Impact of gene expression noise on organismal fitness and the efficacy of natural selection

Zhi Wang[1] and Jianzhi Zhang[2]

Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109

Edited by Masatoshi Nei, Pennsylvania State University, University Park, PA, and approved February 24, 2011 (received for review January 5, 2011)

Gene expression noise is a universal phenomenon across all life forms. Although beneficial under certain circumstances, expression noise is generally thought to be deleterious. However, neither the magnitude of the deleterious effect nor the primary mechanism of this effect is known. Here, we model the impact of expression noise on the fitness of unicellular organisms by considering the influence of suboptimal expressions of enzymes on the rate of biomass production and the energetic cost associated with imprecise amounts of protein synthesis. Our theoretical modeling and empirical analysis of yeast data show four findings. (*i*) Expression noise reduces the mean fitness of a cell by at least 25%, and this reduction cannot be substantially alleviated by gene overexpression. (*ii*) Higher sensitivity of fitness to the expression fluctuations of essential genes than nonessential genes creates stronger selection against noise in essential genes, resulting in a decrease in their noise. (*iii*) Reduction of expression noise by genome doubling offers a substantial fitness advantage to diploids over haploids, even in the absence of sex. (*iv*) Expression noise generates fitness variation among isogenic cells, which lowers the efficacy of natural selection similar to the effect of population shrinkage. Thus, expression noise renders organisms both less adapted and less adaptable. Because expression noise is only one of many manifestations of the stochasticity in cellular molecular processes, our results suggest a much more fundamental role of molecular stochasticity in evolution than is currently appreciated.

flux balance analysis | metabolic network

Many cellular processes are subject to substantial stochastic variation, because they depend on interactions among a small number of molecules. For example, the transcriptional initiation of a gene in a haploid cell relies on the binding of the transcriptional machinery to a single DNA molecule. The stochastic nature of this and other molecular interactions results in a large variation in the expression of the same gene by different isogenic cells under the same environment (1–3). How and to what extent such molecule-level stochasticity affects the well-being of organisms and their evolution is not well-understood. Recent experimental characterization of gene expression noise in multiple species (4–9) provides necessary empirical data for addressing these fundamental but wide-open questions.

In this work, gene expression noise refers to the stochastic variation in the protein expression level of a gene among isogenic cells in a homogenous environment (10). Expression noise arises from both intrinsic and extrinsic variations (2, 6, 8, 10, 11). Stochastic events in gene expression, including those in transcriptional initiation, mRNA degradation, translational initiation, and protein degradation, generate intrinsic noise (10). Differences among cells, either in local environment or the concentration or activity of any factor influencing gene expression, generate extrinsic noise (10). Extrinsic noise arising from variations in local environments will not be considered in this paper unless otherwise noted, because this fraction of expression noise is not caused by molecular stochasticity in gene expression.

Although the expression noise of a small number of genes may be important for cell fate determination or may be beneficial and selected under certain circumstances (2, 12, 13), there is accumulating evidence that gene expression noise is generally harmful to an organism. For example, increased expression noise is associated with disease (14, 15). There is also evidence for reduced expression noise of genes that are important to cell growth or sensitive to dosage (5, 16–18). Furthermore, various molecular mechanisms, regulatory network structures (e.g., negative feedbacks), and genomic organizations (e.g., operons) that could alleviate expression noise or its adverse fitness effect have been observed (2, 19–22). However, to what extent expression noise reduces cellular fitness is unknown. Furthermore, the relative contributions of various mechanisms to the fitness effect of expression noise are unclear.

As a first step to understanding the adverse fitness impact of gene expression noise, we consider unicellular organisms in which the Darwinian fitness can be approximated by the growth rate. Gene expression noise lowers the cellular growth rate through at least three main mechanisms. First, the expression levels of various enzymes in the metabolic network need to be optimized to ensure maximal biomass production; suboptimal enzyme concentrations owing to expression noise constrain metabolic fluxes and result in a decreased rate of biomass production. Second, because a cell devotes a substantial fraction of the total energy budget to transcription and translation (23), expression noise leads to extra and unnecessary protein production that wastes energy. Third, concentrations of multiple components of a system (e.g., stable protein complexes) often need to be balanced; expression noise breaks this balance and thus, may be detrimental (16, 18). In addition, expression noise may affect some non-biomass aspects of fitness that are generally hard to model (e.g., stress response, cell cycle, and mating). Note that the fluctuation time scale for global noise factors in protein production rate is similar to the observed cell-cycle length, indicating that the effects of expression noise do not average out for a cell (9, 10, 24).

Modeling the aforementioned first mechanism is most challenging because of the complicated nonlinear relationships among the concentrations of all enzymes, fluxes of all metabolic reactions, and rate of biomass production. We first analyze simple models of metabolic networks in which metabolic fluxes and fitness can be mathematically calculated from enzyme concentrations. Using insights gained from the analysis of the simple models, we study the impact of expression noise on biomass production using flux balance analysis (FBA) of the yeast metabolic network. The second of the three mechanisms is relatively simple to model as long as we know the fraction of the total energy budget devoted to protein production. We ignore the third mechanism mentioned above because of the lack of empirical knowledge about its prevalence and magnitude of impact. By ignoring this and other

Author contributions: Z.W. and J.Z. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

[1]Present address: Sage Bionetworks, Seattle, WA 98109.

[2]To whom correspondence should be addressed. E-mail: jianzhi@umich.edu.

See Author Summary on page 6345.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1100059108/-/DCSupplemental.

mechanisms, we provide a conservative estimate of the fitness impact of expression noise. We then use these estimates and analyses to address a number of biological questions, including amount of fitness loss caused by expression noise, cellular strategies to mitigate the harm of noise, strength of selection against noise, and adaptability of organisms in the presence of noise.

Results

Fitness Effects of Expression Noise in Linear Pathway Models. We first model the impact of gene expression noise on the maximal biomass output without considering the energetic cost of protein production. We start by analyzing models of small metabolic networks with simple topologies for which the metabolic flux of each reaction can be calculated according to the metabolic control theory (25, 26). In these models, we assume that a cell absorbs nutrients from its environment and converts them into biomass through the metabolic network. The fitness of the cell is equivalent to the rate of biomass production. Let us consider a cell in which the metabolic network contains only one linear pathway consisting of n biochemical reactions, each of which converts the metabolite produced by its immediate upstream reaction into the metabolite exclusively used by its immediate downstream reaction with the biomass produced by the final reaction (Fig. 1*A*). When a cell reaches the steady state, the fluxes of all reactions in the pathway are identical. The flux of the final reaction (i.e., the biomass production flux F) can be calculated by the formula (Eq. **1**)

$$F = \frac{a}{\sum_{i=1}^{n} \frac{M_i}{k_i K_{0,i-1} C_i}}, \qquad [1]$$

where a is a constant related to environmental parameters, C_i is the concentration of enzyme i (which catalyzes reaction i), k_i is the catalytic constant of enzyme i (commonly written as k_{cat}), M_i is the Michaelis constant of enzyme i (commonly written as K_M), and $K_{0,i-1}$ is the equilibrium constant between metabolite 0 (i.e., the nutrient) and metabolite $i-1$ (i.e., the substrate of reaction i; $K_{0,0} = 1$) (25). Let us first consider a simple situation in which $M_i/(k_i K_{0,i-1}) = b$ and $E[C_i] = c$ for all reactions, where $E[]$ refers to expectation. When there is no expression noise, $C_i = c$ and the biomass production flux $F_0 = ac/(bn)$. Gene expression noise is commonly measured by coefficient of variation CV_E, which is the SD of the expression level among isogenic cells divided by the mean expression level. In yeast, the median CV_E for enzyme genes is ~0.2 (*Materials and Methods*). The fitness of a cell with expression noise, relative to that without noise, is calculated by the ratio of F in the presence of expression noise to F_0. When the nutrient uptake rate is unlimited, F/F_0 for a cell may be higher or lower than one, depending on the actual enzyme concentrations. Assuming that expression noise is approximately normally distributed (5) and CV_E is identical among all reactions, our simulation shows that the distribution of F/F_0 is approximately normal (Fig. S1). The expectation of F/F_0, $E[F/F_0]$, can be derived analytically or estimated by numerical simulations (*SI Text* and Fig. S1). We found that $E[F/F_0]$ is a decreasing function of the pathway length n, decreasing from one when $n = 1$ to $1 - CV_E^2$ when n is infinity (Fig. S1*B*). For instance, when $n = 20$ and $CV_E = 0.2$, $E[F/F_0]$ is estimated to be 0.958 from numerical simulation and 0.963 from our analytical formula (Fig. S1*B*).

In nature, the nutrient uptake rate is capped by the environment and/or the cell. Under this constraint, the optimal concentration of each enzyme that maximizes F_0 is $c = bnF_0/a$ in the absence of expression noise. Because now F/F_0 cannot exceed one, even when all enzymes are overexpressed, the distribution of F/F_0 is truncated at one (Fig. 2*A*). We derived an approximate analytical formula for $E(F/F_0)$ under the nutrient uptake constraint and examined its accuracy numerically (*SI Text* and Fig. S1). In Fig. 2*A*, we present the numerical results obtained from simulating gene expression noise by randomly assigning C_i after a normal distribution, with the mean equal to c and the SD equal to $0.2c$ (i.e., $CV_E = 0.2$) for a linear pathway of $n = 20$ reactions. We repeated this process 10^4 times to obtain the distribution of the relative fitness among cells at any given moment (Fig. 2*A*). Because of the truncation of F/F_0 at one, the distribution of F/F_0 is heavily left-skewed with a mean of 0.952, which is close to that estimated by our analytical equation (0.959).

In the linear pathway model with relatively small CV_E and an unlimited rate of nutrient uptake, it can be shown mathematically that (Eq. **2**)

$$CV_F \approx CV_E/\sqrt{n}, \qquad [2]$$

where CV_F is the coefficient of variation of the biomass production flux (or fitness) and n is the number of reactions in the pathway (*SI Text* and Fig. S1). This approximation also applies when the nutrient uptake rate is constrained (*SI Text* and Fig. S1). For instance, the observed CV_F from our numerical examination with $CV_E = 0.2$ and $n = 20$ is 0.045 when the nutrient uptake is limited and 0.052 when it is unlimited, close to that calculated from our formula $(0.2/\sqrt{20} = 0.047)$. The reduction in flux noise compared with the expression noise is because of the buffering effect of linear pathways of multiple enzymes that diminishes the impact of expression fluctuations of individual enzymes (25).

In the above analyses, we used $M_i/(k_i K_{0,i-1}) = b$ and $E[C_i] = c$ for all reactions, which is equivalent to assuming that all enzymes in the pathway are the same. Analytical formulas for the expectation and CV of the relative fitness can also be obtained when the enzymes are different (*SI Text*). In all cases, substantial reduction of the expected fitness caused by expression noise is observed (Fig. S2).

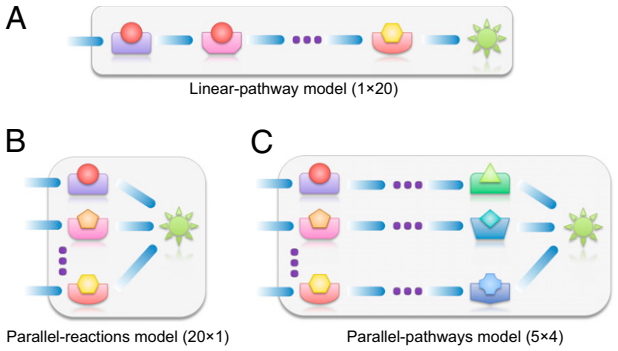

Fig. 1. Three simple models of metabolic networks, each with 20 biochemical reactions. (*A*) A linear pathway model in which the 20 reactions form a linear pathway to convert the absorbed nutrient into the biomass, which is at the right-hand end of the pathway. Corresponding to Eq. **1** in the text, the reaction entering the box from the left is labeled #1, and its substrate and product are labeled metabolites #0 and #1, respectively. (*B*) A parallel reactions model in which the 20 reactions transform the absorbed nutrients into 20 constituents that form the biomass. (*C*) A parallel pathways model that includes five parallel pathways, each with four reactions. It transforms the absorbed nutrients into five constituents that form the biomass.

Fitness Effects of Expression Noise in Parallel Reactions Models. Let us turn to another simple metabolic network that is composed of n parallel reactions (Fig. 1*B*). We assume that the n reactions convert nutrients to n metabolites that make up the biomass in equal amounts. The flux of reaction i is $F_i = C_i[S_i]k_i/M_i$ when $[S_i] << M_i$, where $[S_i]$ is the concentration of the substrate of reaction i (27). Because of the stoichiometry of the biomass constituents, the biomass production flux F equals the minimum of F_i of all reactions. For simplicity, let us assume $E(C_i) = c$ and

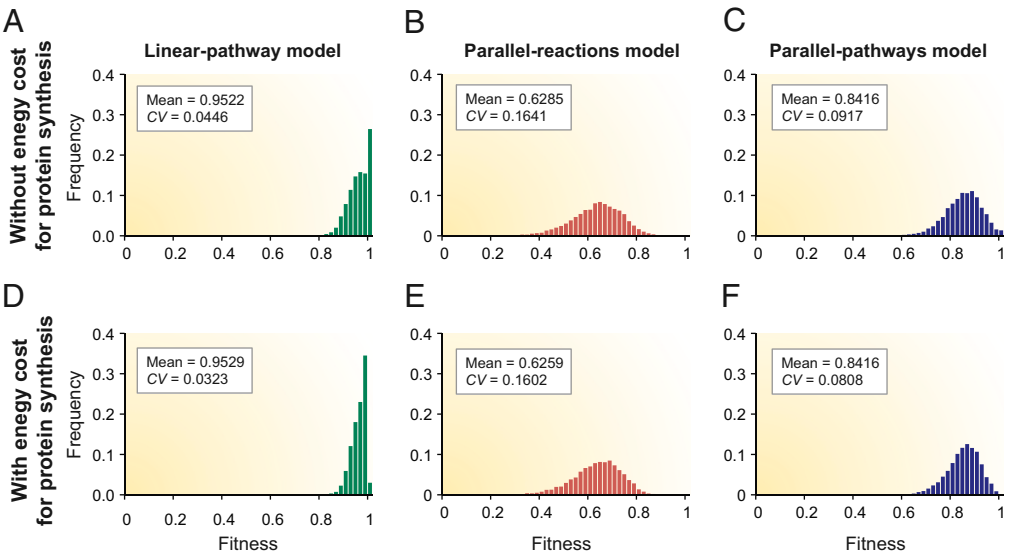

Fig. 2. Variation of the fitness in simple models of metabolic networks caused by gene expression noise when the nutrient uptake rate is constrained and the energy cost of protein synthesis is ignored (*A*–*C*) or considered (*D*–*F*). The fitness is relative to the optimal state with no expression noise. The three simple models depicted in Fig. 1 are used here. All enzymes in a network have the same properties and expected concentrations and have $CV_E = 0.2$. Each distribution is generated from 10^4 simulations of expression noise. The mean and *CV* of the fitness are presented in boxes.

$[S_i]k_i/M_i = d$ for all i. In the absence of expression noise, $F_0 = cd$. There is no analytical formula with good approximation for the mean and *CV* of F/F_0 when expression noise is present (*SI Text*), and therefore, we numerically examined these properties. Using the number of parallel reactions $n = 20$ and expression noise $CV_E = 0.2$ for all reactions, we simulated 10^4 cells. The resulting $E[F/F_0]$ is 0.629 (Fig. 2*B*), which is much lower than that in the linear pathway model (Fig. 2*A*). This is because, under the parallel reactions model, the stoichiometric relationships among biomass constituents require that the lowest flux of all parallel reactions determines the fitness of the cell (28). The frequency distribution of F/F_0 (Fig. 2*B*) is also flat compared with that in the linear pathway model (Fig. 2*A*), suggesting a greater variance of fitness. After examining a large parameter space by computer simulation, we found that CV_F is virtually independent of the number of parallel reactions when $n > 10$ and is only slightly lower than CV_E (Fig. S3). In other words, the buffering effect found in the linear pathway model is absent in the parallel reactions model. In the parallel reactions model, whether the nutrient uptake rate is constrained or not has virtually no effect on the expectation and *CV* of F/F_0 as long as $n > 3$ (*SI Text* and Fig. S3). The heterogeneity in catalytic activity and concentration among enzymes also has no effect on F/F_0 (*SI Text*). When the noise level varies among reactions, the expectation and *CV* of F/F_0 are determined largely by the noisiest reaction (*SI Text* and Fig. S4).

Fitness Effects of Expression Noise in Parallel Pathways Models. Because real metabolic networks contain both linear pathways and parallel reactions, we now consider a model that is a combination of these two models. Our combined model contains five parallel pathways, each of which is a linear pathway with four reactions that convert environmental nutrients into a biomass constituent (Fig. 1*C*). Under nutrient uptake constraint and a CV_E of 0.2 for each enzyme with identical kinetics and expected concentration, we simulated the metabolic network 10^4 times. We found that, in this model, the fitness reduction and fluctuation caused by expression noise are at an intermediate level between those of the linear pathway model and the parallel reactions model (Fig. 2*C*), a result that is expected given the combinatory nature of the present model. Thus, we expect that a more complicated metabolic network will have fitness reduction and fluctuation level bounded by those from the linear pathway and parallel reactions models of the identical size (i.e., total number of reactions). In other words, a complex metabolic network may be studied by transforming it to an equivalent network with a simpler structure. In addition, we found that, if the pathway length varies among the parallel pathways, the mean and *CV* of the fitness are determined primarily by the shortest pathway because of its largest flux variation (*SI Text* and Fig. S5).

Fitness Effects of Expression Noise on Yeast Metabolism. The above results from the analyses of simple models provide important insights that allow the analysis of cellular metabolic networks. Here, we analyze a yeast metabolic network resulting from a high-quality genome-wide reconstruction (29). This network is suitable for FBA (29–31), a powerful computational tool for connecting genotypes, phenotypes, and the environment in evolutionary studies (28, 32–38). Assuming a steady state of metabolism, one can maximize the biomass production rate or Darwinian fitness of the cell by balancing all fluxes simultaneously. In this way, FBA predicts the flux of each reaction as well as the biomass production rate. Extensive experimental validations have shown that FBA predictions are reliable (29, 39–45). The yeast metabolic network, consisting of 642 biochemical reactions after the removal of dead-end reactions that do not carry flux (*Materials and Methods*), produces the biomass that is composed of 43 constituents with fixed stoichiometric relationships. Expression noise has been measured in yeast for over 2,000 genes (5), permitting an empirical investigation of the impact of expression noise on yeast metabolism. There are 553 enzyme genes in the yeast metabolic network, and the majority (55.2%) has experimentally measured expression noise data. For the genes that do not have such data, we assign the median noise level of enzyme genes ($CV_E = 0.2046$) to each of them, which makes our estimates of fitness reduction more conservative, because the fitness reduction is primarily determined by the highest noise of all reactions (Figs. S2, S4, and S5). For the reactions that have not been associated with any known gene in the genome, we assume that they are autocatalyzing without fluctuation, which also makes our results conservative.

Although the level of enzyme expression noise (CV_E) has been quantified, we do not know the corresponding level of reaction flux noise. Nonetheless, the results obtained from the simple

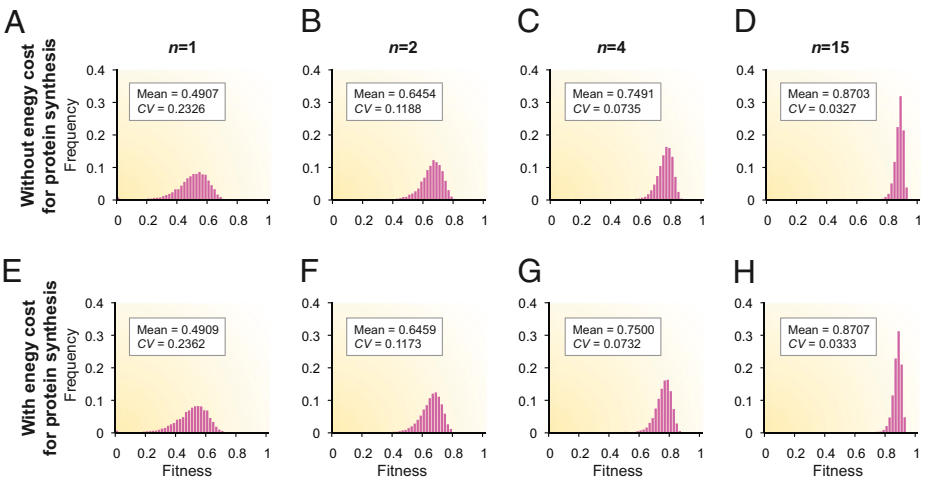

Fig. 3. Variation of the relative fitness in the yeast metabolic network caused by gene expression noise when the energy cost of protein synthesis is ignored (*A–D*) or considered (*E–H*). The fitness is relative to that under the optimal state with no expression noise, and each distribution is generated from 10^4 simulations of expression noise. The mean and *CV* of the relative fitness are presented in boxes. The parameter n can be considered as the length of the shortest pathway among the pathways for generating each of the biomass constituents. There is evidence that n must be ≤15 and is most likely ≤4 (in the text).

models may be applied, because the real network can be studied in the framework of the parallel pathways model. For a linear pathway, we have derived Eq. **2**, but the pathway length n is unknown here. Although we cannot separate distinct pathways in the yeast metabolic network because of the interconnection of all reactions and thus, cannot determine the length of each linear pathway, we may use a single n for all reactions, because the biomass fluctuation in a network of multiple parallel pathways is mainly determined by the shortest linear pathway, which tends to have the largest flux fluctuation (Fig. S5). Motivated by this idea, we simulated gene expression fluctuations in the yeast metabolic network and converted them to flux fluctuations using an assumed n value. More specifically, we first used FBA to calculate the optimal flux of each reaction (*OptF*) under the assumption of no expression noise. Because of the buffering effect of linear pathways, the flux of the reaction is constrained to $F \le OptF \cdot Y$, where Y is a random number drawn from a normal distribution with the mean equal to one and the SD equal to $CV_E/\sqrt{n}$. Here, CV_E is the expression noise level of the enzyme catalyzing the reaction. After obtaining the F value for each reaction, we used FBA to calculate the corresponding cellular fitness. We repeated this process 10^4 times and obtained the distribution of the fitness relative to that under no expression noise with different n values (Fig. 3 *A–D*). One can see that both the fitness reduction and fluctuation are smaller when n is larger, a consequence of larger buffering effects of longer linear pathways.

We found that the minimal mean fitness reduction because of expression noise is ∼0.13 (Fig. 3*D*); this occurs when we assume that the 642 reactions in the yeast metabolic network are evenly distributed to the productions of the 43 biomass constituents ($n = 642/43 \sim 15$). Because the pathway lengths for the 43 constituents vary and the shortest pathway largely determines the overall impact of expression noise on fitness, the mean fitness reduction is expected to be greater than 0.13. The mean fitness reduces to below 0.50 when $n = 1$ (Fig. 3*A*). The decrease of n also makes the *CV* of the cellular fitness much larger (Fig. 3 *A–D*). Thus, the smaller the n value, the lower the mean fitness and the greater the among-cell heterogeneity in fitness.

The gene expression noise data used here were measured in haploid cells (5). Recently, it was experimentally determined that the *CV* of fitness in genetically identical haploid yeast cells is 0.18 ± 0.02 (46) (Fig. 4*A*). It is currently unknown what fraction of the experimentally measured fitness noise is contributed by the fluctuation in metabolic rate caused by enzyme expression noise, because there may be other contributing factors (e.g., expression noise of cell-cycle regulators) (46). We calculated that, if at least 17% of the fitness variance is because of the metabolic rate noise caused by the enzyme expression noise, n cannot be greater than four. When $n = 4$, the mean fitness reduction caused by expression noise is 25% (Fig. 3*C*). In diploid cells, because both the mean and variance of the amount of expression of each gene are expected to be doubled because of the presence of two alleles, CV_E in diploid cells is reduced to $\sqrt{2}/2 = 71\%$ that in haploid cells. We found that, when $n = 4$, the reduction in gene expression noise to 71% increases the mean fitness by 8.8 ± 0.1% (Fig. 4*B*), which is not significantly different from the observed 6.9 ± 1.7% fitness increase from haploid to diploid yeasts (Fig. 4*A*); this suggests that the observed fitness increase from haploids to diploids may be largely owing to the benefit of expression noise reduction (46).

Fitness Effects of Noise Associated with the Protein Production Cost. Yeast uses a substantial fraction of its total energy for protein synthesis (23); more than 51% of the biomass is composed of RNAs and proteins, and 76% of the total cellular energy budget is devoted to protein synthesis (29). Based on these observations, we hypothesize that gene expression noise has additional fitness effects related to the energy expenditure in protein synthesis. On the one hand, energy is saved when expression noise causes a protein to be underexpressed. On the other hand, energy is wasted when a protein is overexpressed because of noise. We first study the fitness effect related to the energy expenditure in the simple models of metabolic networks introduced earlier. After considering the impact of expression noise on both the biomass production rate and the energy expenditure of protein production, we assume that the fitness of a cell with expression noise relative to that of an ideal cell without noise is given by (Eq. **3**)

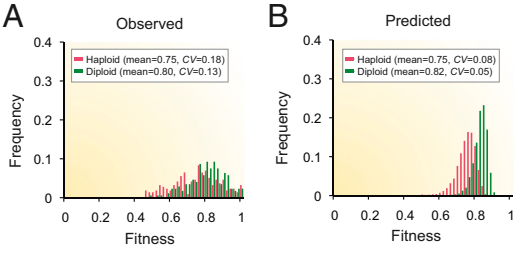

Fig. 4. Variations of observed (*A*) and predicted (*B*) fitness values of haploid and diploid yeast cells. The observed fitness values are obtained from ref. 46 and are rescaled so that the mean fitness of haploid cells is equal between the observed and predicted. Note that the fitness noise is contributed by multiple factors, including the metabolic rate noise caused by the enzyme expression noise (46). Hence, the observed fitness noise is greater than that predicted by considering the metabolic rate noise only.

$$f = F / F_0 - \sum_{i=1}^{n} [t_i (C_i/c_i - 1)], \quad [3]$$

where t_i is a weighting coefficient, n is the total number of enzymes in the metabolic network, F and F_0 are the biomass production fluxes of the cell with noise and without noise, respectively, and C_i and c_i are the actual and expected expression levels of enzyme i, respectively. To be conservative, we assume that the total energy cost for protein synthesis constitutes 30% of fitness, but our later analysis will show that the result is insensitive to this assumption. We further assume that the cost of synthesizing an enzyme molecule is the same for all enzymes in the simple models, and thus, $t_i = 0.3/n$ for all i. We simulated gene expression noise and calculated f for the three simple models considered earlier. We found that, for all three models, the distribution of f and its mean and CV are virtually unaffected by considering the energy expenditure of protein synthesis (Fig. 2 *D–F*).

We similarly used computer simulations to evaluate the effect of expression noise on yeast fitness by considering both the biomass production rate and the energy expenditure in the synthesis of each enzyme. The relative fitness of a yeast cell is determined by Eq. **3**, with the t_i values for each enzyme taken from previous estimates (23, 47). Similar to what was found in the simple models, the relative fitness of yeast is not affected by considering the energy expenditure (Fig. 3 *E–H*). These results suggest that noise-induced reduction in biomass production impacts fitness much more than noise-induced increase in the energy expenditure of protein synthesis.

The above finding is not unexpected and can be explained as follows. Although the fraction of total energy spent on protein synthesis in a cell is substantial, the net change in energy cost caused by expression noise is small, because gene expression fluctuations of multiple genes are uncorrelated. If we assume that all genes have the same energy cost and same noise level, the total contribution of energy cost to the CV of fitness is $h\frac{CV_E}{\sqrt{n}}$, where h is the total energy cost for all enzymes in the metabolic network and CV_E is the expression noise of each enzyme. For the yeast metabolic network (553 enzyme genes), h is estimated to be 0.13 (23), and the average CV_E is 0.20 in rich media. Thus, the CV of the relative fitness contributed by the energy cost variation is 1.1×10^{-3}, much smaller than that contributed by biomass production variation, which is on the order of 10^{-2} to 10^{-1} (Fig. 3).

Reduction of Expression Noise by Natural Selection. Natural selection can detect fitness differentials greater than the inverse of the effective population size, which is estimated to be 10^7 to 10^8 for yeast (23). Because the reduction of cellular fitness by expression noise is several orders of magnitude greater than 10^{-7} (Fig. 3) and because the level of gene expression noise is genetically determined, mutations that reduce the level of expression noise can be selectively favored (3, 5, 48, 49). Furthermore, a single noise-reduction mutation may have a fitness effect that is sufficiently large to be detected by natural selection, because it has been shown that a point mutation in the yeast *GAL1* promoter affects the noise level by greater than threefold (48). The selective strength against the expression noise of a gene is presumably determined by the sensitivity of the cellular fitness to the expression variation of the gene. In the yeast metabolic network, we can measure the fitness sensitivity to the expression noise of an enzyme by (Eq. **4**)

$$z = \frac{\mathrm{d(fitness)}}{\mathrm{fitness}} \Big/ \frac{\mathrm{d(expression)}}{\mathrm{expression}}, \quad [4]$$

where d(expression) is the reduction in enzyme expression level and d(fitness) is the consequent reduction in fitness. Because $\frac{\mathrm{d(flux)}}{\mathrm{flux}} \approx \frac{\mathrm{d(expression)}}{\mathrm{expression}}/\sqrt{n}$, we have (Eq. **5**)

$$z' = \frac{\mathrm{d(fitness)}}{\mathrm{fitness}} \Big/ \frac{\mathrm{d(flux)}}{\mathrm{flux}} = z\sqrt{n}, \quad [5]$$

where z' is the fitness sensitivity to the flux noise of a reaction and n is the equivalent length of the linear pathway that includes the reaction. Because n varies among reactions and is generally unknown, we can use FBA to estimate z' but not z. To estimate z', we constrain the flux of a reaction by a small amount (e.g., 5% of the optimal flux) and calculate the fitness reduction by FBA. We predict that a reaction with a greater z' is subject to a stronger selection against its flux noise. Consequently, it should have a lower flux noise, which may be reflected by a lower expression noise. Indeed, we observed a significant negative correlation between z' and the level of enzyme expression noise (CV_E; $r = -0.13$, $P = 0.004$) (Fig. 5*A*). This observation also verifies our FBA model with expression noise.

It has been reported that essential genes, which cause organismal death or infertility when deleted, have lower expression noise than nonessential genes in yeast (18). We also confirmed the negative correlation between the enzyme expression noise (CV_E) and the fitness effect of deleting the enzyme gene ($r = -0.11$, $P = 0.02$) (Fig. 5*A*), both of which were measured in the same rich media (5, 50). Although such a correlation is commonly interpreted as showing lower noise for more important genes (16, 18), this explanation is not self-evident, because the importance of a gene is measured on gene deletion, which is a much larger expression alteration than stochastic noise. Interestingly, we found a very strong positive correlation between the fitness effect of deleting a gene and the z' associated with the gene ($r = 0.86$, $P < 10^{-10}$) (Fig. 5*A*). This correlation can be explained by a simple model in which the fitness is a power function of the gene expression level, although the power function is not necessary to explain the correlation (Fig. 5*B*). Thus, it is likely that the anticorrelation between enzyme gene importance and the level of expression noise is because of their covariation with the fitness sensitivity to flux fluctuation (Fig. 5*A*). That is, fitness sensitivity to flux fluctuation is higher for more important enzyme genes, creating stronger selection against their expression noise, which results in noise reduction through evolution. Indeed, the partial correlation between the expression noise level and the fitness effect of gene deletion disappears after the control of the fitness sensitivity to flux noise ($r = 0.02$, $P = 0.70$).

Mitigation of the Harm of Expression Noise by Gene Overexpression. Presumably, gene expression noise has been reduced by natural selection to the extent that a further reduction imposes a higher fitness cost than benefit, possibly because of the energy and other resources required for noise reduction. An alternative strategy that a cell can take to alleviate the adverse fitness effect of expression noise is to increase the mean gene expression levels such that the cell can still maintain its optimal functional state even in the presence of expression fluctuation. However, as the mean expression levels rise, the energy cost of synthesizing the proteins will also rise. Based on this idea, we examine whether the harm of expression noise can be substantially alleviated by gene overexpression.

In all previous calculations, we assumed that the mean expression level of an enzyme in the presence of noise equals the expression level required by the maximal metabolic rate in the absence of noise. We now assume that the mean expression level in the presence of noise has been raised to mitigate the harm of expression noise and that the observed CV_E is the noise level in the excess of mean expression. We first implement gene overexpression in the simple models of metabolic networks. Under the assumption that all enzymes have the same catalytic parameters, expected concentrations, and noise levels ($CV_E = 0.2$), the optimal

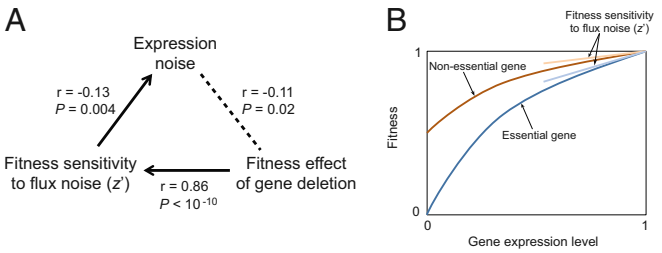

Fig. 5. Reduction of gene expression noise in the yeast metabolic network by natural selection. (*A*) Correlations among gene expression noise, fitness sensitivity to flux noise, and fitness effect of gene deletion. The arrows indicate proposed causal relations, whereas the dotted line indicates the previously reported correlation that lacks a direct causal relation. (*B*) A conceptual model explaining why the fitness sensitivity to expression (or flux) noise is greater for essential genes (or reactions) than for nonessential genes (or reactions). We assume that the fitness function of the expression level x is $g_E(x) = x^a$ for an essential gene and $g_N(x) = (bx + 1 - b)^a$ for a nonessential gene when x is no greater than the optimal level, which is arbitrarily set at one. Here, $a > 0$ and $0 < b = 1 - [g_N(0)]^{1/a} < 1$, where $g_N(0)$ is the fitness of the cell on deletion of the nonessential gene. Thus, the fitness sensitivity to the expression noise of an essential gene is the derivative of $g_E(x)$ at $x = 1$, which equals a. The fitness sensitivity to the expression noise of a nonessential gene is the derivative of $g_N(x)$ at $x = 1$, which equals ab. Because $a > ab$, fitness sensitivity to the expression noise of essential genes is greater than that to the expression noise of nonessential genes. The same model applies to the comparison among nonessential genes with different fitness effects of gene deletion.

strategy is to overexpress all genes by the same fraction. We numerically examined the cellular fitness when using different mean expression levels. The results from examining 10^4 cells per condition showed that the relationship between the relative fitness and gene expression level is not monotonic and that a fitness peak exists when the level of gene overexpression is intermediate (Fig. 6). This is understandable, because when the overexpression level is low, the availability of additional enzyme molecules relieves the constraint in biomass production from the stochastic shortage of the enzymes. However, when the overexpression level is high enough, additional overexpression confers little benefit but high synthesis cost. We found that the maximal increase of relative fitness by overexpression is largest for the parallel reactions model (from 0.63 to 0.85) (Fig. 5*B*), and this increase also requires the highest amount of overexpression (by 40%) (Fig. 5*B*). By contrast, in the linear pathway model, the maximal fitness increase is from 0.95 to 0.98, which requires 10% overexpression. As expected, the result from the parallel pathways model is intermediate between the other two models.

We next examined if gene overexpression can substantially alleviate the damage of expression noise on yeast metabolism. As in the analysis of the simple models, we assume that the current mean expression level in yeast has been raised to mitigate the harm of expression noise and that the observed CV_E is the noise level in the excess of mean expression. Because each yeast enzyme has its own enzyme activity, expression level, expression noise level, and fitness sensitivity to fluctuation, the optimal level of overexpression varies among enzymes. We here explore three different heuristic optimization strategies, because the global optimization is almost impossible because of the large parameter space and the large number of simulation replications required for evaluating each point in the parameter space. The first strategy is to overexpress all enzymes by the same fraction. In the second strategy, the degree of overexpression of an enzyme is weighted by the fitness sensitivity (z') to the flux fluctuation of the reaction catalyzed by the enzyme (i.e., more sensitive genes have higher overexpressions). In the third strategy, we allow different enzymes to have different levels of overexpression that are unconstrained by z' (*Materials and Methods*). We also tried several other heuristic search strategies, but the observed fitness peaks were either similar to or lower than those observed from the above three strategies. In all three strategies used, we assumed $n = 4$ when we converted the expression noise to flux noise. We found that the mean fitness increases from 0.750 to a maximum of 0.831, 0.844, and 0.842, respectively, in the three overexpression strategies used (Fig. 7). Interestingly, to achieve this moderate amount of fitness increase (~12%), the required level of gene overexpression is ~90%. Thus, (*i*) overexpression only moderately alleviates the fitness reduction caused by expression noise, and (*ii*) a high level of overexpression, which may cause other harm (*Discussion*) in addition to wasting energy, is required for the alleviation.

Biological Noise Weakens the Efficacy of Natural Selection. In previous sections, we showed that expression noise substantially reduces cellular fitness, because it causes enzyme shortage, which constrains metabolic fluxes. Because expression noise and other types of biological noise (e.g., stochastic variation in cell division timing or progeny size) generate a fitness variation among isogenic individuals in a population, in this section, we examine how such fitness noise influences the efficacy of natural selection. Here, we consider biological noise arising from both intrinsic and extrinsic sources, including variation in local environment.

Let us consider two alleles, A_1 and A_2, with frequencies of p and q, respectively, in a haploid population with an effective population size N. When diploid populations without dominance are considered, we can simply modify all formulas by substituting N with $2N$. In classical population genetics (51), where no fitness noise is considered, if A_1 and A_2 are equally fit, changes in allele frequencies over generations occur by stochastic sampling or genetic drift. This stochastic sampling is a binomial sampling process, and the SD of the frequency of A_1 in the next generation is $SD_{drift}(p') = \sqrt{\frac{pq}{N}}$. If changes in allele frequencies are caused

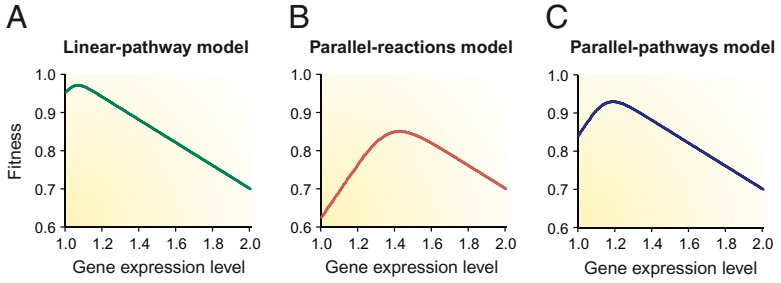

Fig. 6. Gene overexpression moderately lessens the fitness cost of expression noise in the linear pathway (*A*), parallel reactions (*B*), and parallel pathways (*C*) models. The mean gene expression level and fitness shown here are both relative to those in the optimal state with no expression noise.

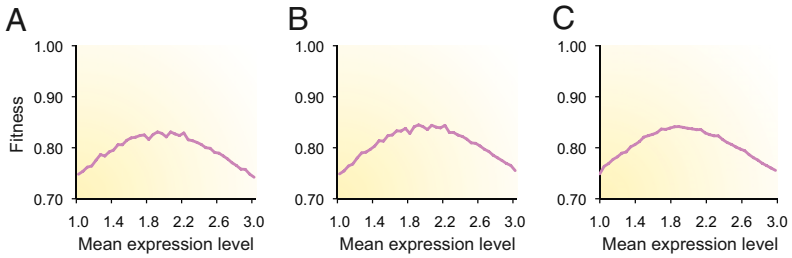

Fig. 7. Gene overexpression moderately lessens the fitness cost of expression noise in the yeast metabolic network, which is shown by a heuristic optimization with equal fractions of overexpression for all genes (*A*), a heuristic optimization with the degree of overexpression proportional to the fitness sensitivity to flux fluctuation (*B*), and a dynamic heuristic optimization (*C*; *Materials and Methods*). The mean gene expression level and fitness shown here are both relative to those in the optimal state with no expression noise.

only by a stochastic fitness variation among isogenic individuals, the SD of the frequency of A_1 in the next generation is $SD_{noise}(p') = e\sqrt{\frac{pq}{N}}$, where e is the level of fitness noise measured by *CV* (*Materials and Methods*). When both fitness noise and genetic drift are considered, the SD of the frequency of A_1 in the next generation is $SD_{total}(p') = \sqrt{\frac{pq}{N/(1+e^2)}}$. Thus, the impact of fitness noise e on the stochastic variation of allele frequencies is equivalent to reducing the effective population size from N to $\frac{N}{1+e^2}$. From this result, we predict that fitness noise weakens the efficacy of natural selection. Fig. 8 shows the fixation probabilities of mutations in the presence of various levels of fitness noise relative to those in the absence of fitness noise (*Materials and Methods*). If a mutation is neutral, its fixation probability is unaffected by the presence/absence of the fitness noise (Fig. 8, blue lines). The fixation probability increases with the level of fitness noise for deleterious mutations (Fig. 8, red lines) but decreases for beneficial mutations (Fig. 8, green lines). This change in fixation probability is much greater when selection is stronger. For example, when the fitness noise rises from $e = 0$ to $e = 1$, the fixation probability increases by 86% for a deleterious mutation with $Ns = -1$ (Fig. 8*A*) but by 1.1×10^4-fold for a deleterious mutation with $Ns = -10$ (Fig. 8*B*), where s is the selection coefficient. By contrast, when the fitness noise rises from $e = 0$ to $e = 1$, the fixation probability decreases by 32% for a beneficial mutation with $Ns = 1$ (Fig. 8*A*) and by 50% for a beneficial mutation with $Ns = 10$ (Fig. 8*B*). Fitness noise also affects an allele's time to fixation just like population shrinkage.

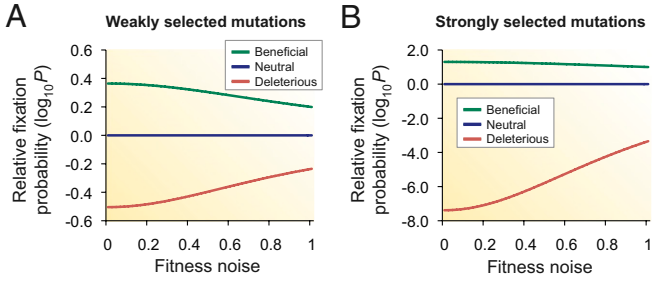

Fig. 8. Fitness noise reduces the efficacy of natural selection. The fixation probability of a newly arisen mutation, relative to that of a newly arisen neutral mutation, is presented (in log_{10} scale) as a function of fitness noise measured by *CV*. Note the difference in the *y* axis scale between the two panels. We use $Ns = 1$ or -1 for weakly selected mutations (*A*) and $Ns = 10$ or -10 for strongly selected mutations (*B*).

Discussion

In this study, we investigated the impact of gene expression noise on the fitness of unicellular organisms by considering the influence of suboptimal expressions of enzymes on the rate of biomass production and the energetic cost associated with imprecise amounts of protein synthesis. Our results suggest that expression noise reduces the yeast fitness by at least 25% and that this damage cannot be substantially alleviated by other means. Our analyses were based on a number of simplifying assumptions that are worth discussion. First, we converted enzyme gene expression noise to flux noise by Eq. **2**, a formula derived on the basis of homogeneous enzymes with small CV_E (*SI Text*). Numerical studies showed that this formula is accurate when $CV_E < 0.25$ but underestimates CV_F when $CV_E > 0.25$ (Fig. S1). Furthermore, the formula underestimates CV_F for heterogeneous enzymes (*SI Text*), making our estimate of the fitness reduction overall conservative. Second, different reactions in the yeast metabolic network have different n, but we used Eq. **2** with the same n to transform CV_E to CV_F for all reactions. This simplification, however, is unlikely to have a big impact. This is because it is the shortest pathway that determines the biomass production rate; as long as n is correct for the shortest pathway, the result should be reliable. Third, our analysis of yeast metabolism relied on FBA, which may contain errors. Nevertheless, FBA predictions have been extensively verified by experiments (29, 39–45). For example, FBA predicts yeast essential and nonessential genes with an accuracy of nearly 90% (33). With regard to quantitative fitness effects of gene deletion, FBA predictions are also well-correlated with experimental measures (Pearson's $r = 0.562$, $P < 10^{-41}$) (32). Our recent study further showed the power of FBA in predicting epistasis among metabolic reactions (28). Most importantly, our FBA results regarding yeast are qualitatively similar to the results obtained from the simple models without the use of FBA. Thus, it is unlikely that our yeast results have been substantially affected by potential errors in FBA. Fourth, we used a set of t_i values for the energy cost of gene expression that was previously estimated (23). This set of estimates may not be accurate. However, because the energy cost associated with expression noise does not affect the fitness much, our results are insensitive to potential errors in t_i. Fifth, in modeling the fitness effect of expression noise, we ignored potential effects through nonmetabolic pathways such as signal transduction and regulatory networks. For instance, it is possible that underexpression of an enzyme gene induces the up-regulation of another enzyme gene. Such effects, equivalent to correlated expression variation among genes, may alleviate the harm of expression noise to some extent. However, the alleviation, much like that from gene overexpression, is expected to be mild (Fig. 7). Furthermore, feedbacks only have limited power in reducing stochastic noise (52).

Biological systems such as cells are generally believed to be robust against genetic perturbations, supported by the observation that even deleting a gene frequently results in little to no fitness

effect in yeast (33, 38, 50). This apparent genetic robustness against gene deletion is not contradictory to our finding, because the disturbance caused by expression noise occurs to every gene rather than to only one gene. Although the impact of noise on the protein concentration of each gene is relatively small, the total effect of expression noise on the entire metabolic system can be much larger than deleting one gene because of the nonlinear relationship between enzyme concentrations and the biomass production rate.

Because of the large fitness reduction imposed by expression noise, minimizing expression noise or the impact of expression noise must be highly beneficial and important in evolution (with the exception of a minority of genes whose elevated noise is apparently beneficial) (13). Indeed, we observed a significant anticorrelation between the sensitivity of fitness to the flux fluctuation of a reaction and the expression noise of the enzyme catalyzing the reaction, which likely results from stronger selection against the flux noise of sensitive reactions. Our estimates of z' also provide potential targets for effective engineering of yeast cells to minimize the negative fitness impact of expression noise. Our calculation showed that, compared with a haploid cell, a diploid cell has a reduced CV_E for every gene, leading to an increase of the expected metabolic rate by 8.8%. Thus, even in the absence of sexual reproduction, diploidy outperforms haploidy by a large fitness margin. Our prediction of the 8.8% fitness increase from haploids to diploids by the reduction in expression noise is not significantly different from the experimentally observed fitness difference between them ($6.9 \pm 1.7\%$), suggesting that the reduction in expression noise can explain, in large part or in total, the fitness difference between haploid and diploid yeast. Nonetheless, we note that the fitness of tetraploid yeast is 6% lower than diploid yeast (46), although tetraploids have even lower expression noise than diploids. The fitness reduction in tetraploids could be because of the fact that yeast is maladapted for the tetraploid life (e.g., in surface/volume ratio). Because yeast is normally diploid in nature (53), it would be important to study whether the fitness difference between haploid and diploid yeasts is because of the adaptation of yeast to a diploid life or the fitness effect of noise reduction by diploidy.

We observed lower expression noises of genes associated with reactions that have higher fitness sensitivities to flux fluctuations, suggesting that expression noise has been reduced by natural selection. Mechanistically, noise reduction may be achieved by an increase in transcriptional rate and a decrease in translational rate when the protein concentration is unaltered (16). It may also be achieved by using certain genetic control circuits, such as negative feedback loops (54, 55). However, noise reduction does not imply a complete elimination of noise, because the cost associated with noise reduction may exceed its benefit. Based on control and information theory, a recent study showed that the minimum level of expression noise decreases with the quartic root of the number of signaling events, making it extremely expensive to reduce noise (52). We showed that gene overexpression can partially alleviate the detrimental effect of expression noise. However, we considered only the energy cost of overexpression but not other potential costs of gene overexpression. Gene overexpression can have at least two additional costs. First, it may cause unnatural protein–protein interactions that are deleterious (56). Second, it increases the number of misfolded proteins from both mistranslated and correctly translated molecules, which imposes a fitness cost (57, 58). Thus, gene overexpression is less effective than what was shown in Fig. 7 in minimizing the adverse effect of expression noise. If gene overexpression is indeed part of the cellular strategy to lessen the harm of expression noise, the fitness loss owing to expression noise under the assumption of no elevation of mean expression (Fig. 3) must have been underestimated, because CV_E under no mean expression elevation would have been greater than what is currently observed. In any case, even with an increase of the mean expression level, there is still a substantial loss of fitness (>15%) caused by gene expression noise.

We found that the efficacy of natural selection is reduced in the presence of fitness noise, which is generated by expression noise and other types of biological noise. This reduction in selection is equivalent to a reduction in effective population size from N to $\frac{N}{1+e^2}$. Because the experimentally measured fitness noise in yeast is $CV \sim 0.2$ (46), on average, two isogenic cells differ in their fitness by 0.23. Our analysis indicates that this large variation in individual fitness is equivalent to only 4% (0.2^2) of the power of genetic drift in causing random changes of allele frequencies. This is because, although the fitness variation among cells is large, the mean fitness of an allele has a small variation, and it is the mean fitness of an allele that determines the change of its frequency over generations (*Materials and Methods*). Nonetheless, fitness noise can be greater than $CV = 1$ in certain species, such as *Drosophila* flies (59). In these organisms, the efficacy of natural selection is at least halved by the fitness noise.

Taken together, our study showed that expression noise renders unicellular organisms both less adapted and less adaptable. Because we did not consider all harms that expression noise causes (e.g., dosage imbalance) and were conservative in our calculations, the actual fitness reduction owing to expression noise is likely much greater. Because expression noise is only one of many manifestations of the high stochasticity in cellular molecular processes, our results imply a much more fundamental role of molecular randomness in evolution than is currently appreciated (60). The facts that many fundamental cellular processes involve only a few molecules and that the cellular concentrations of many molecules are extremely low (≤1 per cell) determine that unicellular systems are inevitably imperfect and hard to perfect. Cell to cell variation in gene expression seems greater in multicellular organisms than in yeast (1), but whether the adverse fitness effect of noise is lowered in multicellular organisms because of the existence of many cells of each tissue type remains an open question. It is currently difficult to model this fitness effect in multicellular organisms, because the biomass production rate may not be the primary fitness determinant in these organisms. Nevertheless, multicellular organisms may have other aspects of their physiology that are more susceptible to expression noise (e.g., intercellular communication). Regardless, like genetic drift (61, 62), molecular stochasticity must be included in a comprehensive evolutionary theory that potentially explains the origins of all biological observations.

Materials and Methods

Yeast Gene Expression Noise Data. Our gene expression noise data were from Newman et al. (5), who measured the protein concentrations of 2,213 GFP-tagged genes in individual *Saccharomyces cerevisiae* cells by fluorescence-activated cell sorting. Newman et al. (5) controlled cell sizes in their measurement and thus, largely excluded the extrinsic noise caused by variations in local environment or cell-cycle stage.

Yeast Metabolic Network. The metabolic network model iND 750 of *S. cerevisiae* (29) was downloaded from the BiGG database (http://bigg.ucsd.edu) and parsed by the COBRA toolbox (63). The network is composed of 1,149 reactions associated with 750 known genes. Some reactions do not have associated genes, either because the reactions are spontaneous such that no enzymes are required or because the enzyme genes have yet to be identified. The metabolic network model also provides information about stoichiometry, direction of biochemical reactions, isoenzymes, and enzymatic protein complexes.

FBA. Details of FBA have been described in the literature (30, 31, 64). Briefly, FBA can be used to analyze a metabolic network at the steady state under the constraint of stoichiometry. The biomass reaction describes the relative contributions of metabolites to the cellular biomass. The steady-state flux distribution is determined by maximizing the rate of biomass production. The formulated linear programming problem is shown as maximize object

$Z = \sum c_i \cdot v_i$ subject to $S \cdot v = 0$ and $\alpha \le v \le \beta$. Here, Z is the biomass production rate, c is the biomass stoichiometric vector, S is the stoichiometric matrix, v is the metabolic flux vector, and vectors α and β represent the lower- and upper-bound constraints of the fluxes, respectively. We used the optimization package CLPEX (www.ilog.com) to solve the linear programming problem. We simulated the yeast metabolic network under the yeast extract peptone dextrose (YPD) condition using previously defined nutrient specifications (65). To delete a reaction, we constrained the flux of the reaction to zero and obtained the maximal biomass production. To constrain a reaction, we set the maximal allowable flux of the reaction according to the constraint. The fitness of a genetically perturbed strain, relative to the WT strain, is the maximal biomass production rate of the perturbed strain divided by that of WT.

We followed a published protocol (66) to identify dead-end reactions, which must have zero flux under steady state. These reactions are involved in the generation of metabolites that are neither included in biomass nor transported outside the cell, and they may reflect the incompleteness of the metabolic network models. They were identified by maximizing and minimizing in turn each flux under the condition that all nutrients are provided. If both the maximization and minimization result in zero flux, this reaction is considered a dead-end reaction. Because neither active transportation that requires ATP nor ionic transportation is modeled in FBA, these reactions were also not considered in our analysis. After the removal of all such reactions, the yeast metabolic network used in our analysis contains 642 reactions with 553 associated enzyme genes.

Expression Noise in the Metabolic Network. We used experimentally measured gene expression noise data (5) and annotated gene reaction associations (29) to calculate the expression noise of the enzymes catalyzing metabolic reactions. For genes that are annotated to catalyze reactions but have no experimentally measured expression noise, we assume that the noise level is $CV_E = 0.2046$, which is the median experimentally measured expression noise for enzyme genes in yeast. This is a conservative treatment, because lowly expressed genes, which have high CV_E, tend to be missing in the experimental noise data. For reactions that are not associated with any known yeast gene, we also conservatively assume $CV_E = 0$.

The gene reaction association is annotated as Boolean logic relationships. For example, a reaction is catalyzed by enzyme [(A and B) or C], meaning that A and B form a protein complex that can catalyze the reaction, whereas isoenzyme C can also catalyze the reaction independently. This Boolean logic can be recursively applied for more complicated gene reaction relationships. Thus, the reaction expression noise is also calculated recursively. Because one enzyme may catalyze multiple reactions and the stoichiometry of multiple components of a protein complex is usually unknown, we assume that all of the Boolean logic components have 1:1 ratio in protein concentration. For a protein complex (A and B and C . . .), the noise is determined by the protein with the largest noise because of the stoichiometric requirement for protein complexes. Because the stochastic gene expression variation is assumed to be independent among genes, the total noise for n isoenzymes (A or B or C . . .) that catalyze the same reaction is calculated by $CV_{isoenzymes} = \sqrt{\frac{CV_A^2 + CV_B^2 + CV_C^2 + \ldots}{n^2}}$ under the assumption that the expected activities (catalytic activity/mol × expression level) of the isoenzymes are similar. To simulate the expression noise of a reaction, we generate random numbers from a normal distribution with the mean equal to zero and SD equal to CV_E, and then, we apply the Boolean logic to calculate the reaction expression noise. The reaction expression noise is then converted to the reaction flux noise using Eq. **2**.

Energy Cost of Yeast Gene Expression. The energy cost of yeast gene expression under the YPD condition has been estimated recently by considering a number of factors, including gene expression level, amino acid composition, and protein degradation (23, 47). We downloaded this dataset from http://www.bioc.uzh.ch/wagner/data/Wagner_MolDevEvol_Vol308B_p322_2007.txt. The dataset included 2,097 genes, with their total energy cost (sum of t_i) equal to 0.34, which is relative to the total fitness of one. For the genes that are not estimated in the study, we used the median energy cost of all enzyme genes (7.38×10^{-5}).

Heuristic Optimization of Overexpression to Reduce the Harm of Expression Noise. We implemented three different heuristic algorithms to optimize gene overexpression such that the fitness is maximized in the presence of expression noise. In the first algorithm, we overexpress every enzyme gene by the same fraction. We set the expected expression level of each gene, relative to the optimal expression level determined by FBA, at r ($r > 1$). Thus, the actual flux of the reaction is constrained by $F \le OptF \cdot Y$, where Y is a random number drawn from a normal distribution with the mean equal to r and SD equal to $r \cdot CV_E/\sqrt{n}$. Here, CV_E is the observed expression noise level of the enzyme(s) catalyzing the reaction, and $n = 4$ is used. We then look for the r value that maximizes the expected fitness.

In the second algorithm, we optimize the gene overexpression according to the fitness sensitivity to flux noise (z'). The flux of reaction i is constrained by $F \le OptF \cdot Y$, where Y is a random number generated from the normal distribution with mean equal to $\frac{rz'_i}{\sum_{i=1}^{n} z'_i}$ and SD equal to $\frac{rz'_i}{\sum_{i=1}^{n} z'_i} \cdot CV_E/\sqrt{n}$. Again, we look for the r value that maximizes the expected fitness.

In the third algorithm, we optimize gene overexpression by a dynamic heuristic search. In this process, we first overexpress every gene by 5% and calculate the mean fitness. We then pick a gene and adjust its expression to the original level to see if the fitness is improved. If it does increase the fitness, the original expression is accepted. Otherwise, the overexpression is accepted. This examination for more fitness improvement is done for all of the enzyme genes in a random order. This entire process of a 5% expression increase for every gene followed by the examination of individual genes is repeated many times.

Changes in Allele Frequencies in the Presence of Fitness Noise. Let us consider a haploid population in which two equally fit alleles, A_1 and A_2, segregate at a locus. Let us assume that both alleles have an expected fitness of one and that the degree of fitness variation among individuals that is caused by biological noise is $CV = e$. Let f_1 be the mean fitness of A_1 individuals, and let f_2 be the mean fitness of A_2 individuals. Thus, the SDs of f_1 and f_2 are (Eq. **6**)

$$SD(f_1) = \frac{e}{\sqrt{Np}} \text{ and } SD(f_2) = \frac{e}{\sqrt{Nq}}, \quad [6]$$

respectively, where p and q are the allele frequencies of A_1 and A_2, respectively, and N is the effective population size. Because $E(f_1) = E(f_2) = 1$ and $\mathrm{Cov}(f_1, f_2) = 0$, we have (Eq. **7**)

$$SD(f_1/f_2) = \sqrt{SD^2(f_1) + SD^2(f_2)} = \frac{e}{\sqrt{N_e pq}}. \quad [7]$$

Thus, although $E(f_1) = E(f_2)$, f_1 may differ from f_2. Let $f = f_1/f_2$. Because of selection, the allele frequency for A_1 in the next generation becomes (Eq. **8**)

$$p' = \frac{pf_1}{pf_1 + qf_2} = \frac{pf_1/f_2}{pf_1/f_2 + q} = \frac{pf}{pf + q}. \quad [8]$$

Because f is a random variable, p' is also a random variable. The SD of p' owing to the variation of f or fitness noise is (Eq. **9**)

$$\begin{aligned} SD_{noise}(p') &= SD\left(\frac{pf}{pf+q}\right) \approx \frac{d\left(\frac{pf}{pf+q}\right)}{df} \cdot SD(f) \\ &= \left(\frac{p(pf+q) - p^2 f}{(pf+q)^2}\right) \cdot SD(f) \\ &= \frac{pq}{(pf+q)^2} \cdot \frac{e}{\sqrt{Npq}} \approx e \cdot \sqrt{\frac{pq}{N}} \end{aligned} \quad [9]$$

Because the variation in allele frequency in the next generation caused by genetic drift is (Eq. **10**)

$$SD_{drift}(p') = \sqrt{\frac{pq}{N}}, \quad [10]$$

the total variation caused by fitness noise and drift is (Eq. **11**)

$$\begin{aligned} SD_{total}(p') &= \sqrt{\left(e\sqrt{\frac{pq}{N}}\right)^2 + \left(\sqrt{\frac{pq}{N}}\right)^2} \\ &= \sqrt{1+e^2}\sqrt{\frac{pq}{N}} = \sqrt{1+e^2}\, SD_{drift}\, t(p') \end{aligned} \quad [11]$$

The above equations can also be applied to diploids when N is substituted with $2N$.

In a haploid population of size N, the fixation probability is $\mu = \frac{2s}{1 - e^{-2Ns}}$ for a newly arisen and weakly selected allele of selective coefficient s (62). We showed that, in the presence of fitness noise, the behavior of the allele is equivalent to that without fitness noise but in a smaller population with the

effective size of $\frac{N}{1+e^2}$. Therefore, we only need to substitute N with $\frac{N}{1+e^2}$ in the above formula to calculate fixation probabilities.

Our model of natural selection in the presence of fitness noise differs from earlier population genetic (67) and quantitative genetic (68) models. In these models, two alleles have different fitness variances, and selection on fitness variance is considered. By contrast, we consider two alleles with the same fitness variance, and therefore, fitness variance is not under selection.

ACKNOWLEDGMENTS. We thank Meg Bakewell, Chungoo Park, Wenfeng Qian, and Jian-Rong Yang for valuable comments on the manuscript. This study was supported by research grants from the US National Institutes of Health (to J.Z.).

1. Raj A, van Oudenaarden A (2008) Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell* 135:216–226.
2. Kaern M, Elston TC, Blake WJ, Collins JJ (2005) Stochasticity in gene expression: From theories to phenotypes. *Nat Rev Genet* 6:451–464.
3. Maheshri N, O'Shea EK (2007) Living with noisy genes: How cells function reliably with inherent variability in gene expression. *Annu Rev Biophys Biomol Struct* 36:413–434.
4. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. *Nat Genet* 31:69–73.
5. Newman JR, et al. (2006) Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* 441:840–846.
6. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297:1183–1186.
7. Ramsey S, et al. (2006) Transcriptional noise and cellular heterogeneity in mammalian macrophages. *Philos Trans R Soc Lond B Biol Sci* 361:495–506.
8. Raser JM, O'Shea EK (2004) Control of stochasticity in eukaryotic gene expression. *Science* 304:1811–1814.
9. Taniguchi Y, et al. (2010) Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 329:533–538.
10. Raser JM, O'Shea EK (2005) Noise in gene expression: Origins, consequences, and control. *Science* 309:2010–2013.
11. Volfson D, et al. (2006) Origins of extrinsic variability in eukaryotic gene expression. *Nature* 439:861–864.
12. Acar M, Mettetal JT, van Oudenaarden A (2008) Stochastic switching as a survival strategy in fluctuating environments. *Nat Genet* 40:471–475.
13. Zhang Z, Qian W, Zhang J (2009) Positive selection for elevated gene expression noise in yeast. *Mol Syst Biol* 5:299.
14. Bahar R, et al. (2006) Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature* 441:1011–1014.
15. Kemkemer R, Schrank S, Vogel W, Gruler H, Kaufmann D (2002) Increased noise as an effect of haploinsufficiency of the tumor-suppressor gene neurofibromatosis type 1 in vitro. *Proc Natl Acad Sci USA* 99:13783–13788.
16. Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB (2004) Noise minimization in eukaryotic gene expression. *PLoS Biol* 2:e137.
17. Batada NN, Hurst LD (2007) Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat Genet* 39:945–949.
18. Lehner B (2008) Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Mol Syst Biol* 4:170.
19. Becskei A, Serrano L (2000) Engineering stability in gene networks by autoregulation. *Nature* 405:590–593.
20. Pedraza JM, van Oudenaarden A (2005) Noise propagation in gene networks. *Science* 307:1965–1969.
21. Ramsey SA, et al. (2006) Dual feedback loops in the GAL regulon suppress cellular heterogeneity in yeast. *Nat Genet* 38:1082–1087.
22. Kovács K, Hurst LD, Papp B (2009) Stochasticity in protein levels drives colinearity of gene order in metabolic operons of Escherichia coli. *PLoS Biol* 7:e1000115.
23. Wagner A (2005) Energy constraints on the evolution of gene expression. *Mol Biol Evol* 22:1365–1374.
24. Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB (2005) Gene regulation at the single-cell level. *Science* 307:1962–1965.
25. Kacser H, Burns JA (1981) The molecular basis of dominance. *Genetics* 97:639–666.
26. Kacser H, Burns JA (1973) The control of flux. *Symp Soc Exp Biol* 27:65–104.
27. Berg JM, Tymoczko JL, Stryer L (2002) *Biochemistry* (Freeman, New York), 5th Ed.
28. 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.
29. Duarte NC, Herrgård MJ, Palsson BO (2004) Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res* 14:1298–1309.
30. Price ND, Reed JL, Palsson BO (2004) Genome-scale models of microbial cells: Evaluating the consequences of constraints. *Nat Rev Microbiol* 2:886–897.
31. Orth JD, Thiele I, Palsson BO (2010) What is flux balance analysis? *Nat Biotechnol* 28: 245–248.
32. Wang Z, Zhang J (2009) Why is the correlation between gene importance and gene evolutionary rate so weak? *PLoS Genet* 5:e1000329.
33. Papp B, Pál C, Hurst LD (2004) Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* 429:661–664.
34. Pál C, Papp B, Lercher MJ (2005) Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat Genet* 37:1372–1375.
35. Wagner A (2007) From bit to it: How a complex metabolic network transforms information into living matter. *BMC Syst Biol* 1:33.
36. Matias Rodrigues JF, Wagner A (2009) Evolutionary plasticity and innovations in complex metabolic reaction networks. *PLoS Comput Biol* 5:e1000613.
37. Pál C, et al. (2006) Chance and necessity in the evolution of minimal metabolic networks. *Nature* 440:667–670.
38. Wang Z, Zhang J (2009) Abundant indispensable redundancies in cellular metabolic networks. *Genome Biol Evol* 1:23–33.
39. Snitkin ES, et al. (2008) Model-driven analysis of experimentally determined growth phenotypes for 465 yeast gene deletion mutants under 16 different conditions. *Genome Biol* 9:R140.
40. Papp B, Pál C, Hurst LD (2003) Dosage sensitivity and the evolution of gene families in yeast. *Nature* 424:194–197.
41. 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.
42. 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.
43. Fong SS, Palsson BO (2004) Metabolic gene-deletion strains of Escherichia coli evolve to computationally predicted growth phenotypes. *Nat Genet* 36:1056–1058.
44. Famili I, Forster J, Nielsen J, Palsson BO (2003) Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proc Natl Acad Sci USA* 100:13134–13139.
45. Harrison R, Papp B, Pál C, Oliver SG, Delneri D (2007) Plasticity of genetic interactions in metabolic networks of yeast. *Proc Natl Acad Sci USA* 104:2307–2312.
46. Di Talia S, Skotheim JM, Bean JM, Siggia ED, Cross FR (2007) The effects of molecular noise and size control on variability in the budding yeast cell cycle. *Nature* 448: 947–951.
47. Wagner A (2007) Energy costs constrain the evolution of gene expression. *J Exp Zoolog B Mol Dev Evol* 308:322–324.
48. Blake WJ, et al. (2006) Phenotypic consequences of promoter-mediated transcriptional noise. *Mol Cell* 24:853–865.
49. Ansel J, et al. (2008) Cell-to-cell stochastic variation in gene expression is a complex genetic trait. *PLoS Genet* 4:e1000049.
50. Steinmetz LM, et al. (2002) Systematic screen for human disease genes in yeast. *Nat Genet* 31:400–404.
51. Hartl DL, Clark AG (1997) *Principles of Population Genetics* (Sinauer, Sunderland, MA), 3rd Ed.
52. Lestas I, Vinnicombe G, Paulsson J (2010) Fundamental limits on the suppression of molecular fluctuations. *Nature* 467:174–178.
53. Tsai IJ, Bensasson D, Burt A, Koufopanou V (2008) Population genomics of the wild yeast Saccharomyces paradoxus: Quantifying the life cycle. *Proc Natl Acad Sci USA* 105:4957–4962.
54. Dublanche Y, Michalodimitrakis K, Kümmerer N, Foglierini M, Serrano L (2006) Noise in transcription negative feedback loops: Simulation and experimental analysis. *Mol Syst Biol* 2:41.
55. Paulsson J (2004) Summing up the noise in gene networks. *Nature* 427:415–418.
56. Vavouri T, Semple JI, Garcia-Verdugo R, Lehner B (2009) Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell* 138: 198–208.
57. Drummond DA, Wilke CO (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134:341–352.
58. Yang JR, Zhuang SM, Zhang J (2010) Impact of translational error-induced and error-free misfolding on the rate of protein evolution. *Mol Syst Biol* 6:421.
59. Crow JF, Morton NE (1955) Measurement of gene frequency drift in small populations. *Evolution* 9:202–214.
60. Zhang J (2010) Evolutionary genetics: Progress and challenges. *Evolution Since Darwin: The First 150 Years*, eds Bell MA, Futuyma DJ, Eanes WF, Levinton JS (Sinauer, Sunderland, MA), pp 87–118.
61. Lynch M (2007) *The Origins of Genome Architecture* (Sinauer, Sunderland, MA).
62. Kimura M (1983) *The Neutral Theory of Molecular Evolution* (Cambridge University Press, Cambridge, UK).
63. Becker SA, et al. (2007) Quantitative prediction of cellular metabolism with constraint-based models: The COBRA toolbox. *Nat Protoc* 2:727–738.
64. Edwards JS, Covert M, Palsson BO (2002) Metabolic modelling of microbes: The flux-balance approach. *Environ Microbiol* 4:133–140.
65. Förster J, Famili I, Palsson BO, Nielsen J (2003) Large-scale evaluation of in silico gene deletions in Saccharomyces cerevisiae. *OMICS* 7:193–202.
66. Burgard AP, Nikolaev EV, Schilling CH, Maranas CD (2004) Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res* 14:301–312.
67. Gillespie JH (1975) Natural selection for within-generation variance in offspring number II. Discrete haploid models. *Genetics* 81:403–413.
68. Hill WG, Zhang XS (2004) Effects on phenotypic variability of directional selection arising through genetic differences in residual variability. *Genet Res* 83:121–132.