

Analysis of Sexual vs Asexual Reproduction Using Four Measurements

Feng Lin,^{1*} Robert J. Lin²

¹Department of ECE, Wayne State University,
Detroit, MI 48202, USA

²Lin Group, Houston, TX 77098, USA

*To whom correspondence should be addressed; E-mail: flin@wayne.edu.

The prevalence of sexual reproduction has long been an outstanding problem of evolutionary biology. In accordance with the mathematical approach employed by past researchers, we propose a mathematical framework to address this problem. We define and derive four measurements, diversity measure DVM, diffusion measure DFM, optimality measure OPM, and survivability measure SVM to compare sexual reproduction with asexual reproduction. We show that DVM increases exponentially in sexual reproduction, while only linearly in asexual reproduction. Hence, sexual reproduction allows species more opportunity to adapt. We also show that DFM is bounded in sexual reproduction and OPM is inversely related to DFM. Thus, sexual reproduction leads to smaller DFM and hence a larger OPM. We further show that SVM is a monotonic increasing function of OPM. Hence, sexual reproduction is better by virtue of producing a more homogeneous population.

Introduction

Why most species reproduce sexually is a question investigated extensively, but not yet answered satisfactorily in biology (1–4). The prevalence of sexual reproduction suggests that there are major benefits provided by this mode of reproduction. The benefits, however, are not well-understood. In many ways, asexual reproduction seems to be a better evolutionary strategy: only one parent is required, and all of the parent’s genes are passed on to its progeny (5–12). In a sexual population, the males are unable to produce progeny of their own and females only transfer half their genes to progeny, hence the so called problem of the “two-fold cost of males.” Furthermore, sexual reproduction must also overcome obstacles that do not exist in asexual reproduction. Sexually reproducing organisms must spend a great deal of time and energy to find and attract mates. Also, copulation in sexually reproducing organisms leaves both organisms vulnerable to predation.

Despite these and other major drawbacks to sexual reproduction, it remains a very prevalent form of reproduction in most higher order organisms. Biologists have put forth numerous hypotheses for why sexual reproduction is so prevalent. The main hypotheses to explain sexual reproduction typically focus on the benefits of the inherent ability of sexual reproduction to recombine and shuffle genetic information (13–20). Generally, these hypotheses have relied on fusing observational zoological studies with our understanding of genetics to build their foundations; a very understandable approach given the inherent methodological constraints on the field of evolutionary biology

Whereas hypotheses in applied physics or chemistry can be directly tested in experiments with controlled variables in real-time, evolutionary biology deals with “evolutionary time” where the timescales necessary for experimentation far exceed the human lifespan, or human civilization itself for that matter. Such controlled experiments are largely unfeasible for the field

of evolutionary biology. As such, even Darwin was forced to use the inductive reasoning approach of accumulating large reservoirs of circumstantial evidence in combination with critical thinking to propound his ideas (21). The field of evolutionary biology has largely inherited this inductive reasoning methodology.

With the dearth of ability to conduct concrete controlled experiments to prove evolutionary principles by a deductive approach, we believe that turning back to a mathematical approach, as done by such past luminaries as JBS Haldane (22) or Kimura and Maruyama (23) can be valuable for addressing the problem of sexual reproduction. We believe that reproductive biology generally follows consistent principles that are amenable to mathematical modeling; as such, a mathematical framework can be used to provide further support for existing ideas, as well as illuminate novel ideas that could be overlooked if not for a deductive mathematical approach. With a mathematical methodology in mind, we introduce several mathematical constructs to analyze population changes via sexual reproduction or asexual reproduction. We first introduce a diversity measure DVM defined as the number of distinct individuals/genotypes in a population divided by the total number of individuals/genotypes in the population. We derive dynamic equations of DVM for both asexual reproduction and sexual reproduction. We show that DVM increases exponentially in sexual reproduction rather than linearly as in asexual reproduction. This exponential increase in sexual reproduction's diversity DVM is consistent with current hypotheses that posit genetic recombinatorial shuffling from sexual reproduction allows for more rapid adaptation.

We further introduce a diffusion measure DFM defined as the average “distance” of individuals in a population. Here, the distance between two individuals is defined as the percentage difference in their genes. Consequently, DFM measures how homogeneous a population is. We derive dynamic equations of DFM for both asexual reproduction and sexual reproduction. We show that, despite its diversity, sexually reproducing species are more homogeneous

(smaller DFM) than asexually reproducing species. In other words, sexual reproduction increases diversity (by increasing DVM) while decreases diffusion (by reducing DFM). This is less intuitive, but true.

We show that a more homogeneous population has evolutionary advantages over a less homogeneous population (smaller DFM is better). This is because the diffusion measure DFM is inversely related to optimality measure OPM , which measures the average distance of individuals in a population to the optimal individual/genotype who is most suitable in a given environment. More precisely, we show that $OPM = 1 - DFM$. Therefore, a smaller DFM implies a larger OPM . Hence, a more homogeneous population is more suitable for the environment. Since sexual reproduction leads to smaller DFM versus asexual reproduction, it has advantage over asexual reproduction, in addition to the advantage of having a larger DVM .

The last performance measure we define in this paper is the survivability measure SVM which is the percentage of individuals having progeny. We show that the survivability measure increases as the optimality measure increases (and that survivability measure decreases as the diffusion measure increases).

By formally defining four performance measures, DVM , DFM , OPM , and SVM , we develop a formal mathematical framework, in which advantages of sexual reproduction can be precisely addressed. All the above discussions can be verified by simulations of the dynamical equations for DVM , DFM , OPM , and SVM derived from basic biology.

Some theories/hypotheses that have been proposed to address the advantages of sexual reproduction over asexual reproduction, such as genetic recombination, Muller's Ratchet, and the Red Queen Hypothesis, can all be addressed by the mathematical framework using the performance measures proposed in this paper.

Diversity, Diffusion, Optimality, and Survivability Measures

In this section, we present the framework of our research. We define diversity measure, diffusion (or inhomogeneity) measure, optimality measure, and survivability measure to describe a species. We also define selective pressure to describe an environment. We show how survivability measure depends on diffusion measure, optimality measure, and selective pressure.

To do this formally and mathematically, let us first introduce the following notations.

N - the number of genes

$g_i, i \in \mathcal{N} = \{1, 2, \dots, N\}$ - the i -th gene

In asexual reproduction, a progeny inherits its genes from one progenitor.

In sexual reproduction, a progeny inherits its genes from two progenitors.

g_i^w - the i -th gene from the maternal progenitor.

g_i^d - the i -th gene from the paternal progenitor

A progeny will randomly pick the i -th gene from the maternal or paternal progenitor with equal probability ($=0.5$).

From one generation to the next, the probability of mutation of g_i is p .

A mutation causes $h\%$ change in g_i .

Change of mutation is additive: k mutations cause $kh\%$ change in g_i .

Q - the total population of a species.

$q \in Q$ - an individual/genotype in the species Q .

g_i^q - the i -th gene of $q \in Q$.

$|g_i^{q_1} - g_i^{q_2}|$ - distance between $g_i^{q_1}$ and $g_i^{q_2}$, defined as the percentage difference between $g_i^{q_1}$ and $g_i^{q_2}$.

If the distance between g_i^w and g_i^d exceeds a bound b , that is, $|g_i^w - g_i^d| > b$, then either the mating will not take place or the resultant progeny will not survive.

Diversity measure

Diversity measure describes how diverse a population is. We will argue that a more diverse population has an evolutionary advantage over a less diverse population, as more diversity increases the chance to adapt to the environment.

Diversity measure of a species Q is defined as the number of distinct individuals/genotypes in Q divided by the total number of individuals/genotypes in Q , that is,

$$DVM(Q) = \frac{\text{the number of distinct } q}{|Q|}, \quad (1)$$

where $|Q|$ is the number of elements (cardinality) of Q . Clearly, larger $DVM(Q)$ means that the species is more diverse.

Diffusion measure

Distance between $q_1, q_2 \in Q$ is defined as

$$DS(q_1, q_2) = \frac{1}{N} \sum_{i=1}^N |g_i^{q_1} - g_i^{q_2}|.$$

Diffusion measure of a species Q is defined as the average or expected value of $DS(q_1, q_2)$:

$$DFM(Q) = E(DS(q_1, q_2)) = \frac{1}{|Q|^2} \sum_{q_1, q_2 \in Q} DS(q_1, q_2), \quad (2)$$

where $E(\cdot)$ denotes the expectation (mean). Clearly, larger $DFM(Q)$ means that the species is less similar.

Optimality measure

In a given environment, the optimal individual is denoted by $o \in Q$, whose gene is denoted by $g_i^o, i \in \mathcal{N}$. The goal of evolution is for the population to evolve around o , that is, the distance of Q to o reduces as time passes.

The distance of Q from o is defined as the average or expected distance of individuals $q \in Q$ to o . One minus this distance is called the optimality measure OPM . Hence, optimality

measure of a species Q is one minus the expected value of $DS(q, o)$:

$$OPM(Q) = 1 - E(DS(q, o)) = 1 - \frac{1}{|Q|} \sum_{q \in Q} DS(q, o). \quad (3)$$

The reason to use $1 - E(DS(q, o))$ rather than $E(DS(q, o))$ as optimality measure is that, intuitively, we would like to see that the larger OPM is, the closer Q is to o .

As to be shown in Section V, optimality measure is related to diffusion measure as follows.

$$OPM(Q) \approx 1 - DFM(Q).$$

Selective pressure

For a given environment, its selective pressure is measured as a threshold $SP \in [0, 1]$ such that, if $DS(q, o) > 1 - SP$, then individual q will not have progeny. The larger SP is, the strong selective pressure is.

Survivability measure

In a given environment, survivability measure of a species Q is defined as the percentage of individuals having progeny, that is,

$$\begin{aligned} SVM(Q) &= \frac{\text{the number of } q \text{ having progeny}}{|Q|} \\ &= \text{percentage of } q \text{ such that } DS(q, o) \leq 1 - SP. \end{aligned} \quad (4)$$

Survivability measure depends on selective pressure, optimality measure, and diffusion measure as shown in the following theorem, whose proof is in Appendix.

Theorem 1

- (1) As the selective pressure SP increases, the survivability measure $SVM(Q)$ decreases.
- (2) As the optimality measure $OPM(Q)$ increases, the survivability measure $SVM(Q)$ increases.
- (3) As the diffusion measure $DFM(Q)$ increases, the survivability measure $SVM(Q)$ decreases.

In the next two sections, we derive the diversity and diffusion measures for both asexual reproduction and sexual reproduction. Without loss of generality, we assume that all individuals in Q are of the same generation. Denote the m -th generation of the species as Q_m . Since $DVM(Q)$, $DFM(Q)$, $OPM(Q)$, and $SVM(Q)$ are all defined as a percentage, we further assume that, without loss of generality, the number $|Q_m|$ of individuals in Q_m is unchanged from one generation to the next.

Asexual Reproduction

For asexual reproduction, each individual in the m -th generation $q' \in Q_m$ is produced by an individual in the $m-1$ -th generation $q \in Q_{m-1}$, denoted as $q \rightarrow q'$.

Diversity measure

For asexual reproduction, each mutation produces an individual who is different from its parent. The probability that this individual is distinct in Q is proportional to $1 - DVM(Q_{m-1})$. The expected number of mutations in one generation is given by $|Q| \times N \times p$. Hence,

$$\begin{aligned} DVM(Q_m) &= DVM(Q_{m-1}) + (1 - DVM(Q_{m-1})) \frac{|Q|Np}{|Q|} \\ &= (1 - Np)DVM(Q_{m-1}) + Np. \end{aligned} \quad (5)$$

It can be shown (Section V) that $DVM(Q_m) \rightarrow 100\%$ as $m \rightarrow \infty$. Assume that $DVM(Q_m)$ is small initially, then $1 - DVM(Q_{m-1}) \approx 1$ and the above equation can be approximated as

$$DVM(Q_m) \approx DVM(Q_{m-1}) + \frac{|Q|Np}{|Q|} = DVM(Q_{m-1}) + Np.$$

In other words, $DVM(Q_m)$ increases linearly.

Diffusion measure

To calculate $DS(q'_1, q'_2)$, where $q_1 \rightarrow q'_1$ and $q_2 \rightarrow q'_2$, let us consider their i -th genes, $g_i^{q'_1}$ and $g_i^{q'_2}$. Since mutations are independent with probability (w.p.) p , we have the following

$$g_i^{q'_j} = \begin{cases} g_i^{q_j} & \text{w.p. } 1 - p \\ g_i^{q_j} + \delta_j & \text{w.p. } p \end{cases}, j = 1, 2.$$

where δ is used to denote a mutation. Hence, the possible pairs $(g_i^{q'_1}, g_i^{q'_2})$ and probabilities of their occurrences are as follows.

$$(g_i^{q'_1}, g_i^{q'_2}) = \begin{cases} (g_i^{q_1}, g_i^{q_2}) & \text{w. p. } (1-p)(1-p) \\ (g_i^{q_1}, g_i^{q_2} + \delta_2) & \text{w. p. } (1-p)p \\ (g_i^{q_1} + \delta_1, g_i^{q_2}) & \text{w. p. } (1-p)p \\ (g_i^{q_1} + \delta_1, g_i^{q_2} + \delta_2) & \text{w. p. } p^2 \end{cases}$$

A mutation causes $h\%$ change. The probability that this mutation increases $|g_i^{q'_1} - g_i^{q'_2}|$ is proportional to $1 - |g_i^{q_1} - g_i^{q_2}|$. Thus,

$$|g_i^{q'_1} - g_i^{q'_2}| = \begin{cases} |g_i^{q_1} - g_i^{q_2}| & \text{w.p. } (1-p)^2 \\ |g_i^{q_1} - g_i^{q_2}| + (1 - |g_i^{q_1} - g_i^{q_2}|)h & \text{w.p. } 2p(1-p) \\ |g_i^{q_1} - g_i^{q_2}| + (1 - |g_i^{q_1} - g_i^{q_2}|)2h & \text{w.p. } p^2 \end{cases}$$

Hence¹,

$$E(|g_i^{q'_1} - g_i^{q'_2}|) = (1 - 2hp)E(|g_i^{q_1} - g_i^{q_2}|) + 2hp. \quad (6)$$

Therefore,

$$DFM(Q_m) = (1 - 2hp)DFM(Q_{m-1}) + 2hp. \quad (7)$$

It can be shown (Section V) that $DFM(Q_m) \rightarrow 100\%$ as $m \rightarrow \infty$. In other words, in asexual reproduction, individuals in a species will become less and less similar as time passes.

Sexual Reproduction

In sexual reproduction, each individual in the m -th generation $q' \in Q_m$ is produced by two individuals in the $m-1$ -th generation $q_w, q_d \in Q_{m-1}$, denoted as $q_w, q_d \rightarrow q'$.

Diversity measure

While in asexual reproduction, one new mutation will increase the number of distinct new individuals by 1, in sexual reproduction, one new mutation has the potential to double the

¹Derivations of all equation can be found in Appendix.

number of distinct new individuals due to genetic recombination. We assume that c percentage of genetic recombination will be accomplished. Since the expected number of mutations in one generation is $|Q| \times N \times p$, the expected number of new individuals due to genetic recombination is $|Q|cNp \times DVM(Q_{m-1})|Q|$. The probability that the new individuals are distinct in Q_m is proportional to $1 - DVM(Q_{m-1})$. Hence,

$$\begin{aligned} DVM(Q_m) &= DVM(Q_{m-1}) + (1 - DVM(Q_{m-1})) \frac{|Q|cNpDVM(Q_{m-1})|Q|}{|Q|} \\ &= (1 + (1 - DVM(Q_{m-1}))|Q|cNp)DVM(Q_{m-1}). \end{aligned} \quad (8)$$

It can be shown (Section V) that $DVM(Q_m) \rightarrow 100\%$ as $m \rightarrow \infty$. Assume that $DVM(Q_m)$ is small initially, then $1 - DVM(Q_{m-1}) \approx 1$ and the above equation can be approximated as

$$DVM(Q_m) \approx DVM(Q_{m-1}) + \frac{|Q|NpDVM(Q_{m-1})|Q|}{|Q|} = (1 + |Q|cNp)DVM(Q_{m-1}).$$

In other words, $DVM(Q_m)$ increases exponentially with respect to m . It is well-known that exponential increase is much faster than linear increase.

Diffusion measure

To calculate $DS(q'_1, q'_2)$, where $q_{1,f}, q_{1,d} \rightarrow q'_1$ and $q_{2,f}, q_{2,d} \rightarrow q'_2$, let us consider their i -th gene, $g_i^{q'_1}$ and $g_i^{q'_2}$. Since mutations are independent with probability p , we have

$$g_i^{q'_j} = \begin{cases} g_i^{q_{j,f}} & \text{w.p. } 0.5(1-p) \\ g_i^{q_{j,f}} + \delta_j & \text{w.p. } 0.5p \\ g_i^{q_{j,d}} & \text{w.p. } 0.5(1-p) \\ g_i^{q_{j,d}} + \delta_j & \text{w.p. } 0.5p \end{cases}, j = 1, 2.$$

A mutation causes $h\%$ change. The probability that this mutation increases $|g_i^{q'_1} - g_i^{q'_2}|$ is proportional to $1 - |g_i^{q_{1,e}} - g_i^{q_{2,e}}|$, where $e \in \{f, d\}$. Thus,

$$|g_i^{q'_1} - g_i^{q'_2}| = \begin{cases} |g_i^{q_{1,e}} - g_i^{q_{2,e}}| & \text{w.p. } (1-p)^2 \\ |g_i^{q_{1,e}} - g_i^{q_{2,e}}| + (1 - |g_i^{q_{1,e}} - g_i^{q_{2,e}}|)h & \text{w.p. } 2p(1-p) \\ |g_i^{q_{1,e}} - g_i^{q_{2,e}}| + (1 - |g_i^{q_{1,e}} - g_i^{q_{2,e}}|)2h & \text{w.p. } p^2 \end{cases}$$

Since, $E(|g_i^{q_{1,f}} - g_i^{q_{2,f}}|) = E(|g_i^{q_{1,f}} - g_i^{q_{2,d}}|) = E(|g_i^{q_{1,d}} - g_i^{q_{2,f}}|) = E(|g_i^{q_{1,d}} - g_i^{q_{2,d}}|) = E(|g_i^{q_{1,e}} - g_i^{q_{2,e}}|)$, let us denote them by $E(|g_i^{q_1} - g_i^{q_2}|)$. Then,

$$E(|g_i^{q'_1} - g_i^{q'_2}|) = \min\{b, (1 - 2hp)E(|g_i^{q_1} - g_i^{q_2}|) + 2hp\}. \quad (9)$$

The reason for min is that, if $|g_i^{q'_1} - g_i^{q'_2}| > b$, then the progeny will not survive. Therefore,

$$\begin{aligned} DFM(Q_m) &\approx \min\{b, (1 - 2hp)DFM(Q_{m-1}) + 2hp\} \\ &\approx -\frac{1}{\tau} \ln(e^{-\tau b} + e^{\tau((1-2hp)DFM(Q_{m-1})+2hp)}), \end{aligned} \quad (10)$$

where $\tau > 0$ is the parameter that determines the softness of the approximation of min.

It can be shown (Section V) that $DFM(Q_m) \rightarrow b$ as $m \rightarrow \infty$. This is very different than asexual reproduction, where $DFM(Q_m) \rightarrow 100\%$ as $m \rightarrow \infty$. In other words, while asexual reproduction can make individuals in a species completely dissimilar, to the point that a taxonomist could be forced to concede that the population of individuals are in fact multiple “species,” whereas, sexual reproduction maintains a species as a related and coherent group.

Simulation Results

There are two tasks for simulations. The first task is to verify

$$OPM \approx 1 - DFM \quad (11)$$

The second is to compare sexual reproduction with asexual reproduction by means of the diversity, diffusion, and optimality measures.

We use MATLAB, a popular engineering software to perform simulations. For the first task, we calculate DFM and OPM using Equations (2) and (3).

Since $DS(q_1, q_2)$ and $DS(q, o)$ are percentages, we represent $q \in Q$ by a number between 0 and 1. We generate K random numbers, denoted, with a slight abuse of notation, as $Q = \{q_1, q_2, \dots, q_K\}$, and then calculate

$$\begin{aligned} DFM(Q) &= \frac{1}{K^2} \sum_{i=1}^K \sum_{j=1}^K |q_i - q_j| \\ OPM(Q) &= 1 - \frac{1}{K} \sum_{i=1}^K |q_i - q_1|, \end{aligned} \quad (12)$$

where, without loss of generality, we let $o = q_1$.

We run the above simulations R times and take the average. For different K and R , the results are shown in Table 1.

K	R	DFM	OPM	SUM	ERROR
3000	200	0.3331	0.6679	1.0010	0.1%
1000	100	0.3337	0.6610	0.9947	-0.5%
1000	300	0.3331	0.6716	1.0047	0.5%
5000	100	0.3336	0.6590	0.9926	-0.7%
5000	300	0.3332	0.6683	1.0015	0.2%

Table 1: Simulation results to verify $OPM \approx 1 - DFM$ or $OPM + DFM \approx 1$.

For the second task, we simulate the derived dynamic equations, which are summarized below, to compare sexual reproduction with asexual reproduction. Since it is a comparison, only the relative values (rather than the absolute values) of the diversity, diffusion, and optimality measures are of importance.

Diversity measure: Diversity measure for asexual reproduction is given in Equation (5). Diversity measure for sexual reproduction is given in Equation (8).

Diffusion measure: Diffusion measure for asexual reproduction is given in Equation (7). Diffusion measure for sexual reproduction is given in Equation (10).

Optimality measure Diffusion measures for both asexual and sexual reproductions are given by Equation (11).

We conducted several simulations with different parameter values. The results are all similar. Typical parameters used in the simulations are as follows. We do not claim that these parameters are biologically representative. Note that since we are only interested in the relative values of the diversity, diffusion, and optimality measures, absolute values of parameters are not critical.

$$p = 0.001, \quad h = 25\%, \quad N = 10, \quad |Q| = 200, \quad b = 30\%, \quad c = 10\%, \quad \tau = 10.$$

The initial conditions are $DVM(Q_0) = 10\%$ and $DFM(Q_0) = 10\%$. Results of simulations are shown in Figures 1 and 2.

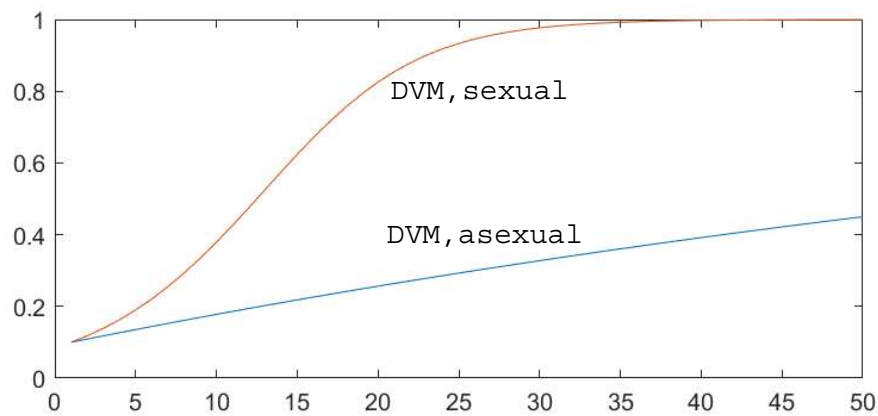


Figure 1: Diversity measure DVM of sexual vs asexual reproduction. Horizontal axis shows the number of generations.

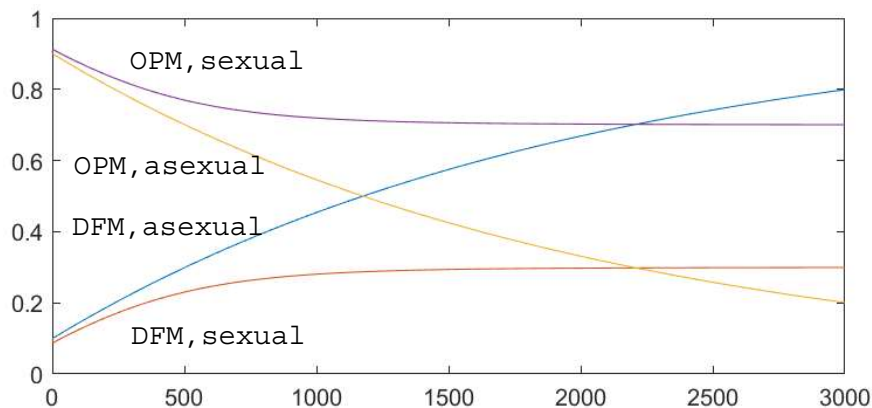


Figure 2: Diffusion measure DFM and optimality measure OPM of sexual vs asexual reproduction. Horizontal axis shows the number of generations.

From Figure 1, it is clear that the diversity measure of sexual reproduction is larger than and increases faster than that of asexual reproduction. Larger diversity measure allows a species to

evolve faster. Hence, in terms of diversity measure, sexual reproduction has an advantage over asexual reproduction.

On the other hand, from Figure 2, diffusion measure of sexual reproduction is bounded and smaller than that of asexual reproduction. By Theorem 1, smaller diversity measure leads to larger survivability measure. Hence, in terms of diffusion measure, sexual reproduction also has an advantage over asexual reproduction.

Comparing the horizontal axis of Figure 1 with the horizontal axis of Figure 2, we note that the advantage of sexual reproduction over asexual reproduction in terms of diffusion measure takes a longer time to show than the advantage in terms of diversity measure. This probably explains why the effects of diversity measure have been so readily latched onto and investigated by biologists, whereas the potential evolutionary consequences of diffusion measure's effects have been overlooked and not been investigated in any comprehensive manner.

Diversity vs Diffusion

By the definitions, DVM is the number of distinct individuals in a population Q divided by the total number of individuals in Q , while DFM is the average distance between any two individuals in Q . Hence, DVM describes how diverse a population is, while DFM describes how similar or homogenous a population is. At first glance, one may think that DVM and DFM depend on each other. However, they are actually independent of each other. This is illustrated in Figure 3, where two populations have the same DVM , but the population on the left has a much smaller DFM than the one on the right. Hence, it is entirely possible to keep DVM large while keeping DFM small. We argue that, for a population to best adapt to an environment, it is best for the population Q to have a large DVM and a small DFM (see Theorem 1).

Sexual reproduction allows genetic recombination to produce recombinant types via mech-

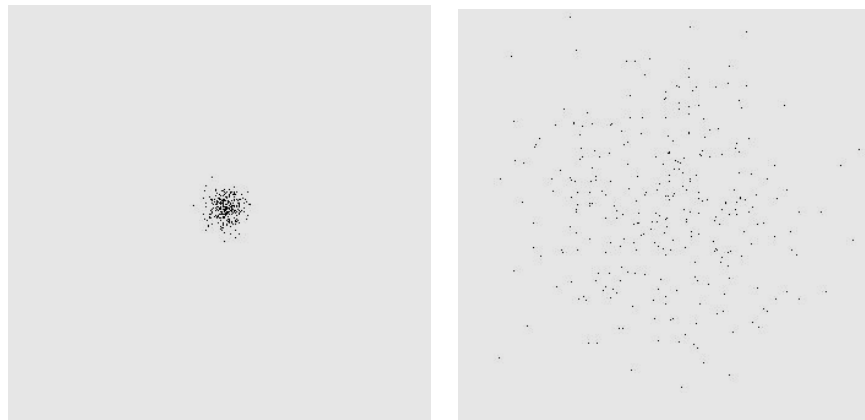


Figure 3: Diversity measure DVM and diffusion measure DFM are independent of each other. The above two populations have the same DVM , but the population on the left has a much smaller DFM than the one on the right.

anisms such as chromosome shuffling, and this process occurs with each successive progeny. Subsequently, through this probabilistic process, every progeny is forced to be distinct from both its progenitors and sibling progeny (assuming the progenitors do not have exactly identical genomes, and excluding atypical phenomena such as twins, etc.). In other words, sexual reproduction results in a high DVM . This point has been well established. Our added contribution is the insight that such a chromosomal reshuffling phenomenon could only occur if both progenitors are not radically dissimilar, such as a paternal progenitor with 46 chromosomes contributing 23 chromosomes and a maternal progenitor with 44 chromosomes contributing 22 chromosomes. This is an example of a high DFM value precluding the generation of progeny. Asexual reproduction has no such constraints. A progenitor of 46 chromosomes could produce 10 progenies: 9 with identical 46 chromosomes, and 1 with 47 chromosomes as a mutational accident. Such a population would have low DVM (90% of the population are 100% identical) with high DFM (10% of the population is radically different in chromosome number).

Our mathematical models and simulation results are consistent with the evidences that sexual reproduction promotes genetic similarity/homogeneity, while asexual reproduction leads to

genetic diffusion. As seen from our simulations, asexual reproduction results in an increase of DFM over time, while sexual reproduction results in bounded DFM , implying the formation of a tight cluster of similar individuals. The tight cluster is maintained by sexual reproduction that prevents the diffusion seen in asexual reproduction. Theorem 1 states that a smaller DFM means a large OPM and hence a large SVM .

To illustrate this benefit of genetic similarity and maintenance of adaptational advantages in a biological context, we take the example of a developed ecosystem that is near its “climax community”. Such a system can be considered relatively stable with high biodiversity. The high biodiversity means there is more competition for the same limited resources. In order for all these species to successfully live amongst each other, selective pressures have caused each species to develop specific adaptations that allow it to occupy an exclusive niche. Deviating away from such adaptations, which means deviating away from the species’ niche, results in competition with other species that occupy other niches. These other species are highly adapted for their niches; consequently, the “deviant organism” is unlikely to survive the competition. Sexual reproduction maintains adaptational advantages and minimizes the conversion of precious resources to the production of deviant organisms that are unlikely to survive.

In summary, both DVM and DFM are distinct principles that can provide a more nuanced approach to investigating the advantages of sexual reproduction over asexual reproduction. (1) Sexual reproduction leads to a more diverse population and hence allows more rapid adaption to environments. (2) Sexual reproduction is a species stabilization mechanism that naturally maintains genetic homogeneity and species identity. (3) Asexual reproduction, which does not have this inherent species stabilization mechanism, leads to genetic inhomogeneity and no definitive species identity. (4) Sexual reproduction is beneficial because the maintenance of species identity maintains desired adaptational advantages, which is important when selective pressures are strong. In the next section, we use DVM and DFM to explain some theories/hypotheses that

have been proposed to address the advantages of sexual reproduction over asexual reproduction.

Advantages of Sexual Reproduction

Various theories/hypotheses have been proposed to address the advantages of sexual reproduction over asexual reproduction. The main theories/hypotheses are as follows. (1) *Genetic recombination*: Sexual reproduction allows genetic recombination to produce recombinant types that can make the population better able to adapt changes in the environment. In other words, sexual reproduction allows greater exploration of genotype possibilities hence providing the raw fuel for natural selection to act upon. (2) *Muller's Ratchet*: Continued accumulation of deleterious mutations leads to a degradation of a species' average fitness over time, without any method of correction, the species will likely face extinction (24–26). Muller's Ratchet posits that sexual reproduction allows good genes in separate loci to be recombined to restore genotypes to optimal fitness. (3) *Red Queen hypothesis*: In this hypothesis, parasites and hosts engage in a constant arms battle with hosts evolving resistances to parasites and parasites evolving ways to get past those resistances (10, 27, 28). Sexual reproduction generates new genotypes at a much faster rate than would be possible with asexual reproduction. The faster generation of new genotypes then allows for rapid adaptation of resistances against parasites for the hosts. Likewise, sexually reproducing parasites can have rapid adaptation against host resistances. The mathematical models we presented fits well with these theories/hypotheses.

(1) *Genetic recombination*: The diversity measure DVM describes genotype possibilities of a population. Comparing Equation (5) with Equation (8), we can see that initially (when DVM is small), DVM increases linearly in asexual reproduction, but exponentially in sexual reproduction. Simulation results of Figure 1 clearly show that DVM increases much faster in sexual reproduction than in asexual reproduction.

(2) *Muller's Ratchet*: Muller's Ratchet is reflected in the optimality measure OPM . As

shown in Figure 2, *OPM* decreases in both sexual reproduction and asexual reproduction as time goes by. This is exactly what Muller's Ratchet anticipates. However, in sexual reproduction, *OPM* will reach a limit and no longer decreases. In other words, sexual reproduction can restore genotypes to optimal fitness.

(3) *Red Queen hypothesis*: The Red Queen Hypothesis can also be explained by *DVM*. Sexually reproduced parasites and hosts have much larger *DVM* than asexually reproduced parasites and hosts as shown in Figure 1. Larger *DVM* allows parasites and hosts to better adapt in the battle with hosts evolving resistances to parasites and parasites evolving ways to get past those resistances.

References

1. G. Bell, *The masterpiece of nature: the evolution and genetics of sexuality* (Routledge, 2019).
2. N. Colegrave, *Nature* **420**, 664 (2002).
3. L. D. Hurst, J. R. Peck, *Trends in Ecology & Evolution* **11**, 46 (1996).
4. D. Misevic, C. Ofria, R. E. Lenski, *Proceedings of the Royal Society B: Biological Sciences* **273**, 457 (2006).
5. A. F. Agrawal, *Current Biology* **16**, R696 (2006).
6. D. Bachtrog, *Current opinion in genetics & development* **16**, 578 (2006).
7. J. Chasnov, *Genetics* **156**, 1419 (2000).
8. J. A. G. de Visser, S. F. Elena, *Nature Reviews Genetics* **8**, 139 (2007).
9. E. S. Dolgin, S. P. Otto, *Genetics* **164**, 1119 (2003).

10. R. J. Ladle, *Trends in ecology & evolution* **7**, 405 (1992).
11. S. P. Otto, *Genetics* **164**, 1099 (2003).
12. S. P. Otto, T. Lenormand, *Nature Reviews Genetics* **3**, 252 (2002).
13. A. F. Agrawal, *Nature* **411**, 692 (2001).
14. A. F. Agrawal, J. Chasnov, *Genetics* **158**, 913 (2001).
15. I. Arkhipova, M. Meselson, *Bioessays* **27**, 76 (2005).
16. N. H. Barton, B. Charlesworth, *Science* **281**, 1986 (1998).
17. B. Charlesworth, *Genetics Research* **61**, 205 (1993).
18. S. F. Elena, R. E. Lenski, *Nature* **390**, 395 (1997).
19. M. W. Feldman, F. B. Christiansen, L. D. Brooks, *Proceedings of the National Academy of Sciences* **77**, 4838 (1980).
20. A. S. Kondrashov, *Nature* **336**, 435 (1988).
21. C. Darwin, *On the Origin of Species by Means of Natural Selection* (London: John Murray, 1859).
22. J. B. S. Haldane, *Mathematical Proceedings of the Cambridge Philosophical Society* **23**, 363 (1926).
23. M. Kimura, T. Maruyama, *Genetics* **54**, 1337 (1966).
24. H. J. Muller, *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **1**, 2 (1964).

25. W. Gabriel, M. Lynch, R. Bürger, *Evolution* **47**, 1744 (1993).
26. I. Gordo, B. Charlesworth, *Genetics* **154**, 1379 (2000).
27. W. D. Hamilton, R. Axelrod, R. Tanese, *Proceedings of the National Academy of Sciences* **87**, 3566 (1990).
28. W. P. Barnett, M. T. Hansen, *Strategic management journal* **17**, 139 (1996).

1 Appendix

Proof of Theorem 1

Let us enumerate individuals in Q based on their distance to o , that is, let

$$Q = \{q_1, q_2, \dots, q_K\},$$

where $K = |Q|$, such that

$$DS(q_1, o) \leq DS(q_2, o) \leq \dots \leq DS(q_K, o).$$

Define a distance function $\rho : \mathcal{K} \rightarrow [0, 1]$, where $\mathcal{K} = \{1, 2, \dots, K\}$, as

$$\rho(k) = DS(q_k, o).$$

Then ρ is a monotonic increasing function, that is,

$$k_1 < k_2 \Rightarrow \rho(k_1) \leq \rho(k_2),$$

where \Rightarrow denotes “implies”.

From the definition of survivability, we have

$$q_k \text{ having progeny} \Leftrightarrow \rho(k) \leq 1 - SP,$$

where \Leftrightarrow denotes “if and only if”.

Therefore,

$$SVM(Q) = \frac{|\{q_k \in Q : \rho(k) \leq 1 - SP\}|}{K} = \frac{\max\{k : \rho(k) \leq 1 - SP\}}{K}$$

To prove Part (1) of Theorem 1, consider two selective pressures SP_1, SP_2 with $SP_1 \leq SP_2$.

We have

$$SP_1 < SP_2$$

$$\begin{aligned}
 &\Rightarrow (\rho(k) \leq 1 - SP_2 \Rightarrow \rho(k) \leq 1 - SP_1) \\
 &\Rightarrow \max\{k : \rho(k) \leq 1 - SP_2\} \leq \max\{k : \rho(k) \leq 1 - SP_1\} \\
 &\quad (\text{because } \rho \text{ is a monotonic increasing function}) \\
 &\Rightarrow \frac{\max\{k : \rho(k) \leq 1 - SP_2\}}{|Q|} \leq \frac{\max\{k : \rho(k) \leq 1 - SP_1\}}{|Q|} \\
 &\Rightarrow SVM_2(Q) \leq SVM_1(Q),
 \end{aligned}$$

where $SVM_i(Q)$ is the survivability measure under selective pressure SP_i , $i = 1, 2$.

To prove Part (2) of Theorem 1, consider two optimality measures $OPM(Q_1), OPM(Q_2)$ with $OPM(Q_1) \leq OPM(Q_2)$. By the definition of $OPM(Q_i)$, $i = 1, 2$,

$$OPM(Q_i) = 1 - E(DS(q, o)) = 1 - \frac{1}{K} \sum_{k=1}^K DS_i(q_k, o) = 1 - \frac{1}{K} \sum_{k=1}^K \rho_i(k),$$

where $\rho_i(k) = DS_i(q_k, o)$, for $k \in \mathcal{K}$, corresponds to $OPM(Q_i)$. Hence, for any SP

$$\begin{aligned}
 &OPM(Q_1) \leq OPM(Q_2) \\
 &\Rightarrow 1 - \frac{1}{K} \sum_{k=1}^K \rho_1(k) \leq 1 - \frac{1}{K} \sum_{k=1}^K \rho_2(k) \\
 &\Rightarrow \sum_{k=1}^K \rho_1(k) \geq \sum_{k=1}^K \rho_2(k) \\
 &\Rightarrow \rho_1(k) \geq \rho_2(k) \text{ with probability 1} \\
 &\Rightarrow \rho_1(k) \leq 1 - SP \Rightarrow \rho_2(k) \leq 1 - SP \\
 &\Rightarrow \max\{k : \rho_1(k) \leq 1 - SP\} \leq \max\{k : \rho_2(k) \leq 1 - SP\} \\
 &\Rightarrow \frac{\max\{k : \rho_1(k) \leq 1 - SP\}}{K} \leq \frac{\max\{k : \rho_2(k) \leq 1 - SP\}}{K} \\
 &\Rightarrow SVM(Q_1) \leq SVM(Q_2).
 \end{aligned}$$

Part (3) of Theorem 1 is the direct result of Part (2) of Theorem 1 and the relation $OPM(Q) \approx 1 - DFM(Q)$.

■

Derivation of Equation (6)

$$\begin{aligned}
 E(|g_i^{q'_1} - g_i^{q'_2}|) &= E(|g_i^{q_1} - g_i^{q_2}|) + (1 - E(|g_i^{q_1} - g_i^{q_2}|))h(2p(1 - p) + 2p^2) \\
 &= E(|g_i^{q_1} - g_i^{q_2}|) + 2(1 - E(|g_i^{q_1} - g_i^{q_2}|))hp \\
 &= (1 - 2hp)E(|g_i^{q_1} - g_i^{q_2}|) + 2hp.
 \end{aligned}$$

Derivation of Equation (7)

$$\begin{aligned}
 DFM(Q_m) &= E(DS(q'_1, q'_2)) = E\left(\frac{1}{N} \sum_{i=1}^N |g_i^{q'_1} - g_i^{q'_2}|\right) = \frac{1}{N} \sum_{i=1}^N E(|g_i^{q'_1} - g_i^{q'_2}|) \\
 &= \frac{1}{N} \sum_{i=1}^N ((1 - 2hp)E(|g_i^{q_1} - g_i^{q_2}|) + 2hp) \\
 &= \frac{1}{N} \left(\sum_{i=1}^N (1 - 2hp)E(|g_i^{q_1} - g_i^{q_2}|) + \sum_{i=1}^N 2hp \right) \\
 &= \left(\frac{1}{N} \sum_{i=1}^N (1 - 2hp)E(|g_i^{q_1} - g_i^{q_2}|) \right) + \left(\frac{1}{N} \sum_{i=1}^N 2hp \right) \\
 &= (1 - 2hp)E\left(\frac{1}{N} \sum_{i=1}^N |g_i^{q_1} - g_i^{q_2}|\right) + 2hp \\
 &= (1 - 2hp)E(DS(q_1, q_2)) + 2hp \\
 &= (1 - 2hp)DFM(Q_{m-1}) + 2hp.
 \end{aligned}$$

Derivation of Equation (9)

$$\begin{aligned}
 E(|g_i^{q'_1} - g_i^{q'_2}|) &= \min\{b, E(|g_i^{q_1} - g_i^{q_2}|) + (1 - E(|g_i^{q_1} - g_i^{q_2}|))h(2p(1 - p) + 2p^2)\} \\
 &= \min\{b, E(|g_i^{q_1} - g_i^{q_2}|) + 2(1 - E(|g_i^{q_1} - g_i^{q_2}|))hp\} \\
 &= \min\{b, (1 - 2hp)E(|g_i^{q_1} - g_i^{q_2}|) + 2hp\}.
 \end{aligned}$$

Derivation of Equation (10)

$$DFM(Q_m) = E(DS(q'_1, q'_2)) = E\left(\frac{1}{N} \sum_{i=1}^N |g_i^{q'_1} - g_i^{q'_2}|\right) = \frac{1}{N} \sum_{i=1}^N E(|g_i^{q'_1} - g_i^{q'_2}|)$$

$$= \frac{1}{N} \sum_{i=1}^N \min\{b, (1 - 2hp)E(|g_i^{q_1} - g_i^{q_2}|) + 2hp\}.$$

Mathematically,

$$\begin{aligned} & \frac{1}{N} \sum_{i=1}^N \min\{b, (1 - 2hp)E(|g_i^{q_1} - g_i^{q_2}|) + 2hp\} \\ & \leq \min\{b, \frac{1}{N} \sum_{i=1}^N ((1 - 2hp)E(|g_i^{q_1} - g_i^{q_2}|) + 2hp)\}. \end{aligned}$$

However, we expect that $E(|g_i^{q_1} - g_i^{q_2}|) \approx E(|g_j^{q_1} - g_j^{q_2}|)$, for $i, j \in \mathcal{N}$. Therefore,

$$\begin{aligned} DFM(Q_m) &= \frac{1}{N} \sum_{i=1}^N \min\{b, (1 - 2hp)E(|g_i^{q_1} - g_i^{q_2}|) + 2hp\} \\ &\approx \min\{b, \frac{1}{N} \sum_{i=1}^N ((1 - 2hp)E(|g_i^{q_1} - g_i^{q_2}|) + 2hp)\} \\ &= \min\{b, \frac{1}{N} (\sum_{i=1}^N (1 - 2hp)E(|g_i^{q_1} - g_i^{q_2}|) + \sum_{i=1}^N 2hp)\} \\ &= \min\{b, (\frac{1}{N} \sum_{i=1}^N (1 - 2hp)E(|g_i^{q_1} - g_i^{q_2}|)) + (\frac{1}{N} \sum_{i=1}^N 2hp)\} \\ &= \min\{b, (\frac{1}{N} \sum_{i=1}^N (1 - 2hp)E(|g_i^{q_1} - g_i^{q_2}|)) + 2hp\} \\ &= \min\{b, (1 - 2hp)(\frac{1}{N} \sum_{i=1}^N E(|g_i^{q_1} - g_i^{q_2}|)) + 2hp\} \\ &= \min\{b, (1 - 2hp)E(\frac{1}{N} \sum_{i=1}^N |g_i^{q_1} - g_i^{q_2}|) + 2hp\} \\ &= \min\{b, (1 - 2hp)E(DS(q_1, q_2)) + 2hp\} \\ &= \min\{b, (1 - 2hp)DFM(Q_{m-1}) + 2hp\}. \end{aligned}$$

For better simulations, we approximate min by a continuous function as follows.

$$\begin{aligned} DFM(Q_m) &\approx \min\{b, (1 - 2hp)DFM(Q_{m-1}) + 2hp\} \\ &\approx -\frac{1}{\tau} \ln(e^{-\tau b} + e^{\tau((1-2hp)DFM(Q_{m-1})+2hp)}). \end{aligned}$$