Why Phenotype Robustness Promotes Phenotype Evolvability

Xinzhu Wei and Jianzhi Zhang*

Department of Ecology and Evolutionary Biology, University of Michigan

*Corresponding author: Email: jianzhi@umich.edu.

Abstract

Robustness and evolvability are fundamental characteristics of life whose relationship has intrigued generations of biologists. Studies of several genotype–phenotype maps (GPMs) such as the map between short DNA sequences and their bindings to transcription factors showed that phenotype robustness (PR) promotes phenotype evolvability (PE), but the underlying reason is unclear. Here, we show mathematically that the expected PE is a monotonically increasing function of the expected PR in random GPMs. Population genetic simulations confirm that increasing PR raises the probability that a target phenotype appears in a population within a given time, under empirical as well as randomly rewired GPMs. These and other results demonstrate that the positive correlation between PR and PE is mathematical rather than biological. Hence, it is unsurprising to observe this correlation in every empirical GPM investigated, although the magnitude of the correlation may vary due to influences of various biological factors.

Key words: evolution, genotype–phenotype map, neutral network, transcription factor binding sequences.

Robustness and Evolvability

Genetic robustness refers to phenotypic invariance in the face of mutation and is a widespread phenomenon at multiple levels of biological organization (de Visser et al. 2003; Kitano 2004; Wagner 2005b; Masel and Trotter 2010; Yang et al. 2014; Ho and Zhang 2016). Evolvability is the ability to produce (adaptive) phenotypic variation (Wagner and Altenberg 1996; Kirschner and Gerhart 1998; Wagner 2005b; Masel and Trotter 2010). Although robustness and evolvability are both fundamental characteristics of life, their relationship has been a long-standing controversy (Kitano 2004; Wagner 2005b; Masel and Trotter 2010). On the one hand, they are apparently antagonistic to each other, because the higher the robustness, the lower the probability with which a mutation results in a new phenotype (Ancel and Fontana 2000; Carter et al. 2005). On the other hand, robustness has been suggested to promote evolvability, not least because robustness allows the accumulation in a population of cryptic genetic variations that may be exposed and adaptive in a new environment (Aldana et al. 2007; Elena and Sanjuan 2008; Masel and Trotter 2010). Experimental evolution of RNA enzymes (Hayden et al. 2011), RNA viruses (McBride et al. 2008), and bacteria (Stiffler et al. 2015) showed that robustness can indeed enhance evolvability under certain conditions, but the generality of these findings is unknown.

Theoretical analysis of the robustness–evolvability relationship is often conducted in the context of a genotype–phenotype map (GPM; Fig. 1A), where each node is a genotype, each edge connects two genotypes that differ by one mutation, and nodes are colored based on their phenotypes (Wagner 2012). The set of connected nodes with the same color is commonly referred to as a neutral network (Schuster et al. 1994), because wandering in this network alters the genotype but not the phenotype. Note, however, that phenotypes are defined qualitatively in this context.

A decade ago, Wagner revolutionized the study of the robustness–evolvability relationship by distinguishing between genotype robustness (GR) and phenotype robustness (PR) and between genotype evolvability (GE) and phenotype evolvability (PE) (Wagner 2008). GR is the probability with which a random mutation occurring in a given genotype does not change its phenotype. By contrast, PR is the mean GR of all genotypes exhibiting a given phenotype. GE is the fraction of all phenotypes reachable by one mutation from a given genotype. By contrast, PE is the fraction of all phenotypes reachable by one mutation from any genotype exhibiting a given phenotype. Wagner and colleagues found that, within a GPM, GR and GE are negatively correlated but PR and PE are positively correlated for the phenotypes of RNA structure (Wagner 2008), protein structure (Ferrada and Wagner

2008), and DNA binding to transcription factors (TFs) (Payne and Wagner 2014). However, the broader generality and the underlying cause of the positive PR–PE correlation are unclear.

PE Is Expected to Increase Monotonically with PR in Random GPMs

That a positive PR–PE correlation is observed in every GPM investigated (Ferrada and Wagner 2008; Wagner 2008; Payne and Wagner 2014) prompts us to investigate the possibility that this correlation is mathematical rather than biological. To this end, we consider a random GPM between G DNA sequences (genotypes) and their binding to K TFs (phenotypes). Each node represents a genotype of an l-nucleotide DNA sequence, and each phenotype represents the binding of the DNA to a TF. Let the number of genotypes showing phenotype i (i.e., the number of binding sequences of TF_i) be g_i. With a single-nucleotide replacement, each genotype can change to one of $m = 3l$ other genotypes, which are collectively called the neighborhood of the focal genotype. In this random GPM, under the assumption that $1 \ll g_i \ll G$ for any i, it can be shown (see Materials and Methods) that the expected PR of binding to TF_i is

$$E(PR_i) \approx g_i/G, \tag{1}$$

whereas the corresponding expected PE is

$$E(PE_i) \approx 1 - \sum_{j \neq i} e^{-m g_j g_i/G}/(K-1). \tag{2}$$

Hence,

$$E(PE_i) \approx 1 - \sum_{j \neq i} e^{-m g_j E(PR_i)}/(K-1). \tag{3}$$

Equation (3) shows that the expected PE_i is a monotonically increasing function of the expected PR_i. In other words, the expected PR and PE are intrinsically positively correlated in random GPMs. Importantly, equation (3) does not rely on any specific distribution of g_i.

To evaluate the accuracy of the above formulas that were derived with approximations, we simulated a random GPM with $K = 80$ TFs that all use 8-mer binding sequences. We chose these parameters because the empirically determined yeast and mouse TF-DNA binding GPMs have 89 and 105 TFs, respectively, and their binding sequences inferred from microarray data all contain 8 nucleotides (see Materials and Methods). To examine the variations of PR and PE in the entire range of possible g_i values, we chose the g_i values to be 15, 25, 35, ..., and 805. We repeated the simulation 100 times and calculated the mean empirical PR and PE of binding to each TF. We found that E(PR) (fig. 1B), E(PE) (fig. 1C), and their relationship (fig. 1D) based on the analytical formulas are indistinguishable from the corresponding average values observed from the

simulation. This was also the case when g_i follows a normal (supplementary fig. S1A–C, Supplementary Material online), bimodal (supplementary fig. S1D–F, Supplementary Material online), or exponential (supplementary fig. S1G–I, Supplementary Material online) distribution, suggesting that our analytical formulas are sufficiently accurate and general.

The PR–PE Correlation Is Stronger in Empirical than Randomly Rewired GPMs

We noticed from the analytical and simulation results of random GPMs that PE becomes virtually independent of PR when PR exceeds a certain value (fig. 1 and supplementary fig. S1, Supplementary Material online). This phenomenon is much less pronounced in the empirical TF-DNA binding GPMs of mouse (fig. 2A–C) and yeast (supplementary fig. S2A–C, Supplementary Material online). To quantitatively compare empirical with random GPMs, we analytically computed the expected PR and PE for each TF in a randomly rewired mouse GPM, where the number of genotypes for each phenotype is unchanged but the genotype–phenotype relationships are randomized. Relative to a randomly rewired GPM, the actual GPM has higher PR and lower PE values for most TFs (fig. 2A and B). This result is similar to that of Payne and Wagner (2014), although they computed PR and PE for a TF by randomly rewiring the binding sequences of the focal TF instead of those of all TFs simultaneously. Furthermore, they did not examine the relationship between PR and PE in any random or randomly rewired GPM. We found that the positive rank correlation between PR and PE is greater in the actual GPM than in each of 100 randomly rewired GPMs (fig. 2D). Similar results were found when the yeast GPM was compared with corresponding randomly rewired GPMs (supplementary fig. S2, Supplementary Material online).

The Increase in the PR–PE Correlation Is Related to Large Neutral Networks

We hypothesize that the differences between the empirical GPMs and their randomly rewired GPMs in PR, PE, and PR–PE correlation are primarily related to the existence of large neutral networks in the former (i.e., genotypes of the same phenotypes tend to be connected) but not the latter. On average, the largest connected network for a mouse (or yeast) TF contains 81% (or 79%) of its binding sequences. This number drops to 1.2% in the randomly rewired GPMs of both species. Based on the definitions of PR and PE, it is obvious that, given g_i values, the presence of large neutral networks raises PR but reduces PE. As a result, PE increases with g_i in almost the full range of g_i values in the empirical GPMs (fig. 2B and supplementary fig. S2B, Supplementary Material online), but saturates even in the bottom tenth of g_i values in the randomly rewired GPMs (fig. 2B and supplementary fig. S2B, Supplementary Material online).
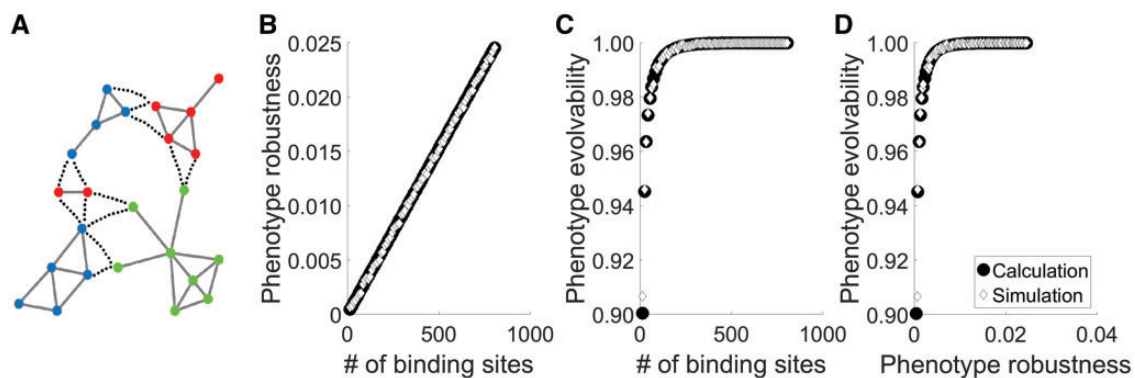
FIG. 1.—PR and PE are positively correlated in random GPMs. (A) A hypothetical GPM. Each node represents a genotype, while its color represents its phenotype. Two genotypes that are one mutational step away from each other are connected by an edge, where a solid edge connects genotypes of the same phenotype and a dotted edge connects genotypes of different phenotypes. (B) The expected PR increases with the number of binding sequences in random TF-DNA binding GPMs. Each symbol represents one TF. Solid circles show analytically calculated values while open diamonds show corresponding means observed from 100 simulations of random GPMs. The observed standard deviation of PR (average 0.0016) is not correlated with the number of binding sequences. See main text for the parameters of the GPMs used. (C) The expected PR increases with the number of binding sequences in these random GPMs. The observed standard deviation of PE (maximum 0.0304) is negatively correlated with the number of binding sequences. (D) The expected PE is a monotonically increasing function of the expected PR in these random GPMs.
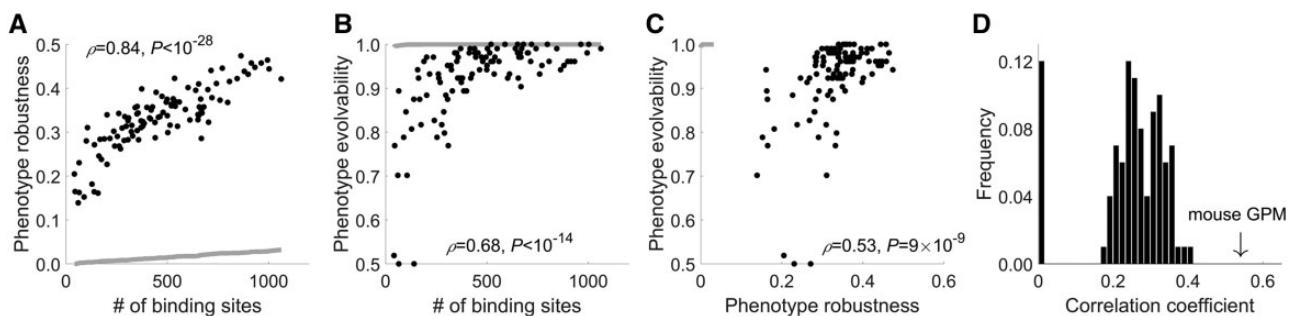

FIG. 2.—PR–PE relationships in the mouse TF-DNA binding GPM and corresponding randomly rewired GPMs. (A) PR increases with the number of binding sequences in the mouse GPM. Each dot is a TF. (B) PE increases with the number of binding sequences in the mouse GPM. (C) PE is an increasing function of PR in the mouse GPM. In (A–C), the analytically computed results in corresponding random GPMs are presented by the grey curves. (D) Frequency distribution of the rank correlation between PR and PE in 100 randomly rewired mouse GPMs. The arrow points to the observed correlation in the mouse GPM.

To further demonstrate that the differences in PR, PE, and PR–PE correlation between empirical GPMs and their randomly rewired GPMs are due primarily to neutral networks instead of other properties of empirical GPMs, we created randomized GPMs with large neutral networks (see Materials and Methods). Indeed, patterns of PR, PE, and PR-PE correlation in these GPMs closely resemble those in empirical GPMs (supplementary fig. S3, Supplementary Material online).

The Biophysics of TF-DNA Binding Creates Large Neutral Networks

It is interesting to note that, if the binding sequences of a TF were randomly distributed in a GPM, a population starting with a weak binding sequence would have to cross deep binding affinity valleys to reach a strong binding sequence,

which is improbable except in very small populations. Thus, the presence of strong TF-DNA binding per se implies the existence of large (qualitatively) neutral networks of its binding sequences. But what forces have led to the large neutral networks? It is known that the genotypes for a phenotype tend to form a large neutral network simply by chance when the genotype number is sufficiently large. This phenomenon of percolation is, however, irrelevant here, because the phenotype with the largest number of genotypes contains only 2–3% of all genotypes in the GPMs studied here, much lower than the lower bound required for percolation (6.25%) (Gravner et al. 2007).

TF-DNA binding is known to be primarily determined by specific base pair recognition (von Hippel and Berg 1986), and at different amino acid binding positions, different base pairs are preferred due to interaction with hydrogen bonds provided by appropriately positioned amino acids and peptide

functional groups (von Hippel and Berg 1986; Stormo and Fields 1998; Afek et al. 2014). The biophysical property of TF-DNA binding dictates that the binding energy between a TF and a segment of DNA is largely the sum of the interaction energies of individual couples of an amino acid residue and a base pair. Only at 5% of sites does the binding strength deviate from the multiplicative expectation by more than 2-fold (Jolma et al. 2013). The scarcity of epistasis means that the one-mutation neighborhood of a strong binding sequence of a TF is likely filled with the binding sequences of the same TF, because a single-nucleotide change cannot drastically reduce the TF-DNA binding strength. Indeed, binding sequences with higher binding affinities tend to have higher GR (Payne and Wagner 2014). This property leads to the creation of large neutral networks. A recent extensive analysis of TF-DNA binding affinities generally supports this notion (Aguilar-Rodríguez et al. 2017).

PR Facilitates Adaptation in Population Genetic Simulations under Randomly Rewired GPMs

Because Wagner's definition of PE does not explicitly consider the population genetic process of adaptation, we turn to another, arguably more relevant measure of evolvability—the probability that a target phenotype appears in a population within a given time, which we will refer to as PE'. We start with a haploid adult population with a homogenous genotype corresponding to phenotype i, which is optimal in the current environment. All other phenotypes are lethal. In each generation, genetic drift occurs such that N offspring are produced and their genotype frequencies may differ from those of the parental population. Each offspring has a probability of μ to become a neighboring genotype due to mutation, and only those with viable phenotypes mature and reproduce (i.e., some of the N individuals may not mature). Based on theory (Nei et al. 1975) and our pilot simulation, we repeat this process for $1/\mu$ generations to allow the population to reach an equilibrium level of genetic diversity. An environmental shift then occurs, which renders phenotype i suboptimal, phenotype j ($\neq i$) optimal, and all other phenotypes still lethal. We repeat the process of mutation, purifying selection, and drift over many generations until an individual with phenotype j appears in the population or the number of generations after the environmental shift reaches a preset limit T, whichever occurs first. We examine each and every new phenotype j ($\neq i$) and calculate the fraction of phenotypes that can be reached from i within time T, which is PE'. We repeat the evolutionary simulation 50 times, each starting from a randomly picked genotype of the phenotype i and present the average result from these 50 simulations. We consider the first appearance of the adaptive phenotype rather than the first fixation of the adaptive phenotype, because the fixation probability and expected fixation time is the same given N, μ,
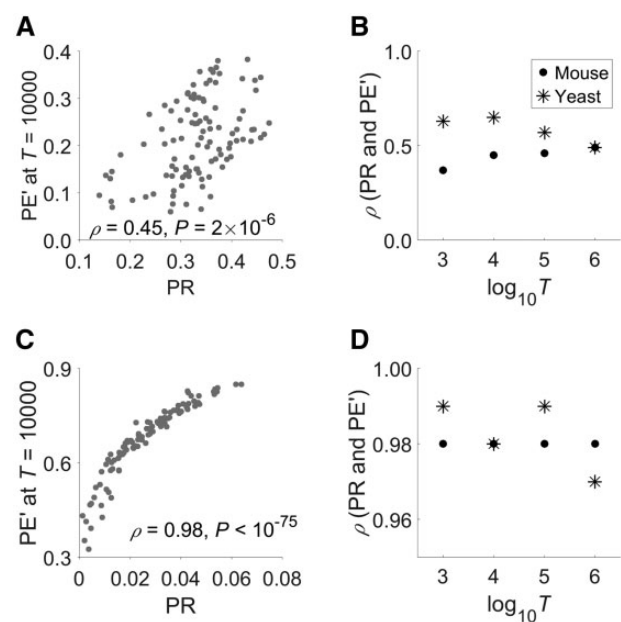

FIG. 3.—Population genetic simulations show that PR promotes PE', which is the probability that a target phenotype appears in a population within time T. (A) Positive correlation between PR and PE' under the mouse GPM when $T = 10{,}000$ generations. ρ, Spearman's rank correlation coefficient. (B) Rank correlation between PR and PE' under mouse (asterisks) and yeast (dots) GPMs, respectively. (C) Positive correlation between PR and PE' under a randomly rewired mouse GPM when $T = 10{,}000$ generations. (D) Rank correlation between PR and PE' under randomly rewired mouse (asterisks) and yeast (dots) GPMs, respectively. In panels (B) and (D), all correlations significantly exceed 0 ($P < 10^{-4}$). For mouse, our simulation used $N\mu = 0.004$ per generation per motif, based on the motif length of 8 nucleotides, mutation rate of 5.4×10^{-9} per generation per site (Uchimura et al. 2015), and effective population size of 10^5 (Phifer-Rixey et al. 2012). For yeast, our simulation used $N\mu = 0.016$ per generation per motif, based on its motif length of 8 nucleotides, mutation rate of 2×10^{-10} per generation per site (Zhu et al. 2014), and effective population size of 10^7 (Wagner 2005a).

and selective strength. In all simulations, we use $N = 100$ to speed up the process.

We first conducted the population genetic simulation under the mouse TF-DNA binding GPM using mouse-appropriate $N\mu$. When $T = 10{,}000$ generations is the upper limit in waiting time for the target phenotype, we found a positive correlation between the PR of the starting phenotype and PE' ($\rho = 0.45$, $P < 10^{-5}$; fig. 3A). Similar results were obtained (fig. 3B) when T is 1,000 ($\rho = 0.37$, $P < 10^{-4}$), 100,000 ($\rho = 0.46$, $P < 10^{-5}$), or 1,000,000 generations ($\rho = 0.49$, $P < 10^{-6}$). Thus, increasing PR raises the chance of adaptation upon an environmental shift.

We similarly conducted the population genetic simulation under the yeast TF-DNA binding GPM using yeast-appropriate $N\mu$. We again observed that the higher the PR of the starting phenotype, the higher the probability of appearance of a target phenotype in the population (fig. 3B).

Interestingly, the correlation between PR and PE' becomes even stronger when we conducted simulations under randomly rewired mouse and yeast GPMs, respectively (fig. 3C and D). These results indicate that PR promotes PE' and that this property is intrinsic rather than biological.

Implications

Our mathematical and empirical results showed that (1) the expected PR and PE are intrinsically positively correlated even in random GPMs; (2) compared with the corresponding randomly rewired GPMs, the mouse and yeast TF-DNA binding GPMs show stronger PR–PE correlations, likely because of their large neutral networks; and (3) these large neutral networks are explainable by the biophysical nature of TF-DNA binding. While (1) is a general finding for GPMs of all classes of phenotypes, (2) and (3) are derived from the analysis of TF-DNA binding GPMs. Nonetheless, for any phenotype that can be improved by natural selection, its genotypes must form some neutral networks such that quantitatively better phenotypes are reachable by mutation; otherwise, the phenotype could not be improved by natural selection. Hence, we expect (2) to be true in the GPM for any adaptable phenotype (when adaption occurs primarily via mutation rather than recombination). Note, however, that our finding that the expected PR and PE are positively correlated in random GPMs does not imply that PR and PE cannot have a negative correlation even in hypothetical GPMs. For instance, one could imagine a GPM where the genotypes of some phenotypes form large neutral networks whereas those of other phenotypes are largely unconnected. Compared with the latter group of phenotypes, the former group is expected to have higher PR but lower PE. Consequently, a negative correlation between PR and PE would result when the two groups of phenotypes are analyzed together. Nevertheless, such GPMs should be the exception rather than the rule. Hence, observing a positive PR–PE correlation in an empirical GPM is expected and does not offer any specific biological insight, as far as Wagner's definitions are concerned.

Our population genetic simulations showed that PR promotes PE' under real and randomly rewired GPMs. PE' is similar to Wagner's definition of PE except that PE' is defined in a population genetic framework and hence is more realistic and more relevant to actual adaptation. Our population genetic simulation differs from a previous treatment of the same subject by Draghi et al. (2010), who found PR to promote PE' under some but not all circumstances. However, their study contained a number of simplifying assumptions. For instance, they assumed that any genotype has a non-zero probability to show any phenotype by a minimum of one mutation, which is untrue. In addition, no GPM was explicitly modeled and only genotypes of the starting phenotype were assumed to form a neutral network. They also unrealistically assumed that all genotypes of the same phenotype have equal robustness. Furthermore, although the robustness of a phenotype

correlates with the number of neighboring phenotypes, they neglected this correlation in their model. Hence, our analysis, based on actual and randomly rewired GPMs, coupled with more realistic assumptions, is biologically more relevant than theirs. Note that, Draghi et al. observed a decrease in PE when PR is very high, which we did not observe in our study. Because such high PR values are not observed in our data, our analysis cannot confirm or invalidate their finding. Together, our findings on the impacts of PR on PE and PE' demonstrate that observing a positive correlation between PR and evolvability in an empirical GPM requires no biological explanation. This said, the magnitude of the positive correlation is certainly impacted by some biological factors, as in the TF-DNA binding GPMs studied here.

Compared with phenotypes without large neutral networks, those with large neutral networks (but the same numbers of genotypes) have two apparent benefits. First, mutations are less likely to alter these phenotypes qualitatively. Second, they are more selectable, meaning that mutations could lead to quantitatively fitter but qualitatively unchanged phenotypes. One drawback is that they have a reduced evolvability. Nevertheless, it is clear by comparing the mouse (or yeast) TF-DNA binding GPM with its randomly rewired GPM that the PE and PE' reduction in the empirical GPM is moderate while the PR increase is substantial (figs. 2 and 3 and supplementary fig. S2, Supplementary Material online).

Kitano contended that there are architectural requirements for complex systems to be evolvable and that such requirements also give rise to robustness (Kitano 2004). If his "evolvable" meant "selectable," our results strongly support his hypothesis, because having a large neutral network given the number of genotypes is necessary for a phenotype to be selectable and is also the reason behind its high robustness. If his "evolvable" is in the sense of PE or PE', our findings refute his hypothesis, because the architecture that confers high evolvability—a lack of neutral networks (given the number of genotypes)—reduces robustness.

In the case of TF-DNA binding GPMs, large neutral networks arise naturally from the biophysics of TF-DNA binding. It seems likely that, in other systems such as RNA secondary structures or protein structures, large neutral networks can also result from physical and/or chemical properties of the systems. If this conjecture proves to be generally true, it would mean that simple physical and chemical laws not only permit the origin of life but also provide life with robustness and selectability while allowing reasonably high evolvability. This intriguing possibility is worth exploration in the future.

Materials and Methods

Expected PR and PE in a Random GPM

Let us consider a random GPM, where each node represents a genotype of l nucleotides and the GPM contains

$G = (4^l - 4^{0.5l})/2 + 4^{0.5l} = (4^l + 4^{0.5l})/2$ unique genotypes and K phenotypes. The above formula of G was derived by considering that each sequence is equivalent to its reverse complement and that there are $4^{0.5l}$ palindromic l-mers (when l is an even number) (van Helden et al. 1998). Because palindromic sequences constitute a tiny fraction ($< 0.5^l$) of all genotypes, we ignored their palindromic effects in the following modeling. As shown in the numerical examples (fig. 1 and supplementary fig. S1, Supplementary Material online), this approximation is acceptable. Let the number of unique binding sequences of TF$_i$ be g_i. With a single-nucleotide replacement, each genotype can change to one of $m = 3l$ other genotypes, which are collectively called the one-step neighborhood of the focal genotype. We assume that $1 \ll g_i \ll G$ for any i. The expected GR of a binding sequence of TF$_i$ is the expected number of other binding sequences of TF$_i$ that fall in the one-step neighborhood of the focal binding sequence, divided by m. Because the number of other binding sequences of TF$_i$ is $g_i - 1$ and the probability for any one of them to fall in the one-step neighborhood of the focal binding sequence is $m/(G - 1)$, the expected GR is $E[GR] = [(g_i - 1)m/(G - 1)]/m = (g_i - 1)/(G - 1) \approx g_i/G$. Because PR is the mean GR of all binding sequences of TF$_i$, the expected PR is $E[PR] = E[\text{mean GR}] = E[GR] \approx g_i/G$.

Now let us consider another TF (TF$_j$), which has g_j binding sequences. The probability that a particular binding sequence of TF$_i$ is in the one-step neighborhood of a particular binding sequence of TF$_j$ is approximately m/G. Hence, the probability that a particular binding sequence of TF$_i$ is in the neighborhood of any binding sequence of TF$_j$ (or more precisely the expected number of edges between a particular binding sequence of TF$_i$ and all binding sequences of TF$_j$) is approximately mg_j/G. The expected number of edges between all binding sequences of TF$_i$ and all binding sequences of TF$_j$ is approximately mg_ig_j/G. Because the number of edges between two phenotypes follows a binomial distribution (with g_ig_j trials each having a success rate of m/G), the probability that the phenotype of TF$_j$ binding is reachable from the phenotype of TF$_i$ binding by one mutation from at least one binding sequence of TF$_i$ equals $q_{ij} = 1 - (1 - m/G)^{g_ig_j} \approx 1 - e^{-\frac{mg_ig_j}{G}}$. Thus, PE$_i$, the fraction of all phenotypes reachable from the phenotype of TF$_i$ binding by one mutation, is expected to be $\sum_{j \neq i} q_{ij}/(K - 1) = \sum_{j \neq i} (1 - e^{-mg_jg_i/G})/(K - 1) = 1 - \sum_{j \neq i} e^{-mg_jg_i/G}/(K - 1)$. One can substitute g_i/G in the above formula by $E(PR_i)$ to obtain $E(PE_i) = 1 - \sum_{j \neq i} e^{-mg_jE(PR_i)}/(K - 1)$, which indicates that $E(PE)$ is an increasing function of $E(PR)$.

Microarray Data

The TF-DNA binding microarray data for mouse and yeast were downloaded from UniPROBE (http://the_brain.bwh.harvard.edu/uniprobe/downloads.php; last accessed December 12, 2017) (Newburger and Bulyk 2009). We defined binding sequences using the same data and enrichment score (E-score)

cutoff (0.35) as in Payne and Wagner (2014); this cutoff corresponds to a low false discovery rate (Payne and Wagner 2014).

PR and PE Calculation

We considered only single-nucleotide substitutions in computing PR and PE. This is slightly different from a previous study (Payne and Wagner 2014), in which insertions and deletions (indels) were also considered. While considering indels should in theory make the analysis better, Payne and Wagner (2014) assumed that indels are one nucleotide long and are restricted to the two ends of a binding sequence, which are unrealistic. Contemplating the complication of indels and the problem with the assumption, we decided not to consider indels. Note that our mathematical model has a variable m that measures the number of one-step neighbors per node that in theory takes into account all kinds of mutations. Hence, ignoring indels in the empirical analysis does not impact our mathematical analysis. Unlike the previous study (Payne and Wagner 2014), we considered all binding sequences of a TF rather than only those belonging to the largest neutral network (giant component). Because sequences that do not belong to the giant component can also bind to its TF and has potentials to evolve to a binding sequence of other TFs, including all binding sequences makes our analysis more complete. This change in methodology does not qualitatively affect the results on empirical (fig. 2A–C and supplementary fig. S2A–C, Supplementary Material online) or randomly rewired GPMs (supplementary fig. S4, Supplementary Material online). A binding sequence of TF$_i$ can be zero mutational steps away from a binding sequence of TF$_j$ if they share the same binding sequence.

Generation of Randomly Rewired GPMs

Given the g_i values of all TFs, we randomly picked genotypes from the 8-mer genotype space (with replacement) and assigned the genotypes to each TF. This was done with replacement, because both mouse and yeast GPMs contain genotypes that map to multiple phenotypes and because the sum of g_i exceeds G in both mouse and yeast. A genotype can map to multiple phenotypes but it cannot occur twice for the same phenotype.

PR, PE, and PR–PE Correlation in Random GPMs with Large Neutral Networks

The ensemble of all binding sequences of a TF is often represented by a position weight matrix (PWM), which shows the frequencies of A, T, G, and C at each nucleotide position of all binding sequences of the TF. Because potential epistasis is ignored in constructing PWMs from microarray-based TF-DNA binding data, when PWMs are used, all binding sequences of a TF are connected to form one large neutral network in the GPM. We downloaded PWMs for mouse and yeast

from UniPROBE (http://the_brain.bwh.harvard.edu/uniprobe/downloads.php; last accessed December 12, 2017) (Newburger and Bulyk 2009). For microarray data, we defined binding sequences using the same data and same enrichment score (E-score) cutoff (0.35) as previously used (Payne and Wagner 2014); this cutoff corresponds to a low false discovery rate (Payne and Wagner 2014). To convert PWMs back to binding sequences, we calculated the probability of each genotype for each TF, and used the cutoff of 0.0000469 in yeast and 0.00023885 in mouse to define binding sequences. Using these cutoffs led to similar total numbers of binding sequences as in the microarray data. We considered all binding sequences passing our cutoff to have equal binding affinities to the TF of concern.

We then constructed a random GPM with large neutral networks. Specifically, to remove the evolutionary relationships among the PWMs (and those among their corresponding TFs), we constructed a new set of PWMs by randomly shuffling all nucleotide positions among all existing PWMs of the species. We then used these scrambled PWMs to construct the GPM. In this GPM, large neutral networks are still present (albeit different from those in the empirical GPMs).

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

Literature Cited

Afek A, Schipper JL, Horton J, Gordân R, Lukatsky DB. 2014. Protein-DNA binding in the absence of specific base-pair recognition. Proc Natl Acad Sci USA. 111(48):17140–17145.

Aguilar-Rodríguez J, Payne JL, Wagner A. 2017. A thousand empirical adaptive landscapes and their navigability. Nat Ecol Evol. 1(2):0045.

Aldana M, Balleza E, Kauffman S, Resendiz O. 2007. Robustness and evolvability in genetic regulatory networks. J Theor Biol. 245(3):433–448.

Ancel LW, Fontana W. 2000. Plasticity, evolvability, and modularity in RNA. J Exp Zool. 288(3):242–283.

Carter AJ, Hermisson J, Hansen TF. 2005. The role of epistatic gene interactions in the response to selection and the evolution of evolvability. Theor Popul Biol. 68(3):179–196.

de Visser JA, et al. 2003. Perspective: evolution and detection of genetic robustness. Evolution. 57(9):1959–1972.

Draghi JA, Parsons TL, Wagner GP, Plotkin JB. 2010. Mutational robustness can facilitate adaptation. Nature. 463(7279):353–355.

Elena SF, Sanjuan R. 2008. The effect of genetic robustness on evolvability in digital organisms. BMC Evol Biol. 8:284.

Ferrada E, Wagner A. 2008. Protein robustness promotes evolutionary innovations on large evolutionary time-scales. Proc Biol Sci. 275(1643):1595–1602.

Gravner J, Pitman D, Gavrilets S. 2007. Percolation on fitness landscapes: effects of correlation, phenotype, and incompatibilities. J Theor Biol. 248(4):627–645.

Hayden EJ, Ferrada E, Wagner A. 2011. Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. Nature. 474(7349):92–95.

Ho WC, Zhang J. 2016. Adaptive genetic robustness of Escherichia coli metabolic fluxes. Mol Biol Evol. 33(5):1164–1176.

Jolma A, et al. 2013. DNA-binding specificities of human transcription factors. Cell. 152(1-2):327–339.

Kirschner M, Gerhart J. 1998. Evolvability. Proc Natl Acad Sci USA. 95(15):8420–8427.

Kitano H. 2004. Biological robustness. Nat Rev Genet. 5(11):826–837.

Masel J, Trotter MV. 2010. Robustness and evolvability. Trends Genet. 26(9):406–414.

McBride RC, Ogbunugafor CB, Turner PE. 2008. Robustness promotes evolvability of thermotolerance in an RNA virus. BMC Evol Biol. 8(1):231.

Nei M, Maruyama T, Chakraborty R. 1975. The bottleneck effect and genetic variability in populations. Evolution. 29(1):1–10.

Newburger DE, Bulyk ML. 2009. UniPROBE: an online database of protein binding microarray data on protein–DNA interactions. Nucleic Acids Res. 37(Database):D77–D82.

Payne JL, Wagner A. 2014. The robustness and evolvability of transcription factor binding sites. Science. 343(6173):875–877.

Phifer-Rixey M, et al. 2012. Adaptive evolution and effective population size in wild house mice. Mol Biol Evol. 29(10):2949–2955.

Schuster P, Fontana W, Stadler PF, Hofacker IL. 1994. From sequences to shapes and back: a case study in RNA secondary structures. Proc Biol Sci. 255(1344):279–284.

Stiffler MA, Hekstra DR, Ranganathan R. 2015. Evolvability as a function of purifying selection in TEM-1 beta-lactamase. Cell. 160(5):882–892.

Stormo GD, Fields DS. 1998. Specificity, free energy and information content in protein-DNA interactions. Trends Biochem Sci. 23(3):109–113.

Uchimura A, et al. 2015. Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice. Genome Res. 25(8):1125–1134.

van Helden J, Andre B, Collado-Vides J. 1998. Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. J Mol Biol. 281(5):827–842.

von Hippel PH, Berg OG. 1986. On the specificity of DNA-protein interactions. Proc Natl Acad Sci USA. 83(6):1608–1612.

Wagner A. 2005. Energy constraints on the evolution of gene expression. Mol Biol Evol. 22(6):1365–1374.

Wagner A. 2005b. Robustness and Evolvability in Living Systems. Princeton (NJ): Princeton University Press.

Wagner A. 2008. Robustness and evolvability: a paradox resolved. Proc Biol Sci. 275(1630):91–100.

Wagner A. 2012. The role of robustness in phenotypic adaptation and innovation. Proc Biol Sci. 279(1732):1249–1258.

Wagner GP, Altenberg L. 1996. Perspective: complex adaptations and the evolution of evolvability. Evolution. 50(3):967–976.

Yang JR, Ruan S, Zhang J. 2014. Determinative developmental cell lineages are robust to cell deaths. PLoS Genet. 10(7):e1004501.

Zhu YO, Siegal ML, Hall DW, Petrov DA. 2014. Precise estimates of mutation rate and spectrum in yeast. Proc Natl Acad Sci USA. 111(22):E2310–E2318.

Associate editor: Bill Martin