Evolution of genetic redundancy: the relevance of complexity in genotype–phenotype mapping

Nen Saito, Shuji Ishihara and Kunihiko Kaneko
Graduate School of Arts and Sciences The University of Tokyo 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan
E-mail: saito@complex.c.u-tokyo.ac.jp, shuji@meiji.ac.jp and kaneko@complex.c.u-tokyo.ac.jp

Received 17 October 2013, revised 11 April 2014
Accepted for publication 29 April 2014
Published 9 June 2014
New Journal of Physics 16 (2014) 063013
doi:10.1088/1367-2630/16/6/063013

Abstract
Despite its ubiquity among organisms, genetic redundancy is presumed to reduce total population fitness and is therefore unlikely to evolve. This study evaluates an evolutionary model with high-dimensional genotype–phenotype mapping (GPM) by applying a replica method to deal with quenched randomness. From the method, the dependence of fitness on genetic redundancy is analytically calculated. The results demonstrate that genetic redundancy can have higher population fitness under complex GPM, which tends to favor gene duplication in selection processes, further enhancing the potential for evolutionary innovations.

Keywords: evolutionary model, spin glass, fitness landscape

1. Introduction

All living organisms experience selection pressures, which act on their phenotypes, while only their genotypes are heritable. Therefore, a link between genotype and phenotype, referred to as genotype–phenotype mapping (GPM), is essential for fully understanding evolutionary

---

1 Present address: Department of Physics, Meiji University, Higashi-Mita 1-1-1, Tama-ku, Kawasaki-shi, Kanagawa, 214-8571, Japan
processes. In quantitative genetics, the theory of polygenic traits is based on the assumption that each gene linearly affects a specific trait [1]. However, GPM is generally non-linear and many traits are the result of complex processes involving interactions between many proteins, RNAs, and genes. For example, development is largely regulated by transcriptional networks that are modeled by high-dimensional and non-linear [2, 3]. Inherent complexities in GPM obscure genotypes associated with a high-fitness phenotype, which can impact evolution [4–6]. While evolution under a complex fitness landscape has been studied from various perspectives, including error catastrophe [7, 8], mutational robustness [4, 5, 9] and evolvability [10], the relevance of GPM complexity to evolutionary processes has not yet been fully understood.

An important characteristic of GPM is genetic redundancy, i.e. the encoding of a phenotypic trait by two or more genes. Numerous examples of genetic redundancy are found in higher organisms [11, 12] and even in microorganisms [13]; however, a classical premise of evolutionary theory is that genetic redundancy lowers fitness at the population level, and is therefore evolutionarily suppressed [14–17]. According to this argument, genetic redundancy is expected to weaken sensitivity to deleterious mutations, which could persist and reduce the total fitness of the population. Conversely, in the absence of genetic redundancy, there is greater susceptibility to the deleterious effects of mutations; since most mutations are lethal, mutants are effectively removed from the population, thereby maintaining population fitness. This suggests that the presence of redundant genes increases mutational load and weakens purifying selection, making genetic redundancy evolutionarily unstable. Similar arguments have been made for the loss of one functional copy of a gene after a duplication event [14, 18–21]. The suppression of redundancy is especially pronounced in asexual populations. However, Ohno [14] discussed that genetic redundancy can be retained if one copy of a duplicated gene acquires a novel function prior to becoming a pseudogene, a process known as neofunctionalization [22]; while this has been extensively investigated [21, 23], it requires a much higher rate of beneficial mutations than what is expected. Several studies have proposed an alternative explanation for the ubiquity of genetic redundancy [16, 24–27], but it is still not yet well understood. Specifically, the relevance of complexity in GPM to the evolution of genetic redundancy has not been evaluated.

In the present study, an asexual evolutionary model is used to examine the relationship of genetic redundancy to evolution. Genetic redundancy is defined here as the ratio of the number of genotypes to the number of phenotypes, and is analogous to genome size under a fixed number of phenotypes. The role of GPM complexity, represented by quenched randomness, in the evolution of genetic redundancy is examined for both simple and complex GPM. In contrast to the prevailing views, it is found that populations with higher genetic redundancy can have higher fitness under complex GPM. This preference for redundancy is characteristic of high-dimensional complex GPM and is independent of previously reported mechanisms [16, 24–27]. These findings provide a novel explanation for the ubiquity and persistence of genetic redundancy in biological systems.

2. Model

An adiabatic spin system is introduced in order to model both simple and complex GPM, where the latter is represented as non-trivial mapping with quenched randomness. A similar spin
model was adopted in another study [9] that examined evolution (but not genetic redundancy) under complex GPM. In this model, the configuration of \(N\) sites of a locus is described by \(g = (g_1, g_2, ..., g_N)\), where each \(g_i\) can take one of two allelic states, i.e., \(g_i = \pm 1\). Genotype \(g\) determines \(M\) phenotypic traits, represented by \(p = (p_1, p_2, ..., p_M)\), where each \(p_i\) can take \(p_i = \pm 1\), indicating whether or not the \(i\)th trait is expressed. Genetic redundancy is characterized by the ratio between the number of genotypes and the number of phenotypes, which is described by the parameter \(\gamma \equiv N/M\). For a given genotype \(g\), phenotype \(p\) is determined in a stochastic manner, with probability depending on stochastic GPM \(P(p|g)\). Assuming that each \(p_i\) is independent of any other \(p_j\), and the contribution from \(g_j\) to \(p_i\) is multiplicative and independent of any other \(g_j\), namely \(P(p|g) = \prod_{ij} P(p_i|g_j)\), \(P(p|g)\) is given by the following Boltzmann distribution (a detailed explanation can be found in appendix I):

\[
P(p|g) = \exp\left(\beta \sum_{ij} J_{ij} p_i g_j\right)/Z_p(g),
\]

where \(J_{ij}\) determines the mapping from \(g_j\) to \(p_i\) and is typically a quenched random variable. This \(J_{ij}\) represents the physical or environmental constraints that depend neither on phenotype nor on genotype, and is thus fixed over the evolutionary timescale. Here \(\beta^{-1} \equiv T_p\) is the temperature for the stochasticity in GPM, representing the strength of the phenotypic fluctuation in isogenic individuals, and \(Z_p(g)\) is the partition function for \(p\). The average phenotypic value of \(p\) under fixed \(g\) is given by \(\langle p_i \rangle = \tanh(\beta \sum_{i} J_{ij} p_i g_j)\). Thus, the dependence of \(\langle p_i \rangle\) upon \(g\) is non-linear and non-additive, indicating that the model includes epistasis.

Changes in genotype distribution are governed by mutation and selection processes under a given fitness, which is a function of phenotype \(p\) and is described here as \(\Phi(p)\). Rather than introducing a complex fitness landscape, a simple fitness function \(\Phi(p) = \sum p_i\) is adopted in order to focus on the relevance of GPM to evolution. At the population level, detailed dynamics of individual phenotypes are adiabatically eliminated, so that only the phenotypic distribution given by equation (1) is concerned with the selection of subpopulations with genotype \(g\). Thus, only the average fitness \(\phi(g) = \sum_p \Phi(p)p(p|g)\) contributes to the evolutionary dynamics of the genotype \(g\). The steady state of the evolutionary process is addressed, rather than transient adaptation dynamics. The equilibrium distribution \(P(g)\) of genotype \(g\) in the population is determined only by the effective fitness \(\phi(g)\) and ‘genotypic temperature’ \(T_g\), which represents the selection pressure. The distribution of genotypes is approximated by the Boltzmann distribution as

\[
P(g) = \exp\left(\beta'\phi(g)\right)/Z_g,
\]

where \(\beta' = T_g^{-1}\) and \(Z_g\) is a partition function for \(g\).

This Boltzmann distribution of genotype has been derived in several earlier studies using standard evolutionary models, and the derivation is verified for an asexual population [28–32].
Although $\beta^\prime$, the inverse of genetic temperature, depends on details of a model, it always represents the strength of selection, and as it approaches infinity (i.e., in the limit of zero genotype temperature) only the fittest genotype remains; for example, $\beta^\prime \approx 2N_e$, where $N_e$ is effective population size, for the Fisher–Wright process [29]. Precise interpretation of $P(g)$ in the following calculation is the probability that a homogeneous population with genotype $g$ appears, rather than the distribution of genotypes within a population at a given time. However, when the population size is sufficiently large, the latter interpretation is possible and thus is adopted in the following calculations.

In this model, the accumulation of mutations in the population is represented by the entropic contribution to the free energy $-\beta^\prime \ln Z_g$, and thus the mutation–selection balance at the steady state of the evolution corresponds to the balance between entropic and energetic contributions to free energy.

The evolutionary steady states under complex and simple GPMs, representing two extreme cases, are compared. In the simple GPM, $J_{ij}$ is chosen to be $J_{ij} = J_{ij}/N$ without randomness, whereas the complex GPM is represented by quenched random variables $J_{ij}$ drawn from $P(J_{ij}) = \exp\left[-J_{ij}^2/2(\sigma^2/N)\right]/\sqrt{2\pi\sigma^2/N}$. In the latter, it is reasonable to use an ensemble average with respect to $J_{ij}$ to find the behavior of a system with the typical realization of $J_{ij}$. It can also be assumed that $J_{ij}$ depends on the environment, which changes over a sufficiently slower timescale than evolution. In this case, an ensemble average with respect to $J_{ij}$ represents a time average over the environmental timescale.

3. Result I: simple GPM

First, we consider the behavior of simple GPM $J_{ij} = J_{ij}/N$. The explicit form of effective fitness $\phi(g)$ can be calculated as $\phi(g) = \text{Tr}_p P(p|g) \phi(p) = M \tanh\left(\beta J_0 \sum_j g_{ij}/N\right)$. For sufficiently large $N$, $Z_g$ is thus obtained as

$$Z_g \approx \exp\left[N\left(-\frac{1+m}{2} \ln \frac{1-m}{2} - \frac{1+m}{2} \ln \frac{1+m}{2} + \frac{\beta^\prime}{\gamma} \tanh \beta J_0 m\right)\right],$$

where $m$ is obtained from the following saddle point equation:

$$\frac{1}{2} \ln \frac{1-m}{2} - \frac{1}{2} \ln \frac{1+m}{2} + \frac{\beta^\prime}{\gamma} \frac{\beta J_0}{\cosh \beta J_0 m} = 0. \quad \text{(4)}$$

The mean fitness per unit phenotype $\gamma \left< \phi \right>$ is given by $\gamma \left< \phi \right> = \frac{\partial}{\partial \phi} \ln Z_g = \tanh \beta J_0 m$. Here, we adopt the fitness per unit phenotype $\gamma \left< \phi \right>$, rather than the fitness per locus $\phi$, in order to evaluate the contribution of genetic redundancy to fitness, because $\gamma \left< \phi \right>$ provides a natural measure for comparing systems between high and low redundancy (large $N$ or small $N$), each with a fixed number of phenotypes $M$. In figure 1(a), the dependence of $\gamma \left< \phi \right>$ on $T_s$ and $T_p$ for $\gamma = 1$ is represented. This figure illustrates that $\gamma \left< \phi \right>$ decreases by the increase of either $T_s$ or $T_p$, i.e., by the decrease in selection pressure or increase in phenotypic fluctuation. Figures 1(b) and (c) show...
the dependence of $\gamma \langle \phi \rangle$ on $\gamma$ with fixed $T_p$ and $T_g$, showing that $\gamma \langle \phi \rangle$ decreases as $\gamma$ increases. Since the dependence of $\gamma \langle \phi \rangle$ on $\gamma$ appears in equation (4) with the form $\beta' / \gamma$, one may find that increasing $\gamma$ corresponds to an increase in $T_g$ and thus leads to a decrease in fitness. This implies that more mutations tend to accumulate in the presence of redundant genes, leading to a reduced fitness. This is consistent with the suggestions that genetic redundancy weakens the effectiveness of selection in eliminating deleterious mutations from a population, thus causing a decline in mean population fitness [15–17].

To quantify the diversity of genotypes in a population, we introduce entropy $s \equiv - \sum \log P(g) / N$, which is calculated from $s = \sum Z_g - \beta' \langle \phi \rangle$. Although the entropy represents the genetic diversity of an ensemble of independent homogeneous populations in the canonical argument, it could also be taken as the genetic diversity within a population at equilibrium by assuming a large population size. As shown in figure 1(d), entropy per unit phenotype $s$ increases with increasing $\gamma$, indicating that genetic redundancy enhances genetic diversity as a consequence of the accumulation of mutations. It follows from

**Figure 1.** Fitness and entropy for simple GPM. (a) A color map of the mean fitness per phenotype $\gamma \langle \phi \rangle$ for $\gamma = 1$ is described on the $T_p-T_g$ plane. (b) Dependence of the mean fitness per phenotype $\gamma \langle \phi \rangle$ on $\gamma$ with fixed $T_p = 1$ and $T_g = 0.1, 0.5, 1.0, \text{ and } 1.5$. (c) Dependence of the mean fitness per phenotype $\gamma \langle \phi \rangle$ on $\gamma$ with fixed $T_g = 1$ and $T_p = 0.1, 0.5, 1.0, \text{ and } 1.5$. (d) Dependence of the mean entropy per phenotype $s$ on $\gamma$ with fixed $T_p = 1$ and $T_g = 0.1, 0.5, 1.0, \text{ and } 1.5$. 
these results that, in the case of simple GPM, genetic redundancy $\gamma$ decreases fitness per unit phenotype $\langle \phi \rangle$ and increases entropy per unit phenotype $\gamma s$.

4. Result II: complex GPM

In complex GPM, where $J_{ij}$ is a quenched random variable drawn from $P (J_{ij}) = \exp \left[ -J_{ij}^2 / 2 \left( \sigma^2 / N \right) \right] / \sqrt{2\pi \left( \sigma^2 / N \right)}$, the explicit form of effective fitness $\phi (g)$ is calculated as $\phi (g) = \sum_{m} \tanh \left( \beta \sum_{j} J_{ij} b_{ij} \right)$. In this case, we need to calculate the expectation value of free energy over quenched random variables, rather than that of $Z_i$. The replica method \cite{33} is useful for calculating a quenched average of the free energy $\left[ F (T_g) \right]_J = \int F (T_g) dJ P (J)$, where $\left[ Z^n \right]_J$ is first calculated for an integer $n$, before $\left[ \log Z_i \right]_n$ is obtained as

\[ \left[ \log Z_i \right]_n = \lim_{n \to 0} \left[ \frac{Z^n - 1}{n} \right]. \] (5)

From the calculation in appendix II, $\left[ Z^n \right]_J$ for an integer $n$ is obtained as

\[ [Z] = \left[ \Pi_{p>q} \int \frac{Ndq^{\mu,\nu}}{2\pi} \right] \cdot \left[ \Pi_{p>q} \int \frac{N\omega^{\mu,\nu}}{2\pi} \right] \cdot \exp \left[ M \log \Psi \left( q^{\mu,\nu} \right) + NS \left( q^{\mu,\nu}, \omega^{\mu,\nu} \right) \right] \] (6)

where $q^{\mu,\nu}$ and $\omega^{\mu,\nu}$ are the spin-glass order parameter and its conjugate, respectively. In addition, $\Psi \left( q^{\mu,\nu} \right)$ and $S \left( q^{\mu,\nu}, \omega^{\mu,\nu} \right)$ are given by

\[ \Psi \left( q^{\mu,\nu} \right) = \int \Pi_{p} \frac{dm_{\mu}}{\sqrt{2\pi}} \exp \left( -\frac{1}{2\sigma^2} \left( m^{\mu} Q^{-1} m^{\mu} + \beta \sum_{p} f_{p} (m^{\mu}) \right) \right), \] \] (7)

and

\[ S \left( q^{\mu,\nu}, \omega^{\mu,\nu} \right) = \frac{1}{N} \log \text{Tr} \left[ e_{N} \sum_{\mu,\nu} e_{N}^{\mu,\nu} - \sum_{\mu,\nu} q^{\mu,\nu} \omega^{\mu,\nu} \right] \] (8)

where $m_{\mu} = (m_{\mu}^{1}, ..., m_{\mu}^{n})$, $f_{p} (m_{\mu}) = \tanh \beta m_{\mu}$ and $Q$ is an $n \times n$ matrix with diagonal elements equal to 1 and off-diagonal elements $q^{\mu,\nu}$. For sufficiently large $N$, the integral with respect to $q^{\mu,\nu}$ and $\omega^{\mu,\nu}$ is calculated by the saddle point method as

\[ [Z] \approx 1 + N \left( \gamma^{-1} \log \Psi \left( q^{\mu,\nu} \right) + S \left( q^{\mu,\nu}, \omega^{\mu,\nu} \right) \right), \] (9)

where the order parameters $q^{\mu,\nu}$ and $\omega^{\mu,\nu}$ are determined by

\[ q^{\mu,\nu} = \left\langle \delta_{i}^{\mu} \delta_{i}^{\nu} \right\rangle_{L}, \] (10)
Here, $\langle \cdot \rangle_L$ indicates the average over the distribution
\[
\exp \left( \sum_{\mu>\mu'} g_{\mu}^a g_{\mu'}^a \omega_{\mu}^a \right) / \text{Tr}_{\omega_{\mu}^a} \exp \left( \sum_{\mu>\mu'} g_{\mu}^a g_{\mu'}^a \omega_{\mu}^a \right).
\]

For further calculation, we assume the replica symmetry (RS) ansatz as $q_{\mu}^a = q$ and $\omega_{\mu}^a = \omega$ for all $\mu \neq \mu'$. After taking the limit $n \to 0$, this ansatz provides equations for $q$ and $\omega$ as
\begin{align}
q &= \int Dz \left( \tanh \left( \sqrt{\omega} z \right) \right)^2, \\
\omega &= \frac{\sigma^2 (\beta')^2}{\gamma} \int Dz \left[ f_0'(m) \right]_{\omega},
\end{align}
where $Dz = e^{-z^2} dz / \sqrt{2\pi}$ and $[.]_{\omega}$ indicates taking an average over the distribution
\[
\exp \left( \Theta (m, z; \beta') \right) / \int dm \exp \left( \Theta (m, z; \beta') \right)
\]
with
\[
\Theta (m, z; \beta') \equiv \beta' f_0 (m) - \left( m + \sqrt{\sigma^2 q z} \right)^2 / 2\sigma^2 (1 - q).
\]
For all the regions on the $T_g - T_p$ plane, $q$ is always positive except for when $T_g = \infty$ or $T_p = \infty$, indicating that the model has no paramagnetic phase. By using equations (5), (9), (12) and (13), the free energy per locus for the RS solution with $n \to 0$, $f_{\text{RS}}$, is calculated as
\[
-\beta' f_{\text{RS}} = -\frac{(1 - q) \omega}{2} + \int Dz \log 2 \cosh \left( \sqrt{\omega} z \right) + \int Dz \log \int dm \frac{e^{\Theta (m, z; \beta')}}{\sqrt{2\pi \sigma^2 (1 - q)}}.
\]

The RS solution is invalid when the de Almeida–Thouless (AT) condition [33, 34] is violated, which is accompanied by replica symmetry breaking (RSB). Thus, we perform Monte Carlo simulations (MCSs) to estimate the fitness and entropy for the RSB phase, while theoretical estimates based on the RS ansatz are used in the RS phase. As shown by the dashed yellow line in figure 2(a), the phase boundary between the RS and RSB phases on the $T_g - T_p$ plane obtained by AT conditions indicates that RSB occurs when either $T_g$ or $T_p$ is small. Details of the derivation of AT conditions and phase boundary are given in appendix II.

By differentiating $-\gamma \beta' f_{\text{RS}}$ with respect to $\beta'$, the mean fitness per unit phenotype of the RS solution, $\gamma \langle \phi \rangle$, is obtained as
\[
\gamma \langle \phi \rangle = \int Dz \int dm \frac{e^{\Theta (m, z; \beta')}}{\sqrt{2\pi \sigma^2 (1 - q)}} \tanh \beta m.
\]

In the RS phase, the estimates obtained from the MCS are in agreement with the theoretical estimates. A color map in figure 2(a) shows the dependence of the fitness per unit phenotype $\gamma \langle \phi \rangle$ on $T_p$ and $T_g$ with fixed $\gamma$ and illustrates that $\gamma \langle \phi \rangle$ decreases as either $T_p$ or $T_g$ increases. This decrease of $\gamma \langle \phi \rangle$ against $T_p$ and $T_g$ is consistent with the results from the analysis of simple
The dependence of $\gamma \langle \phi \rangle$ on $\gamma$ with fixed $T_p$ and $T_g$ is shown in figures 2(b) and (c). In contrast to simple GPM, $\gamma \langle \phi \rangle$ increases with increasing $\gamma$, contradicting the classical view that genetic redundancy decreases fitness at evolutionary equilibrium.

Entropy per phenotype is calculated as $\gamma s = -\gamma \beta (\langle \phi \rangle + f_{RS})$ for the RS phase; $\gamma$ is evaluated for the RS and RSB phases numerically by using a multicanonical Monte Carlo method [35–37], which is often used in spin-glass systems and other fields [38, 39].
Figure 2(d) shows that the dependence of $\gamma_s$ on $\gamma$ with fixed $T_p$ and $T_g$ is demonstrated by the near-linear increase with increasing $\gamma$, which is similar to the case for simple GPM, where the presence of redundant genes permits the accumulation of mutations, with a consequent increase in genotypic diversity in a population. Thus, for complex GPM, genetic redundancy $\gamma$ increases both fitness per unit phenotype, $\gamma \langle \phi \rangle$, and entropy per unit phenotype, $\gamma s$.

Finally, to exclude the possibility that the above conclusion is dependent on a specific choice of the fitness function, a numerical simulation is performed with an alternative fitness function $\Phi (p) \equiv \sum_i s_i p_i$, in which $\{ s_i \}$ (target sequence) are quenched random variables and each $s_i$ takes $s_i = 1$ (or $-1$) with probability $\rho$ (or $1 - \rho$). The results are similar to those described above; the redundancy always evolves for complex GPM, but not for simple GPM (data not shown).

5. Discussion

In the present study, an adiabatic spin model with high-dimensional GPM is investigated, where both the complexity in GPM (represented by random quenched variables $J_{ij}$) and genetic redundancy ($\gamma$) are controllable parameters. We compared evolutionary steady states between for simple and complex GPM cases. For simple GPM, fitness per unit phenotype $\gamma \langle \phi \rangle$ decreases as $\gamma$ increases (figures 1(b) and (c)), indicating that, consistent with the classical view [15–17], redundancy is suppressed by selection pressure. As shown in figure 1(c), when $T_p$ is close to zero, $\gamma \langle \phi \rangle$ is almost independent of $\gamma$, because, for any $g$, nearly all phenotypes that can exist are either $p_i = 1$ for all $i$ or $p_i = -1$ for all $i$ at $T_p \approx 0$. Nevertheless, it is presumed that $T_p$ cannot be 0 due to unavoidable phenotypic fluctuation, and thus evolution of genetic redundancy should be suppressed. As is shown in figure 1(d), the decrease in redundancy reduces entropy per unit phenotype $\gamma s$; since less genetic diversity in the population hinders accessibility to a novel genotype and therefore reduces evolvability, selection pressure towards less genetic redundancy as in the case of simple GPM will lead a population to an evolutionary dead end.

Remarkably, this is not true for a complex GPM. In this case, a population with higher $\gamma$ exhibits higher fitness $\gamma \langle \phi \rangle$, as shown in figures 2(b) and (c), indicating that genetic redundancy, i.e., a system with higher $\gamma = N/M$, can evolve, contrary to the classical view. A possible explanation is that a larger $N$ provides greater degrees of freedom in $g$ to optimize the fitness, and thus enables realization of $g$ that provides higher fitness than the highest fitness found in systems with smaller $N$. At the same time, a system with smaller $M$ decreases the variety of connections from $g_i$ to $p_i$ and diminishes frustrations (e.g. evolutionary constraints) in $g_i$ to optimize the fitness. Thus, a system with larger $N$ and smaller $M$ provides higher fitness, implying that genetic redundancy can evolve under a complex GPM. It is interesting to note the possible connection between this finding and so called subfunctionalization [22], where a gene with different functions is separated into several genes by gene duplication with a subsequent accumulation of mutations. This process is thought to facilitate the preservation of redundant genes, and presents an alternative scenario to ‘neofunctionalization’ [14]. The present results suggest that the presence of frustrations in GPM facilitates the
evolution of genetic redundancy, where the ancestral (pre-duplication) gene is not required to have multiple functions.

In contrast to the simple GPM, $\gamma \langle \phi \rangle$ depends on $\gamma$, even when $T_p$ is close to zero, because the phenotype $p$ is critically dependent on $g$ and $J_{ij}$ for $T_p \sim 0$. Although the model disregards the costs associated with redundant genes, for instance the cost to DNA replication, incorporating these into the model is straightforward, in which case redundancy would not indefinitely increase but would instead be counterbalanced by the costs of redundancy.

As is shown in figure 2(d), a larger $\gamma$ accompanies increases in entropy per unit phenotype $\gamma s$ and therefore in growth in genetic diversity. This allows a population to access a variety of novel genotypes, which could accelerate the emergence of evolutionary innovations. Thus, a population with complex GPM can have evolvability. It is worth noting that genetic diversity may enable a population to rapidly respond to large environmental changes [27, 41, 42]; as such, the present results also suggest that a population with complex GPM can potentially adapt successfully to new environments.

The model with complex GPM exhibits an RS–RSB transition, which was also reported in a different model by Sakata et al [9], asserting that the RSB phase is biologically unfavorable. Similar to their study, here it is demonstrated that phenotypic fluctuations suppress the appearance of the RSB phase. This RS–RSB transition can appear in many evolutionary models with complex GPM, and its biological significance merits further investigation.

In summary, we have demonstrated that complexity in GPM can promote the evolution of genetic redundancy. The mechanism is general, and the recombination process in a sexual population does not weaken the proposed preferences in genetic redundancy. Selection processes with complex GPM favor gene duplication, which can further enhance the potential for evolutionary innovations.

Acknowledgments

We would like to acknowledge helpful discussions with Koji Hukushima. This work was partially supported by a Grant-in-Aid for Scientific Research (no 21120004) on Innovative Areas ‘Neural creativity for communication’ (no 4103), the Grant-in-Aid MEXT/JSPS (no 24115503) and the Platform for Dynamic Approaches to Living System from MEXT, Japan.

Appendix I. Derivation of the Boltzmann distribution equation (1)

We assume that $p_i$ and $p_j$ are mutually independent, namely $P(p|g) = \prod_i P(p_i|g)$, and that $P(p|g)$ is given by a product of contributions from each $g_j$, namely $P(p|g) = \prod_j P(p_j|g_j)$. Now, let us denote $P(p_j|g_j)$ for $p_j = g_j$ by $a_j$, and that for $p_j = -g_j$ by $b_j$. Then, $P(p|g)$ is written as
where $C$ is the normalization factor that satisfies $\sum_p P(p|g) = 1$, and is given by $C = \prod_g \left( a_g + b_g \right)^{-1}$. This $P(p|g)$ can always be transformed into the Boltzmann distribution equation (1) by introducing $\beta J_g \equiv \log \sqrt{a_g/b_g}$.

Appendix II. Replica calculation

$$Z^n = \sum_{[g^n]} \exp \left( \beta' \sum_{\mu} \phi(g^n) \right)$$

$$= \sum_{[g^n]} \prod_{i,\mu} \int dm_i^{\mu} \cdot \delta \left( m_i^{\mu} - \sum_j J_{g_j}^{n,\mu} \right) \exp \left( \beta' \sum_{\mu} \sum_i f_0 \left( \beta' J_{g_i}^{n,\mu} \right) \right)$$

$$= \sum_{[g^n]} \prod_{i,\mu} \int dm_i^{\mu} \int \frac{dr_i^{\mu}}{2\pi} \exp \left( \beta' \left( m_i^{\mu} - \sum_j J_{g_j}^{n,\mu} \right) + \beta' f_0 \left( \beta m_i^{\mu} \right) \right),$$

where a Fourier transform of the Dirac delta function is used. Taking the average of $Z^n$ with respect to $J_g$ over $P(J_g) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -J_g^2 / 2\sigma^2 \right)$ yields

$$[Z]^n = \sum_{[g^n]} \left( \prod_{\mu, i} \int \frac{dr_i^{\mu}}{2\pi} \int dm_i^{\mu} \right) \times \exp \left[ -\frac{N\sigma^2}{2} \sum_{\mu} \sum_i \left( r_i^{\mu} \right)^2 - \sigma^2 \sum_{\mu>\mu'} \left( \sum_i r_i^{\mu} r_i^{\mu'} \right) \left( \sum_j g_j^{n,\mu} g_j^{n,\mu'} \right) + \sum_{\mu} \sum_i \left( ir_i^{\mu} m_i^{\mu} + \beta f_0 \left( \beta m_i^{\mu} \right) \right) \right].$$

By using the identity:

$$1 = \left[ \prod_{\mu, i} \int \frac{dr_i^{\mu}}{2\pi} \int dm_i^{\mu} \right] \left[ \prod_{\mu>\mu'} \int \frac{Ndq^{\mu,\mu'}}{2\pi} \delta \left( Nq^{\mu,\mu'} - \sum_j g_j^{n,\mu} g_j^{n,\mu'} \right) \right],$$

we obtain

$$[Z]^n = \left( \prod_{\mu, i} \int \frac{dr_i^{\mu}}{2\pi} \int dm_i^{\mu} \right) \cdot \left[ \prod_{\mu>\mu'} \int \frac{Ndq^{\mu,\mu'}}{2\pi} \right] \left[ \prod_{\mu>\mu'} \int \frac{dq^{\mu,\mu'}}{2\pi} \right] \times \exp \left[ -\frac{N\sigma^2}{2} \sum_{\mu} \sum_i \left( r_i^{\mu} \right)^2 - \sigma^2 \sum_{\mu>\mu'} \left( \sum_i r_i^{\mu} r_i^{\mu'} \right) \left( Nq^{\mu,\mu'} \right) + \sum_{\mu} \sum_i \left( ir_i^{\mu} m_i^{\mu} + \beta f_0 \left( \beta m_i^{\mu} \right) \right) \right.$$
Using multiple Gaussian integrals:
\[
\int_{-\infty}^{\infty} dx \exp \left( -\frac{1}{2} x C x + i x \cdot q \right) = \frac{(2\pi)^{N/2}}{\det(C)^{1/2}} \exp \left( -\frac{1}{2} (q C^{-1} q) \right),
\]
and by defining \( \omega^{\mu \nu} = -i \tilde{q}^{\mu \nu} \), we obtain equation (6).

**Appendix III. Stability analysis of the replica symmetry (RS) solution**

To examine the validity of the RS solutions, we performed a stability analysis using the de Almeida–Thouless (AT) condition [33, 34], where the RS solution is considered as stable when the Hessian matrix \( H \) of \( f_{RS} \) with respect to the order parameters, is positive semidefinite. The Hessian matrix \( H \) is described as
\[
H = \begin{pmatrix}
H^{qq} & H^{q\omega} \\
H^{\omega q} & H^{\omega \omega}
\end{pmatrix},
\]
using submatrices \( H^{qq} = \partial^2 f_{RS} / \partial q^a \partial q^c \), \( H^{q\omega} = -\partial^2 f_{RS} / \partial q^a \partial \omega^c \) and \( H^{\omega q} = \partial^2 f_{RS} / \partial q^a \partial (i \omega^d) = (1/\beta') I \), where \( i \) is the unit complex number and \( I \) is the unit matrix. Although each submatrix \( H^{qq} \) and \( H^{\omega \omega} \) has three eigenvalues, e.g., \( H^{qq} \) has \( \lambda_1^{qq}, \lambda_2^{qq} \) and \( \lambda_3^{qq} \), two of the three eigenvalues degenerate to \( n \rightarrow 0 \) as \( \lambda_1^{qq} = \lambda_2^{qq} \). Therefore, only two eigenvalues for each \( H^{qq} \) and \( H^{\omega \omega} \) have to be considered as
\[
\lambda_2^{qq} = -\frac{\beta' \sigma^4}{\gamma} \int Dz \left( \left[ f''(m) \right]_a + \beta' \left[ f'^2(m) \right]_a - 3\beta' \left[ f'(m) \right]_a^2 \right)
\times \left( \left[ f''(m) \right]_a + \beta' \left[ f'^2(m) \right]_a - \beta' \left[ f'(m) \right]_a^2 \right)
\]
\[
\lambda_3^{qq} = -\frac{\beta' \sigma^4}{\gamma} \int Dz \left( \left[ f''(m) \right]_a + \beta' \left[ f'^2(m) \right]_a - \beta' \left[ f'(m) \right]_a^2 \right)^2
\]
\[
\lambda_2^{\omega \omega} = \frac{1}{\beta'} \int Dz \left( (\tanh \sqrt{\omega} z)^2 - 1 \right) \left( 3 (\tanh \sqrt{\omega} z)^2 - 1 \right)
\]
\[
\lambda_3^{\omega \omega} = \frac{1}{\beta'} \int Dz \left( (\tanh \sqrt{\omega} z)^2 - 1 \right)^2.
\]
All the eigenvalues of \( H \) are positive when the following four conditions are satisfied.

**AT2 (a):** \( \lambda_2^{qq} + \lambda_2^{\omega \omega} > 0 \), \hspace{1cm} (A.3)

**AT2 (b):** \( 1 + \beta' \lambda_2^{qq} \lambda_1^{\omega \omega} > 0 \) \hspace{1cm} (A.4)

**AT3 (a):** \( \lambda_3^{qq} + \lambda_3^{\omega \omega} > 0 \), \hspace{1cm} (A.5)

**AT3 (b):** \( 1 + \beta' \lambda_3^{qq} \lambda_3^{\omega \omega} > 0 \). \hspace{1cm} (A.6)
When one of these conditions is violated, the RS solution becomes unstable and replica symmetry breaking (RSB) occurs. As far as we explored the region of parameters $T_s$, $T_p$ and $\gamma$, $AT^2(a)$ is always violated first, triggering the RS–RSB transition.

References

[1] Falconer D S and Mackay T F C 1996 Introduction to Quantitative Genetics 4th edn (Harlow: Longman)
[2] Jaeger J et al 2004 Dynamic control of positional information in the early drosophila embryo Nature 430 368–71
[3] Fujimoto K, Ishihara S and Kaneko K 2008 Network evolution of body plans PLoS One 3 e2772
[4] Kaneko K 2007 Evolution of robustness to noise and mutation in gene expression dynamics PLoS One 2 434
[5] Ciliberti S, Martin O C and Wagner A 2007 Robustness can evolve gradually in complex regulatory gene networks with varying topology PLoS Comput Biol 3 e15
[6] Villarreal C, Padilla-Longoria P and Alvarez-Buylla E R 2012 General theory of genotype to phenotype mapping: derivation of epigenetic landscapes from n-node complex gene regulatory networks Phys. Rev. Lett. 109 118102
[7] Franz S and Peliti L 1997 Error threshold in simple landscapes J. Phys. A: Math. Gen. 30 4481
[8] Sato K and Kaneko K 2007 Evolution equation of phenotype distribution: general formulation and application to error catastrophe Phys. Rev. E 75 061909
[9] Sakata A, Hukushima K and Kaneko K 2012 Replica symmetry breaking in an adiabatic spin-glass model of adaptive evolution Europhys. Lett. 99 68004
[10] Saito N, Ishihara S and Kaneko K 2013 Baldwin effect under multipeaked fitness landscapes: phenotypic fluctuation accelerates evolutionary rate Phys. Rev. E 87 052701
[11] Molin L, Mounsey A, Aslam S, Bauer P, Young J, James M, Sharma-Oates A and Hope I A 2000 Evolutionary conservation of redundancy between a diverged pair of forkhead transcription factor homologues Development 127 4825–35
[12] Baugh L R, Wen J C, Hill A A, Slonim D K, Brown E L and Hunter C P 2005 Synthetic lethal analysis of caenorhabditis elegans posterior embryonic patterning genes identifies conserved genetic interactions Genome Biol. 6 R45
[13] Musso G et al 2008 The extensive and condition-dependent nature of epistasis among whole-genome duplicates in yeast Genome Res. 18 1092–9
[14] Ohno S et al 1970 Evolution by Gene Duplication (New York: Springer)
[15] Haigh J 1978 The accumulation of deleterious genes in a populationmuller’s ratchet Theor. Population Biol. 14 251–67
[16] Thomas J H 1993 Thinking about genetic redundancy Trends Genetics 9 395–9
[17] Lynch M, Bürger R, Butcher D and Gabriel W 1993 The mutational meltdown in asexual populations J. Heredity 84 339–44
[18] Haldane J B S 1933 The part played by recurrent mutation in evolution Am. Naturalist 67 5–19
[19] Fisher R A 1935 The sheltering of lethals Am. Naturalist 69 446–55
[20] Nei M and Roychoudhury A K 1973 Probability of fixation of nonfunctional genes at duplicate loci Am. Naturalist 107 362–72
[21] Walsh J B 1995 How often do duplicated genes evolve new functions? Genetics 139 421–8
[22] Force A, Lynch M, Pickett F B, Amores A, Yan Y-I and Postlethwait J 1999 Preservation of duplicate genes by complementary, degenerative mutations Genetics 151 1531–45
[23] Lynch M, O’Hely M, Walsh B and Force A 2001 The probability of preservation of a newly arisen gene duplicate Genetics 159 1789–804
[24] Nowak M A, Boerlijst M C, Cooke J and Smith J M 1997 Evolution of genetic redundancy Nature 388 167–71
[25] Wagner A 2000 The role of population size, pleiotropy and fitness effects of mutations in the evolution of overlapping gene functions Genetics 154 1389–401
[26] Krakauer D C and Plotkin J B 2002 Redundancy, antiredundancy, and the robustness of genomes Proc. Natl Acad. Sci. 99 1405–9
[27] Wagner A 2005 Robustness and Evolvability in Living Systems (Princeton NJ: Princeton University Press)
[28] Iwasa Y 1988 Free fitness that always increases in evolution J. Theor. Biol. 135 265–81
[29] Sella G and Hirsh A E 2005 The application of statistical physics to evolutionary biology Proc. Natl Acad. Sci. 102 9541–6
[30] de Vladar H P and Barton N H 2011 The contribution of statistical physics to evolutionary biology Trends Ecology & Evolution 26 424–32
[31] Demetrius L 1997 Directionality principles in thermodynamics and evolution Proc. Natl Acad. Sci. 94 3491–8
[32] Mustonen V and Lässig M 2010 Fitness flux and ubiquity of adaptive evolution Proc. Natl Acad. Sci. 107 4248–53
[33] Mézard M, Parisi G and Virasoro M A 1987 Spin Glass Theory and Beyond (Singapore: World Scientific)
[34] Almeida J R L and Thouless D J 1978 Stability of the sherrington-kirkpatrick solution of a spin glass model J. Phys. A: Math. Gen. 11 983
[35] Berg B A and Neuhaus T 1992 Multicanonical ensemble: a new approach to simulate first-order phase transitions Phys. Rev. Lett. 68 9–12
[36] Berg B A and Celik T 1992 New approach to spin-glass simulations Phys. Rev. Lett. 69 2292–5
[37] Wang F and Landau D P 2001 Efficient, multiple-range random walk algorithm to calculate the density of states Phys. Rev. Lett. 86 2050–3
[38] Saito N and Iba Y 2011 Probability of graphs with large spectral gap by multicanonical Monte Carlo Comput. Phys. Commun. 182 223–5
[39] Saito N, Iba Y and Hukushima K 2010 Multicanonical sampling of rare events in random matrices Phys. Rev. E 82 031142
[40] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 1953 Equation of State Calculations by Fast Computing Machines J. Chem. Phys. 21 1087
[41] Landry C R, Lemos B, Rifkin S A, Dickinson W J and Hartl D L 2007 Genetic properties influencing the evolvability of gene expression Science 317 118–21
[42] Hayden E J, Ferrada E and Wagner A 2011 Cryptic genetic variation promotes rapid evolutionary adaptation in an rna enzyme Nature 474 92–95