The Effects of Recombination on Phenotypic Exploration and Robustness in Evolution

Ting Hu*,**
Geisel School of Medicine at Dartmouth
Dartmouth College

Wolfgang Banzhaf†
Department of Computer Science
Memorial University

Jason H. Moore**
Geisel School of Medicine at Dartmouth
Dartmouth College

Abstract Recombination is a commonly used genetic operator in artificial and computational evolutionary systems. It has been empirically shown to be essential for evolutionary processes. However, little has been done to analyze the effects of recombination on quantitative genotypic and phenotypic properties. The majority of studies only consider mutation, mainly due to the more serious consequences of recombination in reorganizing entire genomes. Here we adopt methods from evolutionary biology to analyze a simple, yet representative, genetic programming method, linear genetic programming. We demonstrate that recombination has less disruptive effects on phenotype than mutation, that it accelerates novel phenotypic exploration, and that it particularly promotes robust phenotypes and evolves genotypic robustness and synergistic epistasis. Our results corroborate an explanation for the prevalence of recombination in complex living organisms, and helps elucidate a better understanding of the evolutionary mechanisms involved in the design of complex artificial evolutionary systems and intelligent algorithms.

Keywords
Recombination, robustness, epistasis, evolvability, genotype network, genetic programming

1 Introduction

Mutation and recombination are the two major sources of genetic variation, and are essential components of evolution. Although the majority of genetic variations may be deleterious, complex living systems are able to harness these evolutionary forces to generate long-term phenotypic innovation and adaptation. Mutation and recombination are, however, very distinct from each other in effecting genetic variations. One of the main forms of mutation, referred to as point mutation, changes a single nucleotide in a genome. In contrast, recombination reuses and reorganizes units of two genomes, and is potentially better at preserving well-adapted modules while creating more innovative phenotypic traits.

Recombination is ubiquitous in more complex organisms, evidence of its evolutionary success [3, 24]. However, explanations for its benefits are still largely under debate in evolutionary biology and population genetics [12, 38, 52]. Given the high cost of sexual recombination in living individuals and populations, sexual organisms must derive significant benefit from recombination to explain its

* Contact author.
** Computational Genetics Laboratory, Geisel School of Medicine at Dartmouth, Hanover, NH; Institute for Quantitative Biomedical Sciences, Dartmouth College, Hanover, NH. E-mail: ting.hu@dartmouth.edu (T.H.); jason.h.moore@dartmouth.edu (J.H.M.)
† Department of Computer Science, Memorial University, St. Johns, NL, Canada. E-mail: banzhaf@mun.ca

© 2014 Massachusetts Institute of Technology

Downloaded from http://direct.mit.edu/artl/article-pdf/20/4/457/1664703/artl_a_00145.pdf by guest on 26 September 2021

pervasive existence. It has been conjectured that recombination generates new beneficial alleles faster than sequential mutations by bringing together different individuals, resulting in a speedup of adaptive evolution [14, 36]. Moreover, recombination may accelerate the elimination of deleterious mutations [24, 37]. When the joint negative effects of multiple mutations are stronger than the sum or product of individual effects, recombination helps impose a strong selection pressure towards removing those deleterious mutations. Based on mathematical models developed earlier, more recent empirical studies have investigated the effect of recombination on genetic robustness and epistasis, in order to explain the essential benefits and mechanisms of sexual reproduction/recombination [1, 5, 30, 43].

Robustness is a pervasive property of natural organisms, and is used to understand the resilience of evolutionary systems when confronted with environmental and heritable perturbations [9, 31, 49]. Robustness enables living systems to remain phenotypically intact in the face of constant genetic variations, by allowing genetic variants to expand in neutral spaces. These neutral spaces are genotypic regions in which mutations do not change the phenotype or the fitness of an individual and are the consequence of a redundant genotype-to-phenotype mapping [51]. Such neutrality then augments evolvability, by accumulating genetic variations that might be non-neutral under changes of the environmental context [11, 19, 26, 32, 51, 53].

Epistasis is another ubiquitous property of living organisms [6, 34, 35, 39, 40, 47]. Given the complexity of many phenotypic traits of living organisms, the phenotypic expression of one gene very likely depends on other genes. This dependence, or interaction, among multiple genes or genetic components is described by epistasis. Epistasis can be either synergistic (i.e., the joint effect of multiple genes is stronger than the total individual effects) or antagonistic (i.e., individual genes mitigate each other's effects). Synergistic epistasis amplifies the joint phenotypic effects of cumulative individual mutations, especially when they are negative, and thus is subject to a strong selection pressure towards removing those mutations [8, 10, 37].

Using simulated gene regulatory network models, studies have shown that recombination can create novel phenotypes more efficiently and with less disruption than mutation can [30, 52]. Moreover, it is reported that sexual reproduction/recombination results in stronger selection for genetic robustness than asexual reproduction, causing synergistic epistasis as a by-product of this selection pressure [1, 16, 28, 29, 46, 55]. Using a digital organism system, it is also reported that sexual reproduction influences the evolution of modularity and epistasis [33]. The genomes of sexual organisms were found to be more modular and to evolve weaker antagonistic epistasis than asexual ones. Such observations strengthen the hypothesis that recombination results in high genetic robustness and synergistic epistasis, and consequently accelerates generating new adaptive phenotypes and eliminating negative mutations.

In this study, we use a genetic programming method, linear genetic programming [4], to investigate the effect of recombination on evolution and its interplay with mutation. Linear genetic programming is a branch of the evolutionary computation family of algorithms. Recombination is a commonly used genetic variation operator in evolutionary computation algorithms [18, 25]. In the literature, recombination has long been the center of the discussion about effective genetic operators. Various recombination operators have been proposed to accommodate specific algorithmic representations and to improve optimization and search performance [15, 17, 22, 41, 42, 45].

Similar to many natural and artificial evolutionary systems, linear genetic programming has a redundant genotype-to-phenotype mapping, where many mutational variants of a genotype yield identical phenotypes. This redundant mapping creates robustness at both the genotypic and phenotypic levels. That is, a genotype is robust when randomly mutating its alleles does not alter its phenotypic outcome, and a phenotype is robust when it has a large number of underlying genotypic variants. This property provides a suitable framework for recombination and robustness analysis. The specific model used in this study has a fixed-length genome for the representation, which allows us to define finite and compact spaces for genotypes and phenotypes. This separates the effect of code growth (i.e., program size increase as evolution proceeds) on robustness, which was previously reported as an important element responsible for the increasing robustness in genetic programming [2, 44], from robustness generated by the genotype-phenotype map in certain regions of the search space. We use a linear genetic programming system to investigate the distribution of robustness at

Downloaded from http://direct.mit.edu/artl/article-pdf/20/4/457/1664703/artl_a_00145.pdf by guest on 26 September 2021

both genotypic and phenotypic levels, the effect of recombination on evolution (specifically on robustness and epistasis), and how recombination interacts with mutation during evolution.

The primary goal of this study is to bring new perspectives developed in evolutionary biology to artificial and computational evolutionary systems so as to help us better understand the mechanisms of essential evolutionary system components. This could greatly benefit the future design of complex artificial systems and intelligent algorithms [20]. Additionally, as argued in the community of evolutionary biology and population genetics, many of the proposed hypotheses on recombination greatly rely on specifications and restrictions of evolutionary models. Therefore, testing a simple, yet representative, computational evolutionary model may help elucidate the general role that recombination plays in evolution.

2 Methods

2.1 Evolutionary Program Model in a Boolean Search Problem

The artificial evolutionary system we used in the current study is linear genetic programming (LGP). Evolvable individuals are represented as executable linear computer programs. Each program consists of a set of L instructions, which are structurally similar to those found in register machine languages. Each instruction has an operator, a set of operands, and a return value. In the current study, we restrict the LGP model to a Boolean search problem where each instruction consists of an operator drawn from the Boolean function set {AND, OR, NAND, NOR}, two Boolean operands, and one Boolean return value. The inputs, operands, and return values are stored in registers with differing read/write permissions. Specifically, R_0 and R_1 are calculation registers that can be read and written, whereas R_2 and R_3 are input registers that are read-only. Thus, a calculation register in an instruction can serve as an operand or a return, but an input register can only serve as an operand. An example program of length $L = 3$ is given as follows:

$$\begin{array}{lcl} R_1 = R_2 & \text{AND} & R_3 \\ R_0 = R_2 & \text{OR} & R_1 \\ R_0 = R_3 & \text{NAND} & R_0 \end{array}$$

Instructions are executed sequentially from top to bottom. Prior to program execution, the values of R_0 and R_1 are initialized to FALSE. Registers R_2 and R_3 read two Boolean input values. After program execution, the final value in R_0 is returned as output.

2.2 Implementation of Mutation and Recombination

As is common in the evolutionary computation literature, in this study we employ two major genetic variation operators, point mutation and recombination. Their implementation in the LGP representation is straightforward. Point mutation can be applied to any of the $4 \times L$ loci of an individual, and the mutant locus randomly takes a new allele from the set of all compatible alleles at that locus.

To keep a fixed length of individuals, we use symmetric single-point recombination (see Figure 1). That is, for two parent programs with the same length L, a recombination point is randomly selected

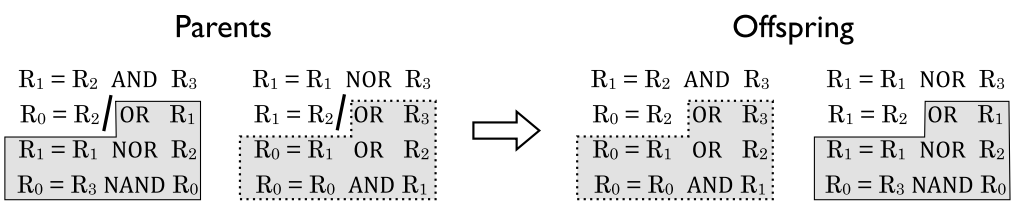

Figure 1. An example of symmetric single-point recombination for programs with length $L = 4$. The recombination point is chosen randomly at any of the possible loci, but it needs to be the same for both parents. Then the two parent programs (left) swap their fractions after the recombination point with each other to form two new offspring (right).

Downloaded from http://direct.mit.edu/artl/article-pdf/20/4/457/1664703/artl_a_00145.pdf by guest on 26 September 2021

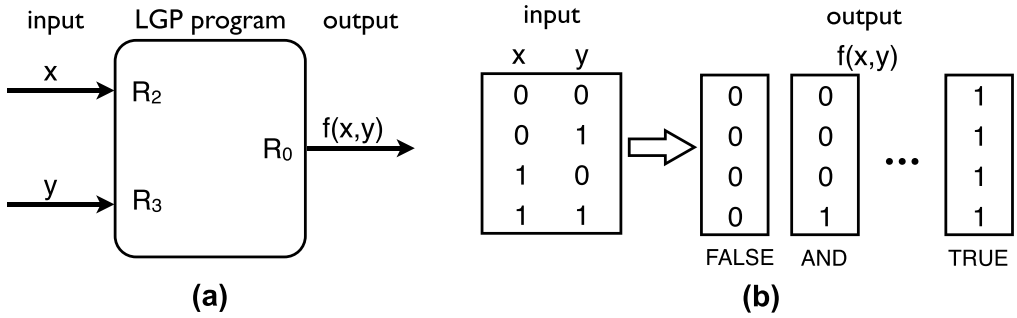

Figure 2. Schematic description of the LGP phenotype. (a) The input and output of a LGP program. Registers R_2 and R_3 read the input values, and the final value stored in register R_0 is returned as a program's computational outcome. (b) The phenotype of a LGP program is defined as the Boolean relationship it encodes, represented as the four-digit output of the ordered two-variable inputs.

with a uniform probability and the two parents swap the code fragments after the recombination point to form two new offspring. This operation allows all resulting offspring to have the same length as their parents, and thus helps, for the current study, to limit recombination dynamics to a finite genotype space.

2.3 Genotype and Phenotype Networks

A *genotype* in our LGP model is a program with a unique set of alleles at a total of $4 \times L$ loci. In the specified system, since the return register can be chosen from $\{R_0, R_1\}$, the two operand registers can be chosen from $\{R_0, R_1, R_2, R_3\}$, and the operation can be chosen from {AND, OR, NAND, NOR}, the total number of different instructions is $2 \times 4 \times 4 \times 4 = 2^7$ and the total number of possible genotypes is $2^{7 \times L}$.

The binary Boolean relationship $f: \mathbf{B}^2 \rightarrow \mathbf{B}$, mapping inputs R_2, R_3 into outputs R_0 that a program encodes (where $\mathbf{B} = \{\text{TRUE}, \text{FALSE}\}$), is defined as its *phenotype*. As shown in Figure 2, when all four possible combinations of two Boolean variables are input into a LGP program, the output can be regarded as a four-digit binary string representing one of the sixteen possible binary Boolean functions. For instance, the output string "0000" represents the FALSE function, and the output string "0001" represents the AND function.

Genotypes can be transformed from one to another through single-point mutations. We use the notion of a *genotype network* [50] to characterize the neighborhood structure of genotypes. For instance, in the genotype network of Figure 3a, each vertex denotes a genotype, and two genotypes are connected if a single-point mutation can change the genotype to another. An edge can stand for either a neutral point mutation (i.e., the phenotype is retained for the two genotypes) or a non-neutral one (i.e., the point mutation leads to a phenotypic change). For a particular genotype, its neutral neighbors include those genotypes that share the same phenotype, and its non-neutral neighbors are all the other neighbors that have different phenotypes.

On top of the genotype network, we consider the *phenotype network* to describe the connectivity among various phenotypes. In the phenotype network of the LGP system with a fixed length $L = 4$ (see Figure 3b), each vertex corresponds to a unique phenotype, and the size of the phenotype is proportional to the total number of genotypes representing it. Our particular LGP system has 16 different phenotypes, shown as the 16 vertices in the network. Two phenotypes are connected by an edge if there exists at least one single-point mutation that can transform genotypes of the two phenotypes into each other. The thickness of an edge is proportional to the total number of single-point mutations that can change one phenotype into the other. As seen in the graph, for that particular LGP system, the phenotype network is fully connected, that is, each phenotype can be reached from any other phenotype by mutating their genotypes. The vertex size is highly heterogeneous, however, ranging from a minimum of 24,832 genotypes (phenotypes XOR and EQUAL) to

Downloaded from http://direct.mit.edu/artl/article-pdf/20/4/457/1664703/artl_a_00145.pdf by guest on 26 September 2021

a maximum of 60,393,728 genotypes (phenotype `FALSE`) that occupy from $\ll 1\%$ up to 23% of the entire genotype space, respectively. Moreover, edge widths are also heterogeneous, indicating the connectivity among phenotypes is not evenly distributed. For instance, phenotype `NOR` is more likely to be mutated to phenotypes `!y`, `!x`, and `FALSE` than to other phenotypes.

It is interesting to note that what Kaufmann calls "the adjacent possible" [23] finds a natural representation in this network: It is all the next nearest neighbors to a particular phenotype as encoded by the edges of the phenotype network. This assumes a definition of the adjacent possible on the phenotypic level, which is probably what Kaufmann envisioned. In our particular LGP example, the complete phenotype network indicates that all phenotypic changes are adjacent possible.

2.4 Genotypic and Phenotypic Robustness

Similar to many evolutionary systems, either natural or artificial, our LGP system has a highly redundant mapping from a large genotype space ($2^{7\times L}$ genotypes) to a relatively small phenotype space (16 phenotypes). The redundant genotype-to-phenotype mapping generates robustness, which means that altering some alleles on a genotype does not change its phenotype. We investigate this robustness at the genotypic and phenotypic levels.

For a genotype g_i, we define its *genotypic robustness* $R_g(g_i)$ as the fraction of its neighbors in the genotype network that are neutral neighbors. This measures the probability of retaining g_i's phenotype after applying single-point mutations, and thus captures the resilience of the genotype g_i in the face of mutational perturbations.

At the phenotypic level, we define the *phenotypic robustness* $R_p(p_i)$ of a phenotype p_i as the total number of its underlying genotypes. The indication is that the more a phenotype is represented among genotypes, the more robust it is. This phenotypic robustness is reflected as the vertex size in the phenotype network (see Figure 3b).

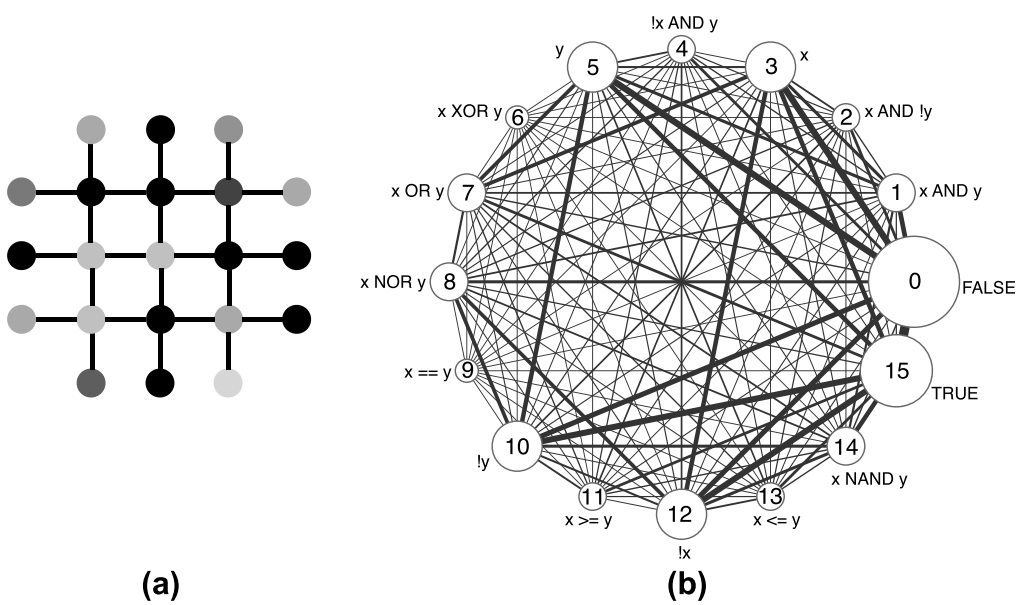

Figure 3. The genotype and phenotype networks. (a) In the genotype network, each vertex represents a LGP genotype, and its grayscale encodes its phenotype. Two vertices are connected by an edge if the two genotypes can be transformed from one to another by a single-point mutation. Two connected vertices are neutral neighbors if the single-point mutation does not change the phenotype; otherwise, they are non-neutral neighbors. (b) In the phenotype network, 16 vertices are the 16 distinguishing phenotypes of the two-input, one-output Boolean system. Phenotypes are numbered from their four-digit binary string and are labeled using the Boolean relationships they represent. The size of a vertex is proportional to its phenotypic robustness, and the edge thickness is proportional to the total number of mutational connections between the two phenotypes.

Downloaded from http://direct.mit.edu/artl/article-pdf/20/4/457/1664703/artl_a_00145.pdf by guest on 26 September 2021

2.5 Measuring Epistasis for Genotypes

In general, the mutational effects on phenotypes are expected to be positively correlated with the total number of mutations. When mutations co-occur, their total phenotypic effect depends on the interactions between them. Without any interactions, mutations are regarded as independent and act on a phenotype individually. A departure from independence could lead to either reinforcing or mitigating the accumulation of mutational effects on a phenotype. Epistasis describes this interaction effect, or the departure from independence.

It is commonly estimated in the evolutionary biology literature that the average fitness at a mutational distance k from a given reference genotype decays approximately exponentially with increasing k [10, 13, 27, 54]. We adopt this estimate in our LGP system. Let's define the genotypic distance D_g as the total number of differing loci between two genotypes. The phenotypic distance D_p is defined in our example as the Hamming distance between two Boolean relationships, or phenotypes, which come in the form of four-digit binary strings (see Figure 2(b)). For instance, the phenotypic distance between phenotypes `FALSE` (0000) and `AND` (0001) is 1. The average fitness w, calculated as $(4 - D_p)/4$ for our LGP system, of a mutational genotypic distance $k = D_g$ is modeled by the relationship

$$w(k) = e^{-\alpha k^{\beta}}, \tag{1}$$

where α and β are constant parameters. This relationship can also be written as

$$\log[-\log(w(k))] = \log(\alpha) + \beta\log(k). \tag{2}$$

The parameter β captures the extent to which the accumulation of multiple mutations amplifies or mitigates individual mutational effects, and expresses the dependence among multiple mutations, that is, epistasis. The value of β indicates the direction of epistasis. If $\beta > 1$, we observe negative (synergistic) epistasis, where multiple mutations amplify individual effects on the phenotypic level. If $\beta < 1$, we observe positive (antagonistic) epistasis, where multiple mutations cancel out each others' harmful phenotypic effects. If multiple mutations are independent of each other, then $\beta = 1$. In general, a larger value of β implies a stronger synergistic interaction among mutations. In our implementation of this epistasis measure, both w and k are directly observable, and we use linear regression on Equation 2 to estimate β.

2.6 Population Evolution

In addition to characterizing the static genotype and phenotype spaces of this LGP system, we perform a dynamic study by using a population to investigate the interplay between mutation and recombination and the population dynamics during evolution. As the target phenotype we choose one of the least-represented phenotypes, `XOR`, in order to allow evolution to proceed over a longer time.

A non-overlapping generational evolution model with fixed population size $|P|$ is used in our simulation. First we initialize a population of LGP programs of fixed length $L = 4$ for a given starting phenotype by randomly sampling $|P|$ genotypes underlying the starting phenotype. This allows us to focus on a finite genotype space for our quantitative study of the genotype and phenotype networks, and to look into the influence of the initial phenotype's properties on evolution. Then new generations of offspring are produced sequentially. In the mutation-only scenario, we randomly choose an individual with replacement, mutate according to a certain mutation probability, and place it into the next generation. This process is repeated $|P|$ times until the next generation of the population is complete. When both mutation and recombination are applied, for each generation, we randomly choose two individuals with replacement, perform a crossover based on a certain recombination probability by randomly choosing a recombination point, mutate them with a certain mutation probability, and place both offspring into the next generation. This is repeated $\frac{|P|}{2}$ times until the next generation of the

Downloaded from http://direct.mit.edu/artl/article-pdf/20/4/457/1664703/artl_a_00145.pdf by guest on 26 September 2021

population is complete. The evolutionary process is terminated once the target phenotype is reached, that is, at least one genotype from the target phenotype is found.

3 Results

3.1 Epistasis and Genotype Complexity

To model the relation of the mutational effect $w(k)$ to the number of mutations k for a given genotype, we applied consecutive five-step point mutations (i.e., $k = 1, 2, \ldots, 5$) to the genotype, and recorded the phenotypic distance from the original reference phenotype after each mutational step. Recall that the phenotypic distance is the Hamming distance between two phenotypes in the form of four-digit binary strings. The value of $w(k)$ ranges from 0 to 1. We repeated this measurement 100 times and obtained the average values to model $w(k)$, $k = 1, 2, \ldots, 5$. Then the epistasis indicator β was estimated by applying a linear regression to $\log[-\log(w(k))]$ and $\log(k)$ (Equation 2).

By randomly sampling genotypes from the entire genotype space, we observed epistasis with all three possible directions: $\beta > 1$, $\beta = 1$, and $\beta < 1$. This indicates the variation of epistasis in our LGP program system. It is interesting to see this diversity in such a simple artificial evolutionary system.

We also looked into the distribution of epistasis for programs with various lengths. From length $L = 3$ to $L = 15$, we randomly sampled 10,000 genotypes of each length and depicted their distributions of epistasis in Figure 4. Most genotypic epistasis observed in our LGP system is antagonistic ($\beta < 1$); however, as the genome complexity increases, this antagonistic epistasis is gradually alleviated and changes its direction to synergistic ($R^2 = 0.978$, $p = 1.119 \times 10^{-10}$). This observation is consistent with findings in a digital organism system [33]. Note that although different from many previously discussed biological models where genotypic epistasis is mostly detected as synergistic, our observation is consistent with others' on the general trend towards a stronger synergistic effect as the genome complexity increases.

3.2 Genotypic Robustness and Epistasis within Different Phenotypes

The notion of phenotype networks (Figure 3b) is a very useful vehicle for characterizing various properties of different phenotypes and their intertwined connections. Besides the heterogeneity of the phenotypic robustness and the mutational connectivity of phenotypes to each other, it is useful to look into the distribution of genotypic properties among different phenotypes.

In order to retain a finite genotype space, we have set a fixed length of $L = 4$ for all genotypes throughout all our simulations. For each of the 16 phenotypes, we enumerated all possible genotypes and calculated their genotypic robustness and epistasis. Figure 5 shows the distributions of genotypic

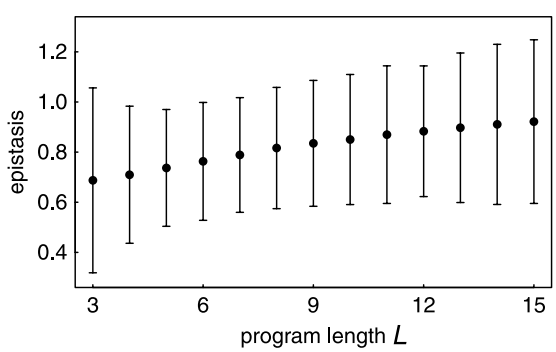

Figure 4. Average genotype epistasis for LGP programs with increasing lengths. Epistasis is calculated and averaged over 10,000 randomly sampled genotypes for each of the given program lengths L. The points represent means, and the error bars show the standard deviations.

Downloaded from http://direct.mit.edu/artl/article-pdf/20/4/457/1664703/artl_a_00145.pdf by guest on 26 September 2021

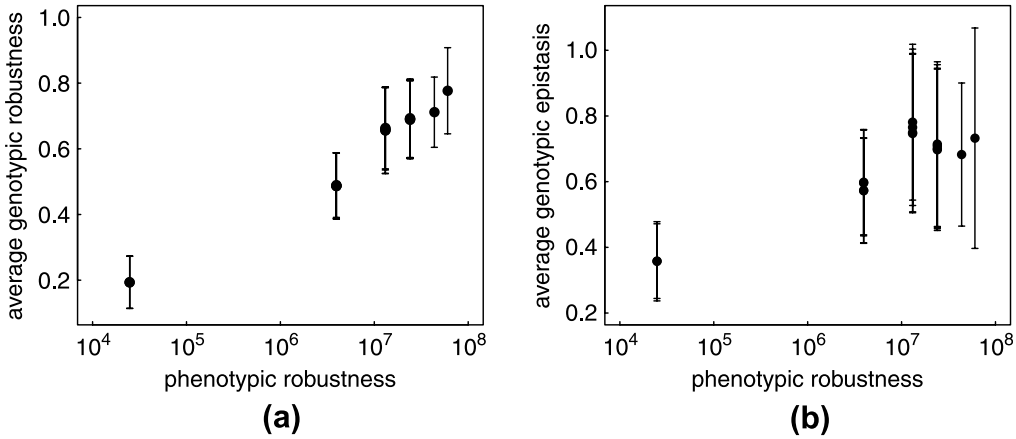

Figure 5. Average genotypic (a) robustness and (b) epistasis as a function of phenotypic robustness. The data in both plots correspond to the average of all genotypes within a given phenotype, and the error bars denote their standard deviations. The scales of measurement on both axes were chosen and displayed according to their best-fitting correlations.

robustness and epistasis as a function of phenotypic robustness. Both genotypic robustness ($R^2 = 0.958, p = 3.308 \times 10^{-11}$) and epistasis ($R^2 = 0.851, p = 2.206 \times 10^{-7}$) are positively correlated with their corresponding phenotypes' robustness. This indicates that more robust phenotypes are also composed of more robust genotypes, and that genotypes of more robust phenotypes tend to have weaker antagonistic epistasis.

3.3 Recombination Has Less Destructive Effects on Phenotype

Recombination changes a genome in a very different way than mutation, by rearranging already existing genetic components. To compare the effects of recombination and mutation on the phenotype is helpful because it gives us insight into how differently these two major genetic variation operators behave. We randomly sampled 1% of the genotypes (≈2.7 million) from the entire finite genotype space, regardless of their phenotypes, and applied mutation and recombination to observe the phenotypic outcome.

For mutation, we applied m-point mutations with m ranging from 1 to 15 to a genotype and repeated this 100 times for each genotype. We recorded the average phenotypic outcome—that is, the phenotypic distance from the original phenotype, which allows us to uncover the relation between genotypic distance m and phenotypic result. For recombination, we chose two parent programs from the genotype pool with a uniform probability, and applied single-point crossover at a randomly picked crossover point. We repeated this 10,000 times and recorded phenotypic and genotypic distances between all four parent-offspring pairs to examine average phenotypic effect related to the parent-offspring genotypic distance.

Figure 6 shows (a) the average percentage of retained phenotypes and (b) the average phenotypic distance between a parent and an offspring if the phenotype is changed, either by mutation or by recombination. It can be observed that, for the same amount of genotypic change, recombination is more likely to retain a phenotype than mutation. In addition, whenever a genetic variation operator changes the phenotype, recombination results in less phenotypic variation than does mutation. The same findings were observed with the setting that recombination was restrained on parent programs from the same phenotype (data not shown).

3.4 Recombination Accelerates Evolution and Elevates Robust Phenotypes

The simulation of evolutionary runs we report here was designed to help understand population dynamics changes produced by recombination relative to those produced by mutation. In our simulation, the population size is set to $|P| = 100$, and for each experimental setting we performed

Downloaded from http://direct.mit.edu/artl/article-pdf/20/4/457/1664703/artl_a_00145.pdf by guest on 26 September 2021

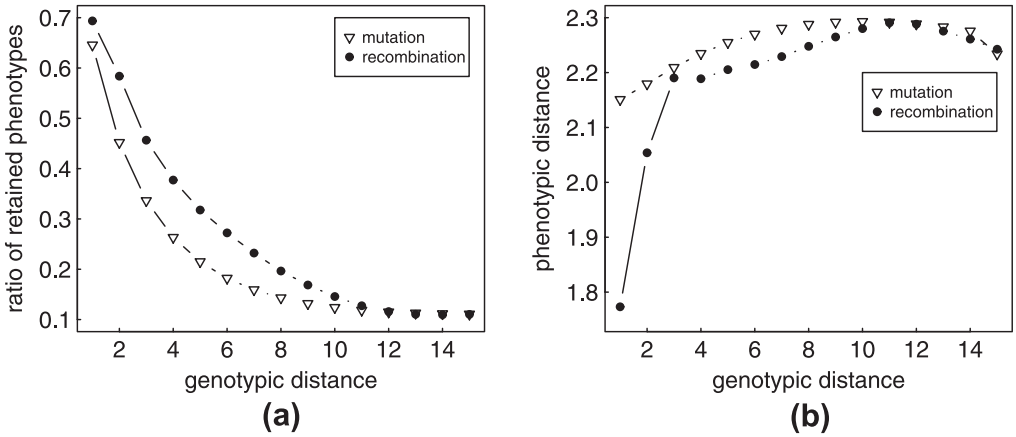

Figure 6. Comparison of mutational and recombinational effects. (a) The ratio of retained phenotypes after applying mutation/recombination to a genotype as a function of the genotypic variation resulted by mutation/recombination. (b) For mutation/recombination events that change the phenotypic outcome, we record the phenotypic variation as a function of the genotypic variation. In both plots, data points are averaged values over a randomly sampled 1% of the entire genotype space: ≈2.7 million sampled genotypes.

10,000 runs to provide a sufficient statistical support for observations. Following the convention on parameter configurations in the evolutionary computation literature, the mutation probability is set to 0.1 and the recombination probability is set to 1. We compared two evolutionary scenarios: Scenario 1 is driven solely by mutation, whereas scenario 2 is driven by both mutation and recombination. First we examined whether recombination helps find the target of evolution faster. Figure 7a shows the evolution time necessary to reach the target with regard to the robustness of the starting phenotype. For any starting phenotype, applying both mutation and recombination reduces the time to find the target significantly.

Second, we investigated whether the acceleration effect of recombination varies with different starting phenotypes. Figure 7b shows the reduced evolution time, obtained by taking the difference of the total evolution times with and without recombination, as a function of the robustness

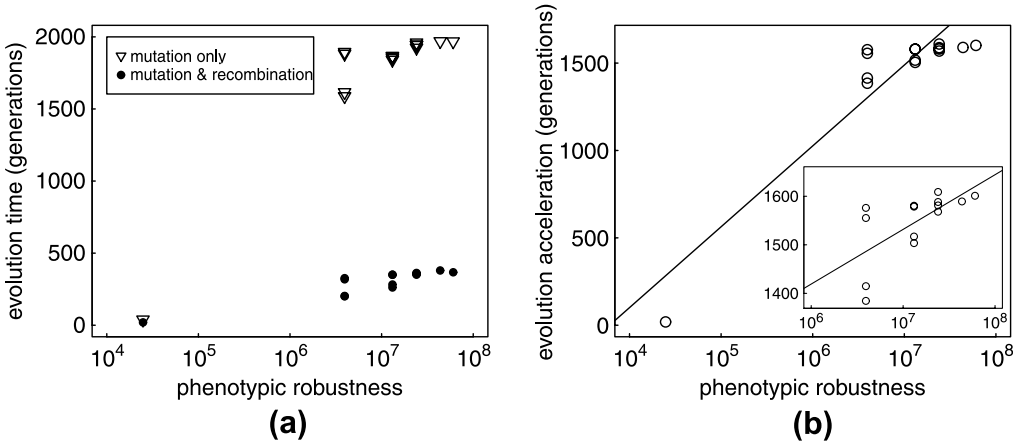

Figure 7. Comparison of evolution time for two evolution scenarios: with mutation only and with both mutation and recombination. (a) The total required time to reach the target as a function of the starting phenotype's robustness. (b) Evolution acceleration (i.e., reduced evolution time) by adding recombination as a function of the starting phenotype's robustness. The straight line is the best linear-log fit to the data and provides a guide for the eye. The inset shows the correlation for starting phenotypes on the upper right corner, that is, excluding the one data point with the lowest robustness.

Downloaded from http://direct.mit.edu/artl/article-pdf/20/4/457/1664703/artl_a_00145.pdf by guest on 26 September 2021

of the starting phenotype. We observed a clear positive correlation ($R^2 = 0.830$, $p = 1.447 \times 10^{-6}$), which indicates that the acceleration effect of recombination is more prominent when a population starts from a more robust phenotype. Due to the large heterogeneity of the distribution of phenotypic robustness, shown in the gap between the lower left single data point (phenotype EQUAL) and the upper right cluster of other data points, the inset zooms in and depicts the correlation of the reduced evolution time and phenotypic robustness for only the upper right cluster of data points (Figure 7b inset). A significant positive correlation is still observable ($R^2 = 0.428$, $p = 0.011$).

3.5 Recombination Evolves Genotypic Robustness and Epistasis

Since we have seen now that recombination accelerates evolution and particularly promotes robust phenotypes, it is intriguing to see whether recombination has any influences on genotypic properties. To investigate this, we evaluated all genotypes of the population at the end of evolution and checked if they differ between the two evolutionary scenarios. We calculated the average genotypic robustness evolved in the final population, and examined its relation to the starting phenotypic robustness (Figure 8a). We can see that populations where both recombination and mutation were applied consist of more robust genotypes at the end of evolution, and this observation is persistent with all starting phenotypes.

Figure 8b shows a comparison of average genotypic epistasis at the end of the runs with and without applying recombination. It can be stated that no matter which phenotype was used to start from, applying recombination with mutation led to an evolved population with higher genotypic epistasis. In other words, recombination selects for genotypes with decreasing antagonistic epistasis during evolution.

To further investigate the population dynamics with and without recombination, we use another two sets of runs, again applying either mutation only or both mutation and recombination, where we let the populations evolve for a relatively long period (100,000 generations), and check if populations reach an equilibrium status and how this status correlates with the starting phenotype's properties. Figure 9 shows the equilibrium robustness and epistasis as a function of the starting robustness and epistasis. Equilibrium robustness (epistasis) is calculated as the average genotypic robustness (epistasis) across all 100,000 generations. Given the long evolution time, such a calculation provides a good estimate of the population dynamics in a stable equilibrium state. As seen in the figure, applying mutation alone leads a population to the same average genotypic robustness and epistasis, independent of the starting phenotype. However, applying both mutation and recombination seems to

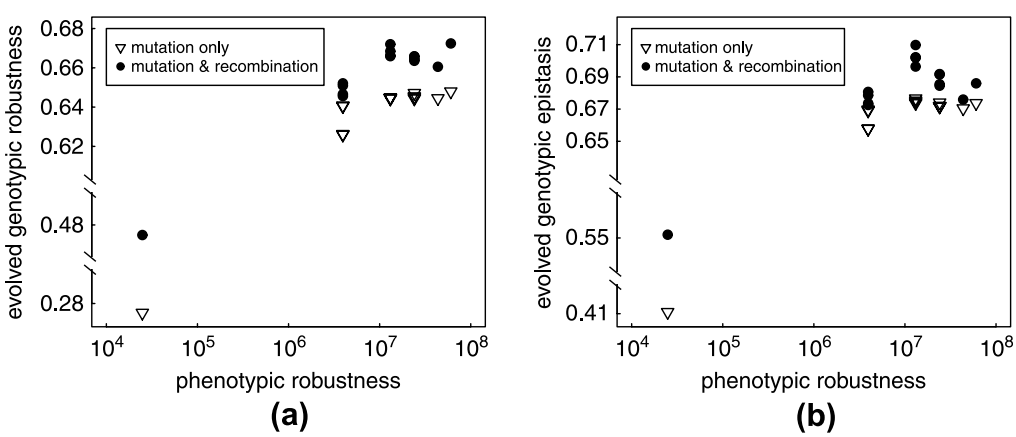

Figure 8. Recombination effects on evolving genotypic robustness and epistasis. Average genotypic (a) robustness and (b) epistasis in the final population after evolution by applying mutation only (triangles) and both mutation and recombination (solid circles) for different starting phenotypes (*x* axis).

Downloaded from http://direct.mit.edu/artl/article-pdf/20/4/457/1664703/artl_a_00145.pdf by guest on 26 September 2021

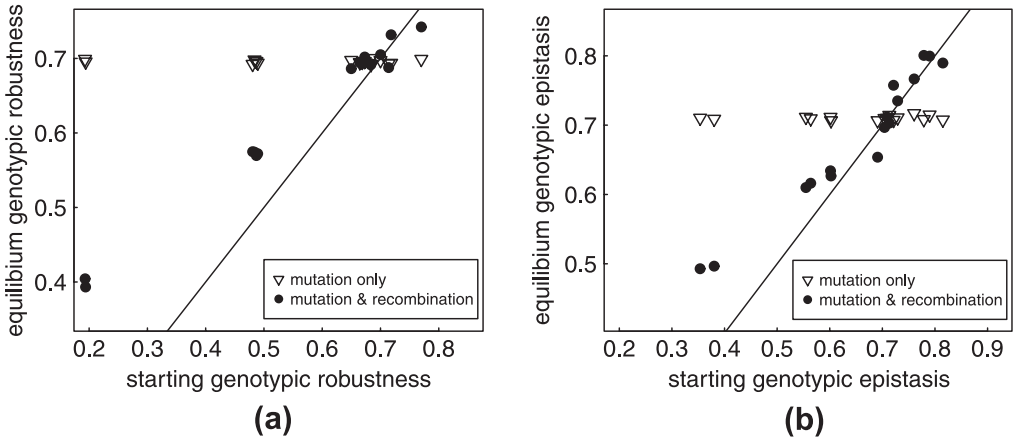

Figure 9. Recombination effects on genotypic robustness and epistasis on the long term of evolution. Equilibrium genotypic (a) robustness and (b) epistasis as a function of starting phenotypic properties. Lines represent function $y = x$ and are to provide a visual guide.

retain the starting phenotypic properties while generally improving robustness and epistasis over (long) evolutionary runs (data points stay mostly on the left side of the line $y = x$).

4 Discussion

Utilizing a genetic programming system, this study has described the general role that recombination plays in evolution, specifically in relation to accelerating the search for novel phenotypes and promoting genetic robustness and epistasis.

The LGP system used in this study has a fixed-length genome representation to define finite and compact genotype and phenotype spaces. This enabled us to exhaustively enumerate all genotypes and phenotypes and to characterize their intertwined connections using networks. It also allowed us to separate other possible elements from the analysis of recombinational effects on evolution. For instance, we showed here that the increasing genome complexity (i.e., code growth) accompanies a change in the directional genotypic epistasis (Figure 4). This indicates that more complex genomes might be better at alleviating the antagonistic epistasis and thus impose a stronger pressure removing deleterious mutations.

Using genotype and phenotype networks, it was shown that robustness is distributed unevenly among both genotypes and phenotypes (Figure 3). This nonuniformity is a pervasive feature of natural and artificial evolutionary systems, and calls for a quantitative analysis of the genotype-to-phenotype map and robustness [7, 21]. Our particular LGP system exhibits a positive correlation between genotypic and phenotypic robustness (Figure 5a); in other words, it supports the intuitive observation that more robust phenotypes are composed of more robust genotypes. In addition, more robust phenotypes are also composed of genotypes that have weaker antagonistic epistasis and stronger synergistic epistasis (Figure 5b). This corroborates the conclusion drawn from a previous experimental study on yeast [48] that the robustness to mutations is not due to genetic redundancy but due to epistatic interactions between genes.

Recombination changes the genetic content of an individual in quite a different manner from mutation, by reusing and rearranging existing units of genomes. This leads to different effects on phenotypes. Compared to mutation, the same amount of genotypic variation resulting from recombination has a higher probability of retaining the phenotype or, if the phenotype is changed, does not move away as much from the reference phenotype (Figure 6). This could be the result of the fact that recombination preserves more well-adapted modules by reorganizing existing components and has less disruptive effects than mutation does [52].

Downloaded from http://direct.mit.edu/artl/article-pdf/20/4/457/1664703/artl_a_00145.pdf by guest on 26 September 2021

Through evolution experiments, we found that applying recombination in addition to mutation significantly accelerates the search for novel phenotypes (Figure 7). More interestingly, this acceleration by recombination is more significant when a population starts from a more robust phenotype. It is expected that more robust phenotypes have a wider-spreading neutral genotypic network, and recombination is very effective in utilizing those vast cryptic genetic variations for further phenotypic innovations. Having both recombination and mutation significantly enhances the ability of evolutionary systems to explore novel phenotypes over having mutation alone. This could also contribute to an explanation of why recombination evolved in the first place in more complex organisms: Recombination benefits and enhances the evolvability of more complex evolutionary systems.

Recombination not only benefits more robust phenotypes, but also produces genotypic robustness and epistasis (Figure 8). A population subject to both recombination and mutation moves towards genotypic regions with higher robustness and epistasis than a population subject to mutation alone. Recombination evolving stronger synergistic epistasis suggests that it imposes a strong selection pressure on purging deleterious mutation, a plausible explanation for the maintenance of recombination in sexual organisms. Moreover, previous studies report that the evolution of mutational robustness is usually accompanied by the evolution of epistasis [1, 43]. These two were also found associated in our study. Mutational robustness is evolved in response to stabilizing adaptive organisms in the face of constant environmental and genetic perturbations. The key mechanism supporting this robustness is not the redundancy in mapping genotypes to phenotypes (traits), but the increased interactions among genetic components and the underlying complex genetic architecture. More interestingly, using a long-term evolution simulation, we observed that applying both recombination and mutation preserves the starting phenotypic properties, in contrast with applying mutation alone, which led to a loss of memory of the starting phenotypes (Figure 9). This result is intriguing but requires further study.

In summary, using a simple, yet representative, artificial evolutionary model, this study demonstrates the effects of recombination in evolution. The findings of our study coincide with many observations from evolutionary biology that use model organisms or simulated gene regulatory circuits. Our study helps to elucidate the general effects of recombination in evolution. We hope that it can inform the future design of genetic operators in artificial evolutionary systems by providing a theoretical foundation to better understand evolutionary mechanisms.

This study also opens the door to several future research avenues. First, our analysis can be extended to both larger-scale LGP systems and alternative evolutionary computation systems. This could help generalize the findings we draw in the current study and support their application to a larger range of problem instances. Second, larger-scale evolutionary systems will generate more complex genotype and phenotype networks, and network topology analysis could then be adopted to explore the dynamics of evolutionary populations. Varying network topology could also influence the evolutionary effects of recombination. In particular, our LGP example has a very special complete phenotype network, and looking into alternative systems without this property, perhaps better resembling real biological systems, is important for our future explorations.

Acknowledgments

This work was supported by National Institute of Health (USA) grants R01-LM009012, R01-LM010098, and R01-AI59694 to J.H.M.; W.B. acknowledges support from National Science and Engineering Research Council (Canada) Discovery Grants, under RGPIN 283304-2012.

References

1. Azevedo, R. B., Lohaus, R., Srinivasan, S., Dang, K. K., & Burch, C. L. (2006). Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. *Nature*, *440*(2), 87–90.

2. Banzhaf, W., & Langdon, W. B. (2002). Some considerations on the reason for bloat. *Genetic Programming and Evolvable Machines*, *3*(1), 81–91.

3. Birky, C. W., Jr. (1996). Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. *Genetics*, *144*, 427–437.

Downloaded from http://direct.mit.edu/artl/article-pdf/20/4/457/1664703/artl_a_00145.pdf by guest on 26 September 2021

4. Brameier, M. F., & Banzhaf, W. (2007). *Linear genetic programming*. New York: Springer.

5. Burch, C. L., & Chao, L. (2004). Epistasis and its relationship to canalization in the RNA virus ϕ6. *Genetics*, *167*, 559–567.

6. Cordell, H. J. (2002). Epistasis: What it means, what it doesn't mean, and statistical methods to detect it in humans. *Human Molecular Genetics*, *11*(20), 2463–2468.

7. Cowperthwaite, M. C., Economo, E. P., Harcombe, W. R., Miller, E. L., & Meyers, L. A. (2008). The ascent of the abundant: How mutational networks constrain evolution. *PLoS Computational Biology*, *4*(7), e1000110.

8. de Visser, J. A. G. M., & Elena, S. F. (2007). The evolution of sex: Empirical insights into the roles of epistasis and drift. *Nature Review Genetics*, *8*, 139–149.

9. de Visser, J. A. G. M., Hermission, J., Wagner, G. P., Meyers, L. A., Bagheri-Chaichian, H., et al. (2003). Evolution and detection of genetic robustness. *Evolution*, *57*(9), 1959–1972.

10. de Visser, J. A. G. M., Hoekstra, R. F., & van den Ende, H. (1997). Test of interaction between genetic markers that affect fitness in Aspergillus niger. *Evolution*, *51*(5), 1499–1505.

11. Draghi, J. A., Parsons, T. L., Wagner, G. P., & Plotkin, J. B. (2010). Mutational robustness can facilitate adaptation. *Nature*, *463*, 353–355.

12. Drummond, D. A., Silberg, J. J., Meyer, M. M., Wilke, C. O., & Arnold, F. H. (2005). On the conservative nature of intragenic recombination. *Proceedings of the National Academy of Sciences of the U.S.A.*, *102*(15), 5380–5385.

13. Elena, S. F., & Lenski, R. E. (1997). Test of synergistic interactions among deleterious mutations in bacteria. *Nature*, *390*, 395–398.

14. Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford, UK: Clarendon Press.

15. Francone, F. D., Conrads, M., Banzhaf, W., & Nordin, P. (1999). Homologous crossover in genetic programming. In W. Banzhaf, J. M. Daida, A. E. Eiben, M. H. Garzon, V. Honavar, M. J. Jakiela, & R. E. Smith (Eds.), *Proceedings of the Genetic and Evolutionary Computation Conference* (pp. 1021–1026).

16. Gardner, A., & Kalinka, A. T. (2006). Recombination and the evolution of mutational robustness. *Journal of Theoretical Biology*, *241*, 707–715.

17. Hansen, J. V. (2003). Genetic programming experiments with standard and homologous crossover methods. *Genetic Programming and Evolvable Machines*, *4*, 53–66.

18. Holland, J. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, MI: University of Michigan Press.

19. Hu, T., & Banzhaf, W. (2009). Neutrality and variability: Two sides of evolvability in linear genetic programming. In F. Rothlauf (Eds.), *Proceedings of the Genetic and Evolutionary Computation Conference* (pp. 963–970).

20. Hu, T., & Banzhaf, W. (2010). Evolvability and speed of evolutionary algorithms in light of recent developments in biology. *Journal of Artificial Evolution and Applications*, *2010*, 568375.

21. Hu, T., Payne, J. L., Banzhaf, W., & Moore, J. H. (2012). Evolutionary dynamics on multiple scales: A quantitative analysis of the interplay between genotype, phenotype, and fitness in linear genetic programming. *Genetic Programming and Evolvable Machines*, *13*, 305–337.

22. Jansen, T., & Wegener, I. (1999). On the analysis of evolutionary algorithms—a proof that crossover really can help. In *Proceedings of the 7th Annual European Symposium on Algorithms* (pp. 184–193). Springer.

23. Kauffman, S. A. (2000). *Investigations*. New York: Oxford University Press.

24. Kondrashov, A. S. (1998). Deleterious mutations and the evolution of sexual reproduction. *Nature*, *336*, 435–440.

25. Koza, J. R. (1992). *Genetic programming: On the programming of computers by means of natural selection*. Cambridge, MA: MIT Press.

26. Landry, C. R., Lemos, B., Rifkin, S. A., Dickinson, W. J., & Hartl, D. L. (2007). Genetic properties influencing the evolvability of gene expression. *Science*, *317*, 118–121.

27. Lenski, R. E., Ofria, C., Collier, T. C., & Adami, C. (1999). Genome complexity, robustness and genetic interactions in digital organisms. *Nature*, *400*, 661–664.

28. Lohaus, R., Burch, C. L., & Azevedo, R. B. (2010). Genetic architecture and the evolution of sex. *Journal of Heredity*, *101*, S142–S157.

Downloaded from http://direct.mit.edu/artl/article-pdf/20/4/457/1664703/artl_a_00145.pdf by guest on 26 September 2021

29. MacCarthy, T., & Bergman, A. (2007). Coevolution of robustness, epistasis, and recombination favors asexual reproduction. *Proceedings of the National Academy of Sciences of the U.S.A.*, *104*(31), 12801–12806.

30. Martin, O. C., & Wagner, A. (2009). Effects of recombination on complex regulatory circuits. *Genetics*, *183*, 673–684.

31. Masel, J., & Siegal, M. L. (2009). Robustness: Mechanisms and consequences. *Trends in Genetics*, *25*(9), 395–403.

32. McBride, R. C., Ogbunugafor, C. B., & Turner, P. E. (2008). Robustness promotes evolvability of thermotolerance in an RNA virus. *BMC Evolutionary Biology*, *8*, 231.

33. Misevic, D., Ofria, C., & Lenski, R. E. (2006). Sexual reproduction reshapes the genetic architecture of digital organisms. *Proceedings of The Royal Society B*, *273*(1585), 457–464.

34. Moore, J. H. (2005). A global view of epistasis. *Nature Genetics*, *37*(1), 13–14.

35. Moore, J. H., & Williams, S. M. (2005). Traversing the conceptual divide between biological and statistical epistasis: Systems biology and a more modern synthesis. *BioEssays*, *27*(6), 637–646.

36. Muller, H. J. (1932). Some genetic aspects of sex. *The American Naturalist*, *66*, 118–138.

37. Otto, S. P., & Feldman, M. W. (1997). Deleterious mutations, variable epistatic interactions, and the evolution of recombination. *Theoretical Population Biology*, *51*, 134–147.

38. Otto, S. P., & Lenormand, T. (2002). Resolving the paradox of sex and recombination. *Nature Review Genetics*, *3*, 252–261.

39. Phillips, P. C. (1998). The language of gene interaction. *Genetics*, *149*, 1167–1171.

40. Phillips, P. C. (2008). Epistasis—The essential role of gene interactions in the structure and evolution of genetic systems. *Nature Review Genetics*, *9*, 855–867.

41. Platel, M. D., Clergue, M., & Collard, P. (2003). Maximum homologous crossover for linear genetic programming. In C. Ryan, T. Soule, M. Keijzer, E. Tsang, R. Poli, & E. Costa (Eds.), *Proceedings of the European Conference on Genetic Programming* (pp. 194–203).

42. Poli, R., & Langdon, W. B. (1997). Genetic programming with one-point crossover and point mutation. In *Soft Computing in Engineering Design and Manufacturing* (pp. 180–189). Springer-Verlag.

43. Siegal, M. L., & Bergman, A. (2002). Waddington's canalization revisited: Developmental stability and evolution. *Proceedings of the National Academy of Sciences of the U.S.A.*, *99*(16), 10528–10532.

44. Soule, T., & Heckendorn, R. B. (2002). An analysis of the causes of code growth in genetic programming. *Genetic Programming and Evolvable Machines*, *3*, 283–309.

45. Spears, W. M. (1995). Adapting crossover in evolutionary algorithms. In *Proceedings of the Fourth Annual Conference on Evolutionary Programming* (pp. 367–384).

46. Szollosi, G. J., & Derenyi, I. (2008). The effect of recombination on the neutral evolution of genetic robustness. *Mathematical Biosciences*, *214*, 58–62.

47. Templeton, A. R. (2000). Epistasis and complex traits. In J. B. Wolf, E. D. Brodie, & M. J. Wade (Eds.), *Epistasis and the evolutionary process* (pp. 41–57). New York: Oxford University Press.

48. Wagner, A. (2000). Robustness against mutations in genetic networks of yeast. *Nature Genetics*, *24*, 355–361.

49. Wagner, A. (2007). *Robustness and evolvability in living systems*. Princeton, NJ: Princeton University Press.

50. Wagner, A. (2008). Neutralism and selectionism: A network-based reconciliation. *Nature Review Genetics*, *9*, 965–974.

51. Wagner, A. (2008). Robustness and evolvability: A paradox resolved. *Proceedings of The Royal Society B*, *275*(1630), 91–100.

52. Wagner, A. (2011). The low cost of recombination in creating novel phenotypes. *BioEssays*, *33*(8), 636–646.

53. Wilke, C. O. (2001). Adaptive evolution on neutral networks. *Bulletin of Mathematical Biology*, *63*, 715–730.

54. Wilke, C. O., & Adami, C. (2001). Interaction between directional epistasis and average mutational effects. *Proceedings of the Royal Society B*, *268*(1475), 1469–1474.

55. Xia, Y., & Levitt, M. (2002). Roles of mutation and recombination in the evolution of protein thermodynamics. *Proceedings of the National Academy of Sciences of the U.S.A.*, *99*(16), 10382–10387.

Downloaded from http://direct.mit.edu/artl/article-pdf/20/4/457/1664703/artl_a_00145.pdf by guest on 26 September 2021