Adaptive Genetic Robustness of *Escherichia coli* Metabolic Fluxes

Wei-Chin Ho¹ and Jianzhi Zhang^{*,1}

¹Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor

*Corresponding author: E-mail: jianzhi@umich.edu.

Associate editor: Meredith Yeager

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.

Key words: environmental robustness, metabolic network, flux balance analysis, evolution, natural selection.

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 and Lindquist 1998; Fares et al. 2002; Ciliberti et al. 2007; Ho and Zhang 2014; Yang et al. 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; Félix and 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; Sanjuán 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 and Hartl 2002; de Visser et al. 2003). The intrinsic property hypothesis asserts that genetic robustness is an intrinsic property of biological systems (Siegal and 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 and 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 and 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 et al. 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). In 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 and Kawecki 1994; Stearns et al. 1995; Houle 1998; Proulx et al. 2007). We recently applied the same strategy to 220 yeast morphological traits and found evidence for the adaptation hypothesis (Ho and Zhang 2014). Although 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 and 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 and 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 et al. 2010) to measure trait importance. We then use the method of minimization of metabolic adjustment (MOMA) (Segrè et al. 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.

Results

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 et al. 2010). FBA predictions for model organisms such as the bacterium *E. coli* have been extensively validated experimentally (Ibarra et al. 2002; Fong and 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 et al. 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 (Segrè et al. 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 (Segrè et al. 2002). After individually constraining the 362 fluxes to 0, we found that the frequency distribution of mutant fitness is U-shaped (supplementary fig. S1A, Supplementary Material online), as was previously found (Wloch et al. 2001; Sanjuán et al. 2004; Wang and Zhang 2009a). Note that some f values are extremely small (see supplementary fig. S1B, Supplementary Material online, 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 $s = 1 - f$. In other words, the greater the drop in fitness upon the block of a reaction, the more important the reaction is.

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 $\Delta F = 0.05, 0.10, 0.15, \dots$, 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. 1A). The same pattern was observed when essential reactions were examined (red lines in fig. 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 nonfocal 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 (phosphofructokinase), 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. 1B). That is, no matter which reaction is constrained, the

FIG. 1. Changes in metabolic reaction fluxes from the wild-type levels predict fitness decreases. (A) FBA-predicted fitness decreases with the rise in the extent to which the flux of a reaction is constrained (ΔF). Each line represents a series of constraints imposed on the flux of one reaction. (B) Fitness (f) is negatively correlated with the fractional flux change (γ) for the focal reaction PFK 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.

Fig. 2. Comparison of genetic and environmental robustness between *Escherichia coli* and random networks. (A) Distribution of the average fractional flux change across all focal reactions and mutants ($\bar{\gamma}_G$) 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 1,000 random nutritional environments ($\bar{\gamma}_E$) in 500 random networks. The arrow indicates the corresponding value observed in *E. coli*.

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. 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. 1A). For all ΔF values considered and all k , all ρ values except three are negative (fig. 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.

Genetic Robustness of *Escherichia 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 and Wagner 2009; Barve and 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 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 3 of the 500 random networks have a $\bar{\bar{\gamma}}_G$ smaller than that of *E. coli* (fig. 2A), indicating that the average genetic robustness of metabolic fluxes is significantly greater in *E. coli* than in comparable random networks ($P = 0.006$).

Environmental Robustness of *Escherichia 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 1,000 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. 2B), indicating that the average environmental robustness of *E. coli* metabolic

FIG. 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 in 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 *Escherichia coli*.

fluxes is only marginally significantly greater than the chance expectation ($P = 0.062$).

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 reactions 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. 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 et al. 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. 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.

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, $\bar{\gamma}_E$ 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, $\bar{\gamma}_G$ and $\bar{\gamma}_E$ should be positively correlated. Finally and most critically, because the hypothesis claims that genetic robustness is entirely a byproduct of selection for environmental robustness, $\bar{\gamma}_G$ should no longer correlate with s after the control of $\bar{\gamma}_E$.

We examined whether our data are consistent with the above three predictions of the congruence hypothesis. Consistent with the first prediction, $\bar{\gamma}_E$ is negatively correlated with s ($\rho = -0.890$, $P = 5.5 \times 10^{-37}$; fig. 4A) and the correlation is significantly stronger than those observed in the 500 comparable random networks ($P < 0.002$; fig. 4B). Consistent with the second prediction, we found $\bar{\gamma}_G$ to be highly and positively correlated with $\bar{\gamma}_E$ (Spearman's $\rho = 0.911$, $P = 2.0 \times 10^{-41}$; fig. 4C). But, contrary to the third prediction, the partial correlation between $\bar{\gamma}_G$ and s after the control of $\bar{\gamma}_E$ 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. 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.

FIG. 4. The congruence hypothesis of genetic robustness is rejected. (A) The average fractional flux change for a focal reaction among all 1,000 environments examined ($\bar{\gamma}_E$) decreases with the rise in 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 *Escherichia coli*. (C) Average fractional flux change for a focal reaction on mutation and that on environmental change is 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*.

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 1,000 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 $\bar{\gamma}_E$ to be 0.53, much larger than the previous $\bar{\gamma}_E$ (fig. 2B) and $\bar{\gamma}_G$ (fig. 2A), suggesting that on average the single-carbon-source environments represent severer challenges to *E. coli* than the 1,000 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. 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 figure 4D is not an artifact of much smaller $\bar{\gamma}_E$ than $\bar{\gamma}_G$.

Environmental Robustness as a Side Effect of Genetic Robustness

Although $\bar{\gamma}_E$ is only marginally significantly smaller in *E. coli* than in the random networks (fig. 2B), s and $\bar{\gamma}_E$ are strongly correlated in *E. coli* (fig. 4A) and this correlation is

FIG. 5. Confirmation of results using single-carbon-source environments. (A) The average fractional flux change for a focal reaction among 257 single-carbon-source environments examined ($\bar{\gamma}_E$) decreases with the rise in 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 *Escherichia coli*. (C) Average fractional flux change for a focal reaction on mutation ($\bar{\gamma}_G$) and that on environmental change ($\bar{\gamma}_E$) is 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*.

stronger than that in any of the 500 random networks (fig. 4B). To examine whether the correlation between s and $\bar{\gamma}_E$ is a byproduct of the adaptive genetic robustness, we examined the partial correlation between s and $\bar{\gamma}_E$ after the control of $\bar{\gamma}_G$. Although 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 ($P = 0.060$; supplementary fig. S2A, Supplementary Material online). This is also true under the single-carbon-source environments ($P = 0.39$; supplementary fig. S2B, Supplementary Material online). These results, coupled with those in figure 4, suggest that the environmental robustness of *E. coli* metabolic fluxes is likely a byproduct of the adaptive genetic robustness.

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 and Lindquist 1998; Levy and Siegal 2008; Takahashi 2013). We measured $\Delta\bar{\gamma}_G$, the difference in $\bar{\gamma}_G$ caused by the removal of a nonessential reaction. The more positive $\Delta\bar{\gamma}_G$ 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 $\Delta\bar{\gamma}_G$ is positive in all 96 cases (fig. 6A). The top 10% of the reactions in terms of the associated $\Delta\bar{\gamma}_G$ values are marked in figure 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). Figure 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 figure 6A and B (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. 6C). In addition, PPC is an anaplerotic reaction replenishing metabolites in the TCA cycle that are largely consumed by anabolism (Nelson and 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 $\Delta\bar{\gamma}_G$ 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 $\Delta\bar{\gamma}_G$ and $\Delta\rho$ are smaller in random networks than in *E. coli* (fig. 7), suggesting that at least these three capacitor reactions indeed contribute to genetic robustness beyond the random expectation.

Discussion

Several previous studies revealed the genetic robustness of metabolic networks by treating the viability as a trait (Edwards and 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 on 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 . Furthermore, 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 and Bergman 2002). Interestingly, we found that the correlation between s and $\bar{\gamma}_G$ 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. 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 $\bar{\gamma}_E$ is significantly negative for 99.6% of the random metabolic networks, suggesting that environmental robustness could arise as an intrinsic property. Additional evidence (supplementary fig. S2, Supplementary Material online), 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 et al. 1997; Meiklejohn and Hartl 2002), based on the report

Fig. 6. Removing reactions from the *Escherichia coli* metabolic network affects how it behaves under genetic perturbations. (A) Frequency distribution of $\Delta \bar{\gamma}_G$ 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).

FIG. 7. Removing capacitor reactions does not affect the genetic robustness in random networks as much as in *Escherichia coli*. (A) Frequency distribution of $\Delta\bar{\gamma}_G$ in 46 analyzed random networks upon removal of FUM. (B) Frequency distribution of $\Delta\rho(s, \bar{\gamma}_G)$ in 46 analyzed random networks upon removal of FUM. (C) Frequency distribution of $\Delta\bar{\gamma}_G$ in 15 analyzed random networks upon removal of PPC. (D) Frequency distribution of $\Delta\rho(s, \bar{\gamma}_G)$ in 15 analyzed random networks upon removal of PPC. (E) Frequency distribution of $\Delta\bar{\gamma}_G$ in five analyzed random networks upon removal of SUCDi. (F) Frequency distribution of $\Delta\rho(s, \bar{\gamma}_G)$ in five analyzed random networks upon removal of SUCDi.

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.

Although 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 and 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 relationship 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 and 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 ($\rho = -0.014$, $P = 0.79$). 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 $\bar{\gamma}_G$ and a stronger correlation between s and $\bar{\gamma}_G$ than those of the random networks (supplementary fig. S3, Supplementary Material online). 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 and 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 and Burns 1981; Wang and 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 and 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 3,000 gene expression traits (Ho and 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 et al. 1997; Ho and Zhang 2014). Our finding not only answers the long-standing 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.

Materials and Methods

Metabolic Models

We used *E. coli* metabolic network model iAF1260 (Feist et al. 2007), which includes 2381 reactions and 1,039 metabolites. Among all reactions, 2,082 are metabolic reactions and 299 are exchange reactions allowing the uptake of nutrients from the environment. The SML 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).

Flux Balance Analysis

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 et al. 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 $S\mathbf{v} = \mathbf{0}$ and $\alpha \leq \mathbf{v} \leq \beta$. Here, $\mathbf{v}$ is a vector of reaction fluxes, S is a matrix describing the stoichiometric relationships among metabolites in each reaction, α is a vector describing the lower bound of each flux, and β is a vector describing the upper bound of each flux.

We used the default α and β 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 irreversible metabolic reaction i , the default $\alpha_i = 0$ and $\beta_i = \infty$. The optimized fluxes ($\mathbf{v}_0$) 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 data set of traits and used the remaining 362 reactions.

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

Minimization of Metabolic Adjustment

The objective of MOMA is to minimize $(\mathbf{v} - \mathbf{v}_0)^2$, where $\mathbf{v}$ is a vector of all reaction fluxes upon a genetic or environmental perturbation, whereas $\mathbf{v}_0$ 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 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 ($\mathbf{v}_0$) 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 $-0.9|\mathbf{v}_0| \leq \mathbf{v} \leq 0.9|\mathbf{v}_0|$; otherwise, we use $0 \leq \mathbf{v} \leq 0.9\mathbf{v}_0$.

When introducing environmental changes, we followed the approach previously published (Wang and 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 $10 \text{ mmol gDW}^{-1} \text{ h}^{-1}$ (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, 1,000 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. 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.

Fractional Flux Change

After calculating the flux of a reaction ν 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 (ν_0). To make increasing and decreasing fluxes comparable, we normalize it in different ways depending on whether ν is larger or smaller than ν_0 . A larger γ means larger difference.

$$\gamma = \begin{cases} 1 - \nu/\nu_0, & \nu < \nu_0 \\ 1 - \nu_0/\nu, & \nu \geq \nu_0 \end{cases}$$

Random Networks

We used a previously published approach to generate 500 random networks (Rodrigues and Wagner 2009; Barve and Wagner 2013). We first compiled “the universe of reactions” by acquiring metabolic reactions listed in the REACTION section of the KEGG database (Kanehisa and Goto 2000) except when 1) a reaction appears in the *E. coli* metabolic network iAF1260; 2) a reaction involves polymer subunits with uncertain number of atoms; 3) a reaction involves glycans; 4) a reaction involves metabolites without information about their structure; and 5) a reaction is unbalanced in mass or charge. Then we combined these reactions with the metabolic reactions in iAF1260 (excluding transport reactions). In total, there are 5,001 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 have been subject to neither selection for genetic robustness nor selection for environmental robustness. During the random walk, we sampled a random network after every 5,000 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 (supplementary fig. S4, Supplementary Material online). 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.

Supplementary Material

Supplementary figures S1–S4 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

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.

References

- Barve A, Rodrigues JF, Wagner A. 2012. Superessential reactions in metabolic networks. *Proc Natl Acad Sci U S A*. 109:E1121–E1130.
- Barve A, Wagner A. 2013. A latent capacity for evolutionary innovation through exaptation in metabolic systems. *Nature* 500:203–206.
- Becker SA, Feist AM, Mo ML, Hannum G, Palsson BO, Herrgard MJ. 2007. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc*. 2:727–738.
- Bordbar A, Monk JM, King ZA, Palsson BO. 2014. Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet*. 15:107–120.
- 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.
- Clark AG. 1991. Mutation-selection balance and metabolic control-theory. *Genetics* 129:909–923.
- 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.
- 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.
- Edwards JS, Palsson BO. 2000. Robustness analysis of the *Escherichia coli* metabolic network. *Biotechnol Prog*. 16:927–939.
- Fares MA, Ruiz-González MX, Moya A, Elena SF, Barrio E. 2002. Endosymbiotic bacteria: groEL buffers against deleterious mutations. *Nature* 417:398.
- 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. *Mol Syst Biol*. 3:121.
- Félix MA, Barkoulas M. 2015. Pervasive robustness in biological systems. *Nat Rev Genet*. 16:483–496.
- Fieller EC, Hartley HO, Pearson ES. 1957. Tests for rank correlation coefficients.1. *Biometrika* 44:470–481.
- Fong SS, Palsson BO. 2004. Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nat Genet*. 36:1056–1058.
- 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 X, Qian W, Wang Z, Li Y, Zhang J. 2010. Prevalent positive epistasis in *Escherichia coli* and *Saccharomyces cerevisiae* metabolic networks. *Nat Genet*. 42:272–276.
- 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.
- Houle D. 1998. How should we explain variation in the genetic variance of traits? *Genetica* 102–103:241–253.

- 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.
- Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási AL. 2000. The large-scale organization of metabolic networks. *Nature* 407:651–654.
- Kacser H, Burns JA. 1981. The molecular basis of dominance. *Genetics* 97:639–666.
- Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28:27–30.
- Levy SF, Siegal ML. 2008. Network hubs buffer environmental variation in *Saccharomyces cerevisiae*. *PLoS Biol.* 6:e264.
- 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.
- Lynch M. 1988. The rate of polygenic mutation. *Genet Res.* 51:137–148.
- Meiklejohn CD, Hartl DL. 2002. A single mode of canalization. *Trends Ecol Evol.* 17:468–473.
- Nelson DL, Cox MM. 2008. Lehninger principles of biochemistry. New York: W. H. Freeman and Company.
- Orth JD, Thiele I, Palsson BO. 2010. What is flux balance analysis? *Nat Biotechnol* 28:245–248.
- Proulx SR, Nuzhdin S, Promislow DE. 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.
- Rodríguez JFM, Wagner A. 2009. Evolutionary plasticity and innovations in complex metabolic reaction networks. *PLoS Comput Biol.* 5:e1000613.
- Rutherford SL, Lindquist S. 1998. Hsp90 as a capacitor for morphological evolution. *Nature* 396:336–342.
- Samal A, Rodríguez JFM, Jost J, Martin OC, Wagner A. 2010. Genotype networks in metabolic reaction spaces. *BMC Syst Biol.* 4:30.
- Sanjuán R, Cuevas JM, Furió V, Holmes EC, Moya A. 2007. Selection for robustness in mutagenized RNA viruses. *PLoS Genet.* 3:939–946.
- Sanjuán R, Moya A, Elena SF. 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, Park JO, Conrad TM, Palsson BO. 2010. BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics* 11:213.
- Segrè D, Vitkup D, Church GM. 2002. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci U S A.* 99:15112–15117.
- 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.
- Takahashi KH. 2013. Multiple capacitors for natural genetic variation in *Drosophila melanogaster*. *Mol Ecol.* 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, Booth G, Bagheri-Chaichian H. 1997. A population genetic theory of canalization. *Evolution* 51:329–347.
- 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–E76.
- Wang Z, Zhang J. 2009a. Abundant indispensable redundancies in cellular metabolic networks. *Genome Biol Evol.* 1:23–33.
- Wang Z, Zhang J. 2009b. Why is the correlation between gene importance and gene evolutionary rate so weak? *PLoS Genet* 5:e1000329.
- 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.
- Wloch DM, Szafraniec K, Borts RH, Korona R. 2001. Direct estimate of the mutation rate and the distribution of fitness effects in the yeast *Saccharomyces cerevisiae*. *Genetics* 159:441–452.
- Yang JR, Ruan S, Zhang J. 2014. Determinative developmental cell lineages are robust to cell deaths. *PLoS Genet.* 10:e1004501.