Heterosis of fitness and phenotypic variance in the evolution of a diploid gene regulatory network

Okubo, Kenji; Kaneko, Kunihiko; Nelson, Karen E

Published in:
PNAS Nexus

DOI:
10.1093/pnasnexus/pgac097

Publication date:
2022

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

Document license:
CC BY

Citation for published version (APA):
Heterosis of fitness and phenotypic variance in the evolution of a diploid gene regulatory network

Kenji Okubo and Kunihiko Kaneko

Abstract

Heterosis describes the phenomenon, whereby a hybrid population has higher fitness than an inbred population, which has previously been explained by either Mendelian dominance or overdominance under the general assumption of a simple genotype–phenotype relationship. However, recent studies have demonstrated that genes interact through a complex gene regulatory network (GRN). Furthermore, phenotypic variance is reportedly lower for heterozygotes, and the origin of such variance-related heterosis remains elusive. Therefore, a theoretical analysis linking heterosis to GRN evolution and stochastic gene expression dynamics is required. Here, we investigated heterosis related to fitness and phenotypic variance in a system with interacting genes by numerically evolving diploid GRNs. According to the results, the heterozygote population exhibited higher fitness than the homozygote population, indicating fitness-related heterosis resulting from evolution. In addition, the heterozygote population exhibited lower noise-related phenotypic variance in expression levels than the homozygous population, implying that the heterozygote population is more robust to noise. Furthermore, the distribution of the ratio of heterozygote phenotypic variance to homozygote phenotypic variance exhibited quantitative similarity with previous experimental results. By applying dominance and differential gene expression rather than only a single gene expression model, we confirmed the correlation between heterosis and differential gene expression. We explain our results by proposing that the convex high-fitness region is evolutionarily shaped in the genetic space to gain noise robustness under genetic mixing through sexual reproduction. These results provide new insights into the effects of GRNs on variance-related heterosis and differential gene expression.

Keywords: heterosis, evolution, phenotypic variance, gene regulatory network, sexual reproduction

Significance Statement:

Heterosis, higher fitness in hybrid populations than in inbred populations, is a long-standing mystery in genetics, evolution, and breeding studies. Heterosis does not necessarily result from a trait determined by a single gene, but likely involves stochasticity and interactions among genes. Through numerical evolution of a gene regulatory network, we demonstrate that heterosis is shaped by evolution to achieve robustness against noise with genetic mixing by sexual recombination in gene expression dynamics. In other words, a mixed population is more robust to noise than an inbred population, as previously revealed experimentally by reduced phenotypic variance. This observed link between heterosis and phenotypic robustness, Mendelian dominance, and the convex single-humped fitness landscape represents a novel avenue in evolution and genetics research.

Introduction

Sexual reproduction generally involves mixing genetic information from multiple parents, which is a strategy that has been adopted for a wide range of species. Many diploid organisms undergo meiosis, which results in the recombination of two parental genomes to produce gametes, and then the gametes combine to produce the genome of the offspring.

One of the most remarkable phenomena in sexual reproduction among populations is heterosis, whereby hybrid populations exhibit higher fitness than inbred populations obtained by repeated sexual reproduction among genetically close individuals. Inbreeding depression is often regarded as two sides of the same phenomenon of heterosis, which indicates lower fitness of the inbred than hybrid population. To date, the origin of both heterosis and inbreeding depression has largely been discussed from a genetics perspective (1–4), converging on a similar argument.

One of the simplest models proposed to explain heterosis and inbreeding depression is the dominance model based on Mendelian dominant inheritance pattern (5). In Mendelian dominance (6), if two diploid parent genomes have the genes AA and aa, and show the corresponding traits “A” and “a,” then the child has the gene Aa and shows trait “A” when trait “A” is dominant...
and trait “a” is recessive. In the dominance hypothesis, mutations accumulated in homozygotes of an inbred population are considered to be deleterious and recessive. However, Mendelian dominance cancels out this deleterious mutational effect in heterozygotes of a hybrid population, contributing to its higher fitness. This explanation corresponds to complementation among products from alleles at the molecular level. Although the dominance or complementation model can effectively explain several observations of inbreeding depression, it cannot explain some observations related to heterosis and inbreeding depression (7). First, inbreeding with selection could not reduce the effect of heterosis, although the deleterious and recessive mutations would be decreased in the course of inbreeding with selection, which should then reduce heterosis according to the dominance or complementation model (8, 9). Second, in tetraploids, the deleterious and recessive mutational effect should be reduced according to the dominance model; however, tetraploid organisms do not show less heterosis even though tetraploid inbreeding decreases the ratio of homozygotes (AAAA or BBBB) (10–13).

In contrast, the overdominance hypothesis has been proposed, in which heterozygotes can acquire a novel trait that confers higher fitness relative to that of both parental homozygotes (5). In the (positive) overdominance model, the heterozygote Aa has higher fitness than AA and aa. Although there is sparse direct evidence for the overdominance model, it can explain some results that cannot be explained by the deleterious and recessive mutation models. For example, the abovementioned increase in heterosis in the artificially evolved species (8, 9) or tetraploid (10–13) can be explained by the overdominance model. At the molecular level, the interaction between alleles is assumed to produce such heterozygote advantage (7). By noting that a quantitative trait studied in the context of heterosis is a result of multiple, potentially interacting, loci or genes (14, 15), the relevance of an interaction effect as a result of gene regulatory hierarchies has been proposed (16). For example, Bar-Zvi et al. (17) focused on two factors, i.e. enhancement of growth pathways and impairment of growth inhibition, to explain relatively higher hybrid fitness. For the latter factor, they pointed out the existence of an incompatibility gene regulatory network (GRN), which is suppressed in the heterozygote, leading to heterosis (18).

Another concept that has been proposed to explain this phenomenon is differential gene expression, which is related to overdominance, and has attracted substantial research attention in the context of studying inbreeding depression and GRNs (19–21). In the differential gene expression model, heterozygous F1 hybrids exhibit different gene expression patterns from those of their inbred parents (22). In contrast to the overdominance model, the different expression pattern itself does not necessarily confer higher fitness. The existence of differential gene expression and heterosis have shown a positive correlation (15, 23); however, the mechanism explaining this apparent connection has not yet been elucidated.

The traditional arguments related to heterosis described above are mainly based on an assumption of a simple genotype-phenotype relationship; that is, the case in which one gene is associated with one phenotype (6, 8, 24). However, recent studies demonstrate the relevance of interactions among genes to determine the ultimate phenotype (25–30). Therefore, it is necessary to construct a model in which the phenotype is determined as a result of genetic interactions to unveil the relevance of the genotype-phenotype relationship to heterosis with respect to fitness and phenotypic variance. Toward this end, we focused on the GRN model (31–36). In a GRN, protein expression depends on the mutual activation or suppression of genes, where a phenotype (e.g. expression of a single protein) depends on the expression levels of multiple genes (31). Recent transcriptome analyses have revealed such interactions among genes (37). Furthermore, interactions of gene regulatory regions (i.e. promoters, transcription factor binding sites, and enhancers) change more frequently throughout the evolutionary course compared with changes in protein-coding genes (38).

The GRN model itself assumes neither dominance nor overdominance. Therefore, there is no apparent Mendelian dominance with respect to gene expression levels in a diploid-GRN model, and whether or not such dominance can emerge through evolution is not evident based on the model alone (36). Although some studies have addressed heterosis using GRN models (39, 40), the relevance of phenotypic robustness to noise and genomic mixing by sexual recombination to heterosis has not yet been considered in this context.

Hence, studying how phenotypic variance changes between hybrid and inbred populations is also important. Indeed, in contrast to the standard context of heterosis, in which the fitness of a fitness-related trait is compared between inbred and hybrid populations, Phelan and Austad (41) examined data relating to the traits of over 100 species, and demonstrated that the variance of traits also shows heterosis; that is, the variance of traits for hybrid populations is smaller than that of inbred populations. Furthermore, gene expression dynamics generally involve stochastic noise either due to environmental fluctuation or internal stochasticity in organisms, and the response against such noise is determined by the genotype corresponding to the GRN. However, the possible link between noise and heterosis in a GRN has rarely been discussed (42), especially from the perspective of quantitative genetics. Nevertheless, stochasticity in phenotypes (i.e. gene expression levels) and noise robustness have recently received substantial attention in the context of homeostasis, cell state selection, and GRNs (26, 43–49). Considering the result of Phelan and Austad (41), it is important to explore the possibility of the heterosis of noise-induced phenotypic variance, because total phenotypic variance contains the non-negligible contribution of noise-induced phenotypic variance (50).

In this study, we analyzed heterosis and variance-related heterosis by simulating the evolution of a diploid-GRN. The organization of this paper is as follows. First, we introduce the diploid-GRN model and its evolutionary simulation. Second, we compare the fitness of homozygote and heterozygote groups generated from evolved populations and the outcome of fitness-related heterosis. Third, we compare the phenotypic variances between homozygote and heterozygote groups, demonstrating that the variance in the latter is markedly reduced. Additionally, we compare the ratio of heterosis related to phenotypic variance obtained in our simulation to previous measurement data (41). We also investigate the correlation between hybrid fitness and the dominance of gene expression patterns or differential gene expression that emerged through evolution. Finally, an intuitive explanation of the results is provided regarding the fitness landscape.

**Description of the model and simulation**

**Evolution simulation based on a diploid-GRN model**

First, we introduce the theoretical GRN model, which is largely based on the model presented in (36). Each gene \( i = 1, 2, \ldots, N \) has an expression level \( x_i(t) \) at time step \( t \). In the model, \( x_i(t) \) is
scaled such that it takes on a value between zero (nonexpression) and one (expression). Each gene interacts with other genes and with itself, with interaction between the jth gene and the ith gene described by the matrix $J_{ji}$ (31–35). $J_{ij}$ can take three values: +1, −1, or 0, which represent the activation, inhibition, and lack of interaction between gene i and gene j, respectively. We adopted a discrete-time model (31, 44, 51, 52), in which the expression level $x_i(t + 1)$ in the next time step is determined by $\sum_{j=1}^{N} J_{ij} x_j(t)$, which represents the gene regulatory effect. According to the sigmoid function $f[x] = 1/(1 + \exp[-\beta x])$ with a large $\beta$ ($\beta = 100$), the expression dynamics of the haploid, where the individual has one set of genome, is given by

$$x_i(t + 1) = f(\sum_{j=1}^{N} J_{ij} x_j(t) - \theta),$$

where $\theta$ is the threshold for the expression level.

Therefore, we extended this model to the diploid context to investigate heterosis. We supposed that the products from each chromosome are indistinguishable since promoter or enhancer evolution is faster than that of the coding gene regions (36). Therefore, we can redefine $x_i(t)$ as an average of two expression levels in two genomes. Since diploid cells have two regulatory matrices, $J_{ij}$ and $J_{ij}^{(2)}$, the averaged expression level $x_i(t)$ affects the promoter or enhancer of two genomes, 1 and 2. The gene expression dynamics of a diploid-GRN are then determined by the sum of two indistinguishable products from two chromosomes; thus, the dynamics are given by modifying $f(\sum_{j=1}^{N} J_{ij} x_j(t) - \theta)$ to

$$\frac{1}{2}(f(\sum_{j=1}^{N} J_{ij}^{(1)} x_j(t) - \theta) + f(\sum_{j=1}^{N} J_{ij}^{(2)} x_j(t) - \theta))$$

as

$$x_i(t + 1) = \frac{1}{2}(f(\sum_{j=1}^{N} J_{ij}^{(1)} x_j(t) - \theta) + f(\sum_{j=1}^{N} J_{ij}^{(2)} x_j(t) - \theta)) + \sqrt{x_i(t)} \eta(0, \sigma).$$

(See the Materials and Methods for details of derivation.)

In (3), Gaussian noise, $\sqrt{\text{Var}(\eta)(0, \sigma)}$ of small magnitude, is added to the dynamics to account for the stochasticity in gene expression. The noise term increases $\sqrt{\text{Var}(x_i)}$, as is often adopted in stochastic gene expression dynamics. (This could avoid $x_i(t)$ crossing over to a negative value for a continuous time model, however, for a discrete-time model, such crossing could happen, in which case $x_i(t)$ is reset to zero.)

The calculation of the dynamics is based on the following settings. The computation starts from the given initial expression levels $x_i(0)$, where $x_i(0) = 0$ for most genes and $x_i(0) = 1$ for several genes. $x_i(0)$ reaches a fixed point after a certain time $T$ in most cases, where $x_i(0) = 0$, $x_i(0) = 1$, or $x_i(0) = 0.5$ is achieved depending on the structure of the network $J_{ij}$ in the diploid-GRN model. However, $x_i(0) = 0.5$ is hardly ever reached for the evolved GRN under the following fitness definition. The noise term $\sqrt{\text{Var}(\eta)(0, \sigma)}$ kicks the state out of a certain fixed point, although the system is usually attracted to the same fixed point or to some other point.

Evolutionary simulation of the diploid-GRN was carried out according to the following procedure, using a meiosis-like genetic algorithm. First, fitness is determined by the distance of the expression pattern of genes $x_i(T)$ from a prescribed target pattern $x_{\text{target}}^{(i)} = 1$ for $i = 1, 2, \ldots, M < N$. In this simulation, $M = 10$ and $N = 100$. Only some genes have this target pattern, so that the expression levels of most genes are not under selection pressure. That is, there is no preference between $x_i(0) = 0$ or $x_i(0) = 1$ for most genes. Next, genetic change is introduced into genomes. After the fitness computation, two parents are chosen with a probability exponential to the fitness. A total of two genomes in each parent, $J_{ij}^{(1)}$ and $J_{ij}^{(2)}$, are mixed by recombination to provide a gamete. Next, two genomes from the two parents provide a new $J_{ij}^{(1)}$ and $J_{ij}^{(2)}$ in the next generation. After meiosis, the elements of the new $J_{ij}^{(1)}$ and $J_{ij}^{(2)}$ are modified by a mutation with the rate $\mu$. This mutation is included as a change in the matrix element $J_{ij}$ to 0 or +1 other than $J_{ij}$ itself with equal probability (see Materials and Methods for details on the evolution procedure).

Defining homozygote and heterozygote groups

The numerical procedure for generating the homozygote and the heterozygote groups was as follows: We prepared the genome population at a particular generation in the evolutionary simulation. First, we extracted 2K genomes $J_{ij}^{(k)}$ (k = 1...2K) from K diploid individuals in the evolutionary simulation. We then made a complete copy for each 2K genome, creating a population of 2K complete homozygotes (e.g. $J_{ij}^{(k)}$), which is defined as the homozygote group (the dynamics of the complete homozygote are identical to those of the $J_{ij}^{(k)}$ haploid). Next, two different genomes, $J_{ij}^{(1)}$ and $J_{ij}^{(2)}$ (≠ $J_{ij}^{(1)}$), were randomly chosen from 2K genomes $J_{ij}^{(k)}$ (k = 1...2K) as heterozygotes, and such 2K heterozygotes constituted the heterozygote group. Note that this procedure is not entirely equivalent to the evolution of the homozygote group or the heterozygote group.

This procedure corresponds to an extreme case of heterosis for long inbreeding generations over many inbreeding populations. One might also expect that such generation of a pure homozygote group would be rare in a natural population with a low mutation rate and would be accompanied by inbreeding depression. Therefore, we also examined the population separation procedure by separating the evolved population into two groups, and conducted the same evolutionary simulation for each group separately to compare our results with the usual experiments of heterosis or inbreeding depression.

We first measured the fitness of the homozygote and heterozygote groups, where the mean values for the groups are given by $W_{\text{homo}}$ and $W_{\text{hetero}}$, respectively. Next, phenotypic variance due to noise ($\text{Var}_{\text{noise}}$) is defined as the magnitude of the variance of $x_i(T)$ for individuals with the same genotype but with different levels of noise added to the gene expression dynamics (see the Material and Methods for the definition). To eliminate the effect of the mean phenotype, the coefficient of variation ($CV_{\text{noise}}$), which is the standard deviation ($\sqrt{\text{Var}_{\text{noise}}}$) divided by the mean value ($\bar{x}_i$), was compared. $CV_{\text{noise}}$ is computed as the average of the CV over genes $i = 1..N$ for the homozygote and heterozygote groups, giving $CV_{\text{noise}}^{\text{homo}}$ and $CV_{\text{noise}}^{\text{hetero}}$, respectively.

Results

We conducted the simulations 30 times with different seeds for random numbers for each parameter set. Each of the 30 runs of the simulation is referred to as a "realization" herein.

Fitness-related heterosis

First, we discuss the case of heterosis related to fitness. The change in the degree of fitness-related heterosis $W_{\text{hetero}}/W_{\text{homo}}$ during evolution is shown in Fig. 1(A). The ratio of $W_{\text{hetero}}/W_{\text{homo}}$ increased for earlier generations and then decreased. After evolution (at the 10,000th generation), inequality $W_{\text{hetero}}/W_{\text{homo}} > 1$ was maintained; therefore, fitness-related heterosis was broadly
observed. (See Figure S1 (Supplementary Material) for the evolution of the distributions of $W^{\text{homo}}$ and $W^{\text{hetero}}$.) In Fig. 1(B), we plotted $W^{\text{hetero}}$ versus $W^{\text{homo}}$ for each realization after evolution. $W^{\text{hetero}} > W^{\text{homo}}$ was satisfied, indicating the evolution of heterosis. Then, across each realization, we computed the average ratio of $W^{\text{hetero}}/W^{\text{homo}}$ and examined its dependence on the mutation rate. As shown in Fig. 1(C), $W^{\text{hetero}}/W^{\text{homo}}$ increased with $\mu$, indicating that a greater degree of heterosis evolved at a higher mutation rate. Additionally, at higher $\mu$, the ratio of heterosis $W^{\text{hetero}}/W^{\text{homo}}$ was larger at a larger $\sigma$.

### Variance-related heterosis

The evolution of the mean value of the degree of variance-related heterosis $CV^{\text{hetero}}/CV^{\text{homo}}$ is shown in Fig. 1(D) (see Figure S2 (Supplementary Material) for the evolution of the distributions of $CV^{\text{homo}}$ and $CV^{\text{hetero}}$). Note that we regard the lower $CV^{\text{noise}}$ value as an advantage as it indicates higher robustness to noise, in this case, noise is reduced by the gene expression dynamics to achieve a stable target pattern. Therefore, the lower $CV^{\text{hetero}}/CV^{\text{homo}}$ value shows larger heterosis. $CV^{\text{hetero}}/CV^{\text{homo}}$ increased from 0.5 initially and then decreased to approximately 0.4 after evolution (at the 10,000th generation). For a random network at the first generation, the estimated ratio is 0.5 (see Supplementary Material Appendix). Because such a random network would be rare in an evolved population, a ratio less than one indicates nontrivial variance-related heterosis. The variance-related heterosis in the random network is generated by the existence of $x_i = 0.5$ as a fixed point in the heterozygotes; however, the variance-related heterosis in the evolved population is not derived from that fixed point because heterozygotes achieved the $x_i = 0$ or $x_i = 1$ state after evolution and had a more robust expression pattern than that of the homozygotes (see Figure S3, Supplementary Material). We also computed the frequency of the expression levels in the homozygote and heterozygote groups. As shown in Figure S3 (Supplementary Material), the frequency in the heterozygote group is almost zero or one, whereas the frequency of taking on intermediate values is higher in the homozygote group. This difference leads to variance-related heterosis and the higher robustness in the heterozygote group.

As shown in Fig. 1(E), we measured $CV^{\text{homo}}/CV^{\text{noise}}$ and $CV^{\text{hetero}}/CV^{\text{noise}}$ after evolution for different realizations. In all realizations, $CV^{\text{homo}}/CV^{\text{noise}}$ was maintained, which indicates that the noise-induced variance of the heterozygotes was smaller than that of the homozygotes. The dependence of $CV^{\text{hetero}}/CV^{\text{homo}}$ on $\mu$ is shown in Fig. 1(F), demonstrating $CV^{\text{hetero}}/CV^{\text{homo}} < 1$ for all $\mu$ and the ratio increased with $\mu$. For $\mu > 1.0 \times 10^{-4}$, $CV^{\text{hetero}}/CV^{\text{homo}} < 0.5$ was obtained after evolution, showing higher variance-related heterosis than obtained for the case of random networks. At higher $\mu$, the ratio of heterosis $CV^{\text{hetero}}/CV^{\text{homo}}$ was higher at a larger $\sigma$.

### Comparison with previous measurement data

By measuring the variance of traits from data of 172 traits drawn from 15 different species, Phelan and Austad (41) reported that hybrid populations exhibited fewer phenotypic fluctuations than
inbred populations. These data were explained by the authors (41) as follows: "Data covering a broad range of taxa, including invertebrate and vertebrate animals as well as plants, demonstrate that the assumption that genetic variability and environmental variability are independent of each other is unwarranted." The values listed are the phenotypic standard deviation of the inbred population and the cross-bred population divided by the mean (CV).

We plotted the distribution of $CV_{\text{homo}}^{\text{noise}}/CV_{\text{hetero}}^{\text{noise}}$ for a low mutation rate in our simulation (Fig. 2). In this case, $CV_{\text{noise}}$ and the total phenotypic CV were positively correlated, as also shown in previous studies (50, 53–56). Note also that the ratio $CV_{\text{homo}}^{\text{noise}}/CV_{\text{hetero}}^{\text{noise}}$ has been inverted from that used in the previous section to more clearly emphasize the comparison for the region $CV_{\text{homo}}^{\text{noise}} > CV_{\text{hetero}}^{\text{noise}}$.

As shown in Fig. 2, $CV_{\text{homo}}^{\text{noise}}/CV_{\text{hetero}}^{\text{noise}}$ was similar to the previous experimental data. These results imply that the variance-related heterosis obtained by the diploid-GRN model would be consistent with field data from actual populations.

Correlation of heterosis with pattern dominance and differential gene expression

In this subsection, we discuss how heterosis is correlated with dominance and differential gene expression, as addressed in (8, 22, 24). Here, we need to extend the definition of dominance to study the collective gene expression in the GRN rather than considering the expression of only a single gene. Thus, we defined and computed pattern dominance (36) (the pattern dominance also takes into consideration the expression of nontarget genes, even though it does not directly relate to the fitness. We have discussed the relevance of the expression pattern dominance to fitness in our previous paper (36). The expression pattern dominance is correlated to the robustness against noise, which is also correlated to fitness. Thus, fitness, dominance, and robustness are highly interconnected (36) and differential gene expression from the gene expression patterns and explored their correlation with heterosis.

To introduce the dominance of gene expression patterns, we defined and computed pattern dominance according to the following procedure. We first computed the gene expression patterns over a group of genes for given homozygotes AA and BB taken from an evolved population, which is represented by $x_{AA}^{0}$ and $x_{BB}^{0}$. Next, we computed the gene expression pattern of heterozygotes AB, given by $x_{AB}^{0}$. For genes satisfying $x_{AA}^{0} \neq x_{BB}^{0}$ (as a numerical contribution, $|x_{AA}^{0} - x_{BB}^{0}| > 0.1$ is adopted in the evolved population, almost all genes have $x_{i} = 0$ or $x_{i} = 1$ expression for maintaining fitness as a network), we examined whether the expression pattern of the heterozygotes is biased toward one of the homozygote expression patterns such as $x_{AB}^{0} = x_{AA}^{0}$ (as a numerical contribution, $|x_{AB}^{0} - x_{AA}^{0}| < 0.1$ is adopted) or $x_{AB}^{0} = x_{BB}^{0}$ (as a numerical contribution, $|x_{AB}^{0} - x_{BB}^{0}| < 0.1$ is adopted), and computed the ratio of the number of genes, $x_{AB}^{0} \geq x_{AA}^{0}$ or $x_{AB}^{0} \geq x_{BB}^{0}$. Pattern dominance is defined by the larger of the two ratios. The value of each gene is an independent one, such bias is not expected, and the pattern dominance is 0.5. The pattern dominance is 1.0 if $x_{i}^{AA} \simeq x_{i}^{0}$ for all $i$ with $x_{i}^{AA} \neq x_{i}^{BB}$.

The ratio of differential gene expression was also computed to investigate the correlation between heterosis and differential gene expression, as discussed in earlier studies (19–22). For a set of genes with $x_{AA}^{0} \simeq x_{i}^{0}$, the ratio of the number of heterozygotes, satisfying $x_{i}^{AB} \neq x_{i}^{AA}$ (as $x_{i}^{0}$), is computed, which gives the degree of differential gene expression (see the Methods section for a detailed description of pattern dominance and differential gene expression calculations).

Here, we remark on the definition of pattern dominance. (Mendelian) dominance is generally defined regardless of fitness. Trait “A” is dominant when AA shows trait “A” even though its fitness is lower than that of “a.” Following this, the definition of pattern dominance adopted herein is also independent of fitness. The expression of nontarget genes (in the present model, the fitness is not assigned for nontarget genes, and they are (quasi)neutral as for $x_{i}(0) = 0$ or $x_{i}(0) = 1$. Therefore, whether or not the gene is expressed directly concerns the fitness in our model, the major aspect of the pattern dominance and differential gene expression pattern is not directly related to fitness. This is in contrast to the “dominance model” for heterosis, which assumes deleterious and recessive mutation. Therefore, our model’s higher pattern dominance value does not directly support the dominance model. The definition of overdominance as well as underdominance is also based on fitness. Hence, the differential gene expression in our model does not necessarily support the overdominance model either.

The correlation between pattern dominance, differential gene expression, and fitness-related heterosis is shown in Fig. 3 (see Figure S4 (Supplementary Material) for the correlation with variance-
related heterosis). Heterosis $W_{hetero}$/$W_{homo}$ exhibited a negative correlation with pattern dominance but a positive correlation with differential gene expression. From these results, we could explain the correlation between differential gene expression and heterosis, as is also discussed in (19–22). In contrast, the dominance of expression pattern was not supported (note that since we did not define deleterious mutation, this does not entirely exclude the possibility of the dominance model). Contrastingly, variance-related heterosis was positively correlated with the pattern dominance but negatively correlated with differential gene expression; however, the absolute values of correlation coefficients were smaller than that for fitness-related heterosis (see Figure S4, Supplementary Material).

The frequencies of the expression levels in the homozygote and heterozygote groups are shown in Figure S3 (Supplementary Material). In the homozygote group, the opposite expression sometimes appears. Therefore, this simulation is reasonable for exploring pattern dominance or differential gene expression.

### Explanation of the results according to the convex fitness region and single-humped fitness function

This section proposes an intuitive explanation of our results with visualization of the fitness landscape. At present, this remains a hand-waiving argument, and further mathematical elucidation and experimental verification are required in the future for validation.

We consider a fitness landscape over a genotype space (57). In particular, we focus on the high-fitness region. After evolution, offspring are produced from the two parents by meiosis within the genome matrices $J$.

$J_{ij} = J_{0} + J_{1} \times s_{ij}$

where $J_{0}$ is the genome matrix of the two parents and $s_{ij}$ is the heterozygosity. The heterozygous genome is not precisely represented by $(J_{0} + J_{1})/2$; however, we use this approximate description for simplicity. The offspring must have high fitness; otherwise, the offspring will not be selected. For example, when the higher-fitness region around the genome population is not convex (Fig. 4A), the offspring may not be selected because of their lower fitness compared with that of the parents. Therefore, the high-fitness region of the genome population shapes the convex region in the genotype space under sexual reproduction and selection (Fig. 4B).

Consequently, the fitness is expected to take on a single-humped structure around the center of the region and will decrease as the genotype moves toward the periphery of the high-fitness region (Fig. 4B). If the fit state is robust, the fitness is less sensitive to genetic change around the center than the periphery. This suggests that the fitness function is flat and becomes steeper when moving farther from the center (i.e. convex, single-humped functions). Thus, the fitness (phenotype) variance will be smaller than on the periphery around the center.

Now, consider the distribution of the homozygote group in a high-fitness region, whose genotype can be broadly located. This population contains individuals with low fitness and large phenotypic variance. Accordingly, a heterozygous group generated from a pair from the homozygote group is biased toward the center of the region. Therefore, the distribution of the heterozygote group is expected to have higher fitness and lower variance than that of the homozygote group, resulting in fitness- and variance-related heterosis. This picture is consistent with the observed correlation between heterosis and differential gene expression.

Thus, we expected the fitness landscape to be single-humped with a flatter top (i.e. convex) around the evolved genome population and examined the validity of this expectation in our model. First, we computed the distance between the two genome matrices $\sqrt{\sum_{i,j} (J_{ij}^{K1} - J_{ij}^{K0})^2}$ for the two genomes (k, l) in the population of the 999th generation. Among the combinations (k, l), we chose combinations of homozygotes (K, L) with the largest distance between genome matrices $\sqrt{\sum_{i,j} (J_{ij}^{K1} - J_{ij}^{K0})^2}$. Next, we produced 100 matrices corresponding to the internal division of $J_{ij}^{K0}$ and $J_{ij}^{K1}$ (i.e. $s_{ij}^{K0} + (1 - s_{ij}^{K1})$) by changing $s$ in a step-by-step manner in the range $0 \leq s \leq 1$. We then computed the mean fitness for homozygotes $s_{ij}^{K0} + (1 - s_{ij}^{K1})$ over 100 matrices of each division point given by $s$.

As shown in Fig. 5, the fitness function of homozygotes exhibited a single-humped structure against the internal division parameter $s$ and had a flatter landscape around the top, as predicted by our proposed explanation. As this type of fitness function is not observed in random matrices $J$ (corresponding to the 0th generation), we suggest that it was acquired by evolution with sexual reproduction and selection. We did not assume this single-humped fitness function, but rather robustness in the fitness function under sexual reproduction might have caused the result.

![Fig. 4. Schematic of the fitness landscape and high-fitness region.](https://academic.oup.com/pnasnexus/article-1/3/pgac097/6620857/6)

The above estimate of the fitness landscape is based on the homozygous diploid $(s_{ij}^{K0} + (1 - s_{ij}^{K1})$ and $s_{ij}^{K0} + (1 - s_{ij}^{K1})$; i.e. haploid). To further analyze the fitness by heterozygotes, we introduced the heterozygote $(s_{ij}^{K0} + (1 - s_{ij}^{K1}) + (1 - s_{ij}^{K1})$ with heterozygosity $d$ for each heterozygote with given $s$. These heterozygotes are only placed between $\frac{d}{2}$ and $1 - \frac{d}{2}$ because we used two genomes in the internal division from 0 to 1 for generating the heterozygotes. As shown in Fig. 5, the fitness with heterozygosity $d > 0$ was higher than that of the homozygotes ($d = 0$) with $s$; that is, the larger heterozygosity $d$ implies a further increase of fitness. The fitness of such heterozygotes also exhibited a single-humped, convex function.

In Fig. 5, one can see two types of heterosis as in genome distribution and heterozygosity. First, even in the homozygote series ($d = 0$), the homozygote (haploid) at a middle internal division genome can achieve higher fitness than those at the edges. For example, the homozygote at the internal division point between $s = 0.2 (w \approx 0.75)$ and $s = 0.8 (w \approx 0.65)$ shows higher fitness ($s = \ldots$)
Black curves are homozygotes. Here, we introduced the heterozygote heterozygosity, as confirmed by including the above-mentioned heterozygotes, whereas a single-humped fitness function was also obtained (see Fig. 5). The heterozygote advantage was further that the distance between genome matrices was the maximum value. The values at \( x = 0, 1 \) can be much lower because the homozygotes at the edge of the region accumulate too many mutations. The horizontal axis represents the internal division parameter \( s \). The average fitness over 300 realizations is plotted for \( \mu = 1.0 \times 10^{-4} \) and \( \sigma = 1.0 \times 10^{-4} \). Black curves are homozygotes. Here, we introduced the heterozygote \((s - \frac{1}{2})_1^{+} + (\frac{1}{2} - s)_2^{+}\) with heterozygosity \( d \). Pink, orange, and blue curves correspond to \( d = 0.31, d = 0.62, \) and \( d = 0.93 \) heterozygotes, respectively. The data for large \( d \) are plotted for a narrower range in \( s \), because we used two genomes in the internal division from 0 to 1 for generating the heterozygotes.

0.5, \( w = 0.85 \) than two. This implies that the event corresponding to heterosis can occur even in a complete homozygote or haploid population if the single-humped fitness function is achieved.

Second, interestingly, the heterozygosity further raises the fitness, as confirmed by including the above-mentioned heterozygosity parameter. For the heterozygosity \( d = 0.31 \), almost all heterozygotes showed higher fitness values than those of the homozygotes, whereas a single-humped fitness function was also obtained (see Fig. 5). The heterozygote advantage was further increased along with the increase in heterozygosity, as shown in the plots in Fig. 5 for \( d = 0.62, 0.93 \). These results imply the existence of a heterozygote advantage beyond the effect of the removal of deleterious mutations. At the top of the single-humped fitness function of homozygotes, the offspring of the highest homozygotes (i.e. the "best-parents") can achieve higher fitness than the original homozygotes, as is also discussed in the overdominance model. Here, generally, even if the population’s average fitness stays constant, there are still variations in genomes, and some heterozygotes with lower fitness may exist, which are placed in the periphery of the convex region. If we make them homozygotes, the fitness will be lower than that of the heterozygote. Here, we chose a pair from the opposite side of the periphery to demonstrate the relevance of the convexity to robustness. Thus, the fitness of the heterozygote from the opposite side of the periphery of fitness is higher than that of the homozygote, which provides one cause for heterosis.

To confirm this scenario, we took the individuals with the worst 5% fitness within the evolved population and generated their offspring either by sexual reproduction without mutation or by mutation only (without sexual reproduction). On the average, the sexual reproduction only showed the average fitness at 0.66, much higher than that by the mutation only, which was 0.35, and the average fitness of the original worst 5% population, which was 0.26. The distributions of fitness of those three cases are shown in Figure SF5 (Supplementary Material). When we make the offspring through sexual reproduction from all individuals in the steady-state population, whose average fitness was 0.844, whereas the average fitness of sexually reproduced individuals, 0.843, was slightly lower than that but distinctively higher than that of asexually reproduced ones, 0.836. These results are consistent with the picture in Fig. 4.

Comparison with the case without sexual reproduction (selfing)

Comparison with selfing is also relevant to conclude that the observed heterosis results from sexual reproduction. Here, we examined the case with a single parent, without sexual reproduction, which can be regarded as corresponding to selfing. A total of two genetic modes, with and without recombination, were numerically investigated. (See the Material and Methods for the definition of selfing genetic modes.)

The evolvability by both selfing populations was slower than that of the case with sexual reproduction involving meiosis (Fig. 6A). This is consistent with earlier studies (36, 58). Genetic variation decreased throughout evolution in all three genetic modes, as shown in Fig. 6(B); however, the result with meiosis showed higher genetic variation than that of the other two cases. For the selfing case without sexual reproduction, \( \text{CV}_{\text{hetero}}/\text{CV}_{\text{homo}} \) was lower and \( \text{CV}_{\text{hetero}}/\text{CV}_{\text{noise}} \) was higher than those of the other cases, as shown in Figs 6(C) and (D). Hence, sexual reproduction with meiosis increases the degree of heterosis compared with selfing, stronger heterosis, or inbreeding depression.

For reference, we also conducted a heterosis test in the neutral evolution scenario to investigate the selection effect on our heterosis result, which had the fitness function of \( w = 1 \) for any genotypes and phenotypes. The result of the neutral simulation in Figure S6 (Supplementary Material) showed that the variance-related heterosis was still achieved, whereas the ratio \( \text{CV}_{\text{hetero}}/\text{CV}_{\text{homo}} \) was larger than that obtained in the case under selection. Comparing Fig. 1(D) with Figure S6 (Supplementary Material), the first increase of \( \text{CV}_{\text{hetero}}/\text{CV}_{\text{noise}} \) would be a result of genetic drift, while the subsequent decrease would be due to selection.

Case of population separation

The numerical procedure we adopted to compare the homozygote and heterozygote groups was based on hybridization among inbreeding individuals. Such a procedure can be achieved in some laboratory experiments by using yeast (59–61). In contrast, most experimental studies on heterosis or inbreeding depression adopt population separation and inbreeding among divided populations. Hence, we also conducted the simulation using such population separation in order to confirm the validity of the results obtained so far. To be specific, we conducted the following simulation to evolve the population. First, we divided the population of 100 individuals into two populations of 50 each. Second, we conducted the same evolutionary simulation with the same fitness conditions over 100 generations, under the constraint that the 50 individuals in each divided population reproduce, thereby keeping the same number of individuals in the divided populations. Third, hybrids between the two separated populations were generated to compare with each inbreeding population. The fitness and noise-induced \( \text{CV} \) were then compared between the inbred and hybrid populations (\( \text{CV}_{\text{hetero}}, \text{CV}_{\text{hybrid}}, \text{CV}_{\text{noise}} \)).
Fig. 6. Evolution of heterosis, fitness, and genetic variation. In all graphs, blue points result from evolution with meiosis in the genetic algorithm, black points are selfing without recombination, and orange points indicate selfing with recombination. The mutation rate per edge was $\mu = 1.0 \times 10^{-4}$ and the noise strength was $\sigma = 1.0 \times 10^{-4}$. The points represent the average of 30 realizations, and error bars range from the first to the third quantiles. All horizontal axes are shown on the logarithmic scale. (A) Fitness increases throughout evolution. $W_{\text{population}}$ is the average fitness of the evolving population. The meiosis population evolved faster than the other two selfing populations, but the fitness of all three populations reached values larger than 0.9. (B) Change in the genetic variation in the population throughout evolution. Genetic variation is defined as the sum of the variance in specific $J_{ji}$ elements in the population. The meiosis population showed higher genetic variation than the other two selfing populations. (C) Change in the degree of fitness-related heterosis $W_{\text{hetero}}/W_{\text{homo}}$ throughout evolution. All three genetic modes showed $W_{\text{hetero}}/W_{\text{homo}} > 1$, but that of meiosis was larger than that with selfing. (D) Change in the degree of fitness-related heterosis $CV_{\text{hetero}}/CV_{\text{homo}}$ throughout evolution. All three genetic modes showed $CV_{\text{hetero}}/CV_{\text{homo}} < 1$ values after the increase of fitness, but those with meiosis were smaller than those with selfing.

Fig. 7. Evolution of heterosis in the population separation simulation. (Parameters and conditions were that same as those in Fig. 1(A) and (D).) (A) Change in the degree of fitness-related heterosis $W_{\text{hybrid}}/W_{\text{inbred}}$ throughout evolution. $W_{\text{hybrid}}/W_{\text{inbred}}$ increased initially and then decreased thereafter. Relaxed $W_{\text{hybrid}}/W_{\text{inbred}}$ value was near one but $W_{\text{hybrid}}/W_{\text{inbred}} > 1$ was maintained. (B) Evolution of the median value of the degree of variance-related heterosis $CV_{\text{hybrid}}/CV_{\text{inbred}}$. The $CV_{\text{hybrid}}/CV_{\text{inbred}}$ increased initially and then decreased through evolution, reaching a median value of approximately 0.5. Figure 7 shows the evolution of heterosis in the population separation simulation. The degree of fitness-related heterosis $W_{\text{hybrid}}/W_{\text{inbred}}$ after evolution satisfied $W_{\text{hybrid}}/W_{\text{inbred}} > 1$, corresponding to fitness-related heterosis. One might still argue that $W_{\text{hybrid}}/W_{\text{inbred}}$ was lower than shown in Fig. 7(A). This would be due to the long-term selection after population separation, which increased the fitness of the inbred population. In fact, $W_{\text{hybrid}}/W_{\text{inbred}}$ was much larger (~1.1) before it reached the stationary value. Moreover, the variance-related heterosis after evolution was also satisfied, as $CV_{\text{hybrid}}/CV_{\text{inbred}} \approx 0.5 < 1$, even though the value was slightly larger than that obtained by the previous procedure, ≈0.4. To sum, the heterosis was also confirmed in the population separation procedure.

Discussion

In the present study, we employed numerical evolution of the diploid-GRN to demonstrate heterosis related to fitness and phenotypic CV and its correlation with differential gene expression, and further compared the simulation results with previously re-
ported measurement data. First, heterozygote groups exhibited higher fitness than homozygote groups, indicating heterosis (or inbreeding depression). Phenotypic CV due to noise was also lower in the heterozygote group than in the homozygote group, indicating variance-related heterosis. The numerically obtained distribution of the ratio of the CV of the homozygote group to that of the heterozygote group showed similarities with that of the previous measurement data (41). Heterosis exhibited a negative correlation with expression pattern dominance and a positive correlation with differential gene expression. We then proposed a possible explanation for these results according to the fitness landscape, which exhibits a convex shape if high fitness is maintained under sexual reproduction and selection. We also conducted selfing and population separation simulations to confirm the validity of our fitness and variance-related heterosis results.

Here, we remark on variance-related heterosis. The traditional quantitative genetics theory explained heterosis with respect to the total phenotypic variance based on the correlation variance; however, there was a discrepancy between the theory and the observed distribution in (41). Furthermore, we compared noise-induced phenotypic variance V_{noise}, which corresponds to the inverse of noise robustness. The heterosis concerning noise-induced variance has scarcely been studied, even though such noise robustness has recently attracted substantial attention in the context of research on homeostasis and cell state selection (26, 43–45, 62); thus, heterosis should also be reexamined in light of noise robustness. Recall that even a random network or selfing can produce considerable variance-related heterosis to some degree; however, sexual reproduction shows quantitative deviation from these scenarios.

The ratio of CV_{fitness}, which contains the genetic variation, showed similarity to the experimental data by Phelan and Austed. Here, however, CV_{noise} occupied the major part of the total phenotypic CV. These results imply that noise robustness and genetic variation would contribute to the variance-related heterosis. Nevertheless, further experiments for given species under fixed environmental conditions will be needed to make a further quantitative comparison, together with ploidy manipulation and direct measurement of the noise. For example, such experiments have been performed in yeast (59–61), in which the GRN has already been analyzed (63–65) in the heterosis context.

We also found a positive correlation between differential gene expression and heterosis. This correlation has also been discussed previously (19–22), whereas the molecular mechanism contributing to heterosis remains unclear (66). According to the present study, such correlation arises from phenotypic robustness due to interactions among genes through the GRN.

We proposed the convex fitness landscape and single-humped fitness function shaped by evolution to explain our results on heterosis. Under evolution with sexual reproduction, it is necessary to create a population of genomes in which the fitness of the offspring is not reduced from those of the parents. A population with a convex fitness landscape and single-humped fitness function will satisfy such a postulate. This viewpoint was also discussed in the context of basic quantitative genetics, in which the fitness landscape or function for a given genome distribution is provided as such. In contrast, we did not explicitly define such a fitness landscape. The fitness landscape does not exist in the first generation but is rather shaped by evolution under sexual reproduction. Furthermore, for the case of selfing, the observed heterosis was much weaker.

The single-humped fitness function leads to an increase in robustness against noise. This explanation based on the fitness landscape and adaptive walk may be similar to the argument proposed on the heterozygote advantages by Fisher’s geometric model (67). However, in the present study, the evolution of robustness to noise in gene expression dynamics shaped such a convex fitness landscape.

The convex fitness landscape and single-humped fitness function may suggest a possible relationship between the dominance and overdominance models. At the periphery of the single-humped fitness function, the dominance model to eliminate the deleterious mutation effect would be valid for explaining heterosis. Around the center of the function, the overdominance model would account for the heterosis result. In the present study, however, further heterozygote superiority arose from the interaction between different GRNs.

We confirmed our result by simulating the divided populations, as adopted in standard heterosis measurements. In the future, we need to examine the dependence on the division ratio or the length of inbreeding to compare with the analysis based on a population genetics approach. In the discussion of heterosis based on gene regulation, the importance of large-scale rewiring in GRNs has been highlighted (17). In fact, in our simulation, heterosis was strengthened under a high mutation rate during evolution. Heterosis, at least to some degree, can be understood by phenotypic robustness under the presence of large-scale rewiring in the GRN by sexual recombination.

To conclude, we elucidated the evolution of heterosis in a system of interacting genes by noting the relevance of robustness to noise as well as sexual recombination (68) with meiosis of genomes (69). This evolution of heterosis also implies the advantage of sexual reproduction, which goes beyond the standard concept of Muller’s ratchet (70) due to the removal of deleterious mutants. Alternatively, the explanation by differential gene expression presented in this study suggests that heterozygosity in diploid organisms goes beyond the range of homozygote phenotypes due to the interactions between the two genomes, which creates a convex fitness landscape through evolution. As the present evolution model does not need to assume inbreeding as adopted in standard heterosis studies, this study will shed novel light on the evolution of heterosis that is applicable to field and laboratory experiments.

Materials and Methods

Derivation of diploid-GRN model

Let us define y_i^m(t) as the concentration of mRNA i from each chromosome (m = 1, 2) and x_i(t) as the concentration of the corresponding protein in a cell. Note that the proteins are synthesized from both chromosomes, where mutations or recombination are mainly introduced in the promoter or enhancer region; thus, the superscript m in x_i(t) is not required. By replicating the regulation in a diploid cell by f_{ij}^m (m = 1, 2) for each gene, we obtain

\[ y_i^m(t + 1) = \frac{1}{2} f\left(\sum_{j=1}^{N} f_{ij}^m(x_j(t) - \theta)\right) \quad (m = 1, 2), \]

where \( \theta = 0.5 \) is the threshold of expression, which makes the states \( x = 0 \) and \( x = 1 \) symmetric for simplicity. As protein i is synthesized from the corresponding mRNA, the total expression level is

\[ x_i(t + 1) = \sum_{m=1}^{2} y_i^m(t + 1). \]
Therefore,

\[ x_i(t + 1) = \frac{1}{2} \sum_{m=1}^{2} \left[ \sum_{j=1}^{N} J_{ij}^m (x_j(t) - \theta) \right]. \]

Details of the evolutionary simulation

Fitness is computed as follows. After a sufficient time, T, the expression level \( x_i(t) \) reaches a stationary state. Here, we define the phenotype as the time average of \( x_i(t) \) for the last 10 steps, \( \bar{x}_i = \frac{1}{100} \sum_{t=9}^{99} x_i(t) \). The fitness is defined by this average \( \bar{x}_i \) for the target genes \( i = 1, \ldots, N \). Positive interactions between target genes can achieve this target pattern. However, there are nontarget genes in the network, which may affect the expression of target genes (the path ratio of the GRN was set to 0.5). By setting the maximum fitness for \( \bar{x}_i - 1 = 0 \) (if \( i = 1, \ldots, N \)), the absolute fitness is defined as \( \exp[-\sum_{i=1}^{N} |\bar{x}_i - 1|] = \exp[\sum_{i=1}^{N} \bar{x}_i] \exp[-SM] \). Then, the relative fitness \( w \) is given by \( w = \exp[\sum_{i=1}^{N} \bar{x}_i] \). The probability that an individual having \( \bar{x}_i \) is selected is proportional to \( w \). In some cases, the expression pattern \( x_i(t) \) oscillates over time and cannot reach a stable state; however, this is rare. Moreover, after evolution, such a case is not observed because a stable expression pattern \( \bar{x}_i = 1 \) (if \( i = 1, \ldots, M \)) can be more advantageous than an oscillatory state under this fitness definition.

Note that the target pattern for fitness is assigned only for some genes \( i = 1, \ldots, M \), so that there is no selection pressure for the expression level of most genes \( i = M + 1, \ldots, N \). In this simulation, \( N = 100 \) and \( M = 10 \). Furthermore, the threshold \( \theta = 0.5 \) so that the states \( x_i(T) = 0 \) and \( x_i(T) = 1 \) are symmetric, which are neutral in the nontarget genes (if \( M = 1, \ldots, N \)). The nontarget genes are neutral as is the case for fitness. However, for each network, the nontarget genes can affect the expression levels of target genes through regulatory interactions, which are also changed by mutation or recombination.

We used the following meiosis-like genetic algorithm for the evolutionary simulation. Two distinct individuals \( k_1 \) and \( k_2 \) are selected as parents with probabilities of \( w_{k_1}/W \) and \( w_{k_2}/W \), from which the next generation is produced according to the above fitness definition (if \( k_1 \neq k_2 \)), where \( W = \sum_{i=1}^{N} w_i \) is the sum of all fitness values in the population. To introduce meiosis, we adopted the following procedure. A single gamete is generated by mixing \( \left[ \begin{array}{c} \bar{x}_i^{(1)} \\ \bar{x}_i^{(2)} \end{array} \right] \) and \( \left[ \begin{array}{c} \bar{x}_i^{(2)} \\ \bar{x}_i^{(1)} \end{array} \right] \) via row vectors with equal probabilities to produce a new genome \( n \), where \( \left[ \begin{array}{c} \bar{x}_i^{(1)} \\ \bar{x}_i^{(2)} \end{array} \right] \) and \( \left[ \begin{array}{c} \bar{x}_i^{(2)} \\ \bar{x}_i^{(1)} \end{array} \right] \) are also used as row vectors corresponding to selfing genetic modes for reference. In these genetic modes, there is one parent that is chosen by the probability \( w_{k_1}/W \). In the first genetic mode, selfing without recombination, \( \left[ \begin{array}{c} \bar{x}_i^{(1)} \\ \bar{x}_i^{(2)} \end{array} \right] \) or \( \left[ \begin{array}{c} \bar{x}_i^{(2)} \\ \bar{x}_i^{(1)} \end{array} \right] \) is copied as the one of the new diploid genome \( n \). A total of two new genomes can be the copies of the same parental genome. In the second genetic mode, selfing with recombination, \( \left[ \begin{array}{c} \bar{x}_i^{(1)} \\ \bar{x}_i^{(2)} \end{array} \right] \) and \( \left[ \begin{array}{c} \bar{x}_i^{(2)} \\ \bar{x}_i^{(1)} \end{array} \right] \) are mixed via row vectors with equal probabilities to produce a new genome \( n \). We have adopted this setting of recombination between the two genomes in one parent to clarify the role of genome mixing between two parents. This setting corresponds to selfing.

The simulation was then performed by changing the mutation rate (\( \mu \)) in the range \([10^{-5}, 10^{-2}] \) and the standard deviation of the noise (\( \sigma \)) in the range \([5 \times 10^{-4}, 10^{-1}] \). The number of genes (\( N \)) was set to 100. The ratio of nonzero elements in \( j_{ij}^{(m)} \) was set to 0.5 so that some \( j_{ij}^{(m)} \) elements changed to zero after mutation or recombination to maintain this ratio. The selection pressure, \( S \), was set to 2 and the number of target genes (\( M \)) was 10. The relaxation time (\( T \)) to examine fitness was 30. The total number of individuals (\( K \)) included in the study was 100. Out of the total number of genes, 90 are not initially expressed.

Definition of \( CV_{\text{noise}} \)

\( CV_{\text{noise}} \) is the phenotypic CV generated by noise over isogenic individuals with a particular genotype. It is computed by running simulations of gene expression dynamics with stochastic noise multiple times for the individual under the same genotype. For a given diplod individual \( k \) with \( j_{ij}^{(k)} \) and \( \bar{x}_i \), after relaxation is computed under a given noise magnitude \( \sigma \neq 0 \). After repeating this process \( L \) times and obtaining \( \bar{x}_i \) for each run, we measured the variance of the expression pattern \( \bar{x}_i \). The CV \( CV_{\text{noise}} \) for an individual \( k \) was computed using the average of the CV over all genes (\( N = 100 \)) and realizations (\( L = 100 \)). \( CV_{\text{noise}} \) was computed by the average of \( CV_{\text{noise}} \) over the individuals for a given genome group.

Pattern dominance: definition and calculation

First, we extracted 2K genomes \( j_{ij}^{k} \) (\( k = 1 \ldots 2K \)) from K diplod individuals in the evolutionary simulation. We then made a complete copy for each 2K genome, creating a population of 2K complete homozygotes (e.g., \( j_{ij}^{k}j_{ij}^{k} \)). Then, we computed the expression dynamics in each of the 2K complete homozygotes to obtain the expression pattern \( x_{AA} \). Next, from the 2K genomes, two \( j_{ij}^{k} \) and \( j_{ij}^{k} \) are randomly selected to create a diplod cell consisting of two heterozygous genomes. From the expression dynamics of the heterozygous genomes, the expression pattern of heterozygous \( x_{AB} \) is obtained.

We could then compare the expression of genes that are differentially expressed between two such homozygotes: those with \( x_{AA} \neq x_{AB} \). By taking into account that \( x_i \) is real, the (in)equality \( x_i \sim y \) (or \( x_i \neq y \)) is defined if \( |x_i - y| < 0.1 \) or \( |x_i - y| > 0.1 \), as a numerical procedure. For most cases, \( x_i \) takes the value of \( 0 \) or \( 1 \); however, the definition is basically the same whether the genes are expressed (\( \sim 0 \)) or not (\( \sim 0 \)). The number of \( i \) genes with \( x_{AA} \neq x_{AB} \) is defined as \( N_{\text{AAif}} \). Next, the number of \( i \) genes with \( x_{AA} \neq x_{AB} \) is defined as \( N_{\text{ABif}} \). Then, the degree of dominance is given by \( \max(N_{\text{AAif}}, N_{\text{ABif}}, N_{\text{AAif}}, N_{\text{ABif}})/N_{\text{AAif}} \).

In the current model, the expression level of each gene \( x \) is almost always very close to zero or one. Therefore, if only one gene is differentially expressed between the two homozygotes, the calculated dominance will always be one. To remove this bias, we calculated the pattern dominance only for homozygote \( x_{AA} \) and \( x_{AB} \) pairs in which more than 10 genes are differentially expressed. For reference, we also show the gene expression pattern of the homozygote and heterozygote groups in the Supplementary Material.

Differential gene expression: definition and calculation

In contrast to pattern dominance that focuses on the expression patterns of genes that differ between homozygotes of the two genomes, differential gene expression focuses on genes expressed in the same manner between the two genomes. Let \( N_{\text{AAif}} \) be the number of genes with the same expression (\( j_{ij}^{AA} - j_{ij}^{BB} < 0.1 \)) as the two homozygotes, and \( N_{\text{AAin}} \) be the number of genes in the heterozygote AB that are expressed differently (\( j_{ij}^{AA} - j_{ij}^{BB} > 0.1 \)) from those in the homozygotes AA and BB. \( N_{\text{AAin}}/N_{\text{AAif}} \) is then defined as the degree of differential gene expression (Fig. 8).
Fig. 8. Definition and calculation of pattern dominance and differential gene expression. (A) Pattern dominance: first, we compute the expression patterns of homozygotes AA and BB. Next, the expression pattern is computed by making the heterozygote AB. We then calculate the percentage of the expression pattern obtained from the heterozygote that matches only one of AA or BB for the differentially expressed genes among the two homozygotes, focusing on the differentially expressed genes in the two homozygotes (blue). The pattern dominance is then computed as a larger percentage of matches between $x_{i}^{AB}$ and $x_{i}^{AA}$ or between $x_{i}^{AB}$ and $x_{i}^{BB}$ (B) Differential gene expression: for a gene showing the same expression in the two homozygotes (blue), we calculate the percentage of the expression pattern obtained from the heterozygote that is different from that of AA and BB. Then, the differential gene expression is computed as the percentage of $x_{i}^{AB} \neq x_{i}^{AA} (= x_{i}^{BB})$.

Acknowledgments

The authors would like to thank Tetsuhiro Hatakeyama, Yuichi Wakamoto, Naoki Irie, Akira Sasaki, and Shuju Ishihara for their stimulating discussions; Hideki Inan and Takahiro Sakamoto for providing information about genetics; and Daniel Hartl for informing us of (41).

Supplementary Material

Supplementary material is available at PNAS Nexus online.

Funding

This research was partially supported by a Grant-in-Aid for Scientific Research (A) 431 (20HO00123) and a Grant-in-Aid for Scientific Research on Innovative Areas (17H06386) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. This research was also supported by Novo Nordisk Fonden.

Authors’ Contributions

K.O. and K.K. conceived and designed the study. K.O. conducted the simulation. K.O. and K.K. analyzed the data and wrote the manuscript. Both authors have read and approved the final version of the manuscript.

References


