BMC
Evolutionary Biology

Proportionality between variances in gene expression induced by noise and mutation: consequence of evolutionary robustness

Kunihiko Kaneko

Abstract

Background: Characterization of robustness and plasticity of phenotypes is a basic issue in evolutionary and developmental biology. The robustness and plasticity are concerned with changeability of a biological system against external perturbations. The perturbations are either genetic, i.e., due to mutations in genes in the population, or epigenetic, i.e., due to noise during development or environmental variations. Thus, the variances of phenotypes due to genetic and epigenetic perturbations provide quantitative measures for such changeability during evolution and development, respectively.

Results: Using numerical models simulating the evolutionary changes in the gene regulation network required to achieve a particular expression pattern, we first confirmed that gene expression dynamics robust to mutation evolved in the presence of a sufficient level of transcriptional noise. Under such conditions, the two types of variances in the gene expression levels, i.e. those due to mutations to the gene regulation network and those due to noise in gene expression dynamics were found to be proportional over a number of genes. The fraction of such genes with a common proportionality coefficient increased with an increase in the robustness of the evolved network. This proportionality was generally confirmed, also under the presence of environmental fluctuations and sexual recombination in diploids, and was explained from an evolutionary robustness hypothesis, in which an evolved robust system suppresses the so-called error catastrophe - the destabilization of the single-peaked distribution in gene expression levels. Experimental evidences for the proportionality of the variances over genes are also discussed.

Conclusions: The proportionality between the genetic and epigenetic variances of phenotypes implies the correlation between the robustness (or plasticity) against genetic changes and against noise in development, and also suggests that phenotypic traits that are more variable epigenetically have a higher evolutionary potential.

Background

Plasticity and robustness are basic concepts in evolutionary and developmental biology. Plasticity refers to the changeability of phenotypes in response to external environmental perturbations. Indeed many important concepts in biology are concerned with the changeability in the system. This changeability depends on each phenotype: some phenotypes are more variable than others. How is such degree of changeability characterized quantitatively?

On the other hand, robustness is another basic concept in evolutionary and developmental biology. Here, phenotypic robustness is defined as the ability of the system to continue to function despite perturbations to it [1-7]. Phenotypes important for survival are expected to be robust, at least to some degree, to enable organisms to survive under such perturbations.

For both plasticity and robustness, there are epigenetic and genetic sources of perturbations to a biological system, which act in different time scales. Epigenetic perturbation works at a faster scale. Phenotypes are changed through noise in gene expression and developmental dynamics. Environmental variation gives another source for variability. Genetic variation, on the other

Correspondence: kaneko@complex.c.u-tokyo.ac.jp
Department of Basic Science, Univ. of Tokyo, and Complex Systems Biology
Project, ERATO, JST, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan

BioMed Central

hand, works at a longer time scale through evolution. Now, is there any relationship between the changes of genetic and epigenetic origins? If a phenotype is changeable easily epigenetically through development or environment, is it also more feasible to change genetically? Similarly, if a phenotype is robust to developmental perturbations, is it also robust to genetic variations through evolution? When we consider generally any dynamical systems, such relationship would not be expected. However, as a biological system is a result of evolution, existence of some relationship between the genetic and epigenetic robustness may be expected.

The relationship between evolution and robustness has been long debated since the pioneering studies by Schmalhausen and Waddington [8,9]. Waddington, then coined the term "genetic assimilation", in which phenotypic changes induced environmentally are then assimilated to genetic changes through evolution. Although important, these pioneering studies have mostly emphasized the qualitative aspects of the relationship between robustness and evolution. However, advances in quantitative studies on cell biology have facilitated the quantitative assessment of this relationship. In particular, fluctuations of phenotypes (e.g., gene expression levels) that have been measured extensively through the fluorescent-protein techniques [10-13] can provide a tool to explore such relationship.

In quantitative terms, robustness can be considered as a measure of the insensitivity of phenotypes to external disturbances and plasticity as a measure of changeability of phenotypes. On the other hand, fluctuation is the degree of phenotypic variation induced by perturbations. Hence, the phenotypic fluctuation (variance) increases with a decrease in robustness, and vice versa. Thus, the variance can serve as an index (inverse) for robustness, and also for plasticity. Now, the question concerning robustness and evolution can also be posed in terms of phenotypic variances. How is the evolution speed correlated with the variance? Does the variance increase or decrease through evolution?

Indeed, findings of previous studies involving artificial selection experiments with bacteria suggest that the rate of evolution, i.e., the increase in the fitness per generation, is proportional to the phenotypic variance among isogenic individuals [14,15]. This relationship, originally defined on the basis of a macroscopic distribution theory, was also confirmed in in-silico experiments by using gene regulation networks (GRNs) and metabolic networks [16,17].

This observed relationship is noteworthy because evolvability is characterized by non-genetic variation of phenotypes. Even if the rate of genetic change (mutation or recombination) is identical, the rate of evolution can differ according to this variation. To elucidate this point,

recall again that there are 2 sources in phenotypic variances, genetic and epigenetic. Quantitatively, the former is characterized by the phenotypic variance in a heterogenic population and is due to genetic modifications, as, denoted as V_g, whereas the latter, denoted here as V_{ip}, is the phenotypic variance in an isogenic population due to noise during the developmental process. The former reflects the structural robustness of the phenotype, i.e., the rigidity of the phenotype against changes induced by genetic mutations, whereas the latter reflects the robustness of the phenotype against the stochasticity encountered during the developmental process or that induced by environmental changes. (Phenotypic variance of nongenetic origin is traditionally discussed as environmental variance V_e. Here, we are concerned with the variance due to fluctuation during developmental process, and thus adopt this notation, but one could regard this variance as a component of V_e [18,19].)

It is obvious that evolution speed, which indicates the change in phenotype due to genetic changes, is correlated with V_g, as was demonstrated by Fisher [18,20]. Thus, the proportionality between the evolution speed and V_{ip} suggests the proportionality between V_g and V_{ip} throughout the course of evolution. Indeed, this relationship was confirmed by the evolutionary stability theory and numerical simulations [14,16], which imply that robustness to noise and to mutation are correlated.

So far the proportionality between V_{ip} and V_g of a given fitness *through the course of evolution (over generations)* was confirmed. Now, let us come back to the original question on a possible relationship between genetic and epigenetic robustness (or changeability) over many phenotypic traits. To discuss such problem, we need to study the relationship of the variances V_{ip} and V_g over many phenotypes or expressions of many genes, for a given individual (not over the evolutionary course). This is the focus of the present paper.

The phenotypes or gene expression levels are generally associated with several genes, as known as epistasis. Even if a fitness value may be directly related to a single phenotype or the expression of a single gene, many genes may modify the expression of each other. These expressions are interrelated through a complex gene regulation network; therefore, the nature of their correlation may change during the course of evolution. This may give some evolutionary constraint on the changes of expression levels of genes. Then, does a gene with a higher fluctuation in its expression level have a higher potentiality in evolution than others? Is there correlation between phenotypic changes by epigenetic noise in gene expression dynamics and by genetic variation? We will demonstrate such proportionality over expressions of many genes, rather than the previously studied proportionality of the variances through the evolutionary course.

Using a numerical model in GRNs and simulating their evolutionary changes required to increase a given level of fitness, we first found that the rate of evolution of the expression levels of several genes was highly correlated with (or roughly proportional to) the respective variances of these genes. Next, we proceeded to present evidence for the proportionality between the 2 types of variances of gene expressions, i.e., of genetic and epigenetic origins *over many genes*. This proportionality was achieved after a selection process under a given fitness condition and is true whenever the phenotype of evolved system is robust to transcriptional noise in gene expression dynamics and genetic mutation. Further, the generality of this relationship was confirmed by studying a variety of models and evolutionary stability theory of multivariate distribution. We will also discuss some experimental evidences for this relationship.

Results
Modeling Strategy
Here, we considered the evolution of a simple model for "development." For this purpose, we postulated the following conditions for development and evolution.

(i) The set of variables x_i ($i = 1, ..., M$) represents the expression levels of M genes. These variables take continuous values, which we set such that if gene i is expressed, then $x_i > 0$ and if not, $x_i < 0$. (Choice of a threshold at zero is a matter of convenience. Indeed, we also carried out simulations of a model in which the expression x_i takes only non-negative values with certain positive threshold values for activation. Results to be discussed were not altered by this change).

(ii) Gene expressions mutually activate or inhibit each other and are regulated by the GRN. The temporal evolution of the gene expression level x_i is generally not simple because the value of M is large. The phenotype of each individual is defined by the set of gene expression level x_i after evolution for "developmental time," which starts from the time point with a given initial gene expression level (set as $x_i < 0$ for all i, i.e., none of the genes is expressed). The developmental time for the system is set to a large period to allow gene expression patterns to settle into an attractor state.

(iii) Gene expression dynamics are noisy. Owing to their dependence on chemical reactions involving molecular collisions, gene expression dynamics are also stochastic in nature. In particular, since the number of molecules (e.g., mRNA and proteins) involved in gene expression is not extremely large, a deviation from the average rate equation for the reaction is possible. This deviation is represented as the noise applied to the average gene regulation expression dynamics. The amplitude of noise is denoted as σ.

(iv) Genotype: Depending on the genotype, the structure of GRN changes, i.e., network paths that determine the genes responsible for activating (or repressing) a given gene. This interaction between genes is represented by a connection matrix J_{ij}, which takes a value of 1 (if activating), -1 (if repressing), or 0 (if there is no influence).

(v) Population: A population of individuals with different genotypes, i.e., with each individual having a slightly different GRN-the matrix J_{ij}. This gives a genotypic distribution.

(vi) Fitness: The pattern of expression in x_i is determined on the basis of gene expression dynamics. Fitness is a function of the expression pattern of a subset of x_i, i.e., the expression of a given set of "target" genes $i = 1, ..., k$. If the pattern of x_i is closer to the prescribed pattern of genes, the fitness is higher. We take the prescribed pattern as "all on" for the target genes, unless otherwise mentioned. If all components of x_i ($i = 1, ..., k$) are positive, the fitness is set at 0, and it is decreased by 1 when the number of negative x_i is increased by 1.

(vii) Mutation and selection: Offspring are produced according to the fitness: individuals with low fitness cannot produce offspring. The selection process according to the fitness thus progresses, while mutation introduces a change in the GRN, represented as 1,-1, or 0, in a few elements of the matrix J_{ij}. The fraction of the elements altered is given by the mutation rate.

We studied numerical models based on the above-mentioned characteristics. (For details see Method). Note that due to the complex nature of gene expression dynamics, GRN with a higher fitness value is generally rare. Further, since the expression dynamics are subjected to perturbations by noise, the "goal" of reaching the highest fitness may not be easily achieved. However, through the evolutionary processes, including natural selection and mutations, a GRN with the highest fitness value is generated. Individuals with the highest fitness value in a population evolve over some generations to reach the highest fitness value, i.e., $x_i > 0$, for the target genes $i = 1, 2, ..., k$.

We previously reported that the fitness distribution in a heterogenic population undergoes a transition when the noise level is increased [17,21]. Before we discuss the main results of the present paper from the next subsection, we briefly summarize the earlier numerical result in this subsection.

When the noise level σ was below a given threshold ($\sigma < \sigma_c$), some individuals within the population might take lower fitness values, thereby inducing a considerable difference in the distribution of fitness over different genotypes. Some mutants derived from individuals with genotypes having high fitness values might take much lower fitness values, even after many generations

of evolution. However, when the noise level was greater than the threshold ($\sigma_c < \sigma$), the distribution of fitness was sharp, with concentration of mutants with high fitness values. Thus, low-fitness mutants were eliminated through the evolution. (If the noise level was too high, the expression levels were not fixed in time and increased or decreased over time. We did not examine such cases).

Accordingly, an increase in the noise strength lead to a transition in the robustness to mutation. GRNs that evolve under a low noise level did not have robustness to mutation, whereby some mutants could not sustain the high fitness value. In contrast, robustness to mutation was achieved for GRNs evolved only under a high noise level $\sigma > \sigma$.

This counterintuitive robustness to mutation for $\sigma > \sigma_c$ can be explained as follows. According to the dynamics of GRN that evolves under higher noise level, a large portion of the initial conditions reach target attractors that give the highest fitness values, thereby achieving robustness against noise, while for those evolved under $\sigma < \sigma_c$, only a tiny fraction reaches target attractors. The developmental landscape for $\sigma > \sigma_c$ gives a global, smooth attraction to the target, whereas the landscape evolved at $\sigma < \sigma_c$ is rugged. Now, consider mutation to a network to slightly change gene expression dynamics. In the smooth landscape with global attraction, such a perturbation will cause little change to the final expression pattern, while under the dynamics with rugged developmental landscape, it often destroys the attraction to the target attractor. In other words, robustness to mutation evolves only under robustness to noise during development.

For $\sigma > \sigma_c$, robustness to both noise and mutation evolved. We computed robustness to both factors using the 2 types of variances, V_{ip} and V_g, where the former was the noise-induced variance in the distribution of the fitness (i.e., the number of "on" target genes), in a population of isogenic individuals and V_g is the variance in the distribution of fitness in a heterogenic population. The latter was obtained by first computing the mean fitness value for each genotype, and then by measuring the variance of these mean values for a heterogenic population. The 2 variances of the fitness decreased proportionally, throughout the course of evolution, for $\sigma > \sigma_c$, in accordance with the principle of evolution of robustness (see [17,21]).

V_g-V_{ip} relationship over genes

Apart from the fitness level, the expression level x_i of each gene i was also distributed, even for isogenic distribution with the same J_{ij}; this was because of the stochasticity in each gene expression dynamics. Similar to the variances for the fitness, the phenotypic variance $V_{ip}(i)$ for each gene i in an isogenic population is defined on

the basis of the variance of the expression of each gene i, with each $X_i = Sign(x_i)$, in an isogenic population. On the other hand, the mean expression level $\overline{X_i}$ over the isogenic population depended on each genotype (i.e., the matrix J_{ij}). The variance computed using the distribution of $\overline{X_i}$ in this heterogenic population, then, gives the genetic variance $V_g(i)$ for each gene i.

As mentioned above, our model also accounted for many non-target genes that do not contribute to fitness. The expression level x_i of such non-target genes i could be either positive or negative because there was no selection pressure directed at fixing their expression level. However, we found that the expression levels of many non-target genes become fixed to positive or negative values over the course of evolution when $\sigma > \sigma_c$. To achieve robust fitness, the expression levels of some of the non-target genes were also fixed to either the on or off status, with consistency across individuals in the gene expression status. Hence, the average expression level $< \overline{X_i} >$ in a heterogenic population increased or decreased to either positive or negative values ($\overline{}$ is the mean over an isogenic population, while $< ... >$ is the average over a heterogenic population).

We computed the rate of evolution for each gene expression over a given number of generations, as the rate of either increase or decrease of the average expression level $< \overline{X_i} >$ in a heterogenic population. It was computed after evolution for some generation to achieve an increase in fitness. We then examined the validity of the relationship between the evolution speed and the fluctuations over genes, by plotting $|< \Delta \overline{X_i} >|$ against $V_{ip}(i)$, where $< \Delta \overline{X_i} >$ is the change in the average expression level $< \overline{X_i} >$ of a given gene over some generations (20, in Figure 1). The proportionality was found to be valid over different genes, both target and non-target, as shown in Figure 1. Genes with higher variances would have a higher potentiality to evolve (In Figure 1, the target genes had less variance and evolution speeds over these generations. This is because the fitness had already increased in this generation, fixing the expression levels to positive values, and there was little room for further increase).

Considering the possible generalization of Fisher's theorem [18,20], which states that the evolution speed is proportional to V_g, and applying it to the expression levels over genes, we may expect that the evolution speed of each gene expression level is proportional to V_g (i). Then, the proportionality between $V_g(i)$ and $V_{ip}(i)$ over the genes can be expected from the proportionality between the evolution speed and $V_{ip}(i)$. Considering this, we examined the relationship between $V_g(i)$ and V_{ip} (i) for several values of noise levels (Figure 2). From the plots, we obtained the following findings:

(i) Proportionality between $V_{ip}(i)$ and $V_g(i)$ was satisfied over many genes, i.e., $r_i \equiv V_{ip}(i)/V_g(i)$ took a common value $\rho(<1)$ for many genes i, when the evolution of robustness progressed at $\sigma > \sigma_c$.

(ii) Target genes always lay on the proportional line, with relatively low values of $V_{ip}(i)$ (and accordingly, $V_g(i)$), while the variances of many non-target genes also lay on the same proportionality line $r_i \sim \rho$. Variances of only a few non-target genes did not exhibit the abovementioned proportional relationship, and the ratios r_i for such genes were scattered between ρ and 1. (See also Additional file 1, Figure S1).

(iii) The fraction of such genes that fitted on the single proportional line increased with the noise strength σ. As the noise level was lowered, an increase was noted in the fraction of genes showing expression variances that deviate from the abovementioned proportional relationship. At around the threshold noise level σ_c, most genes approached $r_i \sim 1$. (See Additional file 1, Figure S1, for noise dependence of the variances $V_{ip}(i)$ and $V_g(i)$).

We then plotted the histogram of r_i over all genes i, sampled over a few sets of generations (see Figure 3a, plotted in log-scale for r_i). The figure showed that the peak at ρ was more prominent with an increase in σ. On the other hand, the broader distribution ranging between $r_i \sim \rho$ and 1 became more prominent as the noise level decreased, until the distribution around $r_i \sim 1$ dominated at $\sigma \sim \sigma_c$.

The proportionality between $V_{ip}(i)$ and $V_g(i)$ was not a property of every gene expression dynamics but was evident only after the system achieved robustness through evolution. The fraction of genes showing a proportional relationship increased during the course of evolution. Indeed, for $\sigma > \sigma_c$, the peak at $r_i \sim \rho$ increased over generations until it approached the distribution shown in Figure 3a (see Figure 3b). Summing up, the evolution of robustness was characterized by the formation of the peak at $\rho < 1$, in the distribution of $r_i = V_{ip}(i)/V_g(i)$.

The proportionality between the 2 variances implies the existence of a correlation between the noise- and mutation-induced changes in the gene expression statuses (see also Additional file 1, Figure S2 for the correlation in variances). Such a correlation was observed by computing the frequency of errors, i.e., changes in the on/off status of gene expression due to noise (without a change in the network) and as a result of mutation to the network (without adding the noise). The frequency of these 2 errors was highly correlated over genes for the GRN evolved at $\sigma > \sigma_c$ (see Figure 4). In other words, genes that were switched on or off more frequently by noise were also switched more frequently by mutation. This was in strong contrast with the GRN evolved at $\sigma < \sigma_c$ where no such correlation was observed (see Additional file 1, Figure S3). To sum up,

the changeability of each gene expression level by noise and mutation was correlated, for a robust evolved system.

Generality

We confirmed that the proportionality between phenotypic variances of genetic and epigenetic origins held true for a system with evolved robustness, by simulating our model and its extended versions over several conditions. For all the conditions below, we confirmed (a) transition to robust evolution with the increase in noise level (b) proportionality between $V_{ip}(i)$ and $V_g(i)$ throughout the course of evolution and over many genes i.

(i) Against the change in k (target set) and M (number of genes): With an increase in the fraction of target genes, evolution to the fittest state became increasingly dicult, and the noise level for robust evolution σ_c was slightly increased, but the proportionality was valid for $\sigma > \sigma_c$ (see Additional file 1, Figure S4a).

(ii) Considering that the density in the connection paths in the actual GRN is rather low, the validity of the results was verified by decreasing the path rates in the model. The conclusion remained unchanged as long as the network was percolated. The fraction of genes deviating from the proportionality slightly increased as the path rate in the network decreased (e.g., to .05 per gene) (see Additional file 1, Figure S4b).

(iii) Even if the noise level depended on each gene i, the conclusion was valid (see Additional file 1, Figure S5). After evolution, the variance in the expression level of each gene was not correlated with the noise level of each gene, implying that the fluctuation of gene expression was mainly controlled by (evolved) gene-to-gene regulation J_{ij}.

(iv) To consider the influence of extrinsic rather than intrinsic noise, we also introduced the same level of noise to all gene expressions. Again, the robustness and the proportionality between the variances persisted as long as the level of the intrinsic noise was larger than σ_c. Although the "extrinsic" (common) noise also contributed to the evolution of robustness, it played only a minor role in this respect. (See Additional file 1, Figure S6).

(v) By the variation in the environmental condition, the fitness condition for the target genes varied accordingly. By switching the condition for the target genes from 'on' to 'off' after some generations, we verified whether the evolution of GRN copes with this environmental variation. When the condition was switched, both the variances of epigenetic and genetic origins, as well as $r_i = V_g(i)/V_{ip}(i)$, increased to adapt novel environmental condition. Once the adaptation was achieved, the variances as well as r_i decreased, to regain robustness (Figure 5).

When the target phenotype was changed periodically, we observed that the increase and decrease of the variances were consistently repeated, when the noise level was near the transition value (σ_c), where rapid adaptation to new environment and robustness in phenotype were compatible.

(vi) We also extended our GRN model to account for diploids with sexual recombination. Here, each individual had a pair of matrices J_{ij}^1 and J_{ij}^2, and the gene expression dynamics were given as a result of summation of the two matrices instead of the equation in the Method. By considering recombination of two matrices from a parent, we evolved GRN to achieve a higher fitness. The proportionality of the 2 variances was again confirmed, while another noteworthy finding was that in the case of heterozygotes, the robustness was further enhanced (suppresses the variances of expression).

Further, the proportionality between the variances was not confined to the present model. Indeed, in a model of catalytic-reaction-network, such a proportionality between the variances of fitness [16] and over chemicals evolved, whereas robustness transition with the increase of noise was confirmed in an abstract spin model for protein folding [22]. We expect that the relationship holds true as long as the fitness is determined through complex developmental dynamics with noise and the high-fitness states are not easily achieved so that error catastrophe appears with the increase in the mutation rate.

A phenomenological distribution theory
Considering the generality of the proportionality between the phenotypic variances of genetic and epigenetic origins, we provide a distribution theory for it without going into detailed setups of the model. We adopt the evolutionary stability argument first introduced for the proportionality between V_{ip} and V_g (of the fitness) through the course of evolution [16,23].

We consider a multivariate distribution function with regard to gene expression level x_i and the genotype a. Considering that multiple genes are involved, we assume that the genotype is represented by a scalar parameter a (e.g., by a Hamming distance from the fittest genetic sequence). Now, we assume evolutionary stability, in which the distribution maintains a single peak through evolution. Then, by a suitable transformation of variables, this peak position is taken to be 0; further, the form is approximated by Gaussian distribution around the origin to give the following equation

$$P(x_i, a) = N_0 exp(-\frac{x_i^2}{2\alpha_i} + C_i x_i a - \frac{a^2}{2\mu}). \quad (1)$$

with N_0, a normalization constant so that $\int P(x:a)dx = 1$. Here, $\alpha_i \equiv V_{ip}(i)$ is the variance of the gene

expression level, while μ is the mutation rate that determines the variance in the genotypes. Only a linear change of x_i with regard to a is considered by neglecting higher order terms. Eq. (1) is rewritten as

$$P(x_i, a) = exp(-\frac{(x_i - C_i a \alpha_i)^2}{2\alpha_i} - \frac{1}{2}(\frac{1}{\mu} - C_i^2 \alpha_i)a^2) \quad (2)$$

Now, recall the stability of the distribution $P(x_i, a)$, i.e., whether it has a peak in the space with x_i and a. This condition is given by $1/\mu - C_i^2\alpha_i > 0$, i.e., $\mu < \mu_{max}^i \equiv \frac{1}{C_i^2\alpha_i}$. For the mutation rate larger than μ_{max}^i, the distribution is flattened. In this case, the peaked distribution concentrated at a certain gene expression level is no longer sustained. This can be interpreted as a kind of error catastrophe originally introduced by Eigen [24], i.e., collapse of the sustenance of the localized distribution of functional genes. The critical mutation rate for the error catastrophe is given by $\mu = \mu_{max}^j$, which can take independent values by any gene i. However, GRNs that have achieved robustness to noise and mutation through evolution may have some constraints among expressions of different genes.

To have higher robustness, the error threshold for the fitness should be postponed to a higher mutation rate. When the system has achieved robustness to noise and mutation through evolution, the fitness level changes only to a small degree (i.e., remains almost neutral) against a considerable amount of change in the GRNs due to mutation [21]. Until the occurrence of this number of mutations, the genes rarely undergo any change in their expression statuses (on or off). This introduces a constraint on the change in gene expression against mutations.

If each of non-target genes were switched on or off independent of each other, the error in the expression of target genes that could be influenced by each switch would occur frequently. Indeed, the evolved robust GRN has some constraint on the errors by noise and mutation, as plotted in Figure 4. To achieve higher robustness, there needs to be some correlation between the changes in the expression statuses of genes. By suitable mutual interaction ($J_{ij} \neq 0$) among genes, the error frequency in the target gene expression can be reduced. This reduction works up to the mutation rate μ_{max}^i, while for $\mu \sim \mu_{max}^i$, errors in the expression status of one gene can be propagated to the expression of many other genes, which in turn will induce changes in the expression statuses of other genes. Hence, for a robust network having higher error threshold mutation rate, many genes may be switched on or off simultaneously

within a GRN, once an error catastrophe in one gene expression occurs. Accordingly, many genes i are expected to share a common critical mutation rate. Hence, μ_{max}^i is roughly equal for many genes, when robustness is evolved, i.e., at $\sigma > \sigma_c$. In fact, we computed the expression pattern of GRNs by adding a larger number of mutations to the evolved GRN, to obtain the variances $V_{ip}(i)$ and $V_g(i)$. When the original network was evolved under high noise conditions, the variances touched with the line $V_{ip}(i) = V_g(i)$ simultaneously over many genes, as the mutation rate was increased (see Figure 6). The flattening of the distribution occurred at similar mutation rates over many genes.

Following this argument, we may expect that

$$\mu_{max}^i = (C_i^2 \alpha_i)^{-1} = independent\ of\ many\ genes \qquad (3)$$

when robustness is evolved (i.e., at $\sigma > \sigma_c$). Note that $\overline{x_i}$, the mean of x_i for a given genotype a, is given by $C_i \alpha_i a$ according to eq. (2). The variance of $\overline{x_i}$ due to this "genetic change" is given by the distribution of a. Thus, we get

$$V_g(i) = <(\delta x_i)^2> = C_i^2 \alpha_i^2 <(\delta a)^2>. \qquad (4)$$

Since $V_{ip}(i) = \alpha_i$, we get

$$r_i = V_g(i) / V_{ip}(i) = C_i^2 \alpha_i <(\delta a)^2>, \qquad (5)$$

but this value is independent of gene i, according to eq.(3). Hence, the proportionality over genes is explained by a common error catastrophe threshold value over different gene expression levels.

As stated earlier, this proportionality over genes is not a general property of (gene expression) dynamical systems, but emerges only when evolution occurs under a sufficient level of noise (see Figure 3b). The use of the distribution function and assumption of stability implies that we are concerned with the stationary distribution that is attained after the progress in evolution. As the noise level reduces resulting in a decrease in robustness, the fraction of genes sharing a common value of $V_g(i)/V_{ip}(i) = \rho$ decreases. This decrease is interpreted as a decrease in the number of genes that have a common error threshold value.

Discussion
The relationship between the variances over genes, rather than the evolutionary course, will be easier to confirm experimentally because variances in this case need not be traced across many generations. By measuring directly the isogenic phenotypic variance and mutational variance over many genes (proteins), the correlation between the 2 can be examined. Although direct experimental support is not yet available, recent studies conducted by Laundry

et al. [25] on such variances in yeast suggest the existence of a correlation between the 2 types of variances (see also [26]). In fact, they measured "expression noise" for each gene as the variance from its expression in isogenic organisms, and "mutational variance" as the variance of the change in the expression levels of the genes after the occurrence of mutations. The former corresponds to $V_{ip}(i)$, while the latter correlates with $V_g(i)$ since both measure the variations in phenotype (gene expression) induced by genetic changes. Although experimental data are scattered, a positive correlation is noted between the expression noise and mutational variance, as is consistent with the inference of the proportionality between $V_{ip}(i)$ and $V_g(i)$. Note that this proportionality holds true only for a set of genes whose expression levels are mutually related and directly or indirectly related with the fitness, whereas the experimental data cover all genes. In fact, by choosing a set of genes that have a stronger mutual relationship, the correlation between the variances is increased [27].

According to the theoretical argument we presented here, the phenotype variable is not restricted to gene expression level, but can represent any trait. Stearns et al. carried out selection experiments of *Drosophila melanogaster* on several fitness conditions such as age at eclosion, weight, and lifespan. They measured the variance of these phenotype traits within lines (i.e., corresponding to $V_{ip}(i)$) and among lines (corresponding to $V_g(i)$). Interestingly, the 2 types of variances (even after being normalized by the mean) showed remarkable proportionality over different phenotypic traits (see Figure 2 of [28]), in agreement with our theory. Note that they used the population selected on the basis of certain phenotypic traits, as in our model and theory. This must explain why the observed proportionality between the 2 variances is clearer than that in [25].

Conclusion
The characterization of robustness, evolvability, and plasticity [1,29-31] is an important issue in the field of evolutionary and developmental biology; however, studies on this issue are often qualitative. In the present study, we have demonstrated that the phenotypic fluctuations provide quantitative measures for these. Consider a population of organisms evolved under a single fitness condition, where the phenotypes that directly or indirectly influence the fitness are given as a result of (gene expression) dynamics under noise (determined by transcriptional networks). Through selection and mutation, the rules for the dynamics (i.e., the transcriptional networks) evolve leading to the achievement of a higher fitness level. Previously, we defined V_{ip} as the variance of the fitness within an isogenic population, and V_g as the variance of the average

fitness within a heterogenic population, and obtained $V_{ip} \propto V_g \propto$ evolution speed through the course of evolution [16,17,23].

In the present paper we defined the variances at each gene expression level i due to noise as $V_{ip}(i)$ and that due to mutation as $V_g(i)$. Then, the conclusion of the present paper is summarized as follows:

(1) For a population of organisms at a given generation evolved after some generations, $V_{ip}(i) \propto V_g(i)$ for most genes i. In other words, $r_i = V_g(i)/V_{ip}(i)$ took a common value ($\rho < 1$) over many genes, and the number of such genes increased as the robustness of fitness increased. (Total phenotype variance is given by $V_{ip}(i) + V_g(i)$ if the 2 variances are added independently. In this case the heritability [18,19], defined as the ratio of $V_g(i)$ to the total phenotypic variance is given by $r_i/(1 + r_i)$. Hence, the heritability takes a common value for mutually correlated traits, for evolved population under a fixed, single fitness condition). The previous relationship through the evolutionary course, implies that organisms with larger phenotypic variances have higher rates of evolution. The relationship (1) we found here implies that genes (or phenotypic traits) that have larger fluctuations have higher evolution speed.

(2) As the fraction of genes sharing a common value $r_i = \rho < 1$ increased, there was a decrease in the degree of freedom to change the gene expressions independently by mutation. This increase in correlated change lead to an increase in robustness of the fitness to mutation and noise. On the other hand, the expression of genes with larger $r_i \sim 1$ was easily switched by noise or mutation, and provided plasticity to environmental as well as mutational change (see also Figure 5).

The generality of our results was confirmed by several extensions of the model including environmental fluctuations, gene-dependent noise amplitudes, diploid with recombination, and so forth, as well as a catalytic reaction network model. Here, we should note that a correspondence between the change induced by noise in development and that induced by mutation was a source of correlation between V_{ip} and V_g in our study. Indeed, Waddington coined the term "genetic assimilation" as a process in which environment-induced phenotypic changes are subsequently embedded into genes [9]. The proportionality among phenotypic plasticity, $V_{ip}(i)$, and $V_g(i)$ is regarded as a quantitative expression of this genetic assimilation.

Method

A simplified gene expression dynamics with a sigmoid input-output behavior [32,33] is adopted here, although several simulations in the form of biological networks will give essentially the same result. In this model,

the dynamics of a given gene expression level, x_i, is described by the following:

$$dx_i / dt = \gamma\{\tanh[\beta \sum_{j>k}^{M} J_{ij}x_j] - x_i\} + \sigma\eta_i(t), \qquad (6)$$

where $J_{ij} = -1, 1, 0$, and $\eta_i(t)$ is a Gaussian white noise given by $< \eta_i(t)\eta_j(t') > = \delta_{i,j} \delta(t - t')$. M is the total number of genes, and k is the number of target genes that determine fitness. The summation only for $j > k$ is introduced to eliminate possible influences from the target genes, which might also fix other gene expressions. Without this restriction and just by the summation over all genes, however, conclusions of the present numerical results are invariant. Of course, the matrix J_{ij} is generally asymmetric. The amplitude of noise strength is given by σ that determines stochasticity in gene expression. The initial condition is given by $(-1,-1,...,-1)$; i.e., all genes are off. The fitness F is determined by how many of the "target" genes are on after a sufficient time, i.e., the number of i such that $x_i > 0$ for $i = 1, 2, ..k < M$. Because the model includes a noise component, the fitness can fluctuate at each run, which leads to a distribution in the fitness F and x_i, even among individuals sharing the same gene regulation network. For each network, we compute the average fitness $\bar{F}$ over L runs.

At each generation, there are N individuals with slightly different J_{ij}. Among the networks, we select those with higher fitness values. From the selected networks, J_{ij} is "mutated," i.e., J_{ij} for a certain pair i, j selected randomly with a certain fraction is changed among ± 1,0. For example, each of the $N_s(< N)$ networks with higher values $\bar{F}$ produce N/N_s mutants. (We also used the selection procedure such that the offspring number is proportional to the fitness, with normalization of the total population to N, but the same results presented here were obtained). We repeat this selection-mutation process over generations.

The variance of fitness or gene expression $Sign(x_i)$ of identical networks over L runs gives V_{ip} or $V_{ip}(i)$, while the variances of their mean over different networks J_{ij} give V_g or $V_g(i)$. we chose $N = L = 200$, and $N_s = N/4$, while the conclusion to be shown below does not change as long as these values are sufficiently large. We use $\beta = 7$, $\gamma = .1$, $M = 64$ and $k = 8$, and initially chose J_{ij} randomly with equal probability for ± 1,0, unless otherwise mentioned.

References

1. de Visser JA, *et al*: **Evolution and detection of genetic robustness.** *Evolution* 2003, **57**:1959-1972.
2. Wagner A: **Robustness against mutations in genetic networks of yeast.** *Nature Genetics* 2002, **24**:355-361.
3. Ciliberti S, Martin OC, Wagner A: **Robustness can evolve gradually in complex regulatory gene networks with varying topology.** *PLOS Comp Biology* 2007, **3**:e15.
4. Wagner GP, Booth G, Bagheri-Chaichian H: **A population Genetic Theory of Canalization.** *Evolution* 1997, **51**:329-347.
5. Siegal ML, Bergman A: **Waddington's canalization revisited: Developmental stability and evolution.** *Proc Nat Acad Sci USA* 2002, **99**:10528-10532.
6. Barkai N, Leibler S: **Robustness in simple biochemical networks.** *Nature* 1997, **387**:913-917.
7. Alon U, Surette MG, Barkai N, Leibler S: **Robustness in bacterial chemotaxis.** *Nature* 1999, **397**:168-171.
8. Schmalhausen II: *(1949, reprinted 1986) Factors of Evolution: The Theory of Stabilizing Selection (University of Chicago Press, Chicago)* .
9. Waddington CH: *The Strategy of the Genes* (Allen & Unwin, London); 1957.
10. Elowitz MB, Levine AJ, Siggia ED, Swain PS: **Stochastic gene expression in a single cell.** *Science* 2002, **297**:1183-1187.
11. Furusawa C, Suzuki T, Kashiwagi A, Yomo T, Kaneko K: **Ubiquity of log-normal distributions in intra-cellular reaction dynamics.** *Biophysics* 2005, **1**:25-31.
12. Bar-Even A, *et al*: **Noise in protein expression scales with natural protein abundance.** *Nature Genetics* 2006, **38**:636-643.
13. Kaern M, Elston TC, Blake WJ, Collins JJ: **Stochasticity in gene expression: From theories to phenotypes.** *Nat Rev Genet* 2005, **6**:451-464.
14. Kaneko K: *Life: An Introduction to Complex Systems Biology* (Springer, Heidelberg and New York); 2006.
15. Sato K, Ito Y, Yomo T, Kaneko K: **On the relation between fluctuation and response in biological systems.** *Proc Nat Acad Sci USA* 2003, **100**:14086-14090.
16. Kaneko K, Furusawa C: **An evolutionary relationship between genetic variation and phenotypic fluctuation.** *J Theo Biol* 2006, **240**:78-86.
17. Kaneko K: **Evolution of robustness to noise and mutation in gene expression dynamics.** *PLoS ONE* 2007, **2**:e434.
18. Futuyma DJ: *Evolutionary Biology*. Second edition. (Sinauer Associates Inc., Sunderland); 1986.
19. Hartl DL, Clark AG: *Principles of Population Genetics*. 4 edition. (Sinauer Assoc. Inc., Sunderland); 2007.
20. Fisher RA: *The Genetical Theory of Natural Selection* (Oxford University Press); 1930.
21. Kaneko K: **Shaping Robust system through evolution.** *Chaos* 2008, **18**:026112.
22. Sakata A, Hukushima K, Kaneko K: **Funnel landscape and mutational robustness as a result of evolution under thermal noise.** *Phys Rev Lett* 2009, **102**:148101.
23. Kaneko K: **Relationship among Phenotypic Plasticity, Genetic and Epigenetic Fluctuations, Robustness, and Evolvability;.** *J BioSci* 2009, **34**:529-542.
24. Eigen M, Schuster P: *The Hypercycle* (Springer, Heidelberg); 1979.
25. Landry CR, Lemos B, Dickinson WJ, Hartl DL: **Genetic Properties Influencing the Evolvability of Gene Expression.** *Science* 2007, **317**:118.
26. Lehner B: **Selection to minimize noise in living systems and its implications for the evolution of gene expression.** *Mol Syst Biol* 2008, **4**:170.
27. Lehner B: **Genes Confer Similar Robustness to Environmental, Stochastic, and Genetic Perturbations in Yeast** *PLoS One*. 2010, e9035.
28. Stearns SC, Kawecki TJ: **The differential genetic and environmental canalization of fitness components in Drosophila melanogaster.** *J Evol Biol* 1995, **8**:539-557.
29. Kirschner MW, Gerhart JC: *The Plausibility of Life* (Yale University Press); 2005.
30. Ancel LW, Fontana W: **Plasticity, evolvability, and modularity in RNA.** *J Exp Zool* 2002, **288**:242-283.
31. West-Eberhard MJ: *Developmental Plasticity and Evolution* (Oxford University Press); 2003.
32. Mjolsness E, Sharp DH, Reisnitz J: **A connectionist model of development.** *J Theor Biol* 1991, **152**:429-453.
33. Salazar-Ciudad I, Garcia-Fernandez J, Sole RV: **Gene networks capable of pattern formation: from induction to reaction-diffusion.** *J Theor Biol* 2000, **205**:587-603.