

**Universal Relationship in Gene-Expression Changes for Cells in Steady-Growth State**

Kunihiko Kaneko\*

*Research Center for Complex Systems Biology, Graduate School of Arts and Sciences, University of Tokyo, Komaba, Tokyo 153-8902, Japan*

Chikara Furusawa†

*Quantitative Biology Center (QBiC), RIKEN, 6-2-3 Furuedai, Suita, Osaka 565-0874, Japan*

Tetsuya Yomo‡

*Department of Bioinformatics Engineering, Graduate School of Information Science and Technology, Graduate School of Frontier Biosciences, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan, and Exploratory Research for Advanced Technology, Japan Science and Technology Agency, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan*

(Received 18 June 2014; published 11 February 2015)

Cells adapt to different conditions by altering a vast number of components, which is measurable using transcriptome analysis. Given that a cell undergoing steady growth is constrained to sustain each of its internal components, the abundance of all the components in the cell has to be roughly doubled during each cell-division event. From such steady-growth constraint, expression of all genes is shown to change along a one-parameter curve in the state space in response to the environmental stress. This conclusion leads to a global relationship that governs the cellular state: By considering a relatively moderate change around a steady state, logarithmic changes in expression are shown to be proportional across all genes, upon alteration of stress strength, with the proportionality coefficient given by the change in the growth rate of the cell. This theory is confirmed by transcriptome analysis of *Escherichia coli* in response to several stresses.

DOI: [10.1103/PhysRevX.5.011014](https://doi.org/10.1103/PhysRevX.5.011014)

Subject Areas: Biological Physics, Nonlinear Dynamics, Statistical Physics

**I. INTRODUCTION**

A cell's internal state is now measurable with expression data on a few thousand genes using transcriptome analysis. High-dimensional data on the gene expressions are gathered, depending on cells and environmental conditions. In spite of the increase in the available data, however, it is sometimes difficult to extract biologically relevant characteristics from them, due to the complexity in the gene-expression network and dynamics. Indeed, the common trend in the transcriptome analysis is to uncover a set of genes that specifically respond to specific environmental changes while discarding other high-dimensional data that are gathered. A search for a simple law that governs a global change in expressions across genes has not been seriously attempted.

On the other hand, biologists are traditionally interested in macroscopic quantities such as activity, plasticity, and robustness [1,2], even though these quantities have involved qualitative, rather than quantitative, characteristics so far. At this stage, then, it will be crucial to extract such macroscopic quantities from a vast amount of the expression data available using transcriptome analysis. Here, the simplest candidate for such a macroscopic quantity will be the growth rate in cell population. Then, can we extract some universal relationship on global gene-expression changes and connect it with a macroscopic (population) growth rate of cells?

In searching for such a universal relationship, it will be relevant to restrict cell states of our concern, just as thermodynamics, the celebrated macroscopic phenomenological theory, is established by restricting our concern to thermal equilibrium. Of course, a cell is not in a state of static equilibrium but involves complex dynamics and grows (and divides) in time. Thus, we cannot directly apply the formulation in thermodynamics. However, we can instead follow the spirit in thermodynamics; we restrict our concern to a system with a steady-growth state and intend to extract a common law that should hold globally to such a state. Considering that the cell keeps its internal state

\*kaneko@complex.c.u-tokyo.ac.jp

†chikara.furusawa@riken.jp

‡yomo@ist.osaka-u.ac.jp

Published by the American Physical Society under the terms of the *Creative Commons Attribution 3.0 License*. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

across cell divisions, it is expected that all the components grow with a common rate. As a consequence of such restriction, then, we may hope to uncover a universal relationship across changes in gene expressions. Indeed, in transcriptome analysis data (e.g., Refs. [3–5]), the existence of the correlation in the expression changes across a vast number of genes is suggested [6–10], which are brought about through adaptation and evolution [11–14].

Here, we first analyze the transcriptome data in bacteria undergoing stress, to confirm a general relationship between global changes across the expression of all genes. To explain such a general relationship in a cellular state, we study a general consequence imposed by a constraint of the cellular states achieved by restricting our concern only to cells that maintain steady growth, i.e., those cells that can grow and divide, retaining their state. Within this constraint, we derive a theoretical relationship of the changes in all components (i.e., expression levels of all genes) in response to stress. Following this theoretical framework, we then reanalyze transcriptome data to demonstrate the validity of our theoretical argument.

## II. RESULT

### A. Changes in gene expression under environmental stress conditions: Experimental observations

To explore the relationship between changes of global gene expression and growth rate, we analyze transcriptome data of *Escherichia coli* obtained under three environmental conditions, namely, osmotic stress, starvation, and heat stress, presented in Ref. [9]. In the experiments, the cells are aerobically cultured with 5 ml of M63 minimal medium at 37°C in a shaking water bath. All transcriptome data are obtained from exponentially growing cells: After the transient response to the addition of given stress, the cells are harvested when the growth rate reaches a constant value. It should be stressed that throughout the paper, the point of interest is cellular behavior after recovery of the steady-growth state (which could be termed adaptation, even though it does not necessarily imply the optimization of the growth rate or the genetic change), and we use only experimental data available for the analysis of the steady-growth state. For each of these stress conditions, three levels of stresses ( $s = \text{high, medium, low}$ ) are used, so that the absolute expression levels, represented by  $x_j$  for the  $j$ th gene, are measured over a total of  $3 \times 3$  conditions in addition to the original (stress-free) condition. The exponential growth of cells is repeatedly tested under precisely controlled culture conditions, and the reproducibility of the results is confirmed (see Ref. [9] for details of the data). For each condition, messenger RNA (mRNA) is extracted from approximately  $10^8$  cells at this exponential growth phase, and absolute expression levels are quantified by Affymetrix’s high-density oligonucleotide microarrays by using the finite hybridization method [15]. The microarray

analyses are done in triplicate, and their averaged expression levels are used for the following analysis.

From the transcriptome analysis in different environmental conditions, we calculate the change in gene-expression levels between the original state and that of a system experiencing environmental stress. We investigate the difference in gene expression using a log scale ( $X_j = \log x_j$ ), that is,  $\delta X_j(E) = X_j(E) - X_j^0$  (i.e.,  $\log[x_j(E)/x_j^0]$ ) for genes  $j$ , where  $E$  represents a given environmental condition and  $X_j^0$  represents the log-transformed gene-expression level under the original condition. We adopt a logarithmic scale, as changes in gene expression typically occur on this scale, and also as it facilitates comparison with the theory described below.

To characterize global changes in the expression induced by these environmental stresses, we plot the relationship between the differences in the expression  $[\delta X_j(E_{s_1}^a), \delta X_j(E_{s_2}^a)]$  in Figs. 1(a)–1(c) for  $s_1 = \text{low}$  and  $s_2 = \text{medium}$ , where  $a$  is either osmotic, heat, or starvation stress. The relationships between all possible combinations of stresses and stress strengths are presented in Supplemental Fig. S1 [16]. For the same type of stress,  $[\delta X_j(E_{s_1}^a), \delta X_j(E_{s_2}^a)]$  correlates strongly over all genes, which suggests that the global trend in changes in expression levels can be represented by a small number of macroscopic variables.

### B. Theory for the steady-growth state

To discuss changes in the cellular state in response to environmental changes, we introduce a simple theory assuming a steady-growth state in a cell. When a cell grows at this steady state and reproduces itself, all the components it contains, e.g., the proteins that are expressed, have to be approximately doubled [2,18]. The abundance of each component increases at an almost equal rate over the time scale of cell division; if the growth rates of some components were lower than those of others, the component would become diluted over time, and after some divisions, the component would be “extinct,” so that the cell state would not accommodate steady growth. For a cell to maintain the same internal state, all the components have to be synthesized at the same rate across cell divisions. This steady-growth condition has to be satisfied among the nonlinearity, complexity, and stochasticity of biochemical reactions.

Consider a cell consisting of  $M$  chemical components, of which the synthesis allows it to grow and divide. In a cellular state under steady-growth conditions, the cell number increases exponentially over time, and so does the cell volume  $V$ , as given by  $dV/dt = \mu V$ . In a steady-growth cell, abundances of all the components increase with the same rate, so that the concentration of each component is preserved over a cell cycle.

This steady-growth constraint is formulated as follows: Let us denote  $x_i (> 0)$  of each component  $i = 1, \dots, M$ .

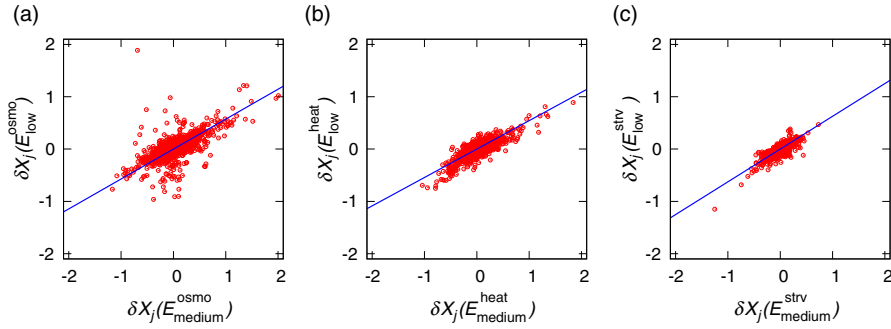


FIG. 1. Examples of the relationship between changes in gene expression  $\delta X_j(E_{s_1}^a)$  and  $\delta X_j(E_{s_2}^a)$  for genes in *E. Coli*.  $\delta X_j$  represents the difference in the logarithmic expression level of a gene  $j$  between the nonstressed and stressed conditions, where  $s_1$  and  $s_2$  represent two different stress strengths, i.e., low and medium. (a)–(c) The plots for  $a =$  osmotic pressure, heat, and starvation stress, respectively. The fitted line is obtained by the major axis method, which is a least-squares fit method that treats horizontal and vertical axes equally and is usually used to fit bivariate scatter data [17]. The slopes are 0.57, 0.54, and 0.62 for (a), (b), and (c), respectively. The expression data are obtained from Ref. [9]. Throughout the paper, we use the expression data of genes of which the expression levels under the three stress conditions as well as the original condition exceed a threshold ( $X_i > -1.5$ ), in order to exclude inaccurate data. (About 10% of the total genes are discarded from the analysis.)

Now, a cellular state is represented as a point in an  $M$ -dimensional state space. Since each component  $i$  is synthesized or decomposed in relationship to other components, the temporal change in the concentration of each component is represented as a function of the concentrations of the component itself and that of others, for instance, by the rate equation in chemical kinetics. Besides, all the concentrations are diluted by the rate  $(1/V)(dV/dt) = \mu$ , so that the time change of a concentration is given by

$$dx_i/dt = f_i(\{x_j\}) - \mu x_i. \quad (1)$$

For the sake of convenience, let us denote  $X_i = \log x_i$  and  $f_i = x_i F_i$ . Then, Eq. (1) can be written as

$$dX_i/dt = F_i(\{X_j\}) - \mu, \quad (2)$$

where we implicitly assume that  $x_i \neq 0$ ; i.e., all components remain to exist. Now, the stationary state is given by a corresponding fixed-point solution

$$F_i(\{X_j^*\}) = \mu \quad (3)$$

for all  $i$ .

In response to environmental changes, the term  $F_i(\{X_j\})$  as well as the growth rate  $\mu$  itself changes, as does each concentration  $x_i^*$ ; however, the  $M - 1$  condition requiring that  $F_i(\{X_j^*\})$  is independent of  $i$  for all  $i$  has to be satisfied. Thus, a cell has to stay at a one-dimensional curve in the  $M$ -dimensional space, under a given change in the environmental conditions [e.g., against changes in stress strength; see Fig. 2(a)]. With an environmental change, all concentrations,  $\mu$ , and  $\{X_j^*\}$  generally change, while the condition that  $F_i(\{X_j^*\})$  is independent of  $i$  is maintained as long as the cells continue steady-state growth [19].

Now, consider intracellular changes in response to environmental changes as being represented by a set of continuous parameters  $E^a$ , which denote environmental changes under the stress condition  $a$ . Here, each environmental change is parametrized by a single continuous parameter  $E^a$  (say, the temperature, osmotic pressure, and so forth), where discontinuous change is not considered. For the moment, we omit the stress type  $a$  in the notation. With this parametrization  $E$ , the steady-growth condition leads to  $F_i(\{X_j^*(E)\}, E) = \mu(E)$ .

We consider the parameter change from  $E_0$  to  $E$ , where each  $X_i^*$  changes from  $X_i^*$  at  $E_0$  to  $X_i^* + \delta X_i$ , which is accompanied by a change from  $\mu$  to  $\mu + \delta\mu$ . Assuming a gradual change in the dynamics  $x_i$ , we introduce a partial derivative of  $F_i(\{X_j^*(E)\})$  by  $X_j$  at  $E = E_0$ , which gives the Jacobi matrix  $J_{ij}$ . Now, considering the condition under

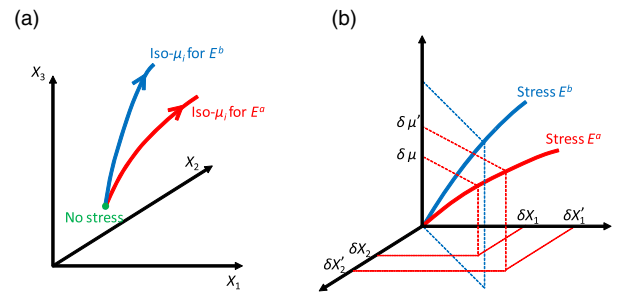


FIG. 2. Schematic representation of our theoretical analysis: (a) Changes in gene expression in a high-dimensional state space, under a given environmental stress, follow a curve satisfying the constraint that the growth rates of all components are identical, i.e., an iso- $\mu$  line  $F_1 = F_2 = \dots = F_M$ . For different environmental conditions, the locus in the state space follows a different iso- $\mu$  line. (b) Changes in expression for each gene  $1, 2, \dots$  are governed by the change in the growth rate  $\delta\mu$ . For different stress types, the change is shifted while governed by  $\delta\mu$ .

which the change is sufficiently small and taking only the linear term in  $\delta X_j$ , we get

$$\sum_j J_{ij} \delta X_j(E) + \gamma_i \delta E = \delta \mu(E) \quad (4)$$

with  $\gamma_i \equiv (\partial F_i / \partial E)$ . Under the linear conditions we are concerned with,  $\delta \mu \propto \delta E$ , so that  $\delta \mu = \alpha \delta E$  holds for a constant  $\alpha$ . Accordingly, we obtain

$$\sum_j J_{ij} \delta X_j(E) = \delta \mu(E) (1 - \gamma_i / \alpha). \quad (5)$$

Hence,  $[\sum_j J_{ij} \delta X_j(E) / \delta \mu(E)] = [\sum_j J_{ij} \delta X_j(E') / \delta \mu(E')]$  so that

$$\frac{\delta X_j(E)}{\delta X_j(E')} = \frac{\delta \mu(E)}{\delta \mu(E')} \quad (6)$$

over all  $j$  [see Fig. 2(b)]. One can confirm this relationship directly by inserting it into Eq. (5), while if the inverse matrix  $L = J^{-1}$  exists [20], the relationship can also be derived from

$$\delta X_j(E) = \delta \mu(E) \times \sum_i L_{ji} (1 - \gamma_i / \alpha). \quad (7)$$

Now, we compare the above formula with the experimental observation mentioned already. We make a remark here: As long as we are concerned with the stationary solution, the above formula can be applied to any component. For example, one can use either the concentration of mRNA or protein, depending on the feasibility in experimental data [21]. Even though the relaxation time scale in the equation (1) as well as the general trend in the concentration against the growth rate may be different between mRNA and protein [22], the general form (6) can be applied to each species as for the stationary growth.

It should be recalled again that  $X_j$  is the logarithmic concentration in any component in a cell. Hence, the change in the expression  $X_j$  in response to external change is proportional over all components  $j$  in this form. This relationship provides a possible explanation for the observed transcriptome analysis shown in Fig. 1. According to our theory, the proportion coefficient in the expression level should agree with the growth rate. Here, for each condition, the change in the growth rate  $\delta \mu(E_s^a)$  is also measured. ( $a$  is either osmotic, heat, or starvation stress.) In Fig. 3, we compare the slope of the changes in gene expression, i.e., the common ratio  $\delta X_j(E_{s_1}^a) / \delta X_j(E_{s_2}^a)$ , with  $\delta \mu(E_{s_1}^a) / \delta \mu(E_{s_2}^a)$ . The plot shows rather good agreement between these two. In this respect, the theory based on the steady-growth state and linearization of changes in stress applies well to the transcriptome change.

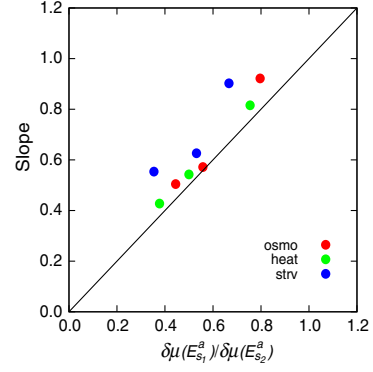


FIG. 3. Relationship between the slope of the change in gene expression and the change in growth rate for the same stress types. The abscissa denotes  $\delta \mu(E_{s_1}^a) / \delta \mu(E_{s_2}^a)$ , while the ordinate is the slope in  $[\delta X_j(E_{s_1}^a), \delta X_j(E_{s_2}^a)]$ . The red, green, and blue dots represent osmotic, heat, and starvation stress data, respectively, while the pair  $s_1, s_2$  runs over different strengths of the same stress type.

### C. Changes in gene expression across different types of stresses

So far, we have compared the expression levels across different strengths for the same type of stress. However, expression changes can also be compared across different stress conditions. Interestingly, the genome-wide correlation of expression levels is not restricted to a change in the same stress condition. In Fig. 4,  $[\delta X_j(E_{\text{high}}^a), \delta X_j(E_{\text{high}}^b)]$ , which plots expression changes across different stress conditions  $a \neq b$  (= either starvation, heat, or osmotic stress), correlation is still observed, even though there are more genes that deviate from the common proportionality, leading to lower correlation coefficients, as compared with the correlations observed under the same stress conditions.

The correlation is also discernible for other choices of  $s_1$  and  $s_2$ , as shown in Supplemental Fig. S1 [16], where all the correlation diagrams of  $[\delta X_j(E_{s_1}^a), \delta X_j(E_{s_2}^b)]$  across all possible stress conditions are plotted. Note that such proportionality across genes has also been suggested for several experiments, over different environmental conditions [6,7]. The finding of correlation, even with reduced proportionality, implies a common trend in changes in expression across many genes, which is not necessarily the result of a given stress condition but is a concept that holds across different environmental conditions.

Since gene-expression dynamics are very high dimensional, this correlation suggests the existence of a strong constraint to adaptive changes in expression dynamics. Below, we discuss the theoretical origin of this correlation. In Eqs. (5) and (6), the environmental change  $E$  is no longer represented by a scalar variable, but the environmental change involves a different direction, so that  $\gamma_i^a$  and  $\alpha^a$  depend on the type of environmental (stress) condition  $a$ . Hence, instead of Eqs. (6) and (7), we get



$$\frac{\delta X_j(E^a)}{\delta X_j(E^b)} = \frac{\delta\mu(E^a) \sum_i L_{ji} (1 - \gamma_i^a / \alpha^a)}{\delta\mu(E^b) \sum_i L_{ji} (1 - \gamma_i^b / \alpha^b)}. \quad (8)$$

Here, the right-hand side (rhs), in general, depends on each gene  $j$ , which could blur the proportionality in  $[\delta X_j(E^a), \delta X_j(E^b)]$  over all genes. In the following case, however, the dependence of the rhs on  $j$  is relaxed, to support approximate proportionality, as indicated in Fig. 4. When  $\gamma_i^a$  and  $\gamma_i^b$  are independent of  $i$ , which we denote as  $\gamma^a$  and  $\gamma^b$ , respectively, the rhs is reduced to

$$\frac{\delta\mu(E^a) (1 - \gamma^a / \alpha^a)}{\delta\mu(E^b) (1 - \gamma^b / \alpha^b)}, \quad (9)$$

so that the common proportionality of the change in expression holds, while the proportion coefficient is shifted from a simple ratio between the growth-rate changes  $\delta\mu$ .

Indeed, environmental changes sometimes globally affect all processes. For example, if temperature or nutrient resources are increased, the synthesis (or decomposition) rates of all reaction processes are amplified across the board.

Of course, there are some genes for which  $\gamma_i^E$  deviates from the above common value. If the number of such genes with a specific response is small (and /or its influence on other genes is small; i.e., the Jacobi matrix is sparse), then the contributions from genes with a common  $\gamma^E$  value make up the major portion of the summation in the rhs of Eq. (9). If we neglect the minor contributions from a few specific genes, common proportionality could generally be maintained. Indeed, only a limited number of specific genes is expected to respond directly to environmental changes.

According to this approximation, the proportion coefficient  $\delta X_j(E^a) / \delta X_j(E^b)$  deviates from  $\delta\mu(E^a) / \delta\mu(E^b)$  by

the factor  $(1 - \gamma^a / \alpha^a) / (1 - \gamma^b / \alpha^b)$ . Note that this correction in the proportion coefficient depends only on the type but not on the strength of each stress.

We examine this point from the transcriptome data analyzed here by plotting the proportion coefficient in  $\delta X_j(E^a) / \delta X_j(E^b)$  versus  $\delta\mu(E^a) / \delta\mu(E^b)$  in Fig. 5. The correlation between  $\delta X_j$  and the growth rate in this figure also exists across different stress conditions. Additionally, the coefficient  $\delta X_j(E^a) / \delta X_j(E^b)$  does not agree with but is roughly proportional to  $\delta\mu(E^a) / \delta\mu(E^b)$  with a proportion coefficient that is roughly common, given the pair of stress types  $a$  and  $b$  over different strengths, as is consistent with the common correction term  $(1 - \gamma^a / \alpha^a) / (1 - \gamma^b / \alpha^b)$  in the approximate theory. As another possible estimate of this factor  $\gamma^a / \alpha^a$ , we directly measure the variance of changes in expression across genes, i.e.,  $\langle [\delta X_j(E) - \langle \delta X_j(E) \rangle]^2 \rangle$ , where the angled brackets represent the average over all genes (see Supplemental Fig. S2 [16]). According to Eqs. (8) and (9), this factor grows in proportion to  $(1 - \gamma^a / \alpha^a)^2 \delta\mu^2$  (in addition to the variances determined by the Jacobi matrices, which are independent of environmental stress  $a$ ). As shown in Supplemental Fig. S2 [16], the variance, and accordingly, the factor  $(1 - \gamma^a / \alpha^a)$  decrease in the order of  $a =$  osmotic stress, starvation, and heat stress. Indeed, the deviation from  $\delta\mu^a / \delta\mu^b$  in Fig. 5 is consistent with the above order of  $(1 - \gamma^a / \alpha^a)$ .

Undeniably, the proportionality over different stress types is not optimal. Indeed, the existence of gene-specific dependence  $\gamma_i^a$  leads to scattering in  $[\delta X_j(E^a), \delta X_j(E^b)]$  (for  $a \neq b$ ) around the common proportionality by genes, and there are more genes that deviate from the common proportionality for  $a \neq b$  than those for  $a = b$  (compare Fig. 4 with Fig. 1), so that the estimation of the proportion coefficient in Fig. 5 is not so reliable, especially for those with a lower correlation coefficient.

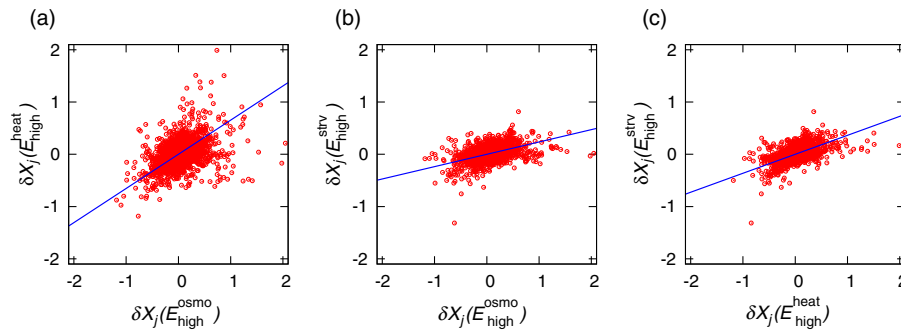


FIG. 4. Examples of the relationship between changes in gene expression  $\delta X_j(E_{s_1}^a)$  and  $\delta X_j(E_{s_2}^b)$  for different stress types. The combination of stresses  $(a, b)$  is (osmotic, heat) for (a), (osmotic, starvation) for (b), and (heat, starvation) for (c), respectively. The strengths of the stress  $s_1$  and  $s_2$  are fixed as high in these figures. The slopes are 0.65, 0.24, and 0.36 for (a), (b), and (c), respectively, while the correlation coefficient for each datum is (a) 0.40, (b) 0.43, and (c) 0.54. The relationships between the changes in gene expression for all possible combinations are presented in Supplemental Fig. S1 [16]. The fitted line is obtained by the major axis method as described in Fig. 1.

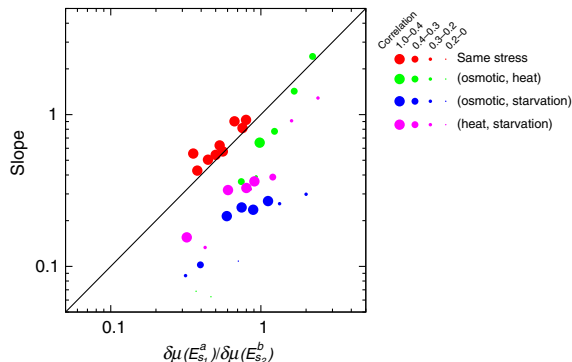


FIG. 5. Relationship between the slope of change in gene expression and change in growth rate for different stress types. The abscissa denotes  $\delta\mu(E_{s_1}^a)/\delta\mu(E_{s_2}^b)$ , while the ordinate is the slope in  $[\delta X_j(E_{s_1}^a), \delta X_j(E_{s_2}^b)]$ . Log-log plot: The red dots represent data for the same stress types, while the green, blue, and purple dots show combinations of different stress types  $(a, b) = (\text{osmotic, heat})$ ,  $(\text{osmotic, starvation})$ , and  $(\text{heat, starvation})$ , respectively. The pair  $s_1, s_2$  runs over different strengths of all stress combinations. The size of the dots represents the correlation coefficient between  $\delta X_j(E_{s_1}^a)$  and  $\delta X_j(E_{s_2}^b)$ . A lower correlation indicates that the fit of the slope may be less accurate. The proportionality between the slope and  $\delta\mu(E^a)/\delta\mu(E^b)$  also holds approximately for different pairs  $a, b$  (i.e., over the same-color dots), while the proportion coefficient decreases in the order of (osmotic, heat), (heat, starvation), and (osmotic, starvation).

### III. DISCUSSION

We have shown here that steady-growth conditions lead to a global constraint over all gene-expression patterns. With a few additional assumptions, the proportionality in the change in expression across genes can be derived, in which the proportion coefficient is mainly governed by the change in the growth rate. These theoretical predictions are compared with several bacterial gene-expression experiments, with approximate agreement.

As an alternative origin of the correlation in gene-expression changes, one might assume the contribution from a coordinated regulatory program. To examine this point, we have adopted the transcriptome data of transient gene expression [10] before the cells reach steady exponential growth upon the application of the heat stress and plotted the correlation in gene-expression changes in Supplemental Fig. S3 [16]. As shown therein, the correlation for the transient response is much worse (Pearson's correlation coefficient  $R \sim 0.3$ ) than that in the steady exponential growth regime ( $R \sim 0.9$ ). This fact demonstrates that the balanced exponential growth, rather than the coordinated regulatory program, mainly contributes to the correlation. Furthermore, the agreement of the slope in the exponential growth regime with the growth-rate change (see Fig. 3) also supports the relevance of the growth-rate constraint.

The correlation with the growth rate is also interpreted by neglecting the direct environment dependence in  $F_i(\{X_j(E)\}, E)$ , i.e., by replacing it with  $F_i(\{X_j(E)\})$  for most genes  $j$ . In other words, external environmental changes trigger changes in the levels of some components  $\{x_m(E)\}$ , which introduces a change in the growth rate  $\mu(E)$ . For the stationary state, only the condition  $F_i(\{x_j(E)\}) = \mu(E)$  is considered. With this approximation, the term  $\gamma_i \delta E$  for direct environmental changes (with  $\gamma_i = \partial F_i / \partial E$ ) is neglected, and Eq. (6) follows directly, so that growth-rate changes determine gene-expression changes globally. Indeed, the experimental data may suggest that the growth rate makes the major contribution to changes in gene expression.

This dominance of growth rate is, however, imperfect, so that the environment-specific term  $\gamma_i^a$  has to be taken into account to compare expression of genes across different stress conditions. In our simple approximation that neglects the gene dependence of  $\gamma_i^a$ ,  $\gamma^a/\alpha^a$  represents the degree of the direct influence of the environment on gene-expression dynamics, as compared with the influence on the growth rate.

Furthermore, the environment-specific response of gene expression  $\gamma_i^a$  generally depends on each gene  $i$ . Here, genes that show specific responses to a given environment  $E^a$  may be few, while most others may not be influenced directly by the environment; their expression levels may mostly be determined by the homeostatic growth condition  $F_i\{X_j(E)\} = \mu(E)$ . Distinguishing such homeostatic genes from those that show specific responses to individual environmental stresses will be important as the next step in the statistical analysis of adaptation.

In Eq. (1), we have not assumed any specific form with particular dependence upon some protein species. In a recent study, the specific dependence of the fraction of ribosomal proteins upon the growth rate is discussed by adopting a description by a few degrees of protein groups [23]. It will be interesting to introduce some specific genes in our formulation while keeping high-dimensional expression dynamics.

It is also interesting to note that gene-expression changes  $\delta X_j$  across genes correlate between environmental and genetic perturbations. In fact, Ying *et al.* [10] measured the changes in gene expression induced by the environmental perturbation  $E^{\text{env}}$  and the genetic perturbation  $E^g$  induced by the external reduction of several genes. Again, they observed a strong correlation between  $\delta X_j(E^{\text{env}})$  and  $\delta X_j(E^g)$  across genes (see Fig. 5 of Ref. [10]).

Indeed, our theory can also be applied to adaptive evolution, in which the growth rate is first reduced by encountering a novel environment  $E^{\text{env}}$  and then is recovered by genetic changes  $E^g$  through evolution, so that  $\delta\mu(E^g)/\delta\mu(E^{\text{env}}) < 1$ . According to our relationship, changes in gene-expression levels introduced by a new environment are reduced through adaptive evolution, i.e.,

$\delta X_i(E^g)/\delta X_i(E^{\text{env}}) \sim \delta\mu(E^g)/\delta\mu(E^{\text{env}}) < 1$ , as discussed. In other words, there is a common homeostatic trend for the expression of most genes to return to the original level, as extensively observed experimentally [8,24].

A few issues should be considered prior to the application of the theory presented here. First, it must be assumed that the components continue to exist and that novel components do not appear (to divide by  $x_j \neq 0$ ). Under this condition, the condition (3) (i.e., the common growth rate for all components) holds, even though the linear approximation does not. However, even if some components do become extinct or novel components emerge, the constraint may still exist for other components, and the proportionality relationship [Eq. (6)] holds approximately, as long as the influence of the extinct or emerging components is limited.

Second, it is assumed that the fixed point of Eq. (1) is not split by bifurcation. When bifurcation occurs, we can apply our theory along each branch, but direct comparisons cannot be made across different branches. Moreover, in some cases, the attractor of the expression level is not a fixed point but is an oscillatory state. However, as long as the oscillation period is shorter than the cell-division time, one can use the average  $\mu_i$  over the period, instead of  $\mu_i$ , leaving the present argument valid.

Third, we adopted a linearization approximation to obtain Eq. (6). For larger changes in external conditions, there will be a gene-specific correction to the linear relationship [Eq. (5)]. However, linearization is adopted after taking the logarithm of gene-expression levels, so that the size of  $\delta X_i$  may not be so restrictive when seen in the original scale of gene expression  $x_i$ . Indeed, the agreement with the theory shown in Figs. 1 and 3 for the same stress indicates that the linearization approximation is valid, even though the growth rate is reduced to less than half of the original.

The present theory facilitates description of a cellular system with only a few macroscopic variables, for the characterization of adaptation and evolution. Furthermore, our theory with regard to common  $\mu_i$  can be applied to any system of stationary growth. As presented, each element  $i$  represents a replicating molecule within a cell, but, similarly, we can apply our theory by using such an element to describe cells of different types within an organism. Alternatively, macroscopically, one can assign an element as a population of each species in a stationary ecosystem. The multilevel constraint of the steady-growth condition across a hierarchy is an important concept for elucidating global relationships in complex-systems biology [2].

#### ACKNOWLEDGMENTS

This work was supported in part by a platform for Dynamic Approaches to the Living Systems from the Ministry of Education, Culture, Sports, Science, and

Technology (MEXT), Japan. This work was also supported in part by a Grant-in-Aid for Scientific Research on Innovative Areas (No. 25128715 and No. 26119719 to C. F.) from MEXT, Japan.

- 
- [1] C. H. Waddington, *The Strategy of the Genes* (Allen & Unwin, London, 1957).
  - [2] K. Kaneko, *Life: An Introduction to Complex Systems Biology* (Springer, Heidelberg, 2006).
  - [3] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, *Cluster Analysis and Display of Genome-Wide Expression Patterns*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14863 (1998).
  - [4] T. S. Gunasekera, L. N. Csonka, and O. Paliy, *Genome-Wide Transcriptional Responses of Escherichia coli K-12 to Continuous Osmotic and Heat Stresses*, *J. Bacteriol.* **190**, 3712 (2008).
  - [5] C. S. Richmond, J. D. Glasner, R. Mau, H. Jin, and F. R. Blattner, *Genome-Wide Expression Profiling in Escherichia coli K-12*, *Nucleic Acids Res.* **27**, 3821 (1999).
  - [6] S. Stern, T. Dror, E. Stolovicki, N. Brenner, and E. Braun, *Genome-Wide Transcriptional Plasticity Underlies Cellular Adaptation to Novel Challenge*, *Mol. Syst. Biol.* **3**, 106 (2007).
  - [7] L. Keren *et al.*, *Promoters Maintain Their Relative Activity Levels under Different Growth Conditions*, *Mol. Syst. Biol.* **9**, 701 (2013).
  - [8] S. M. Carroll and C. J. Marx, *Evolution after Introduction of a Novel Metabolic Pathway Consistently Leads to Restoration of Wild-Type Physiology*, *PLoS Genet.* **9**, e1003427 (2013).
  - [9] Y. Matsumoto, Y. Murakami, S. Tsuru, B. Y. Ying, and T. Yomo, *Growth Rate-Coordinated Transcriptome Reorganization in Bacteria*, *BMC Genomics* **14**, 808 (2013).
  - [10] B. W. Ying, S. Seno, F. Kaneko, H. Matsuda, and T. Yomo, *Multilevel Comparative Analysis of the Contributions of Genome Reduction and Heat Shock to the Escherichia coli Transcriptome*, *BMC Genomics* **14**, 25 (2013).
  - [11] D. Chen, *Global Transcriptional Responses of Fission Yeast to Environmental Stress*, *Mol. Biol. Cell* **14**, 214 (2003).
  - [12] S. Bergmann, J. Ihmels, and N. Barkai, *Similarities and Differences in Genome-Wide Expression Data of Six Organisms*, *PLoS Biol.* **2**, e9 (2004).
  - [13] B. Lehner and K. Kaneko, *Fluctuation and Response in Biology*, *Cell Mol. Life Sci.* **68**, 1005 (2011).
  - [14] S. Marguerat *et al.*, *Quantitative Analysis of Fission Yeast Transcriptomes and Proteomes in Proliferating and Quiescent Cells*, *Cell* **151**, 671 (2012).
  - [15] N. Ono, S. Suzuki, C. Furusawa, T. Agata, A. Kashiwagi, H. Shimizu, and T. Yomo, *An Improved Physico-chemical Model of Hybridization on High-Density Oligonucleotide Microarrays*, *Bioinformatics* **24**, 1278 (2008).
  - [16] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevX.5.011014> for relationships in gene expression change for all combinations [S1], the variance change against the growth rate change

- [S2], and relationship between expression changes in steady and transient states [S3], as well as detailed analysis with mRNA dynamics.
- [17] D. I. Watson, I. J. Wright, D. S. Falster, and M. Westoby, *Bivariate Line-Fitting Methods for Allometry*, Biol. Rev. Camb. Philos. Soc. **82**, 259 (2006).
- [18] C. Furusawa and K. Kaneko, *Zipf's Law in Gene Expression*, Phys. Rev. Lett. **90**, 088102 (2003).
- [19] After changes in the environment, there may be a transient period during which the cells have not yet attained this steady-growth state, during which each component may grow with a different rate  $F_i$ , to adjust the steady-state concentration. After this transient state, the steady state is attained, where over time, the constraint  $F_1 = F_2 = \dots = F_M$  is satisfied, so that the locus of the steady-growth cellular state is on a one-dimensional curve.
- [20] Equation (6) satisfies Eq. (5), even if  $L$  does not exist, while there is subtlety in the uniqueness of the solution in that case, and a nonlinear term may be needed in the selection from degenerated solutions.
- [21] In the case of mRNA, one may expect that the degradation speed is much faster than the dilution. Recall, however, that  $F$  includes both the synthesis and degradation. Even if the degradation may be large, the dilution also exists, so that the present formula can also be applied directly for the concentration of mRNA. For example, the synthesis, degradation, and dilution of mRNA and protein are often modeled as  $dx_i^{\text{mRNA}}/dt = F_i^{\text{mRNA}} - \mu x_i^{\text{mRNA}}$  with  $F_i^{\text{mRNA}} = F_i^{\text{synthesis}}(\{x_j^{\text{protein}}\}) - \gamma_i^{\text{mRNA}} x_i^{\text{mRNA}}$  and  $dx_i^{\text{protein}}/dt = F_i^{\text{protein}} - \mu x_i^{\text{protein}}$  with  $F_i^{\text{protein}} = k_i x_i^{\text{mRNA}} - \gamma_i^{\text{protein}} x_i^{\text{protein}}$ . As our argument can be applied to any form of  $F_i$ , we get  $\delta X_i^{\text{mRNA}}(E) \propto \delta \mu(E)$  as well as  $\delta X_i^{\text{protein}}(E) \propto \delta \mu(E)$ . If the degradation for mRNA (protein) is larger (smaller) than the dilution, respectively, there is a tendency in the steady-state solution that the concentration of mRNA increases with  $\mu$ , while that of protein does not so much [22]. Then, the relationship (6) on the relative change  $\delta X_j(E)/\delta X_j(E')$  generally holds. Indeed, even if the degradation of mRNA is much larger and its dilution is negligible, the relationship (6) is shown to be valid, as is also demonstrated in the Supplemental Material [16].
- [22] H. Bremer and P. P. Dennis, *Escherichia coli and Salmonella: Cellular and Molecular Biology*, edited by F. C. Neidhardt (ASM Press, Washington, DC, 1996), p. 1553–1569.
- [23] M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa, *Interdependence of Cell Growth and Gene Expression: Origins and Consequences*, Science **330**, 1099 (2010).
- [24] T. Kishimoto *et al.*, *Transition from Positive to Neutral in Mutation Fixation along with Continuing Rising Fitness in Thermal Adaptive Evolution*, PLoS Genet. **6**, e1001164 (2010).