Heterosis of Fitness and Phenotypic Variance in the Evolution of Diploid Gene Regulatory Network

Kenji Okubo^a and Kunihiro Kaneko^{a,b,1}

^aDepartment of Basic Science, Graduate School of Arts and Sciences, the University of Tokyo, Meguro, Tokyo; ^bUniversal Biology Institute, the University of Tokyo, Bunkyo, Tokyo

This manuscript was compiled on November 22, 2021

Heterosis describes the phenomenon whereby a hybrid population has higher fitness than an inbred population, and has previously been explained by either Mendelian dominance or overdominance, where it is generally assumed that one gene controls one trait. However, recent studies have demonstrated that genes interact through a complex gene regulatory network (GRN). Furthermore, phenotypic variance due to noise is reportedly lower for heterozygotes, whereas 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 investigate 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, that is, fitness-related heterosis resulting from evolution. In addition, the heterozygote population expressed 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 agreement with previous experimental results. By applying dominance and overdominance to the gene expression pattern rather than only a single gene expression, we confirmed the correlation between heterosis and overdominance. 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.

heterosis | evolution | phenotypic variance | gene regulatory network | genetics | sexual reproduction

Introduction

Sexual reproduction generally involves the mixing of genetic information from multiple parents. Most diploid organisms undergo meiosis, which recombines genomes between parents to produce gametes, then combines the gametes to produce the offspring's genome. One of the most remarkable phenomena in sexual reproduction among populations is heterosis, whereby hybrid populations exhibit higher fitness than inbred populations, which are obtained by repeated sexual reproduction among genetically close individuals. The origin of heterosis has been discussed from a genetics perspective (1–4), in terms of Mendelian dominance.

In Mendelian dominance (5), 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. When the gene Aa expresses the trait over 'A' or 'a', it is called overdominance. For decades, the relationship between dominance (or overdominance) and heterosis has been recognized as a

significant genetics problem; The first is the dominance hypothesis, which explains heterosis by assuming that Mendelian dominance cancels out the phenotypes that reduce fitness in heterozygotes. In contrast, the overdominance hypothesis proposes that heterozygotes can have a novel trait that exhibits higher fitness than both parental homozygotes. However, no solid conclusions have yet been reached(6, 7).

In the standard context of heterosis, the fitness or trait is compared between inbred and hybrid populations. In contrast, Phelan and Austad examined data relating to the traits of over a hundred species, and demonstrated that the variance of traits also shows heterosis, in the sense that the variance of traits for hybrid populations is smaller than that of inbred populations (8). They also discussed the potential relevance of genotype-environment interactions for this variance-related heterosis; however, its theoretical explanation remains elusive.

To discuss the variance-related heterosis, we need to consider the noise or environmental variation, and robustness of traits against such variation(9–13). However, the possible link between noise and heterosis has rarely been discussed (14), despite the fact that stochasticity in phenotypes (gene expression) has recently received substantial attention in the context of homeostasis and cell state selection (15–18).

Moreover, the evolution of gene regulatory networks has been extensively studied (19–23). To date, arguments related to heterosis are based on the simple genotype–phenotype relationship, that is, the case in which one gene determines

Significance Statement

Heterosis, that is higher fitness in hybrid populations than inbred populations, is a long-standing problem in genetics, evolution, and breeding studies. Heterosis is not necessarily the result of a trait uniquely determined by a single gene, but likely involves stochasticity and interactions among genes. Through numerical evolution of the gene regulatory network, we demonstrate that heterosis is shaped by evolution to achieve robustness against noise and genetic mixing by sexual recombination in gene expression dynamics. That is, a mixed population is more robust to noise than inbred populations, as revealed experimentally by reduced phenotypic variance. The 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.

KO and KK conceived and designed the studies. KO conducted a simulation. KO and KK analyzed the data and wrote the manuscript. All authors have read and approved the final version of the manuscript.

None of the authors have any competing interests to declare.

¹To whom correspondence should be addressed. E-mail: kaneko@complex.c.u-tokyo.ac.jp

one phenotype (5–7). However, the existence of a complex genotype–phenotype relationship caused by interactions among genes has garnered substantial attention (16). The importance of gene regulatory networks has also been recognized, in which protein expression depends on the mutual activation or suppression of genes, where a phenotype, for example, expression of a single protein, depends on expression levels of multiple genes (19). Recent transcriptome analysis experiments have revealed that such interactions between genes (24), which constitute a regulatory network, determine the expression or non-expression of genes. Furthermore, regulatory interactions (promoters, transcription factor binding sites, enhancers) change more frequently through the evolutionary course than the respective genes that encode specific proteins (25).

For this purpose, we extend the GRN to a diploid case. Although a few studies have addressed heterosis using GRN models (26, 27), the consequences of GRN dynamics and the robustness of phenotypes to noise and genomic mixing by sexual recombination have not previously been addressed. Therefore, we analyze these topics by simulating the evolution of diploid gene expression dynamics and discuss the mechanism of hybrid fitness and phenotypic variance.

The organization of this paper is as follows. First, we introduce the model of the diploid GRN. Then, we compare the fitness of homozygote and heterozygote populations and reveal the fitness-related heterosis. Next, we compare the phenotypic variances between homozygote and heterozygote populations to demonstrate that the latter takes much smaller values. Additionally, we compare the ratio of heterosis related to phenotypic variance obtained in our simulation to previous measurement data (8). We also reveal the correlation between hybrid fitness and the dominance or overdominance of gene expression patterns. Finally, an intuitive explanation of the results is provided in terms of the fitness landscape.

Brief description of the model

Evolution simulation based on a diploid GRN model. First, we introduce the theoretical GRN model (We basically follow the model in (13)). Each gene $i (= 1, 2, \dots, N)$ has an expression level $x_i(t)$ at time step t . In the model, $x_i(t)$ is scaled such that it takes a value between zero (non-expression) and one (expression). Each gene interacts with others and itself, with interaction between the j th gene and the i th gene described by the matrix J_{ij} (19–23). 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 adopt a discrete-time model (17, 19, 28, 29), 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 GRN. Gaussian noise, $\eta(0, \sigma)$ of small magnitude is added to the dynamics to account for the stochasticity in gene expression. The noise term is multiplied by $\sqrt{x}$ to avoid a negative value of $x_i(t)$. According to the sigmoid function $f[x] = 1/(1 + \exp[-\beta x])$ with a large $\beta (= 100)$, the expression dynamics are given by

$$x_i(t+1) = f\left[\sum_{j=1}^N J_{ij}(x_j(t) - \theta)\right] + \sqrt{x}\eta(0, \sigma), \quad [1]$$

where θ is the threshold for the expression level. Starting from the given fixed initial expression levels $\{x_i(0)\}$, where $x_i(0) = 0$ for most genes and one for several (10) genes. $x_i(t)$

reaches a fixed point after a certain time T in most cases, where x_i is zero or one depending on the structure of the network J_{ij} .

Diploid cells have two genomes that provide two matrices, $J_{ij}^{(1)}$ and $J_{ij}^{(2)}$. The gene expression dynamics of a diploid GRN are determined by the sum of these matrices; 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 follows:

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

Genetic changes progress according to the following procedure. First, fitness is determined by the distance of the expression pattern of genes $\{x_i(T)\}$ from a prescribed target pattern for $i = 1, 2, \dots, M (< N)$. Two parents are chosen with a probability proportional to the fitness. Two genomes from each parent, $J_{ij}^{(1)}$ and $J_{ij}^{(2)}$, were mixed by recombination to provide a gamete. Next, two gametes 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 a mutation rate of μ . This mutation is included as a change in the matrix element J_{ij} of 0 or ± 1 (see Materials and Methods for details on the selection procedure).

Defining homozygote and heterozygote populations. The homozygote population is generated as a diploid population consisting of precisely the same diploid genome: $J_{ij}^{(1)}$ and $J_{ij}^{(2)} (= J_{ij}^{(1)})$. Then, the population of heterozygotes (heterozygote population) is generated as a diploid consisting of two genomes (matrices): $J_{ij}^{(1)}$ and another $J_{ij}^{(2)} (\neq J_{ij}^{(1)})$.

We first measured the fitness of the homozygote and heterozygote populations, where the mean values for the populations are given by W^{homo} and W^{hetero} , respectively. Next, phenotypic variance according to noise V_{noise} is defined as the magnitude of the variance of $x_i(T)$ for individuals with the same genotype but with different noise added to the gene expression dynamics (see Methods for the definition). V_{noise} is computed as the average of the variance over genes i for homozygote and heterozygote populations, giving $V_{\text{noise}}^{\text{homo}}$ and $V_{\text{noise}}^{\text{hetero}}$, respectively.

Results

Fitness-related heterosis. First, we discuss 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. 1A. The ratio of $W^{\text{hetero}}/W^{\text{homo}}$ increased for earlier generations, and then decreased. After evolution (at the 999th generation), inequality $W^{\text{hetero}}/W^{\text{homo}} > 1$ was maintained; therefore, fitness-related heterosis was generally observed. (See Fig. S1 for the evolution of the distributions of W^{homo} and W^{hetero} .)

In Fig. 1B, we plotted W^{hetero} versus W^{homo} for each sample after evolution. Except for three of the 100 samples, $W^{\text{hetero}} > W^{\text{homo}}$ was satisfied, indicating evolution of heterosis. Then, across each sample, we computed the average ratio of $W^{\text{hetero}}/W^{\text{homo}}$ and examined its dependence on the mutation rate. As shown in Fig. 1C, $W^{\text{hetero}}/W^{\text{homo}}$ increased

Fig. 1. Evolution, variation, and mutation-rate dependence of heterosis. (A) Change in the degree of fitness-related heterosis $W^{\text{hetero}}/W^{\text{homo}}$ throughout evolution, calculated for 30 samples in each generation. Points represent the median, and error bars represent the range between the first and third quantiles. $W^{\text{hetero}}/W^{\text{homo}}$ was 0.8 in the 0th generation; it increased initially, then decreased thereafter. The mutation rate per edge was $\mu = 3.0 \times 10^{-4}$ and the noise strength was $\sigma = 1.0 \times 10^{-4}$. (B) Variation between W^{hetero} and W^{homo} for each sample. The noise was set to $\sigma = 1.0 \times 10^{-4}$. Even under the same conditions, variation occurred in W^{hetero} and W^{homo} . For all samples in the $\mu = 9.5 \times 10^{-5}$ condition, and all except three samples out of 50 in the $\mu = 5.5 \times 10^{-5}$ condition, the relationship $W^{\text{hetero}} > W^{\text{homo}}$ was maintained despite sample-to-sample variation. (C) Increase in $W^{\text{hetero}}/W^{\text{homo}}$ with mutation rate μ . Points represent the median, and error bars represent the range between the first and third quantile for 30 samples for different noise magnitudes σ . The ratio $W^{\text{hetero}}/W^{\text{homo}}$ increased with μ and $W^{\text{hetero}}/W^{\text{homo}} > 1$ was maintained. (D) Evolution of the mean value of the degree of variance-related heterosis $V_{\text{noise}}^{\text{homo}}/V_{\text{noise}}^{\text{hetero}}$. The $V_{\text{noise}}^{\text{homo}}/V_{\text{noise}}^{\text{hetero}}$ increased with evolution, reaching a median value of approximately 3.0 ($\mu = 3.0 \times 10^{-4}$ and $\sigma = 1.0 \times 10^{-4}$). (E) $V_{\text{noise}}^{\text{homo}}$ and $V_{\text{noise}}^{\text{hetero}}$ variation for each sample for $\mu = 5.5 \times 10^{-5}$ and $\mu = 9.5 \times 10^{-5}$. The relationship $V_{\text{noise}}^{\text{homo}} > V_{\text{noise}}^{\text{hetero}}$ was maintained despite sample-to-sample variation. (F) Change in $V_{\text{noise}}^{\text{homo}}/V_{\text{noise}}^{\text{hetero}}$ with μ . $V_{\text{noise}}^{\text{homo}}/V_{\text{noise}}^{\text{hetero}}$ increased with μ in this range.

with μ , indicating that a greater degree of heterosis evolved at a higher mutation rate. Additionally, at higher μ , the ratio of heterosis $W^{\text{hetero}}/W^{\text{homo}}$ was larger at a larger σ .

Variance-related heterosis. The evolution of the mean value of the degree of variance-related heterosis $V_{\text{noise}}^{\text{homo}}/V_{\text{noise}}^{\text{hetero}}$ is shown in Fig. 1D (see Fig. S2 for the evolution of the distributions of $V_{\text{noise}}^{\text{homo}}$ and $V_{\text{noise}}^{\text{hetero}}$). $V_{\text{noise}}^{\text{homo}}/V_{\text{noise}}^{\text{hetero}}$ increased from two at the 0th generation to approximately three after evolution (at the 999th generation). Note that, for a random network, the estimated ratio is two (see SI Appendix). Hence, the results indicate that evolution created non-trivial heterosis related to variance. We call this phenomenon variance-related heterosis to distinguish it from fitness-related heterosis. Because a lower variance V_{noise} corresponds to higher robustness to noise, the phenotype of heterozygous diploids is more robust against noise than homozygotes.

As shown in Fig. 1E, we measured $V_{\text{noise}}^{\text{homo}}$ and $V_{\text{noise}}^{\text{hetero}}$ after evolution for different samples under a given mutation rate and noise magnitude. In all samples, $V_{\text{noise}}^{\text{homo}} > V_{\text{noise}}^{\text{hetero}}$, which indicates that fluctuation of the heterozygote was smaller than that of the homozygote. The dependence of $V_{\text{noise}}^{\text{homo}}/V_{\text{noise}}^{\text{hetero}}$ on μ is shown in Fig. 1F. We confirmed that $V_{\text{noise}}^{\text{homo}}/V_{\text{noise}}^{\text{hetero}} > 1$ for all μ , whereas the ratio increased with μ . For $\mu > 5 \times 10^{-5}$,

$V_{\text{noise}}^{\text{homo}}/V_{\text{noise}}^{\text{hetero}} > 2$ was obtained after evolution, showing higher variance-related heterosis than for the case of random networks. At higher μ , the ratio of heterosis $V_{\text{noise}}^{\text{homo}}/V_{\text{noise}}^{\text{hetero}}$ was larger at a larger σ .

Comparison with previous measurement data. By measuring the variance of traits from data of 172 traits drawn from 15 different species, Phelan and Austad reported that hybrid populations exhibited fewer phenotypic fluctuations than inbred populations (8). These data are explained in (8) 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 variance of the inbred population and crossbred population divided by the mean (coefficient of variation, CV). Note that the variance adopted in (8) is not equal to the phenotypic variance by noise (V_{noise}). The variances adopted in their data also include that due to the genetic variation. However, as the V_{noise} and the all phenotypic variance are positively correlated (proportional) in previous studies (30–34). Hence, we compare the results of V_{noise} with the variance data in (8).

We compared the ratio of phenotypic variances for inbred

Fig. 2. Distribution of the ratio of variance (CV) of $V_{noise}^{homo}/V_{noise}^{hetero}$ for our model and previous experimental data. The simulation results (orange dots) are for $\mu = 1.0 \times 10^{-5}$, whereas the measurement data for CV^{inbred}/CV^{hybrid} are measured for inbred and crossbred populations (blue triangles) from (8).

plying that the variance-related heterosis obtained by the diploid GRN model is consistent with the field data.

Correlation of heterosis with pattern dominance and pattern overdominance. In this subsection, we discuss how heterosis is correlated with dominance and overdominance, as defined in (6, 7). Here, we need to extend the definition of dominance and overdominance to adapt to the case in which a collective of expressed genes in a network determines the phenotypic traits, rather than a single gene. Thus, we define and compute pattern dominance and pattern overdominance from the gene expression patterns and explore their correlation with heterosis(13).

To introduce the dominance of gene expression patterns, we first compute the gene expression patterns over a group of genes for given homozygotes AA and BB taken from a given genome pool, which is represented by x_i^{AA} and x_i^{BB} . Next, we compute the gene expression pattern of heterozygotes AB , given by x_i^{AB} , for selecting the genes $x_i^{AA} \neq x_i^{BB}$. Then, we examine whether the expression pattern of the heterozygote is biased toward one of the homozygote expression patterns, for example, $x_i^{AB} = x_i^{AA}$ or $x_i^{AB} = x_i^{BB}$, and compute the fractions. Pattern dominance is defined by a larger between the two. If the expression of each gene is independent, such a bias is not expected, and the pattern dominance is 0.5. The pattern dominance is one if $x_i^{AB} = x_i^{AA}$ or $x_i^{AB} = x_i^{BB}$ for all i with $x_i^{AA} \neq x_i^{BB}$. In contrast, for a set of genes with $x_i^{AA} = x_i^{BB}$, the fraction of such genes i of heterozygotes satisfying $x_i^{AB} \neq x_i^{AA}$ gives the degree of pattern overdominance (see the Methods section for the definitions of pattern dominance and pattern overdominance).

The correlation between pattern dominance, pattern overdominance, and fitness-related heterosis is shown in Fig.3 (see Fig.S3 for the correlation with variance-related heterosis). Heterosis W^{hetero}/W^{homo} exhibited a negative correlation with

Fig. 3. Correlation between heterosis and dominance (overdominance). (A) Correlation between pattern dominance and W^{hetero}/W^{homo} , with a correlation coefficient of -0.80. (B) Correlation between pattern overdominance and W^{hetero}/W^{homo} , with a correlation coefficient of +0.85. Each point corresponds to different mutation rates denoted by different symbols, whereas different points correspond to different noise magnitudes σ of $[5 \times 10^{-5}, 1 \times 10^{-2}]$. Points represent the average of 50 samples, calculated only for the 999th generation out of 50 samples with a fitness average of 0.9 or higher.

pattern dominance but a positive correlation with pattern overdominance. The positive correlation between fitness-related heterosis and pattern overdominance seems natural because the expression of heterozygotes must be different from that of homozygotes in order to achieve heterosis. On the other hand, heterosis was previously explained by dominance, because dominance can compensate for the deleterious mutations in multiple loci. However, the negative correlation between fitness-related heterosis and pattern dominance, observed in our simulation, supports the overdominance explanation rather than the dominance explanation.

Similarly, variance-related heterosis was negatively correlated with pattern dominance but positively correlated with pattern overdominance; however, the correlation coefficient was smaller than that for fitness-related heterosis (see Fig. S3).

Explanation of the results according to the convex fitness region and single-humped fitness function. In this section, we suggest an intuitive explanation of our results using fitness landscape pictures. Nevertheless, further mathematical elucidation is required in the future.

We consider a fitness landscape over a genotype space (35). In particular, we discuss the high-fitness region. After evolution, offspring are produced from the two parents by meiosis within this high-fitness region. The genotypes of the diploid genome $(J_{ij}^{(1)} + J_{ij}^{(2)})/2$ will be located approximately at the internal division of the genotype of the parents. The offspring must have high fitness; otherwise, the offspring will not be selected. For example, when the higher-fitness region is not convex (Fig.4A), the offspring may not be selected because of their lower fitness compared to the parents. Therefore, the high-fitness region for the population under sexual reproduction and selection shapes the convex region in the genotype space (Fig.4B).

The fitness landscape, then, is expected to take a single-humped structure around the center of the region that decreases as the genotype moves toward the periphery of the high-fitness region (Fig.4B). When we consider the robustness of the fitted state, the fitness is expected to be less sensitive to genetic change around the center than at the periphery. This suggests that the fitness function is flat around the center and

Fig. 4. Schematic of the fitness landscape and high-fitness region. (A) When the high-fitness region is not convex, the offspring may have lower fitness than the parents; therefore, the lineage will not be selected. (B) When the high-fitness region is convex over the genotype space, the offspring's fitness is higher than or equal to that of the parents. Thus, the fitness of the heterozygote is higher than that of the homozygote, leading to heterosis.

steeper farther from the center (i.e., convex, single-humped functions). Thus, the variance of the fitness (phenotype) will be smaller around the center.

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

Thus, we expected the fitness landscape to be single-humped with a flatter top (i.e., convex), and examined the validity of this expectation in our model. First, we computed the distance between the two genome matrices $\sqrt{\sum_{i,j} (J_{ij}^{(k)} - J_{ij}^{(l)})^2}$ for the two genomes (k, l) in the population of the 999th generation. Among the combinations, we chose unfitted genomes, that is, the homozygotes of both K, L genomes had fitness ≤ 0.95 and the largest distance between genome matrices $\sqrt{\sum_{i,j} (J_{ij}^{(K)} - J_{ij}^{(L)})^2}$. Next, we produced 100 matrices corresponding to the internal division of $J_{ij}^{(K)}$ and $J_{ij}^{(L)}$, that is, $sJ_{ij}^{(K)} + (1-s)J_{ij}^{(L)}$, by changing s step-by-step in the range $0 \leq s \leq 1$. Then, we computed the mean fitness for homozygotes over 100 matrices of each division point given by s .

As shown in Fig. 5, the fitness function 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_{ij} (corresponding to the 0th generation), we suggest that it was acquired by evolution during sexual reproduction. This fitness increase at the center is further enhanced by introducing differences in the genomes in diploid (heterozygosity) with a few percentages.

Discussion

In the present study, we employed numerical evolution of the diploid GRN to demonstrate heterosis related to fitness and phenotypic variance, as well as its correlation with overdominance, and compared the simulation results with previous measurement data. First, heterozygote populations exhibited

Fig. 5. Fitness according to changes along a given genome pair, parametrized by s , where the internal division genome matrix is given by $sJ_{ij}^{(K)} + (1-s)J_{ij}^{(L)}$ between genome pairs (K, L) . This was chosen so that the fitness for (K, L) was less than 0.95 after evolution (at the 999th generation) and the distance between genome matrices was the maximum value. The horizontal axis represents the internal division parameter s . The average fitness over 30 samples is plotted for $\mu = 3.0 \times 10^{-4}$ and $\sigma = 1.0 \times 10^{-4}$.

higher fitness than homozygote populations, indicating heterosis. Heterosis exhibited a negative correlation with dominance and a positive correlation with overdominance. Moreover, phenotypic variance due to noise was lower in the heterozygote population than in the homozygote population, indicating variance-related heterosis. Furthermore, the numerically obtained distribution of the ratio of homozygote population variance to heterozygote population variance agreed quantitatively with that of previous measurement data (8).

We then proposed a possible explanation for these results according to the convex fitness landscape. Specifically, the observed landscape formation may arise if high fitness is maintained under the conditions of sexual reproduction and noise. Therefore, to increase or maintain fitness during sexual reproduction, it is necessary to create a population of genomes in which the fitness of the offspring is equal to or greater than that of the parents. In addition, the population located inside the high-fitness region will evolve to avoid the periphery of the region in which the phenotype (fitness) is vulnerable to noise. This fitness landscape implies epistasis. However, epistasis alone is not sufficient to support heterosis. A convex high-fitness region with a fitness function that flattens around the top is a stronger hypothesis than epistasis, as it concerns

the global shape of the fitness landscape for the entire genome population. The validity of this hypothesis requires further examination in theoretical models with evolving gene expression dynamics, as well as experimental data.

We also studied the relationship between overdominance and heterosis. Instead of a single locus, the dominance and overdominance of expression patterns at multiple loci were defined and measured in this study. Note that the previous explanation for heterosis by overdominance was based on the locus-specific overdominant effect in (36), whereas its molecular mechanism is still unclear(37). In contrast, pattern overdominance, as defined in this study, is concerned with the robustness of a system with interacting genes(11, 12), whereas future experimental confirmation is needed. Furthermore, because the distribution of the underlying genomic population is diverse, the genomic distribution of the homozygote and heterozygote populations should be measured separately to ensure a careful analysis of heterosis.

Finally, we discussed variance-related heterosis, that is, the relationship between phenotypic variance in heterozygotes and homozygotes. Recall that even a random network can produce considerable variance-related heterosis; however, an evolved population shows stronger variance-related heterosis. The result of variance-related heterosis can also be explained by the single-humped fitness function described above, which increases robustness to noise. Such noise robustness has recently received substantial attention in the context of homeostasis, and cell state selection (15–18, 38); thus, heterosis should also be reexamined in the light of noise robustness.

In this study, our numerical results were compared with data previously compiled by Phelan and Austad (8). We uncovered remarkable similarities in the distributions of the ratio of variance-related heterosis, whereas absolute values were not necessarily the same. The simulation results assumed a given fixed species and fitness condition, whereas the measurements were derived from different organisms under different environmental conditions. In order to make a further quantitative comparison, experiments should be conducted for given species under fixed environmental conditions, with ploidy manipulation and measurement of the noise. For example, both are available in yeast(39–41), in which the GRN has already been analyzed. A quantitative analysis of both heterosis and phenotypic noise is also required in future research.

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(10) and meiosis of genomes(42). This evolution of heterosis also implies advantages of sexual reproduction, which goes beyond the picture of Muller's ratchet(43) that is a removal of deleterious mutants: The overdominance explanation supported in this study suggests that heterozygosity in diploid organisms is advantageous with interactions between the two genomes, by shaping a convex fitness landscape through evolution. As the present evolution does not assume breeding that is often the target of standard heterosis studies, the present study will be applied for the evolution in the field and laboratory experiments as well.

Materials and Methods

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 corre-

sponding protein in a cell. Note that the proteins are synthesized from both chromosomes, where mutations or recombinations are mainly introduced in the promoter or enhancer region, so that the superscript m in $x_i(t)$ is not required. By replicating the regulation in the diploid cell by $J_{ij}^{(m)}$ ($m = 1, 2$) for each gene, we obtain

$$y_i^m(t+1) = \frac{1}{2} f \left[\sum_{j=1}^N J_{ij}^{(m)} (x_j(t) - \theta) \right] \quad (m = 1, 2). \quad [3]$$

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, we then obtain

$$x_i(t+1) = \sum_{l=1}^2 y_i^l(t+1) = \frac{1}{2} \sum_{l=1}^2 f \left[\sum_{j=1}^N J_{ij}^{(l)} (x_j(t) - \theta) \right]. \quad [4]$$

Evolutional simulation. We take the initial condition $x_i(0) = 0$ for $i = 1, 2, \dots, N_{\text{zero}}$ and $x_i(0) = 1$ for $i = N_{\text{zero}} + 1, \dots, N$, where $N_{\text{zero}} = N - 10$. After a sufficient time, the expression pattern $x_i(t)$ reaches a stable state $x_i(T)$. Then, the fitness is determined by the distance between the gene expression levels of the “target genes,” $i = 1, \dots, M$ ($M < N$). The target pattern is defined as one for all genes, and fixed throughout the evolution. Here, the phenotype is defined as the time average of $x_i(t)$ for the last 10 steps, $\bar{x}_i$. By setting the maximum fitness for $\bar{x}_i - 1 = 0$ ($i = 1, \dots, M$), the fitness is defined as $-\sum_{i=1}^M |\bar{x}_i - 1|$. In some cases, the expression pattern $x_i(t)$ oscillates over time and cannot reach a stable state; however, these cases are rare. Moreover, after the evolution, such a case is not observed because a stable expression pattern $x_i = 1$ ($i = 1, \dots, M$) can be more advantageous than an oscillatory state under this fitness landscape. The selection pressure is set to S , so that the probability of obtaining the phenotype $\bar{x}_i$ is proportional to $w = \exp[-S \sum_{i=1}^M |\bar{x}_i - 1|]$.

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 ($k_1 \neq k_2$), where W is the sum of all fitness values in the population. To introduce meiosis, we adopt the following procedure. A single gamete is generated by mixing $[J_{ij}^{(1)}]_{k_1}$ and $[J_{ij}^{(2)}]_{k_1}$ ($[J_{ij}^{(1)}]_{k_2}$ and $[J_{ij}^{(2)}]_{k_2}$) via row vectors with equal probabilities to produce a new genome n , where $[J_{ij}^{(1)}]_n$ ($[J_{ij}^{(2)}]_n$).

The simulation is performed by changing the mutation rate (μ) is taken in $[10^{-5}, 10^{-2}]$ and the standard deviation of the noise (σ) in $[5 \times 10^{-4}, 10^{-1}]$. The number of genes (N) is set to 100, the ratio of non-zero elements in $J_{ij}^{(m)}$, the selection pressure, S , is set to 2, and the number of target genes (M) is 10. The relaxation time (T) to examine the fitness is set to 30. The total number of individuals (K) included in the study is 100.

Definition of V_{noise} . For a given diploid with $J_{ij}^{(1)}$ and $J_{ij}^{(2)}$, x_i after relaxation is computed under noise conditions. After repeating this process L times and obtaining x_i for each run, we measure the variance of the expression pattern x_i . The variance V_{noise} is computed using the average of the variance over all genes ($N = 100$) and trials ($L = 100$).

Pattern dominance: definition and calculation. First, we extract $2N$ genomes $J_{ij}^{(n)}$ ($n = 1, \dots, 2K$) from K individuals in the evolutionary simulation. We then make a complete copy for each $2K$ genome, creating a population of $2K$ complete homozygotes, e.g., $J_{ij}^{(A)} J_{ij}^{(A)}$. Then, we compute the expression dynamics in each of the $2K$ complete homozygotes to obtain the expression pattern x_i^{AA} . Next, from the $2K$ genomes, two $J_{ij}^{(A)}$ and $J_{ij}^{(B)}$ are randomly selected to create a diploid cell consisting of two heterozygous genomes. Then, from the expression dynamics of the heterozygous genomes, the expression pattern of heterozygous x_i^{AB} is obtained.

We now compare the expression of genes that are differentially expressed between two such homozygotes, that is, those with $x_i^{AA} \neq$

Fig. 6. Definition and calculation of pattern dominance and pattern overdominance. (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). Then, the pattern dominance is 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) Pattern overdominance: For the gene that shows the same expression in the two homozygotes, we calculate the percentage of the expression pattern obtained from the heterozygote that is different from that of AA and BB , focusing on the genes that show the same expression in the two homozygotes (blue). Then, the pattern overdominance is computed as the percentage of $x_i^{AB} \neq x_i^{AA} = x_i^{BB}$.

Note that, in the current model, the expression level of each gene x_i is almost always 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 calculate the pattern dominance only for homozygote x_i^{AA} and x_i^{BB} pairs in which more than 30 genes are differentially expressed.

Pattern overdominance: definition and calculation. Whereas pattern dominance focuses on the expression patterns of genes that differ between homozygotes of the two genomes, pattern overdominance focuses on genes expressed in the same manner between the two genomes. Let $N_{\text{same}}^{AA,AB}$ be the number of genes with the same expression as the two pure strains of genomes, and N_{over}^{AB} be the number of genes in heterozygote AB that are expressed differently from the pure strains AA and BB . $N_{\text{over}}^{AB}/N_{\text{same}}^{AA,AB}$ is defined as the degree of pattern overdominance (Fig. 6).

ACKNOWLEDGMENTS. The authors would like to thank Tetsumi Hatakeyama, Yuichi Wakamoto, Naoki Irie, Akira Sasaki, and Shuji Ishihara for their stimulating discussion, and Daniel Hartl for informing us of (8). This research was partially supported by a Grant-in-Aid for Scientific Research (A) 431 (20H00123) 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.

1. C Darwin, The effects of cross and self fertilization in the vegetable kingdom (Murray, London) (1876).
2. GH Shull, The composition of a field of maize. *J. Hered.*, 296–301 (1908).
3. A Bruce, The Mendelian theory of heredity and the augmentation of vigor. *Science* **32**, 627–628 (1910).
4. DF Jones, Dominance of linked factors as a means of accounting for heterosis. *Genetics* **2**, 466 (1917).
5. G Mendel, Versuche über pflanzen-hybriden. *Verhandlungen des naturforschenden Vereins Brunn für* **4**, 3–47 (1866).
6. EM East, Heterosis. *Genetics* **21**, 375 (1936).
7. JA Birchler, H Yao, S Chudalayandi, D Vaiman, RA Veitia, Heterosis. *The Plant Cell* **22**, 2105–2112 (2010).
8. JP Phelan, SN Austad, Selecting animal models of human aging: inbred strains often exhibit less biological uniformity than f1 hybrids. *J. gerontology* **49**, B1–B11 (1994).
9. S Ciliberti, OC Martin, A Wagner, Robustness can evolve gradually in complex regulatory gene networks with varying topology. *PLoS Comput. Biol.* **3**, e15 (2007).
10. OC Martin, A Wagner, Effects of recombination on complex regulatory circuits. *Genetics* **183**, 673–684 (2009).

11. S Nagata, M Kikuchi, Emergence of cooperative bistability and robustness of gene regulatory networks. *PLoS computational biology* **16**, e1007969 (2020).
12. T Kaneko, M Kikuchi, Evolution enhances the mutational robustness and suppresses emergence of a new phenotype. *arXiv preprint arXiv:2012.03030* (2020).
13. K Okubo, K Kaneko, Evolution of dominance in gene expression pattern associated with phenotypic robustness. *BMC Ecol. Evol.* **21**, 1–14 (2021).
14. J Jang, Fmc Amblard, CM Ghim, Heterogeneity is not always a source of noise: Stochastic gene expression in regulatory heterozygote. *Phys. Rev. E* **104**, 044401 (2021).
15. PS Swain, MB Elowitz, ED Siggia, Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci.* **99**, 12795–12800 (2002).
16. J Ou, et al., Dynamic change in promoter activation during lysine biosynthesis in *Escherichia coli* cells. *Mol. biosystems* **4**, 128–134 (2008).
17. MB Elowitz, AJ Levine, ED Siggia, PS Swain, Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
18. C Furusawa, T Suzuki, A Kashiwagi, T Yomo, K Kaneko, Ubiquity of log-normal distributions in intra-cellular reaction dynamics. *Biophysics* **1**, 25–31 (2005).
19. L Glass, SA Kauffman, The logical analysis of continuous, non-linear biochemical control networks. *J. theoretical Biol.* **39**, 103–129 (1973).
20. E Mjolsness, DH Sharp, J Reintz, A connectionist model of development. *J. theoretical Biol.* **152**, 429–453 (1991).
21. I Salazar-Ciudad, S Newman, R Solé, Phenotypic and dynamical transitions in model genetic networks i. emergence of patterns and genotype-phenotype relationships. *Evol. & development* **3**, 84–94 (2001).
22. I Salazar-Ciudad, J García-Fernández, RV Solé, Gene networks capable of pattern formation: from induction to reaction–diffusion. *J. theoretical biology* **205**, 587–603 (2000).
23. K Kaneko, *Life: an introduction to complex systems biology*. (Springer), (2006).
24. D Mercatelli, L Scalambra, L Triboli, F Ray, FM Giorgi, Gene regulatory network inference resources: A practical overview. *Biochimica et Biophys. Acta (BBA)-Gene Regul. Mech.* **1863**, 194430 (2020).
25. NM Luscombe, et al., Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **431**, 308–312 (2004).
26. PMF Emmrich, HE Roberts, V Pancaldi, A gene regulatory network simulation of heterosis in *International Conference on Information Processing in Cells and Tissues*. (Springer), pp. 12–16 (2012).
27. PMF Emmrich, HE Roberts, V Pancaldi, A boolean gene regulatory model of heterosis and speciation. *BMC evolutionary biology* **15**, 1–15 (2015).
28. A Wagner, Does evolutionary plasticity evolve? *Evolution* **50**, 1008–1023 (1996).
29. RB Azevedo, R Lohaus, S Srinivasan, KK Dang, CL Burch, Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. *Nature* **440**, 87–90 (2006).
30. K Kaneko, Evolution of robustness to noise and mutation in gene expression dynamics. *PLoS One* **2**, e434 (2007).
31. CR Landry, B Lemos, SA Rifkin, W Dickinson, DL Hartl, Genetic properties influencing the evolvability of gene expression. *Science* **317**, 118–121 (2007).
32. B Lehner, K Kaneko, Fluctuation and response in biology. *Cell. Mol. Life Sci.* **68**, 1005–1010 (2011).
33. SC Stearns, M Kaiser, TJ Kawecki, The differential genetic and environmental canalization of fitness components in *Drosophila melanogaster*. *J. Evol. Biol.* **8**, 539–557 (1995).
34. C Furusawa, K Kaneko, Global relationships in fluctuation and response in adaptive evolution. *J. The Royal Soc. Interface* **12**, 20150482 (2015).
35. S Wright, Evolution in mendelian populations. *Genetics* **16**, 97–159 (1931).
36. GP Rédei, Single locus heterosis. *Zeitschrift für Vererbungslehre* **93**, 164–170 (1962).
37. ZB Lippman, D Zamir, Heterosis: revisiting the magic. *Trends genetics* **23**, 60–66 (2007).
38. B Lehner, Genes confer similar robustness to environmental, stochastic, and genetic perturbations in yeast. *PLoS one* **5**, e9035 (2010).
39. B Lehner, Conflict between noise and plasticity in yeast. *PLoS genetics* **6**, e1001185 (2010).
40. G Yvert, et al., Single-cell phenomics reveals intra-species variation of phenotypic noise in yeast. *BMC systems biology* **7**, 1–16 (2013).
41. M Mundt, A Anders, SM Murray, V Sourjik, A system for gene expression noise control in yeast. *ACS synthetic biology* **7**, 2618–2626 (2018).
42. DB Goldstein, Heterozygote advantage and the evolution of a dominant diploid phase. *Genetics* **132**, 1195–1198 (1992).
43. HJ MULLER, The relation of recombination to mutational advance. *Mutat. research* **106**, 2 (1964).