Constructing universal phenomenology for biological cellular systems
an idiosyncratic review on evolutionary dimensional reduction
Kaneko, Kunihiko

Published in:
Journal of Statistical Mechanics: Theory and Experiment

DOI:
10.1088/1742-5468/ad1f54

Publication date:
2024

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Constructing universal phenomenology for biological cellular systems: an idiosyncratic review on evolutionary dimensional reduction

To cite this article: Kunihiko Kaneko J. Stat. Mech. (2024) 024002

View the article online for updates and enhancements.
Constructing universal phenomenology for biological cellular systems: an idiosyncratic review on evolutionary dimensional reduction

Kunihiko Kaneko

Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, 2100 Copenhagen, Denmark
Universal Biology Institute, University of Tokyo, 3-8-1 Komaba, Tokyo 153-8902, Japan

E-mail: kaneko@complex.c.u-tokyo.ac.jp

Received 17 October 2023
Accepted for publication 10 January 2024
Published 13 February 2024

Abstract. The possibility of establishing a macroscopic phenomenological theory for biological systems, akin to the well-established framework of thermodynamics, is briefly reviewed. We introduce the concept of an evolutionary fluctuation–response relationship, which highlights the tight correlation between the variance in phenotypic traits caused by genetic mutations and by internal noise. We provide a distribution theory that allows us to derive these relationships, which suggests that the changes in traits resulting from adaptation and evolution are considerably constrained within a lower-dimensional space. We explore the reasons behind this dimensional reduction, focusing on the constraints posed by the requirements for steady growth and robustness achieved through the evolutionary process. We draw support from recent laboratory and
Numerical experiments to substantiate our claims. The universality of evolutionary dimensional reduction is presented and potential theoretical formulations for it are discussed. We conclude by briefly considering the prospects of establishing a macroscopic framework that characterizes biological robustness and irreversibility in cell differentiation, as well as an ideal cell model.

**Keywords:** evolutionary processes, fluctuation phenomena, gene expression and regulation, systems biology

### Contents

1. **Possibility of a phenomenological theory for biological systems: can we learn from the success of thermodynamics?** ............................................ 2
2. **Background—why can we expect macroscopic universality in biology: evolutionary fluctuation–response relationship** .......................... 4
3. **Restriction to steady growth instead of equilibrium** .......................... 8
4. **Dimensional reduction as a consequence of robustness and plasticity** 11
5. **Universality of evolutionary dimensional reduction** .......................... 13
6. **Towards statistical physics theory for evolutionary dimensional reduction** ................................................................. 15
7. **Constructing universal biology** .......................................................... 16
   7.1. **Ideal cell model?** ................................................................. 16
   7.2. **Multicellularity and irreversible differentiation** .......................... 17

**Acknowledgments** ............................................................................. 18

**References** .......................................................................................... 18

1. **Possibility of a phenomenological theory for biological systems: can we learn from the success of thermodynamics?**

Biological systems invariably comprise a multitude of components. Even the simplest cells include thousands of chemical species. The abundances of each component can change over time, and as they increase, cells divide and reproduce themselves. This prompts a fundamental question: can we understand the dynamical systems of such a huge number of chemical species, without going into a detailed description of each molecular process? Note that the hierarchy featuring diversity in micro-level constituents is not restricted to molecules and cells; instead, it recurs at every scale in biological systems. In multicellular organisms, developmental processes lead to diverse cell types.

https://doi.org/10.1088/1742-5468/ad1f54
that coexist and sustain themselves. In ecosystems, diverse species coexist and, in general, it seems hopeless to try to understand such complex and diverse systems without some kind of dimensional reduction process.

In the history of science, however, we find a shining success story: thermodynamics. Despite the immense number of molecules constituting the system (as we now know), a macroscopic phenomenological theory was formulated using only a few degrees of freedom and focussing on stable equilibrium states. This historical triumph provides a hopeful precedent for the prospect of constructing a phenomenological theory for complex biological systems, such as cells consisting of a huge number of chemical species.

Of course, one could reject such optimism: unlike typical thermodynamic systems, a biological cell is not in a static equilibrium. Its state changes over time, often with increases in the abundance of chemical components leading to cell growth and division. Moreover, instead of a huge number of molecules in thermodynamics, the number of molecular species is huge in a cell. Nonetheless, parallels do exist: biological systems are not at equilibrium; instead, they frequently attain stationary or steady-growth states. Both thermal equilibrium states and most biological states exhibit stability. In most biological systems, their internal states have a tendency to return to the original state after perturbations, which is known as adaptation. (Here we do not intend to apply thermodynamics or statistical physics to biology, but borrow its spirit to formulate a universal phenomenology for biology.)


Within the past few decades, quantitative measurements of cellular states have advanced significantly thanks to techniques such as fluorescent imaging, transcriptome analysis, flow cytometry and next-generation sequencing. Furthermore, mounting evidence has emerged supporting the characterization of cellular states in terms of a few degrees of freedom. The time is now ripe to embark on a quantitative exploration of a general theory bridging the microscopic and macroscopic levels, with statistical physics poised to play a pivotal role. In this context, we will survey the possibility of a universal theory for cellular states, with recent experimental progress in mind.

This paper gives an idiosyncratic, brief review of the current status of this evolving field, termed as universal biology. In the subsequent section, we provide a concise overview of the evolutionary fluctuation–response relationship, which hints at the possibility of representing cells, despite their multitude of components, within a few-dimensional space. We then elucidate how the consistency between the molecular and cellular levels, coupled with the robustness of evolved states, drives a dimensional reduction from a high-dimensional phase space. Furthermore, we show how this evolutionary dimensional reduction imposes constraints and directions on phenotype evolution, which is supported
by experimental verification. After discussing the universality of evolutionary dimensional reduction and the potential for related theories, we briefly review several other topics in universal biology that span larger spatial and temporal scales.

2. Background—why can we expect macroscopic universality in biology: evolutionary fluctuation–response relationship

The first insight into the potential application of macroscopic phenomenology to biological systems is gained through the evolutionary fluctuation–response relationship: the phenotypic response throughout the evolutionary course is proportional to the phenotypic variance induced by noise [6, 7].

First, recall that the phenotypes, say the concentrations of chemicals, vary over isogenic cells that share the same genes. For statistical physicists, this is not surprising, as the abundances of each component in a cell are not large and, for cells to grow, components need to be amplified by a positive feedback process. Such isogenic variance of each component, however, does not follow an inheritance pattern similar to that of genes. Hence, if such variance is relevant to evolution is not self-evident.

In a related context, an evolution experiment involving mutation and selection was recently carried out to enhance certain properties. For instance, by inserting genes to synthesize a protein to generate fluorescence, and by mutating the genes, experiments to select those cells with a higher fluorescence were performed [6]. In this case, phenotypic evolution speed was measured as the increase in fluorescence level per each generation upon mutation (genetic change). Here, note that for cells with the same genes, the phenotype (i.e. the level of fluorescence in this case) fluctuates, as mentioned previously. By comparing the variance across generations in the evolution, it was found that the evolution speed was positively correlated with the variance. Furthermore, by using a toy cell model with thousands of components [8, 9] and conducting a numerical evolution to increase the concentration of a certain component, it was found that the evolution speed was proportional to the variance [7, 10].

This result is reminiscent of the fluctuation–response relation in statistical physics. In this relation, the ratio of the response of a state to an external force is proportional to the variance of the state fluctuation in the absence of the force. In the evolution experiment above, the change in phenotype (e.g. fluorescence) due to the mutation-selection process is proportional (correlated) to the variance before the evolution process. The former may correspond to the response to generalized, evolutionary force and the latter to the variance without such a force. Of course, the fluctuation–response relationship in statistical physics is valid in (the vicinity of) thermal equilibrium. Therefore, we cannot apply it directly to evolution. Nevertheless, one may consider the possibility of a phenomenological formulation of the evolutionary fluctuation relation, inspired by the above-mentioned similarity.

Recalling the linear response theory in statistical physics, a possible extension of the fluctuation–response relationship to phenotypic evolution was proposed as follows [6]. Consider a system characterized by a gene parameter $a$ and a phenotypic trait $X$. We can then evaluate the change in $X$ against the change in the value of the gene parameter from $a$ to $a + \Delta a$. We then assume the existence of the distribution $P(X; a)$, given by a

https://doi.org/10.1088/1742-5468/ad1f54
scalar parameter \( a \) for genes. We further assume that it is approximately Gaussian and that the effect of a change in \( a \) on the distribution is represented by a bilinear coupling between \( X \) and \( a \). (In other words, we choose a variable whose distribution is closer to Gaussian, which can be achieved, for example, by appropriate variable transformation, while the bilinear form may be applicable by assuming that the change due to evolution is not so large (this issue will be revisited in later sections).) With this assumption, the distribution is written as

\[
P(X; a) = N_0 \exp \left( -\frac{(X - X_0)^2}{2\alpha} + v(X, a) \right),
\]

with \( N_0 \) a normalization constant so that \( \int P(X : a) dX = 1. \)

Here, \( X_0 \) is the peak value of the variable \( X \) at \( a = a_0 \), which agrees with the average value of \( X \), denoted by \( \bar{X}(a) = \int XP(X; a) dX \), whereas the term \( v(X, a) \) gives a deviation from the distribution at \( a = a_0 \). This deviation \( v(X, a) \) can be expanded as \( v(X, a) = C(a - a_0)(X - X_0) + \ldots \), with \( C \) as a constant, where \( \ldots \), a higher order term in \( (a - a_0) \) and \( (X - X_0) \), which will be neglected in the following analysis.

Assuming this distribution form, \( \bar{X}(a + \Delta a) \) is computed from \( \int X \exp(-\frac{(X - X_0 - C\alpha\Delta a)^2}{2\alpha}) dX \). Then, the change in the average value \( \bar{X} \) following the change of the parameter from \( a_0 \) to \( a_0 + \Delta a \) is represented by

\[
\frac{\bar{X}(a_0 + \Delta a) - \bar{X}(a_0)}{\Delta a} = C\alpha \propto \delta X^2,
\]

by recalling that the variance \( \delta X^2 \) agrees with \( \alpha \).

The above explanation, of course, is not a derivation. We assumed Gaussian-like distributions\(^2\), and that there is a bilinear coupling term between the parameter and variable, which brings about a shift in the average of the corresponding variable. Empirically, however, several model simulations \( [7, 10] \) and experiments support this type of evolutionary fluctuation–response relationship \( [6, 12] \). In this sense, the above distribution provides a phenomenological formulation of the empirical observations, by borrowing the concept from statistical physics. One may also expect the existence of a potential function \( U(X, a) \) such that \( P(X, a) \propto \exp(-U(X, a)) \), as discussed later.

**Mystery** This evolutionary fluctuation–response relationship is associated with the phenotypic variance \( V_{ip} \) of isogenic individuals, which is caused by noise. In standard population genetics, in contrast, the phenotypic variance due to genetic variation, named genetic variance \( V_g \), is considered. In fact, the so-called fundamental theorem of natural selection proposed by Fisher \( [13, 14] \) states that the evolutionary speed is proportional to \( V_g \). This relationship is obtained straightforwardly, as the changes in genes

\(^1\) For the lowest order of the change, we neglect the dependence of \( \alpha \) on \( a \).

\(^2\) Often the distribution is far from Gaussian. If the variance of the distribution in concern is finite, however, the distribution can be transformed to be nearly Gaussian, by suitable choice of transformation of the variable. For example, if a measured variable \( \tilde{X} \) (say the concentration of some protein) follows a log-normal distribution as often observed in biological systems \( [11] \), then, one can just adopt \( X = \log(\tilde{X}) \) as a phenotype variable in concern. In fact, in most cases, log(concentration) or log(fluorescence) is adopted as the variable.
are inherited to the offspring and those from the mother that has a higher fitness for survival can also have a higher fitness\(^3\). In contrast, the aforementioned evolutionary fluctuation relationship concerns the variance caused by noise across isogenic cells. Those cells that happen to take a higher fitness value due to noise cannot necessarily produce offspring with such high fitness. Hence, the evolutionary fluctuation–response relationship is distinct from Fisher’s theorem. If both relationships are valid, however, this implies that \(V_{ip}\) and \(V_g\) are proportional throughout the evolutionary course. Both are variances of phenotype, but the origin for the fluctuation is different: the former is due to noise, the latter due to genetic changes. Is the proportionality valid? If so, how can it be?

We first examined this proportionality relationship between these two variances through evolutionary simulations of a catalytic reaction network and a gene regulation network model [10, 15], which demonstrated its validity.

Then, to formulate a possible relationship between the two variances, a phenomenological theory was proposed. For this, we revised the previous distribution function \(P(X; a)\) to include the distribution of genes \(a\). Here, the phenotype is (stochastically) determined from the gene \(a\), but there is a feedback process from phenotype to genotype through the selection process. Hence, we assume a two-variable distribution \(P(X, a)\) both for phenotype \(X\) and genotype \(a\).

By using this distribution, \(V_{ip}\), the variance of \(X\) of the distribution for a given \(a\), can be written as \(V_{ip}(a) = \int (X - \bar{X}(a))^2 P(X, a) dX\), where \(\bar{X}(a)\) is the average phenotype of a clonal population sharing genotype \(a\), namely \(\bar{X}(a) = \int P(X, a) X dX\). \(V_g\) is defined as the variance of the average \(\bar{X}(a)\), over genetically heterogeneous individuals and is given by \(V_g = \int (\bar{X}(a) - \langle X \rangle)^2 p(a) da\), where \(p(a)\) is the distribution of genotype \(a\) and \(\langle X \rangle\) as the average of \(\bar{X}(a)\) over all genotypes.

We now make another assumption regarding robust and gradual evolutionary processes, and introduce an evolutionary stability hypothesis, as follows: through the course of the (gradual) evolution, the distribution keeps a single peak in the \((X, a)\) space, which leads to a Hessian condition for the derivative of \(P(X, a)\). It can be easily formulated by assuming a Gaussian distribution for \(P(X, a)\) written as

\[
P(X, a) = \hat{N} \exp \left[ -\frac{(X - X_0)^2}{2\sigma^2(a)} + C(X - X_0)(a - a_0) - \frac{1}{2\mu} (a - a_0)^2 \right],
\]

with \(\hat{N}\) as a normalization constant. The Gaussian distribution \(\exp(-\frac{1}{2\mu}(a - a_0)^2)\) represents the distribution of genotypes around \(a = a_0\), whose variance is (in a suitable unit) the mutation rate \(\mu\). The coupling term \(C(X - X_0)(a - a_0)\) represents the change in the

---

\(^3\) Let us consider the fitness \(f\) (degree of adaptiveness of the phenotype, such as the growth rate) and let \(P_n(f)\) be the distribution of the fitness of different genotypes at generation \(n\). Then, the average fitness over genotype distribution is given by \(<f_n> = \int f P_n(f) df\) where \(<\ldots>\) denotes the average over the distribution \(P(f)\). Then, the fitness distribution of the next generation is given by \(P_{n+1}(f) = \int f P_n(f) df = \int f P_n(f) / <f_n> f = \int f P_n(f) f^{<f_n> - 1}\) because the offspring number is proportional to the fitness. From this equation, \(<f_{n+1}> - <f_n> = \int f P_{n+1}(f) df f^{<f_n> - 1} - <f_n> = \int f (P_{n+1}(f) - P_n(f)) f^{<f_n> - 1} = \int \delta f^2 \), where \(<\delta f^2> = \int (f_{n+1} - <f_n>)^2 >\) is the variance of the fitness \(f_n\). Hence the rate of increase in the fitness per generation is proportional to its variance.
Constructing universal Phenomenology for biological cellular systems


phenotype $X$ by the change in the genotype $a$. Recall that the above distribution (5) can be rewritten as

$$P(x,a) = \tilde{N} \exp \left[ -\frac{(X - X_0 - C(a - a_0)\alpha(a))^2}{2\alpha(a)} + \left( \frac{C^2\alpha(a)}{2} - \frac{1}{2\mu} \right) (a - a_0)^2 \right].$$

The average of $X(a)$ for a given genotype $a$ satisfies

$$\overline{X(a)} \equiv \int XP(X,a)\,dX = X_0 + C(a - a_0)\alpha(a).$$

The evolutionary stability condition mentioned above postulates that the multiplication factor in front of $(a - a_0)^2$ is negative. This postulate leads to the condition

$$\frac{\alpha C^2}{2} - \frac{1}{2\mu} \leq 0,$$$$

or, equivalently,

$$\mu \leq \frac{1}{\alpha C^2} \equiv \mu_{\text{max}}.$$

This means that the mutation rate has an upper bound $\mu_{\text{max}}$ beyond which the distribution does not keep a peak in the genotype–phenotype space, and the stability is lost, where the gradual evolution process ceases. In the above formulation, the distribution is symmetric against the peak, as we assumed a Gaussian-type distribution. However, as the distribution is flatter, the deviation from the Gaussian distribution is amplified. In particular, such an assumption with symmetry will be lost. Indeed, the fitted biological state is rarer than the non-fitted one. Then, as the distribution is flatter, non-fitted states are accumulated. In fact, according to the evolution simulation of the catalytic reaction network and gene-regulation network mentioned so far, most phenotypes obtained under large mutation rates have lower fitness [10, 15]. In addition, the evolution to increase fitness cannot progress when the mutation rate is larger than some threshold. This loss of stability of the fitted state under a larger mutation rate is analogous to the error catastrophe described by Eigen [16]. The catastrophe here, however, is a result of genotype–phenotype mapping in contrast to the combinatorial problems in genetic space in the Eigen sense.

Now recalling $V_g = \langle (\overline{X(a)} - X_0)^2 \rangle$ and equation (5), $V_g$ is given by $(C\alpha)^2 < (\delta a)^2$. Here, $(\delta a)^2$ is computed by the average $P(X,a)$ as $\mu/(1 - \mu C^2\alpha)$. Then, it follows that

$$V_g = \mu (C\alpha)^2 \frac{1}{1 - \mu C^2\alpha} = \alpha \frac{\mu}{\mu_{\text{max}}},$$

$$V_g \sim \frac{\mu}{\mu_{\text{max}}} V_{ip},$$

if the mutation rate $\mu$ is small enough to satisfy $\mu \ll \mu_{\text{max}},$

it is obtained by recalling that $V_{ip} = \alpha$. Thus, the proportionality between $V_{ip}$ and $V_g$ is obtained.

An intuitive interpretation of this relationship can be speculated as follows [7, 15]. In general, the dynamics governing the developmental processes that shape phenotypes
are inherently complex and the final state can be swayed by perturbations both in the initial conditions and those arising during the course of these dynamics. Even if the optimal phenotype is achieved through these dynamic processes, the perturbations introduced by noise during the dynamics can lead to diverse non-optimal states, resulting in the sensitivity of the phenotype to noise. However, as evolution proceeds, the system gains robustness of the phenotype against the disruptive effects of noise. To achieve robustness to noise, a global attraction towards the desired target phenotype needs to be shaped through the evolution. On the other hand, genetic changes can also introduce perturbations into these dynamic processes. If the global attraction to the target state is shaped against the imposed noise, the attraction will not be significantly damaged, even if the genetic change alters the parameters or equations in the dynamical systems, up to a certain degree. Thus, the increase in robustness to noise is expected to be accompanied by an increase in robustness to genetic changes. Indeed, it is observed from numerical simulations that the fitted states become increasingly robust to noise as they evolve. In addition, it is expected that they will also become more robust to genetic changes, leading to a correlation between the two types of robustness.

Now, as the robustness to noise is increased, the phenotypic variance $V_{ip}$ will decrease; similarly, an increase in the robustness to genetic mutations leads to a decrease in $V_g$. Hence, throughout evolution, both $V_{ip}$ and $V_g$ decrease in correlation. This cannot prove the proportionality between the two, but a strong correlation between $V_{ip}$ and $V_g$ is expected, as is consistent with the evolution simulations [10, 15] (see also [17] for related, pioneering studies on RNA, and [18, 19] for related studies on gene-regulation networks). For instance, the expression level or concentration of a component may be harder/easier to change by noise if the regulation to maintain the concentration is stronger/weaker. In this case, if a parameter in the system is changed by mutation (e.g. by the change in the enzyme activity for the reaction to the corresponding component), the concentration will also be harder/easier to change. This may provide a possible molecular mechanism for the correlation between phenotypic changes due to noise and mutation.

3. Restriction to steady growth instead of equilibrium

Next mystery Coming back to the distribution theory of $P(x, a)$, there remains another mystery: why can we only consider the distribution of few variables? Recall that biological systems involve a huge number of degrees of freedom for the phenotype; in fact, a bacterial cell contains thousands of components, whereas the models we adopted above involve a thousand components. Next, how are genetic changes represented by just one parameter $a$? Generally, are genes represented by a base sequence of nucleic acids in DNA? How can it be represented by a scalar variable\(^4\)? Furthermore, how can phenotypic and genetic changes, which are on quite different time scales, be treated in the same dimension? In short, how is the description by a few degrees of freedom possible?

\(^4\) One possible candidate for such scalar parameter is the Hamming distance from the fittest sequence or a projection according to the corresponding phenotype [20].

https://doi.org/10.1088/1742-5468/ad1f54
A straightforward origin of such a dimensional reduction could be the condition of the steady state. By restricting our focus to cells with steady growth, some phenomenological laws have been uncovered. They cover the classic, empirical relations by Monod [1] and Pirt [2] as mentioned already, as well as a relation on the change in protein abundances formulated recently [3, 21].

A general formulation to explain restriction due to steady growth has recently been formulated [22]. Consider a cell system with $M$ components and let $\mu_i$ be the growth rate of component $i = 1, 2, \ldots, M$, i.e. the abundances $N_i$ of each component increase with $\exp(\mu_i t)$ over a cell division cycle. Then, the steady growth implies that abundances of every component have to be approximately doubled for cell division, which implies that $\mu_i(N_j)$ is equal for all components $i$, which coincides with the growth rate of cell volume (the whole cell content) denoted by $\mu_g$. This leads to $M-1$ constraints over all concentrations. Then, consider the application of a given environmental stress. After adaptation to the novel environmental condition, the above $M-1$ constraints have to be satisfied so that the change in the cell’s state is restricted along a one-dimensional manifold in the $M$-dimensional compositional space, against the increase in environmental stress (see figure 1).

To formulate this constraint for steady growth, let us denote the concentration of each component $i = 1, \cdots, M$ by $x_i(>0)$. The cellular state is represented as a point in an $M$-dimensional state space. Here, each component is synthesized or decomposed relative to other components at a rate $f_i\{x_j\}$; for instance, by the rate equation in chemical kinetics. Additionally, all concentrations are diluted by the rate $(1/V)(dV/dt) = \mu_g$, so that the time change of a concentration is given by

$$\frac{dx_i}{dt} = f_i\{x_j\} - \mu_g x_i.$$  \hspace{1cm} (9)
For convenience, let us denote $X_i = \log x_i$, and $f_i = x_i F_i$. Then, equation (9) can be written as

$$\frac{dX_i}{dt} = F_i(\{X_j\}) - \mu_g.$$  \hspace{1cm} (10)

in which it is assumed that $x_i \neq 0$, i.e. all components exist. Based on this, the stationary state is given by the fixed-point solution

$$F_i(\{X^*_j\}) = \mu_g$$ \hspace{1cm} (11)

for all $i$.

We now consider intracellular changes in response to environmental changes as represented by a set of continuous parameters $E^a$, the strength of environmental stress of the type $a$. In other words, each environmental change of the stress type $a$ is parameterized by a continuous parameter $E^a$ (such as temperature, nutrient limitation, etc). Using this parameterization, the steady-growth condition leads to $F_i(\{X^*_j(E^a)\}, E^a) = \mu_g(E^a)$.

In response to environmental changes, the growth rate $\mu_g$ changes but the above equations have to be satisfied, so that the $M-1$ conditions $F_1 = F_2 = \cdots = F_M$ must be satisfied. Thus, a cell must follow a one-dimensional curve in the $M$-dimensional space (see figure 1) under a given change in the environmental conditions (e.g. stress strength).

We then linearize the change around the original (non-stressed state) $E_0$. With this, each $X^*_j$ changes from $X^*_j$ at $E_0$, to $X^*_j + \delta X_j$, accompanied by the change from $\mu_g$ to $\mu_g + \delta \mu_g$. Assuming a gradual change in the dynamics $x_j$, we introduce a partial derivative of $F_i(\{X^*_j(E)\})$ by $x_j$ at $E = E_0$, which gives the Jacobi matrix $J_{ij}$. Using this linear approximation, we obtain

$$\sum_j J_{ij} \delta X_j (E) + \gamma_i \delta E = \delta \mu_g (E) = c \delta E.$$ \hspace{1cm} (12)

With $\gamma_i \equiv \frac{\partial F_i}{\partial E}$, where from the linear approximation $\delta \mu_g = c \delta E$ with a constant $c$. Accordingly, we obtain

$$\sum_j J_{ij} \delta X_j (E) = \delta \mu_g (E) \left(1 - \gamma_i / c\right).$$ \hspace{1cm} (13)

By putting $E$ or $E'$ in the above equations and dividing the equation for $E$ by that for $E'$, we get

$$\frac{\delta X_j (E)}{\delta X_j (E')} = \frac{\delta \mu_g (E)}{\delta \mu_g (E')}$$ \hspace{1cm} (14)

across all $j$, against two stress strengths, $E$ and $E'$.

In other words, the change in $\delta X_j$ is proportional across all components $j$ with a common ratio. Now one can examine this relationship by measuring the changes in all the concentrations of mRNAs or proteins by using the transcriptome or proteome analysis. By changing the strength of stresses from the original state to $E$ and $E'$, and $\delta X_j$ was measured across thousands of mRNA species for *Escherichia coli*. Interestingly,
the relationship (14) is satisfied across most components, even against a large degree of environmental changes [22] (based on the data in [23]). Indeed, the linear approximation is valid up to the degree at which the growth rate goes down almost to 20% from the original, below which the steady growth no longer holds. In other words, the common linear relationship is valid across all components, as long as the steady growth requirement is satisfied.

Interestingly, relationship (14) is approximately true for most of a few thousand components (mRNAs or proteins), even across different types of environmental stresses (i.e. heat, osmotic pressure, starvation or other chemical stresses) $E$ [22, 24] (as well as in the protein data based on [25]). This is not explained in the above steady-growth theory because the directions of the steady-growth curves would be different in terms of different stress types $a$ (given by vectors $E^a$, as in figure 1). In other words, $\gamma_i$ depends on each stress type $a$. This experimental observation, then, suggests that drastic dimensional reduction of phenotypic responses occurs in living cells, i.e. even across different environmental perturbations (stresses).

4. Dimensional reduction as a consequence of robustness and plasticity

Next mystery Deep linearity: the above proportionality relationship cannot be explained just by the steady-growth condition. Indeed, when we studied a toy cell model consisting of a large number of components with arbitrarily chosen reaction dynamics, the proportionality in equation (14) was valid only against a tiny change in the environment. Furthermore, if we include different types of environmental changes, equation (14) would not be valid either. Then, what is the origin of such deep linearity?

Here one should recall that a biological system (cell) is shaped as a result of evolution. Such cellular dynamical systems that have a higher fitness (growth rate) are selected under a given environmental condition. Note that such dynamics are under perturbations. Intracellular reactions are noisy, as the number of molecules therein is not so large and the environmental conditions fluctuate. The fitted phenotype state should then be robust to such perturbations. Recalling that the stability of an equilibrium state allows for thermodynamic description by a few macroscopic variables, one may expect that robustness in evolved biological systems similarly allows for a reduction to a few macroscopic variables from a high-dimensional phenotypic space.

In fact, we considered the toy cell model of a huge number of components, in which the abundances of each component change as a result of stochastic intracellular dynamics with catalytic reactions, as well as by taking nutrient chemicals from environment, leading to the cellular growth, as adopted in the study of section 2 [10]. The dynamics thus involve numerous concentration variables, which provide the high-dimensional phenotypic space. The growth rate (fitness) is determined from these phenotypes and, accordingly, we evolved the reaction network so that a fitted phenotype (e.g. higher growth) is realized. After the evolution, we found that relationship (14) is valid across large environmental perturbations and over different types of environmental stresses [24].

The origin of this deep linearity is understood as follows, based on a dynamical-systems analysis. The shaped fitted state from this evolved network is robust against
the inevitable perturbations mentioned above \[24\]. Then, the fitted state (attractor) has to be attracted from most directions in the high-dimensional state space. In terms of dynamical systems, this means that there is a sufficiently strong flow from most directions to the (fixed-point) attractor corresponding to the fitted state. If this attraction occurs from every direction, however, evolution to increase fitness would be difficult, as the state would not be easily changed by the slight alteration in dynamical systems (flow in the state space) induced by genetic changes (mutation). When considering the direction in which evolution has occurred or is ongoing, it is plausible for phenotypes to undergo changes (see figure 2 schematically). All the concentration changes will be constrained along a one-dimensional (or few-dimensional) manifold. If we consider the Jacobi matrix for the relaxation dynamics to a stable state, the magnitude of one (or a few) negative eigenvalue(s) is close to zero, whereas those of others are much more negative.

The above picture of evolutionary dimensional reduction has been confirmed numerically by studying the cell model mentioned above \[24, 26\]. Using numerical evolution, it was found that, after evolution, the global proportionality across a variety of environmental conditions is shaped. In this case, changes in most concentration variables induced by a variety of distinct environmental perturbations are found to be highly restricted along a one-dimensional manifold. One of the (negative) eigenvalues in the Jacobi matrix is close to zero and the principal component matches its eigenvector.

Now, if the response to any environmental perturbations occurs along the above one- (or low-) dimensional manifold, we do not need to consider all the eigenvalues (and eigenvectors), but only the largest one (i.e. a separated slow mode) that dominates the behavior. Then, with straightforward calculations, it is shown that equation (14) holds across different directions of environmental changes. Also, as the magnitude of the largest eigenvalue is close to zero, the change along the eigenvector is extended, so that the region of the linear manifold is extended. Thus, the deep linearity is explained by the evolutionary dimensional reduction.

https://doi.org/10.1088/1742-5468/ad1f54
From the above discussion, it is expected that the change due to genetic variation is also constrained along the same manifold. In fact, the concentration changes $\delta X_i$ by genetic variations, by environmental changes and by internal noise lie on the same curve, according to the simulation of the above cell models [24]. Accordingly, the responses due to environmental and evolutionary (genetic) changes will be correlated, i.e. the relation (14) is satisfied between $\delta X_i$ by the environmental and genetic (evolutionary) changes.

Following this argument, consider the changes in $\delta X_i$ and $\delta \mu_g$ as a function of environmental change $E$ and genetic change $G$. Then one can extend the relationship (14) to the case with genetic evolution after the application of environmental stress. In this case, by representing the changes against $E$ and $G$, we get the relationship

$$\delta X_i(E, G) \approx \frac{\delta \mu_g(E, G)}{\delta \mu_g(E, 0)} \delta X_i(E, 0).$$ (15)

Here, through the applied stress, the growth rate is decreased from the original state, so that $\delta \mu_g(E, 0)$ is negative. Whereas, with evolution, the growth rate is recovered toward the original, so that the change from the original non-stressed state, $|\delta \mu_g(E, 0)|$, should be smaller than $|\delta \mu_g(E, 0)|$. This implies that $\delta X_i(E, G)/\delta X_i(E, 0) < 1$. Hence, as the evolution progresses to recover the growth, the change in each mRNA (or protein) concentration induced by the stress is reduced with evolution. Hence, the long-term response by evolution occurs in the direction of canceling out the environmentally induced change, as is observed in bacterial evolution experiments [27–29]. This homeostasis by adaptation is akin to Le Chatelier's principle in thermodynamics.

To close this section, recall that response and fluctuations are two sides of the same coin in statistical physics, whereas phenotypic fluctuations over isogenic cells ($V_{ip}$) are often large, as recent experiments have demonstrated [11, 30]. As the variations $\delta X_i$ by noise and by mutation are also located on the common low-dimensional manifold (figure 2), it has been shown that the variances by internal noise ($V_{ip}(i)$) and by genetic variation ($V_g(i)$) are proportional across components $i$. This $V_g(i) - V_{ip}(i)$ law was confirmed in the toy cell models [24] numerically and is also consistent with some experimental results [31–34].

5. Universality of evolutionary dimensional reduction

To recapitulate the preceding discussion, we posit the following evolutionary-dimensional reduction hypothesis: over the course of evolution, phenotypic states tend to exhibit robustness, yet they display plasticity in specific dimensions crucial for evolutionary adaptation. Consequently, a limited number of modes, characterized by eigenvalues close to zero (termed slow modes), become isolated from the rest. In its simplest manifestation, this dimensionality reduction reduces to a single dimension. However, in scenarios where multiple survival conditions must be met or when distinct environmental factors exert influence throughout evolution, the reduced dimensionality may be several.

As most variations arising from noise, environmental alterations and genetic mutations are constrained primarily along this slow manifold, these changes from different sources exhibit correlations. This aligns with the proposed evolutionary–fluctuation
response relationship. This phenomenon engenders both predictions and constraints on the evolution of phenotypes.

Moreover, the separation of slow modes fosters enhanced evolvability. When all modes share a similar timescale, mutations affecting each mode would be effectively cancelled out and have minimal impact on the phenotypic dynamics due to the law of large numbers. Hence, they yield only slight fitness alterations from random mutations, akin to the proverb, *too many cooks spoil the broth*. Conversely, when a few slow modes are separated, the faster modes are controlled by these slow modes (as in the adiabatic elimination of fast variables [35, 36]). Consequently, mutational changes induce collective alterations in other variables, thereby amplifying their influence and accelerating the evolution speed.

Given the apparent validity of these premises across a wide array of biological systems, we anticipate that this evolutionary dimensional reduction holds universal significance. In support of this hypothesis, we have conducted evolution simulations under specific fitness conditions and confirmed (a) the evolutionary fluctuation response relationship, (b) the proportionality between genetic variance ($V_g$) and isogenic phenotypic variation ($V_{ip}$) throughout evolution, and (c) the reduction of phenotypic dimensions to one or a few dimensions after evolution, in the following examples:

(i) Reaction network dynamics model: as already mentioned, from the evolution simulation of a few toy cell models with a catalytic reaction network, the aforementioned points are confirmed by adopting the fitness condition to increase the cell-growth speed or to enhance the concentrations of certain components [10, 24, 26].

(ii) Gene regulation network model: gene expressions, involving mRNA and protein synthesis, form a network governed by mutual activation or inhibition. The dynamics of gene expression levels follow threshold-type on-off dynamics, characterized by a matrix of positive and negative elements for activation and inhibition, akin to activity within neural networks. By assigning fitness to specific expression patterns for target genes and evolving the network, we have again verified properties (a)–(c), particularly when introducing a substantial level of noise [7, 15]. Furthermore, the reduced dimensionality may encompass multiple dimensions if the various expression patterns correspond to different external conditions [37].

(iii) Evolving interacting spin model analogous to a spin-glass model: the above dynamics of gene regulation networks show similarities to spin dynamics ($S_i$) influenced by ferro (activating) and anti-ferro (inhibitory) interactions. By formulating a Hamiltonian ($H = \sum J_{ij} S_i S_j$) with an interaction matrix ($J_{ij}$), the distribution of spin configurations under temperature ($T$) is given by $\exp(-H/(kT))$. From this configuration, the fitness is determined, for instance, by the overlap with a predetermined specific configuration of target spins. Accordingly, the matrix $J_{ij}$ undergoes evolution through mutation and selection. It is found that fitted states with robustness to mutation are evolved for a certain temperature range [38], whereas we observe properties (a)–(c) within a certain temperature range [39, 40], indicative of robustness to mutation. This robustness is observed only in the replica symmetric phase. However, for lower temperatures, the robustness collapses due to replica symmetry breaking (RSB).
(iv) Evolution of protein structure: through a combination of data analysis, elastic-network models, and theoretical analysis, we have confirmed the tight correlation between $V_g$ and $V_{ip}$ as well as the dimensional reduction of protein configuration dynamics. Estimated configuration changes resulting from noise and genetic mutations are constrained within a common low-dimensional space, typically around five dimensions; a stark reduction from the hundreds of residues giving rise to the high-dimensional state space [41]. Note that such dimension reduction from high-dimensional state space is also discussed in protein dynamics [42].

Recent experimental research involving the evolution of bacteria exposed to 100 antibiotics by Furusawa’s group has provided further support. Transcriptome analysis of thousands of mRNA expression data indicates that changes in bacterial responses can be predicted by approximately seven principal components [43]. Dimensional reduction has also garnered significant attention in other fields of biology, as has been explored experimentally in the development of *Caenorhabditis elegans* [44], laboratory ecological evolution [45] and neural learning process in monkeys [46]. The insensitivity of phenotypes to most parameters has been conceptualized and discussed as the ‘sloppy parameter hypothesis’ [47].

6. Towards statistical physics theory for evolutionary dimensional reduction

While an argument for dimensional reduction grounded in evolutionary robustness and plasticity appears plausible, it remains somewhat speculative in the absence of a rigorous theoretical foundation. A theory to establish macro–micro consistency is missing at present. Drawing inspiration from established theories in statistical mechanics, several avenues of investigation are expected:

(i) *Projection to the dominant collective mode from high-dimensional phenotype space:* in statistical physics, the method of projection onto the collective mode was established by Mori and Zwanzig [48, 49] to be consistent with thermodynamics. In contrast, in the context of evolutionary dimensional reduction, the space onto which the projection is made emerges as a consequence of evolution itself. Consequently, we need to introduce a self-consistent approach to this projection method. One promising direction will be the marriage of dynamical mean-field theory [50, 51] with evolutionary dynamics [52], although the formulation of the dimension reduction process remains an ongoing endeavor.

(ii) Renormalization group: through evolutionary processes to achieve robustness, microscopic perturbations are wiped out, leaving behind only the low-dimensional aspects of plasticity. This process bears resemblance to the elimination of irrelevant variables in the renormalization group [53–56]. Devising a methodology to integrate renormalization group concepts with evolutionary processes remains a challenging frontier.

(iii) Slow–fast symmetry breaking: as uncovered in the analysis of the eigenvalue spectrum, a few slow modes are separated from many other degrees of freedom, as represented by a few outliers (whose eigenvalues are close to zero). As discussed, this
separation of a few slow modes is expected to facilitate evolution. This phenomenon bears similarity to the origin of the central dogma, wherein symmetry breaking leads to the separation of information molecules with slower changes and functional molecules exhibiting faster changes [57]. A similar separation phenomenon also emerges in neural systems, where separation of areas with slower neural activities is relevant to Bayesian inference [58]. A theoretical framework of slow–fast symmetry breaking for evolution is needed.

(iv) Avoidance of RSB: in the evolving interacting spin model mentioned in the previous section, robustness to noise and dimensional reduction is achieved in the replica symmetric phase, whereas RSB undermines robustness. In the case of evolving interactions, the applicability of replica theory extends to both spins and couplings [40]. On the other hand, when evolution under multiple fitnesses against different environmental conditions is postulated, the region of the fitted replica symmetric phase shrinks towards the RSB side [59], which may suggest that the state achieving both robustness and plasticity against multiple conditions will approach ‘the edge of the glass’. This intriguing phenomenon warrants further investigation.

(v) Macroscopic potential for cellular state: if one assumes dimensional reduction, it raises the possibility of a macroscopic potential theory akin to thermodynamics. Analogous to thermodynamics, where the free-energy potential emerges from dimensional reduction, one might envision a similar description of biological systems employing a limited set of macroscopic variables. This could entail the formulation of a macroscopic potential, in which the growth rate ($\mu_g$) or fitness is expressed as a function of environmental and genetic changes, as $\mu_g(E,G)$ [60]. This framework has the potential to yield linear relationships between evolutionary responses and phenotypic fluctuations caused by noise, as corroborated by bacterial evolution experiments, which were discussed in the previous section. In fact, a general picture akin to potential was proposed by Waddington, with qualitative discussions on correlations between genetic and environmental responses, referred to as genetic assimilation [5], a concept that resonates with the $V_g$–$V_ip$ relationship presented in this paper. With the formalization of macroscopic potential, these discussions can advance to the level of quantitative theory.

7. Constructing universal biology

In the present paper, we discuss the premise of universal biology, by focusing on the consistency between slow evolutionary change and fast adaptive changes. The self-consistency principle between macro–micro scales is generally important.

7.1. Ideal cell model?

At the cellular level, achieving consistency between the replication of numerous molecular species and cellular reproduction is essential. To comprehend this phenomenon, a simplified ideal cell model proves invaluable. In the development of thermodynamics,
the concept of an ideal gas plays a pivotal role, even though thermodynamics itself is not limited to this particular case. Similarly, to facilitate the formulation of a macroscopic theory for cellular states, the introduction of an ideal cell model capable of sustaining robust, steady growth may prove beneficial \[8, 9, 61–63\]. In this context, potential state variables could include the growth rate, nutrient consumption and the energetic costs associated with maintaining a cell, although additional macroscopic parameters might also be necessary. Establishing a possible connection with standard thermodynamic quantities, such as entropy production, becomes a crucial endeavor \[64, 65\].

Furthermore, microbial cells typically undergo a transition from exponential growth to a stationary phase during which cellular growth ceases and the cell enters a sleeping state. Once a cell reaches this dormant state due to nutrient depletion, it generally requires a specific period of time, referred to as the lag time, to resume growth upon the reintroduction of nutrients. Recent experimental observations have unveiled certain empirical relations between lag time, starvation duration and maximal growth rate \[66\]. By extending the ideal cell model to incorporate interactions with components that do not contribute to growth or catalysis, has led to the derivation of empirical laws governing both the transition to the sleeping state and the previously mentioned phenomena \[67\]. This endeavor is akin to the extension of the ideal gas model to the van der Waals model.

7.2. Multicellularity and irreversible differentiation

We have hitherto explored the prospect of a macroscopic description of cells. It is imperative to acknowledge that biological systems encompass a hierarchical structure, spanning from molecules to cells, and further to multicellular organisms, culminating in ecosystems. The resilience of biological systems hinges on their ability to uphold consistency across these diverse levels \[68\]. In multicellular organisms, the cellular community, be it a tissue or an entire organism, is intricately developed to ensure that both the collective ensemble of cells and each individual intra-cellular state remain robust in the face of perturbations. In this case, cells differentiate into distinct types, with the relative abundance of each cell type maintained within a certain range.

The dynamical systems approach for multiple cellular states finds its origins in the pioneering work of Goodwin \[69\] and Kauffman \[70\]. However, comprehending robustness at the multicellular level, as well as the irreversible loss of multipotency (the capacity to generate various cell types during differentiation from embryonic stem cells), necessitates an intricate interplay between intracellular dynamics and cell–cell interactions \[71\]. This prompts the question: can we characterize irreversibility and derive governing principles through the formulation of a macroscopic phenomenological theory? This stands as the next challenge once a phenomenological theory for a cell is established.

In summary, we discuss the possibility of macroscopic universal biology, inspired by thermodynamics, as summarized in table 1.
Table 1. Possible correspondence between thermodynamics and macroscopic phenomenological theory for biological system. Eqb: equilibrium.

<table>
<thead>
<tr>
<th>Thermodynamics</th>
<th>Universal biology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restriction to Eqb systems</td>
<td>Restriction to steady growth</td>
</tr>
<tr>
<td>Stability of Eqb state</td>
<td>Robustness of evolved state</td>
</tr>
<tr>
<td>Irreversibility to Eqb</td>
<td>Return to adapted state</td>
</tr>
<tr>
<td>Irreversibility between Eqb states</td>
<td>Differentiation from stem cells</td>
</tr>
<tr>
<td>Fluctuation–response relation</td>
<td>Evolutionary fluctuation response</td>
</tr>
<tr>
<td>Le Chatelier principle</td>
<td>Homeostasis by adaptation</td>
</tr>
<tr>
<td>Thermodynamic potential</td>
<td>Growth macroscopic potential</td>
</tr>
<tr>
<td>Free energy vs energy</td>
<td>Growth versus nutrient uptake</td>
</tr>
<tr>
<td>Ideal gas</td>
<td>Ideal cell model for growth</td>
</tr>
<tr>
<td>Liquid–gas transition</td>
<td>Exponential–stationary transition</td>
</tr>
</tbody>
</table>

Acknowledgments

This research was partially supported by and the Novo Nordisk Foundation (0065542) and Grant-in-Aid for Scientific Research (A) (20H00123) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. The author would like to thank Chikara Furusawa, Tuan Pham, Takuya Sato, Ayaka Sakata, Qian-Yuan Tang, Tetsuhiro S Hatakeyama, and Yuichi Wakamoto for useful discussions.

References

Constructing universal Phenomenology for biological cellular systems

References:
[16] Eigen M 1971 Selforganization of matter and the evolution of biological macromolecules *Naturwissenschaften* 58 465
[34] Uchida Y, Takeda H, Furusawa C and Irie N 2023 Stability in gene expression and body-plan development leads to evolutionary conservation *EvoDevo* 14 4
[38] Sakata A, Hukushima K and Kaneko K 2009 Funnel landscape and mutational robustness as a result of evolution under thermal noise *Phys. Rev. Lett.* 102 148101

https://doi.org/10.1088/1742-5468/ad1f54
Constructing universal Phenomenology for biological cellular systems

[58] Ichikawa K, and Kaneko K 2022 Bayesian inference is facilitated by modular neural networks with different time scales (arXiv:2210.12294)
[59] Sakata A and Kaneko K 2023 Evolutionary shaping of low dimensional path facilitates robust and plastic switching between phenotypes Phys. Rev. Res. 5 043296
[63] Pandey P P and Jain S 2016 Analytic derivation of bacterial growth laws from a simple model of intracellular chemical dynamics Theory Biosci. 135 121–30

https://doi.org/10.1088/1742-5468/ad1f54