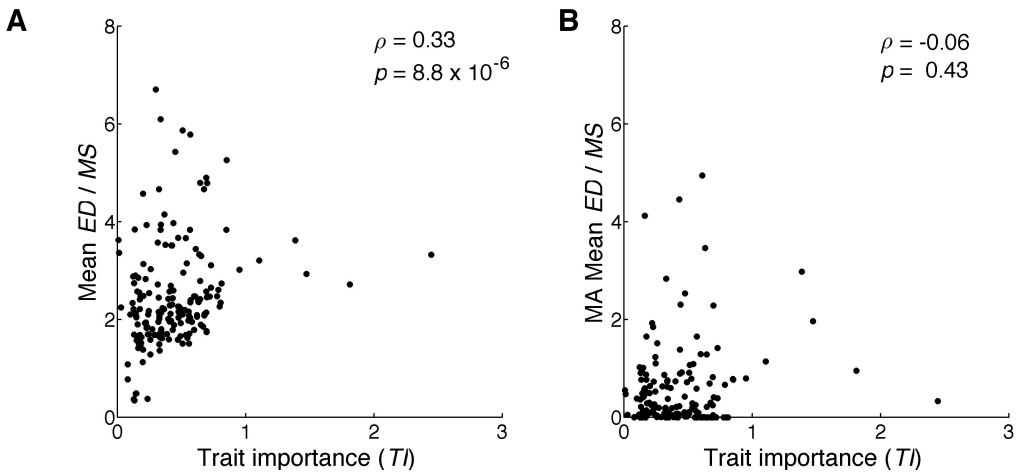

Figure A.3.1 Adaptive evolution of morphological traits among natural strains but not mutation accumulation (MA) lines of *S. cerevisiae*. (**A**) Mean evolution distance (*ED*) among 666 natural strain pairs for a trait relative to its mutation size (*MS*) increases significantly with trait importance (*TI*). The 180 traits for which data are available for both the natural strains and MA lines are used. Each dot represents a morphological trait. ρ, Spearman's rank correlation coefficient. (**B**) Mean *ED* between 89 MA lines and their common ancestor for a trait relative to *MS* is not significantly correlated with *TI*.

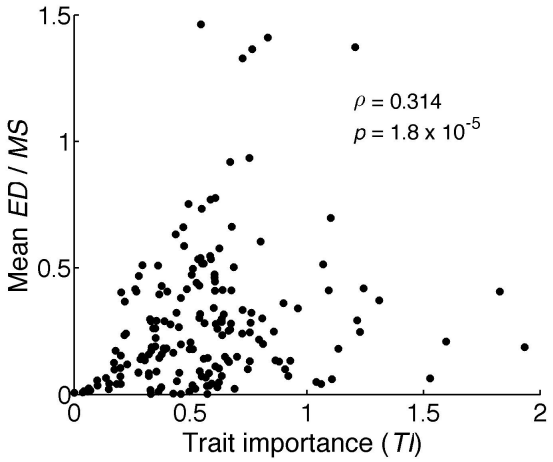

Figure A.3.2 Mean *ED* among 666 natural strain pairs of *S. cerevisiae* for a trait relative to its *MS* increases significantly with *TI*, when *MS* and *TI* are both estimated using diploid deletion strains.

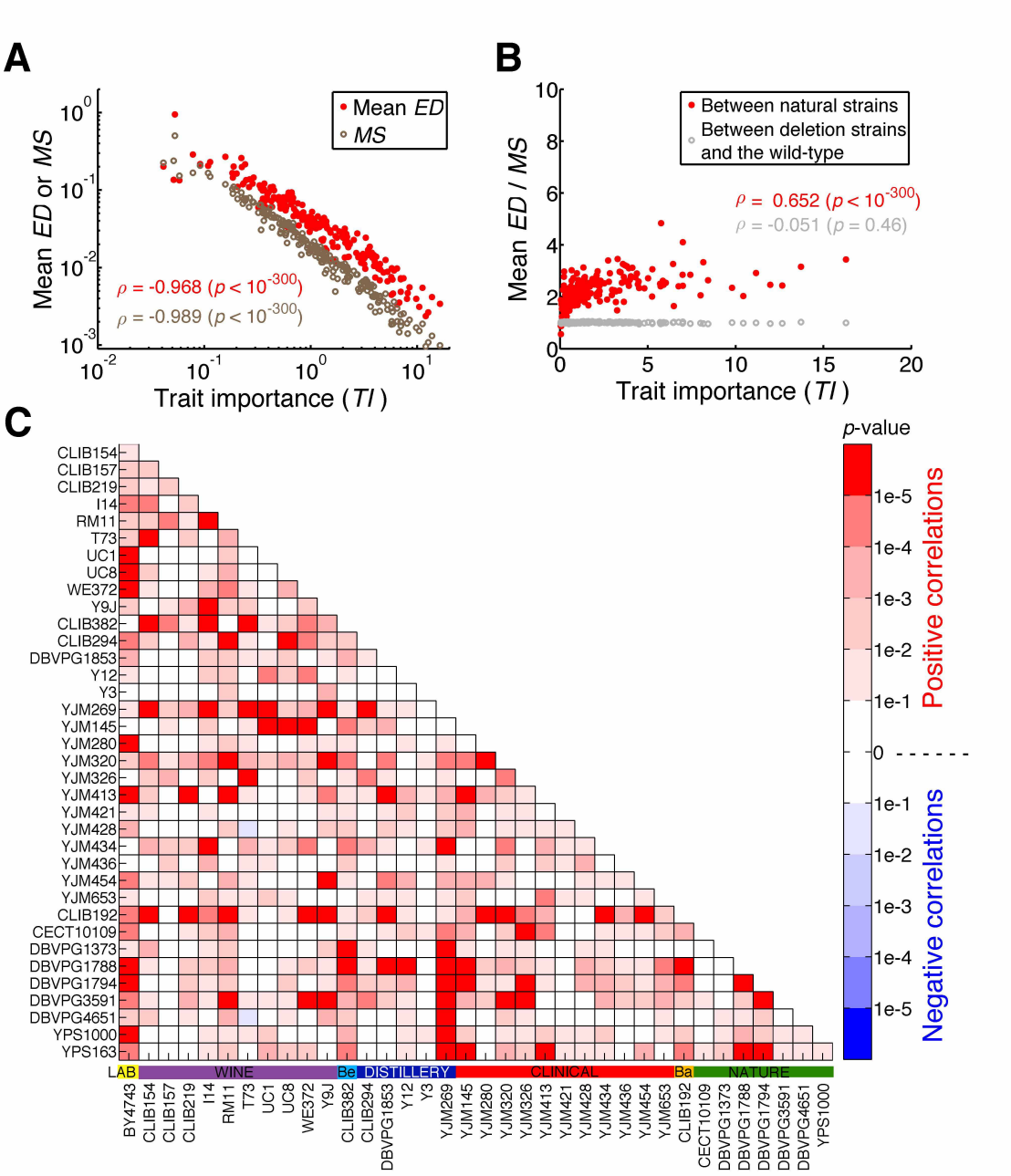

Figure A.3.3 Prevalent adaptive evolution of morphological principal component traits in the yeast *S. cerevisiae*. (**A**) Mean evolution distance (*ED*) of 666 pairs of natural strains for a trait and the mutational size (*MS*) of the trait both decrease with trait importance (*TI*). Each dot represents a morphological principal component trait. ρ, Spearman's rank correlation coefficient. (**B**) Mean *ED* among 666 natural strain pairs relative to *MS* increases significantly with *TI*, while the mean *ED* between 666 gene deletion strains and the wild-type relative to *MS* does not increase significantly with *TI*. (**C**) Nominal *p*-values for Spearman's correlation between *ED*/*MS* and *TI* for all 666 pairs of natural strains.

A.4 Supplementary tables and figures for chapter 5

Table A.4.1 Spearman's correlation coefficients (ρ) between mean mutational correlation (*MC*) and various mean evolutionary differences (*ED*) after the control of mutational size (*MS*) and direct importance (*DI*) in yeast morphological traits with negative coefficients in multivariate regression (n = 110) or positive coefficients in multivariate regression (n = 100)

Source of mean *ED*	Sign of coefficients	Spearman's ρ	p-value
intra-species	negative	0.11	0.28
	positive	-0.12	0.25
inter-species	negative	0.075	0.44
	positive	0.12	0.23

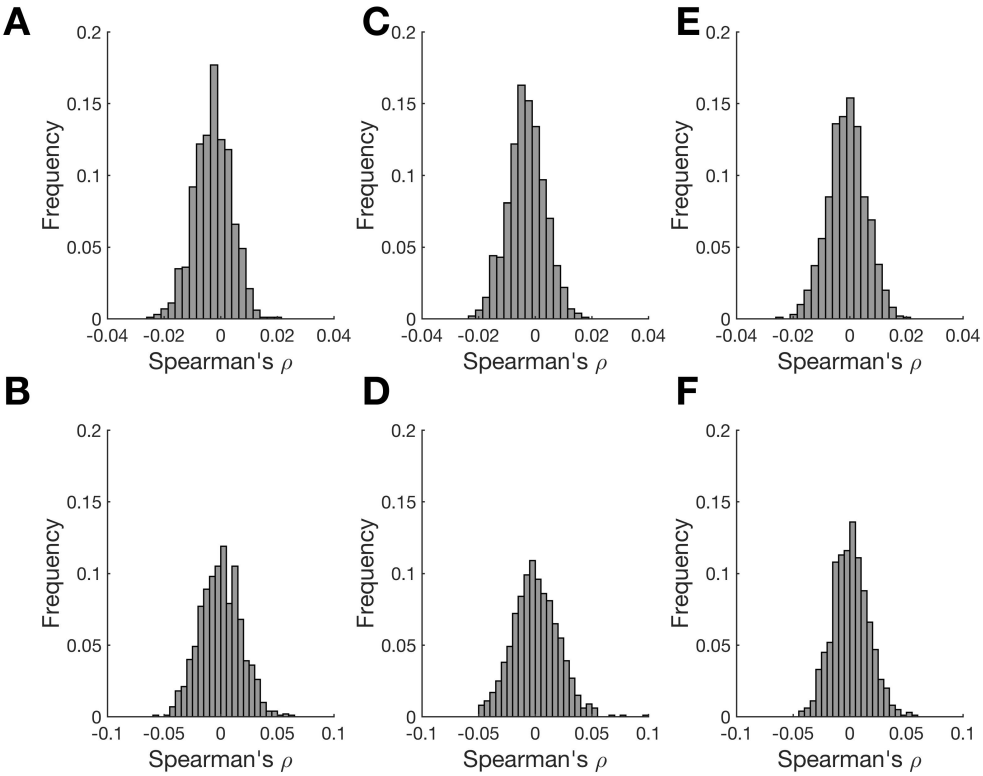

Figure A.4.1 Distributions of Spearman's *ρ* between mutational correlations (*MC*) and evolutionary correlations (*EC*) using shuffled matrix of effect sizes (*ES*). (A)-(B) Spearman's *ρ* between mutational correlations (*MC*) and evolutionary correlations (*EC*) using 37 *Saccharomyces cerevisiae* strains. **(C)-(D)** Spearman's *ρ* between *MC* and phenotypic correlations using 37 *S. cerevisiae* strains. **(E)-(F)** Spearman's *ρ* between *MC* and *EC* between *S. cerevisiae* strains and *S. paradoxus* In **(A)**, **(C)** and **(E)**, for each time of shuffling, the *ES* across different gene deletion lines per each trait was shuffled (randG). In **(B)**, **(D)**, and **(F)**, for each time of shuffling, the *ES* across different traits per each gene deletion line was shuffled (randP). Each distribution represents a result of 1000 times of shuffling.

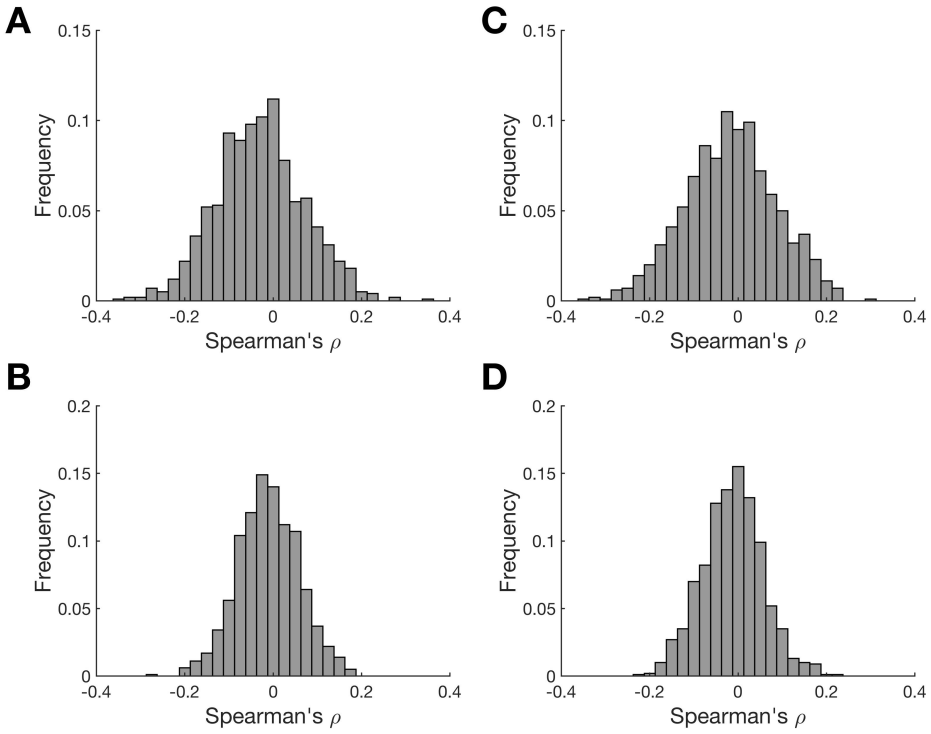

Figure A.4.2 Distributions of Spearman's *ρ* between mean mutational correlations (*MC*) and mean evolutionary differences (*ED*) using shuffled matrix of effect sizes (*ES*). (A)-(B) mean *ED* was calculated using 37 *Saccharomyces cerevisiae* strains. **(C)-(D)** mean *ED* was calculated between *S. cerevisiae* strains and *S. paradoxus* In **(A)** and **(C)**, for each time of shuffling, the *ES* across different gene deletion lines per each trait was shuffled (randG). In **(B)** and **(D)**, for each time of shuffling, the *ES* across different traits per each gene deletion line was shuffled (randP). Each distribution represents a result of 1000 times of shuffling.

A.5 Supplementary tables and figures for chapter 6

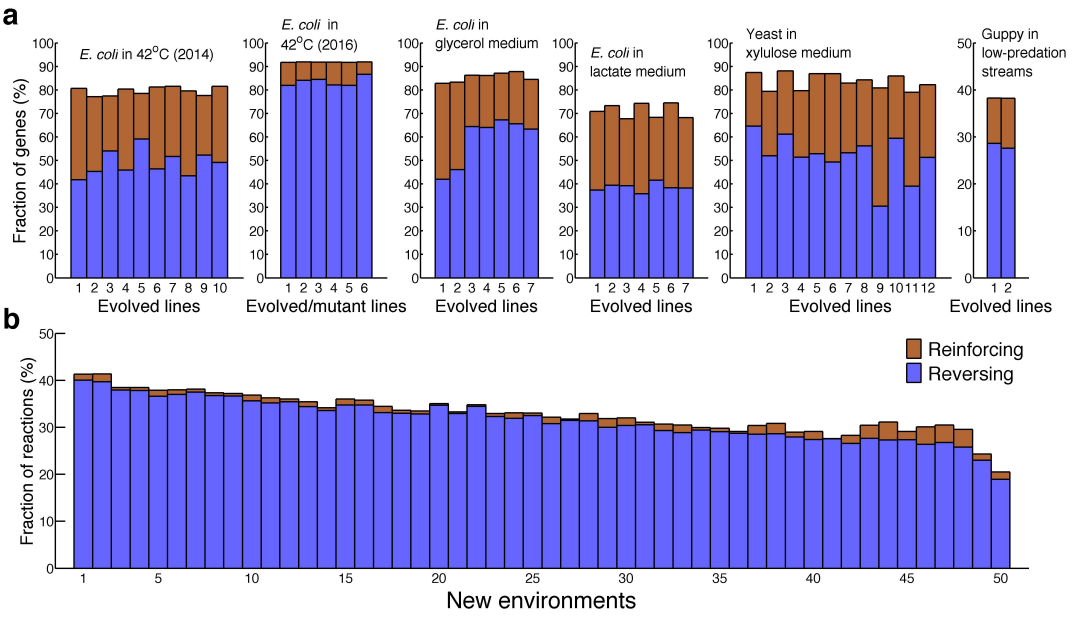

Figure A.5.1 Genetic adaptations more frequently reverse than reinforce plastic phenotypic changes. The cutoff of 0.05 Lo is used in defining plastic and genetic changes. (a) Fractions of genes with reinforcing and reversing expression changes, respectively, in experimental evolution. Organisms as well as the new environments to which the organisms were adapting to are indicated. Each bar represents an adaptation. (b) Fractions of reactions with predicted reinforcing and reversing flux changes, respectively, in *E. coli*'s adaptations to 50 new environments from the glucose environment.

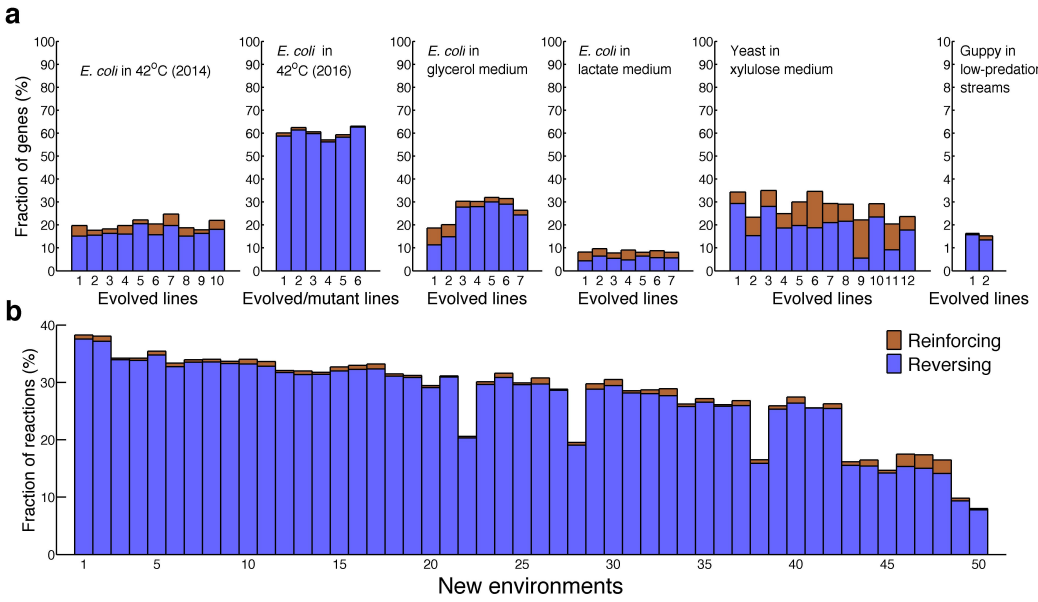

Figure A.5.2 Genetic adaptations more frequently reverse than reinforce plastic phenotypic changes. The cutoff of 0.5 Lo is used in defining plastic and genetic changes. (a) Fractions of genes with reinforcing and reversing expression changes, respectively, in experimental evolution. Organisms as well as the new environments to which the organisms were adapting to are indicated. Each bar represents an adaptation. (b) Fractions of reactions with predicted reinforcing and reversing flux changes, respectively, in *E. coli*'s adaptations to 50 new environments from the glucose environment.

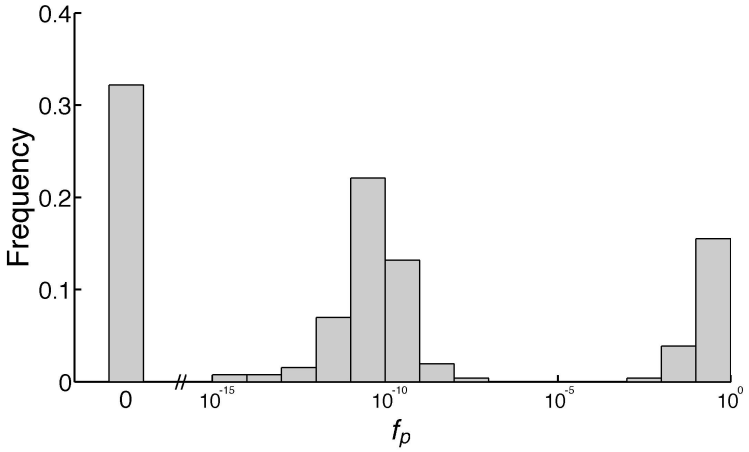

Figure A.5.3 Frequency distribution of the fitness of *E. coli i*AF1260 at the plastic stage in 257 new environments relative to that in the original glucose environment.

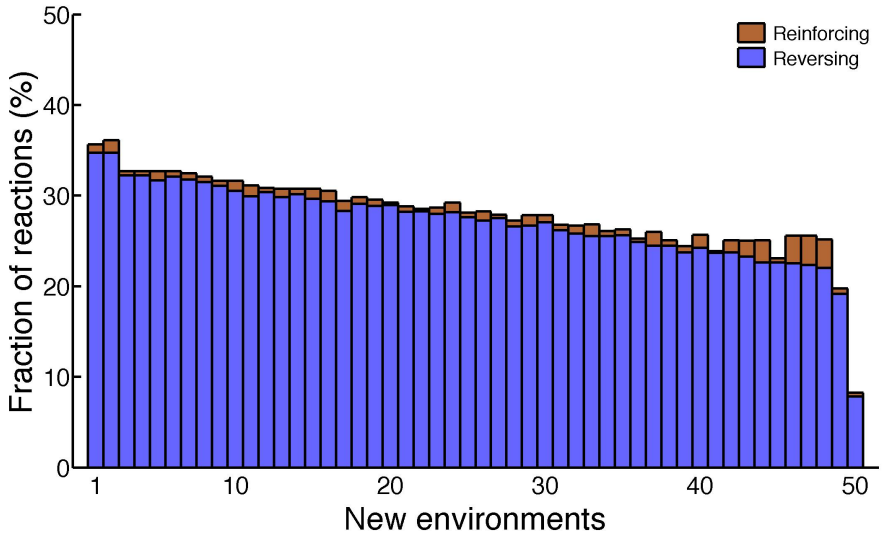

Figure A.5.4 Predominance of flux reversion in the adaptations of *E. coli i*AF1260 to 50 new environments from the glucose environments. The figure.differs slightly from Fig. 6.2a because of the use of a randomized order of reactions in the stoichiometric matrix.

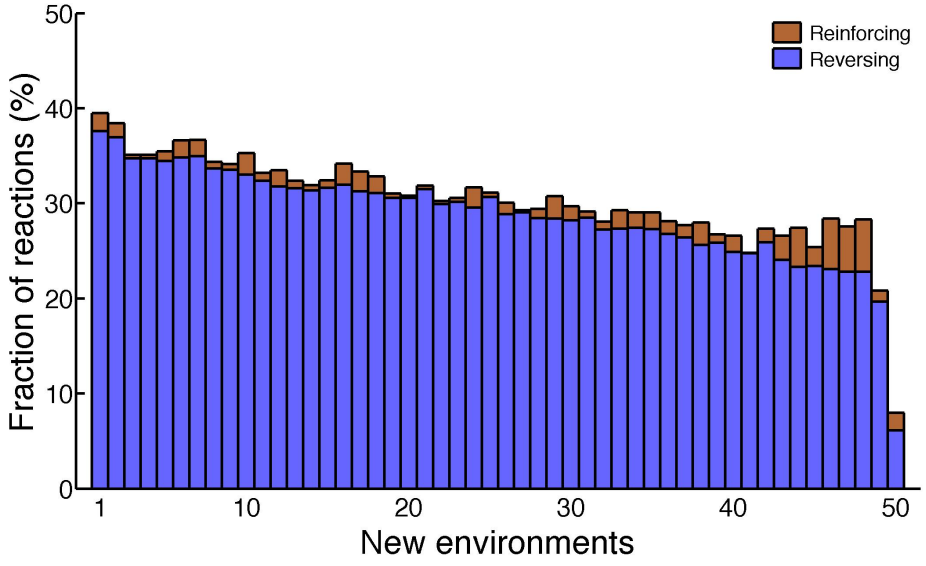

Figure A.5.5 Predominance of flux reversion in the adaptations of *E. coli i*AF1260 to 50 new environments from the glucose environments. The Figure differs slightly from Fig. 6.2a because of the use of MOMA-b to predict fluxes at stage a.

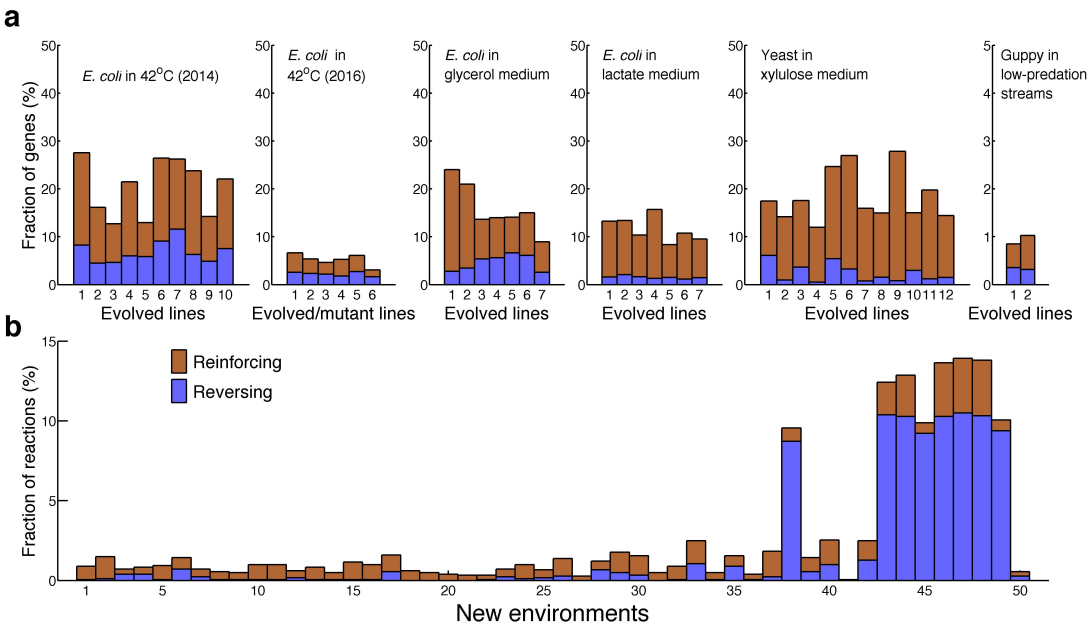

Figure A.5.6 The preponderance of phenotypic reversion disappears after the removal of traits for which the size of the plastic change exceeds that of the total change (PC> TC). (a) Fractions of genes with reinforcing and reversing expression changes, respectively, in experimental evolution. Organisms as well as the new environments to which the organisms were adapting to are indicated. Each bar represents an adaptation. (b) Fractions of reactions with predicted reinforcing and reversing flux changes, respectively, in *E. coli*'s adaptations to 50 new environments from the glucose environment.

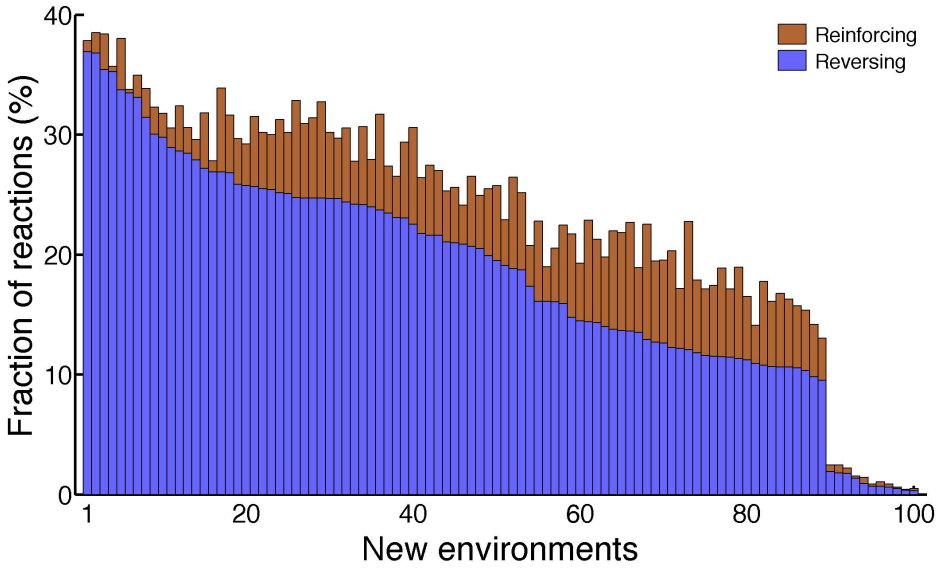

Figure A.5.7 Predominance of flux reversion in the adaptations of *E. coli i*AF1260 to 100 new complex environments from the glucose environment. Each complex environment contains multiple carbon sources.

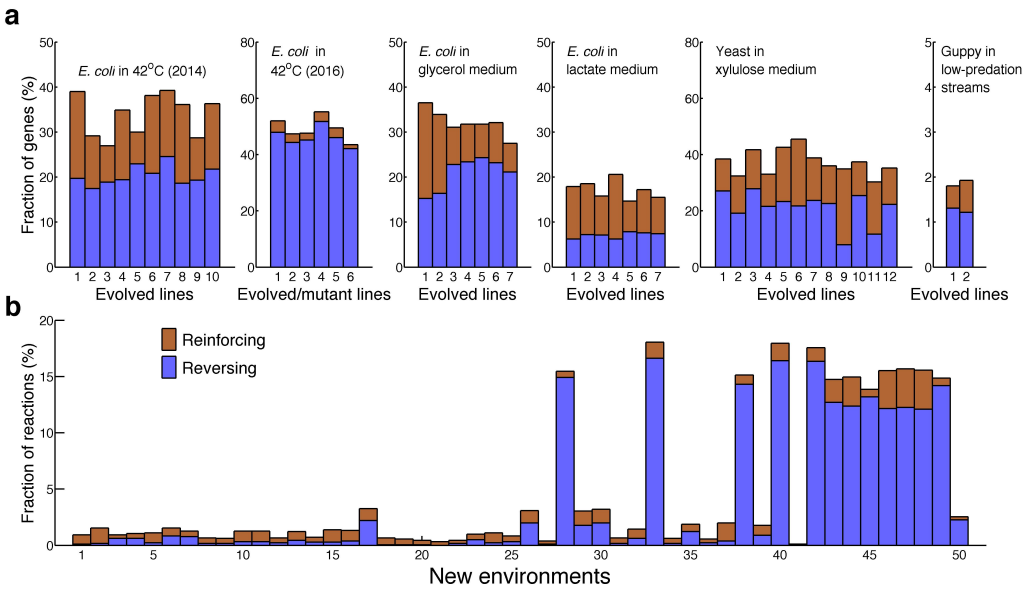

Figure A.5.8 Frequencies of traits showing reinforcing and reversing phenotypic changes in adaptations. Traits satisfying |La - Lo|> 0.2Lo, |Lp - Lo| > 0.2Lo, and |La - Lp| > 0.2Lo are classified into reinforcing and reversing traits based on whether the genetic and plastic changes are of the same direction or opposite directions. (a) Fractions of genes with reinforcing and reversing expression changes, respectively, in experimental evolution. Organisms as well as the new environments to which the organisms were adapting to are indicated. Each bar represents an adaptation. (b) Fractions of reactions with predicted reinforcing and reversing flux changes, respectively, in *E. coli*'s adaptations to 50 new environments from the glucose environment.

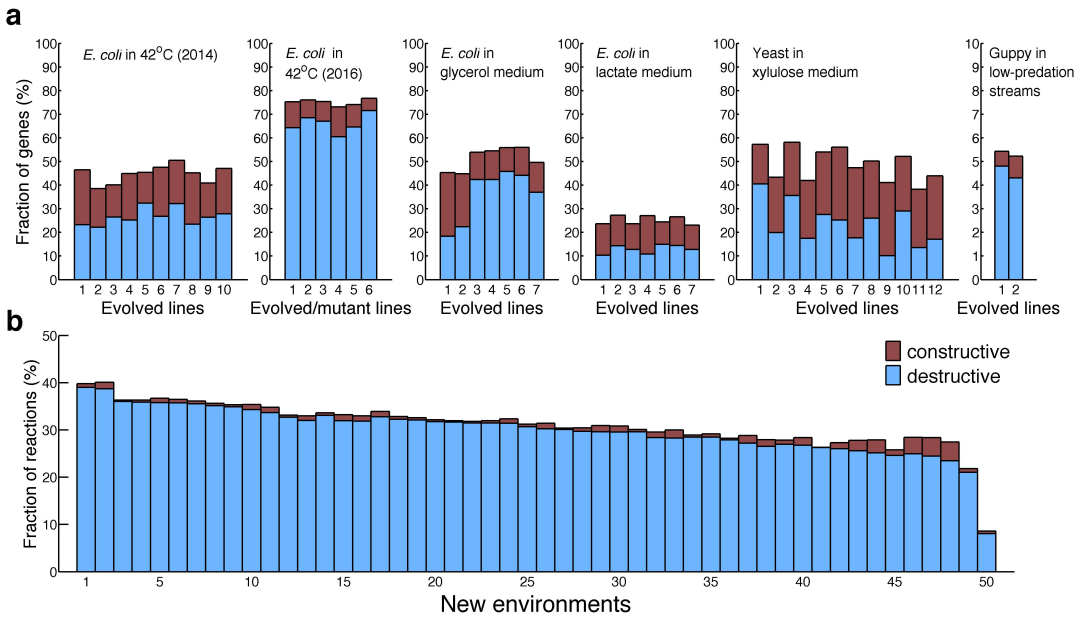

Figure A.5.9 Constructive plasticity is generally less prevalent than destructive plasticity in adaptations. See main text for definitions of constructive and destructive plasticity. (a) Fractions of genes showing constructive and destructive expression plasticity, respectively, in each of the 44 cases of experimental evolution. (b) Fractions of reactions showing constructive and destructive flux plasticity, respectively, in each of the 50 environmental adaptations of the *E. coli* metabolic network from the glucose medium.