Generation of Variation and Mean Fitness Increase: Necessity is the Mother of Genetic Invention

Yoav Ram[a], Lee Altenberg[b], Uri Liberman[c], and Marcus W. Feldman[a]

[a]Department of Biology, Stanford University, Stanford, CA
[b]Information and Computer Sciences, University of Hawai'i at Mānoa, Honolulu, HI
[c]School of Mathematical Sciences, Tel Aviv University, Israel

December 4, 2017

Abstract

Generation of variation may be detrimental in well-adapted populations evolving under constant selection. In a constant environment, genetic modifiers that reduce the rate at which variation is generated by processes such as mutation and migration, succeed. However, departures from this *reduction principle* have been demonstrated. Here we analyze a general model of evolution under constant selection where the rate at which variation is generated depends on the individual. We find that if a modifier allele increases the rate at which individuals of below-average fitness generate variation, then it will increase in frequency and increase the population mean fitness. This principle applies to phenomena such as stress-induced mutagenesis and condition-dependent dispersal, and exemplifies *"Necessity is the mother of genetic invention."*

Introduction

According to the *reduction principle*, in populations at a balance between natural selection and a process that generates variation (i.e. mutation, migration, or recombination), selection favors neutral modifiers that decrease the rate at which variation is generated. The *reduction principle* was demonstrated for modifiers of recombination (Feldman, 1972), mutation (Liberman and Feldman, 1986), and migration (Feldman and Liberman, 1986). These results were unified in a series of studies (Altenberg, 1984; Altenberg and Feldman, 1987; Altenberg, 2009, 2012*a*,*b*; Altenberg et al., 2017).

The latter studies have established the conditions for a *unified reduction principle* by neutral genetic modifiers: (i) effectively infinite population size, (ii) constant-viability selection, (iii) a population at an equilibrium, and (iv) *linear variation* – the equal scaling of transition probabilities by the modifier. A departure from the latter assumption occurs if two variation-producing processes interact (Feldman et al., 1980; Altenberg, 2012*a*). Departures from the *reduction principle* have also been demonstrated when conditions (i)-(iii) are not met, see for example Holsinger et al. (1986) and references therein.

Another departure from the *linear variation* assumption of the *reduction principle* for mutation rates involves a mechanism by which the mutation rate increases in individuals of low fitness – a mechanism first observed in stressed bacteria (Foster, 2007), although not in a constant environment. Ram and Hadany (2012) demonstrated that even in a constant environment, increasing the mutation rate of individuals with below-average fitness increases the population mean fitness, rather than decreases it. Their analysis assumed infinite population size and fitness determined by the number of mutant alleles accumulated in the genotype. In their models, the only departure from the *reduction principle* assumptions was the unequal scaling of mutation probabilities between different genotypes introduced by the correlation between the mutation rate and fitness. A similar result has been demonstrated for conditional dispersal (Altenberg, 2012*a*, Th. 39), fitness-associated recombination (Hadany and Beker, 2003*b*) and for condition-dependent sexual reproduction (Hadany and Otto, 2007), and evidence suggests that both mechanisms are common in nature (Ram and Hadany, 2016).

Ram and Hadany (2012) stated that their result represents a departure from the *reduction principle*, but did not explain this departure. Their analysis was specific to a model that classified individuals by the number of mutant alleles in their genotype, similar to models studied by Kimura and Maruyama (1966) and Haigh (1978). Moreover, their argument was based on the expected increase of the stable population mean fitness, rather than on the invasion success of modifier alleles that modify the mutation rate (i.e., analysis of *evolutionary genetic stability*, see Eshel and Feldman, 1982; Lessard, 1990).

Here, we present an evolutionary model in which the type of the individual determines both its fitness and the rate at which it generates variation. Our results show that the population mean fitness increases if individuals with below-average fitness produce more variation than individuals with above-average fitness, and that modifier alleles that induce below-average individuals to produce more variation are favored by natural selection.

Models

General model. Consider a large population with an arbitrary set of types $A_1, A_2, \ldots, A_n$. The frequency and fitness of individuals of type A_k are f_k and w_k, respectively. The probability that an individual of type A_k will transition to some other type is C_k, and given a transition occurs, the probability that it will transition to type A_j is $M_{j,k}$. Therefore, the change in the frequencies of type A_k is described by the transformation $f \to f'$:

$$\bar{w} f_k' = (1 - C_k) w_k f_k + \sum_{j=1}^{n} C_j M_{k,j} w_j f_j, \tag{1}$$

or in matrix form

$$\bar{w} f' = (I - C + MC) D f, \tag{2}$$

where $f = (f_1, f_2, \ldots, f_n)$ is a frequency vector with $f_k \geq 0$ and $\sum_{k=1}^{n} f_k = 1$; D is a positive diagonal matrix with entries

w_k such that $w_k \neq w_j$ for some $k \neq j$; C is a positive diagonal matrix with entries C_k; M is a primitive column-stochastic matrix: $M_{j,k} \geq 0$ for all j, k, $\sum_{j=1}^{n} M_{j,k} = 1$ for all k, and $(M^\ell)_{j,k} > 0$ for all j, k for some positive integer ℓ; I is the $n \times n$ identity matrix; and $\bar{w}$ is the normalizing factor such that $\sum_{k=1}^{n} f'_k = 1$ and is equal to the population mean fitness $\bar{w} = \sum_{k=1}^{n} f_k w_k$.

The types A_k can represent a single or multiple haploid genetic loci or non-genetic traits. Importantly, type transmission is vertical and uni-parental (the type is transmitted from a single parent to the offspring) and is independent of the frequencies of the other types. This model precludes processes such as recombination, social learning, sexual outcrossing, and horizontal or oblique transmission, as these processes are frequency-dependent (Cavalli-Sforza and Feldman, 1981, pg. 54).

Transition between types is determined by a combination of two effects: (i) the probability of transitioning *out* of type A_k is determined by C_k; (ii) given a transition out of type A_k, the distribution of the destination types A_i is given by $M_{i,k}$ (note the index order). Importantly, different types can have different rates. That is, $C_i \neq C_j$ for some i, j. The case $C_i = C_j$ for all i, j is covered by the *reduction principle* (see Altenberg et al., 2017).

In the following section we present four examples of the model (Eq. 2) that apply to mutation, migration, and learning.

Mutation model 1. Here we consider a large population of haploids and a trait determined by a single genetic locus with n possible alleles $A_1, A_2, \ldots, A_n$ and corresponding fitness values $w_1, w_2, \ldots, w_n$. The mutation rates C_k of individuals with allele A_k are potentially different; specifically, with probability $1 - C_k$, the allele A_k does not mutate, and with probability $\frac{C_k}{n-1}$, the allele A_k mutates to A_j for any $j \neq k$. This is an extension of a model studied by Altenberg et al. (2017) that allows for the mutation rate of A_k, C_k, to depend on properties of the allele A_k.

Let the frequency of A_k in the present generation be f_k with $f_k \geq 0$ and $\sum_{k=1}^{n} f_k = 1$. Then after selection and mutation, f'_k in the next generation is given by

$$f'_k = \big(1 - C_k\big)\frac{w_k}{\bar{w}} f_k + \frac{1}{n-1}\sum_{j \neq k} C_j \frac{w_j}{\bar{w}} f_j, \tag{3}$$

for $k = 1, 2, \ldots, n$, where $\bar{w} = \sum_{k=1}^{n} f_k w_k$ is the population mean fitness.

This model is a special case of the general model (Eq. 2) where

$$M = \frac{1}{n-1}\begin{bmatrix} 0 & 1 & 1 & \ldots & 1 \\ 1 & 0 & 1 & \ldots & \vdots \\ 1 & 1 & 0 & \ldots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \ldots & \ldots & \ldots & 0 \end{bmatrix}, \tag{4}$$

with zeros on the diagonal and $\frac{1}{n-1}$ elsewhere. Note that here M is irreducible and primitive.

Mutation model 2. Again, we consider a large population of haploids, but here individuals with genotype A_k are characterized by the number k of deleterious or mutant alleles in their genotype, where $0 \leq k \leq n$. Specifically, the fitness of individuals with k mutant alleles is w_k ($w_0 > w_1 > \ldots > w_n$), and the probability C_k that a mutation occurs in individuals with k mutant alleles depends on k. When a mutation occurs it is *deleterious* with probability δ, generating a mutant allele and converting the individual from A_k to A_{k+1}, or it is *beneficial* with probability β, converting the individual from A_k to A_{k-1}. Note that such beneficial mutations can be either compensatory or back-mutations, and that mutations are neutral with probability $1 - \delta - \beta$. We assume that both the deleterious and the beneficial mutation rates are low enough that two mutations are unlikely to occur in the same individual in one generation: $C_k(\delta + \beta) \ll 1$ for all $k = 1, \ldots, n$. This model has been analyzed by Ram and Hadany (2012).

Let the frequency of A_k in the present generation be f_k with $f_k \geq 0$ and $\sum_{k=0}^{n} f_k = 1$. Then after selection and mutation f'_k in the next generation is given by

$$\begin{aligned} f'_0 &= \big(1 - \delta C_0\big)\frac{w_0}{\bar{w}} f_0 + \beta C_1 \frac{w_1}{\bar{w}} f_1, \\ f'_k &= \big(1 - (\delta + \beta)C_k\big)\frac{w_k}{\bar{w}} f_k + \\ &\quad \delta C_{k-1}\frac{w_{k-1}}{\bar{w}} f_{k-1} + \beta C_{k+1}\frac{w_{k+1}}{\bar{w}} f_{k+1}, \\ f'_n &= \big(1 - \beta C_n\big)\frac{w_n}{\bar{w}} f_n + \delta C_{n-1}\frac{w_{n-1}}{\bar{w}} f_{n-1}, \end{aligned} \tag{5}$$

for $k = 1, 2, \ldots, n-1$. Here $\bar{w} = \sum_{k=0}^{n} f_k w_k$ is the population mean fitness.

Therefore, setting

$$M = \begin{bmatrix} 1-\delta & \beta & 0 & \cdots & 0 \\ \delta & 1-\delta-\beta & \beta & \ddots & 0 \\ 0 & \delta & 1-\delta-\beta & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \beta \\ 0 & 0 & 0 & \delta & 1-\beta \end{bmatrix}, \tag{6}$$

with the beneficial, neutral, and deleterious mutation probabilities on the three main diagonals and zeros elsewhere, Eq. 5 can be viewed as a special case of Eq. 2. Here, too, M is irreducible and primitive as long as $\delta, \beta > 0$.

Migration model. In this case we consider a large population of haploids that occupy n demes, $A_1, \ldots, A_n$. Let the frequencies of individuals in deme A_k be f_k with $f_k \geq 0$ and $\sum_{k=1}^{n} f_k = 1$. The fitness of individuals in deme A_k is w_k, but the entire population comes together for reproduction, and therefore reproductive success is determined by competition among individuals of all demes – this has been termed *hard selection* (Wallace, 1975; Karlin, 1982).

After reproduction, offspring of individuals from deme A_k return to their parental deme with probability $1 - C_k$, or migrate to a different deme A_j with probability $C_k M_{j,k}$, where the matrix M is primitive, $C_k > 0$, $M_{j,k} \geq 0$, and $\sum_{j=1}^{n} M_{j,k} = 1$ for all $k = 1, \ldots, n$. Therefore, $1 - C_k$ can be interpreted as a *homing rate*.

Following selection and migration the new frequencies f'_k are given exactly by Eq. 1. If the columns of M are identical

$$M = \begin{bmatrix} m_1 & m_1 & \ldots & m_1 \\ m_2 & m_2 & \ldots & m_2 \\ \vdots & \vdots & \vdots & \vdots \\ m_n & m_n & \ldots & m_n \end{bmatrix}, \tag{7}$$

with $m_k > 0$ and $\sum_{k=1}^{n} m_k = 1$, then m_k can be considered the relative population size of deme A_k – this is the non-homogeneous extension of Deakin's *homing model* (Deakin, 1966; Karlin, 1982).

Similarly, if demes are arranged in a circle, for example around a lake, then we can denote the probability p_k of migrating k demes away from the parental deme (conditioned on migration which occurs with probability C_k) and M takes the form

$$M = \begin{bmatrix} p_0 & p_1 & \cdots & p_{n-1} \\ p_{n-1} & p_0 & \cdots & p_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ p_1 & p_2 & \cdots & p_0 \end{bmatrix}, \tag{8}$$

where $p_k > 0$ and $\sum_{k=0}^{n-1} p_k = 1$.

Learning model. In our final example, we consider a large population and an integer phenotype k where $1 \le k \le n$. Individuals are characterized by their initial and mature phenotypes (Boyd and Richerson, 1985, pg. 94). Fitness is determined by the mature phenotype: the fitness of an individual with mature phenotype k is w_k.

An offspring's initial phenotype is acquired by learning the mature phenotype of its parent (assuming uni-parental transmission). In individuals with initial phenotype k, the mature phenotype is the same as the initial phenotype with probability $1 - C_k$, and is modified by individual learning or *exploration* (Borenstein et al., 2008) with probability C_k. Such individual exploratory learning, which can be considered either intentional or the result of incorrect learning, modifies initial phenotype k to mature phenotype j with probability $M_{j,k}$.

Therefore, if the frequency of individuals with mature phenotype k in the current generation is f_k, then the frequency in the next generation f_k' is

$$f_k' = (1 - C_k) f_k \frac{w_k}{\bar{w}} + \sum_{j=1}^{n} C_j M_{k,j} \frac{w_j}{\bar{w}} f_j, \tag{9}$$

where $\bar{w} = \sum_{k=1}^{n} f_k w_k$ is the population mean fitness.

For example, in the case of *symmetric individual learning* (Borenstein et al., 2008), learning is parameterized by its breadth of exploration b and the mature phenotype j is randomly drawn from $2b + 1$ phenotypes symmetrically and uniformly distributed around the initial phenotype k, with the limitation that any "spillover" of phenotypes below 1 or above n is "absorbed" by those boundaries. This "absorption" ensures M is column-stochastic. In other words, given initial phenotype k, the probability of maturation to phenotype j, where $k - b \le j \le k + b$ is $1/(2b + 1)$, but any phenotype $j < 1$ actually becomes $j = 1$ and any phenotype $j > n$ actually becomes $j = n$. The probability for maturation to other phenotypes is 0.

For instance, with $n = 5$ and $b = 1$ we have

$$M = \begin{bmatrix} 2/3 & 1/3 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & 0 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 1/3 & 2/3 \end{bmatrix}, \tag{10}$$

and with $n = 6$ and $b = 2$ we have

$$M = \begin{bmatrix} 3/5 & 2/5 & 1/5 & 0 & 0 & 0 \\ 1/5 & 1/5 & 1/5 & 1/5 & 0 & 0 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 & 0 \\ 0 & 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 0 & 0 & 1/5 & 1/5 & 1/5 & 1/5 \\ 0 & 0 & 0 & 1/5 & 2/5 & 3/5 \end{bmatrix}. \tag{11}$$

Results

Mean fitness principle. We first focus on the stable population mean fitness. We show that if the transition rate from types with below-average fitness increases, then the stable population mean fitness increases, too.

Write the equilibrium frequency vector f in Eq. 2 as $\hat{v}$ and the stable population mean fitness as $\hat{\bar{w}}$, then

$$\hat{\bar{w}}\hat{v} = (I - C + MC)D\hat{v}. \tag{12}$$

Note that (i) the existence and uniqueness of $\hat{\bar{w}}$ and $\hat{v}$ are guaranteed by the *Perron-Frobenius theorem* (Otto and Day (2007)) because $(I - C + MC)D$ is a non-negative primitive matrix; (ii) the global stability of this equilibrium is proven in Appendix C.

The following result constitutes a *mean fitness principle* for the sensitivity of the equilibrium mean fitness $\hat{\bar{w}}$ to changes in C_k, the probability of transition from A_k.

Result 1 (Mean fitness principle)**.** *Let $\hat{\bar{w}}$ be the leading eigenvalue of $(I - C + MC)D$, and $\hat{u}$ and $\hat{v}$ be the corresponding positive left and right eigenvectors, such that $\sum_{k=1}^{n} \hat{v}_k = 1$ and $\sum_{k=1}^{n} \hat{u}_k \hat{v}_k = 1$. Then,*

$$\frac{\partial \hat{\bar{w}}}{\partial C_k} = \frac{\hat{u}_k \hat{v}_k}{C_k} (\hat{\bar{w}} - w_k), \tag{13}$$

or in simpler terms,

$$sign \frac{\partial \hat{\bar{w}}}{\partial C_k} = sign(\hat{\bar{w}} - w_k). \tag{14}$$

Therefore increased transition from type k will increase the stable population mean fitness if the fitness of type k is below the stable population mean fitness.

Proof. Using the formula in Caswell (1978) (see eq. 36 in Appendix A),

$$\frac{\partial \hat{\bar{w}}}{\partial C_k} = \hat{u}^\intercal \frac{\partial (I - C + MC)D}{\partial C_k} \hat{v}. \tag{15}$$

Let e_k and $e_k^\intercal$ be the column and row vectors with 1 at position k and 0 elsewhere, $Z_k = e_k e_k^\intercal$ be the matrix with 1 at position (k, k) and 0 elsewhere, and $[M]_k$ be the k-th column of M.

Then,

$$\begin{aligned} \hat{u}^\intercal \frac{\partial (I - C + MC)D}{\partial C_k} \hat{v} = \hat{u}^\intercal (0 - Z_k + MZ_k) D\hat{v} = \\ -\hat{v}_k \hat{u}_k w_k + \hat{v}_k w_k \hat{u}^\intercal [M]_k = \\ \hat{v}_k w_k (\hat{u}^\intercal [M]_k - \hat{u}_k). \end{aligned} \tag{16}$$

The corresponding equation to Eq. 12 for the left eigenvector u is

$$\hat{u}^\intercal \hat{\bar{w}} = \hat{u}^\intercal (I - C + MC)D, \tag{17}$$

which gives us a relation between $\hat{\bar{w}}$ and the k element of $\hat{u}$:

$$\hat{u}_k\hat{\bar{w}} = (1 - C_k)w_k\hat{u}_k + C_k w_k \hat{u}^\intercal [M]_k. \tag{18}$$

Multiplying both sides by $\hat{v}_k$ and rearranging, we get

$$\frac{\hat{u}_k\hat{v}_k}{C_k}(\hat{\bar{w}} - w_k) = \hat{v}_k w_k(\hat{u}^\intercal [M]_k - \hat{u}_k), \tag{19}$$

which when substituted into Eq. 16 yields:

$$\frac{\partial\hat{\bar{w}}}{\partial C_k} = \frac{\hat{u}_k\hat{v}_k}{C_k}(\hat{\bar{w}} - w_k). \tag{20}$$

Finally, since $\hat{u}_k, \hat{v}_k, C_k > 0$, we have

$$sign\frac{\partial\hat{\bar{w}}}{\partial C_k} = sign(\hat{\bar{w}} - w_k), \tag{21}$$

which completes the proof.

The above result provides a condition for the effect of changing C_k, the probability for transition from A_k, on the stable population mean fitness. Specifically, if A_k individuals have below-average fitness, then increasing C_k will increase the population mean fitness.

We turn our attention to the case where the transition rates from a subset K of the types are correlated, that is, $C_j = C_i$ for $i, j \in K$. In this case, Eq. 13 leads directly to the following.

Corollary 1. *The sensitivity of the stable population mean fitness to change in the rate of transition τ from types in K is*

$$\frac{\partial\hat{\bar{w}}}{\partial\tau} = \hat{u}^\intercal\Big(\sum_{k\in K}\frac{\partial(I - C + MC)D}{\partial\tau}\Big)\hat{v} = \frac{1}{\tau}\sum_{k\in K}\hat{u}_k\hat{v}_k(\hat{\bar{w}} - w_k), \tag{22}$$

and

$$sign\frac{\partial\hat{\bar{w}}}{\partial\tau} = sign\sum_{k\in K}\hat{u}_k\hat{v}_k(\hat{\bar{w}} - w_k) = sign\Big(\hat{\bar{w}} - \frac{\sum_{k\in K}\hat{u}_k\hat{v}_k w_k}{\sum_{k\in K}\hat{u}_k\hat{v}_k}\Big). \tag{23}$$

Therefore, increased transition from types in K will increase the stable population mean fitness if the average fitness of individuals descended from types in K is below the stable population mean fitness. For example, Ram and Hadany (2012, Appendix B) considered individuals that are grouped by the number of their accumulated mutant alleles, k (see Mutation model 2), and the effect of increasing the mutation rate in individuals with at least π mutant alleles. According to Eq. 23, this will result in increased stable population mean fitness if individuals with π or more mutant alleles have below-average fitness.

Reproductive value principle. An interesting interpretation of Eq. 16 is

$$\frac{\partial\hat{\bar{w}}}{\partial C_k} = \hat{v}_k w_k(\hat{u}^\intercal [M]_k - \hat{u}_k) = w_k\Big(\sum_{j=1}^{n}\hat{u}_j M_{j,k}\hat{v}_k - \hat{u}_k\hat{v}_k\Big). \tag{24}$$

Here, $\hat{u}_k$ can be regarded as the *reproductive value* of type k (Fisher, 1930, pg. 27), which gives the relative contribution of type k to the long-term population (see Appendix B). Consequently, $\hat{u}_k\hat{v}_k$ is the *ancestor frequency* of type k (Hermisson et al., 2002), namely the fraction of the equilibrium population descended from type k. The sum $\sum_{j=1}^{n}\hat{u}_j M_{j,k}\hat{v}_k$ can be similarly interpreted as the fraction of the equilibrium population descended from individuals that transitioned from type k to another type (via the k column of the transition matrix M), conditioned on transition occurring.

Since $w_k > 0$, from Eq. 16 we have the following corollary.

Corollary 2 (Reproductive value principle)**.** *In the notation of Result 1,*

$$sign\frac{\partial\hat{\bar{w}}}{\partial C_k} = sign(\hat{u}^\intercal [M]_k - \hat{u}_k), \tag{25}$$

where $[M]_k$ is the k-th column of M.

Therefore, increased transition from type k will increase the stable population mean fitness if the fraction of the population descended from type k is expected to increase due to a transition to another type.

Corollary 2 sheds light on why we require M to be primitive. If M is primitive then individuals of type k can transition into any other type in a finite number of generations. So individuals with below-average fitness can have descendants with above-average fitness, and increased generation of variation in these individuals will increase the stable population mean fitness. In contrast, if M is not primitive, individuals with below-average fitness are "doomed" and increasing the generation of variation in these individuals can only hasten their removal from the population. For example, if we set $\beta = 0$ in Mutation model 2, M becomes triangular and imprimitive, and the stable mean fitness becomes $\bar{w} = (1 - \delta C_0)w_0$, which is not affected by changes in C_k for $k \geq 1$ (see also Agrawal, 2002; Ram and Hadany, 2012, Fig. 1A).

Evolutionary genetic stability. We now focus on a neutral modifier locus completely linked to the types A_k, with no direct effect on fitness, and whose sole function is to determine C_k, the rates of transition from the different types. We will show that modifier alleles that increase the stable population mean fitness in accordance with Result 1 are favored by natural selection.

Modifier model. Consider the case of two modifier alleles, m and M, inducing different transition probabilities $C = \mathrm{diag}\Big[C_1, \ldots, C_n\Big]$ and $\tilde{C} = \mathrm{diag}\Big[\tilde{C}_1, \ldots, \tilde{C}_n\Big]$, respectively. The frequencies of type A_k linked to modifier m or M are f_k and g_k, respectively, where $\sum_{k=1}^{n}(f_k + g_k) = 1$. $\bar{w}$ now ensures that $\sum_{k=1}^{n}(f'_k + g'_k) = 1$, and the rest of the model parameters are the same as in Eq. 2.

The frequencies in the next generation, f' for allele m and g' for allele M, are given by

$$\begin{cases}\bar{w}f' &= (I - C + MC)Df\\ \bar{w}g' &= (I - \tilde{C} + M\tilde{C})Dg\end{cases}. \tag{26}$$

Here, $\bar{w} = \sum_{k=1}^{n}(f_k + g_k)w_k$ is the mean fitness of the entire population. Note that Eq. 2 is the special case of Eq. 26 where allele M is absent, i.e. $g_k = 0$ for all k.

Result 1 provides a condition under which increasing the transition rate C_k from type A_k will increase the stable population mean fitness. Could a modifier allele that increases C_k increase in frequency when initially rare in the population? To

answer this we analyze the stability of resident modifier allele m with transition rates C_k to invasion by a modifier allele M with rates $\tilde{C}_k$ under Eqs. 26.

The equilibrium of Eqs. 26 when modifier allele M is absent from the population is $(\hat{v}, 0)$, where $\hat{v}$ is given in Eq. 12 and $g_k = 0$ for all k. The stability of allele m to invasion by allele M is determined by the leading eigenvalue λ_1 of $\mathbf{L}_{ex}$ the external stability matrix of the equilibrium $(\hat{v}, 0)$, which, in turn, is determined by the Jacobian $\mathbf{J}$ of the system in Eqs. 26 evaluated at the equilibrium $(\hat{v}, 0)$, where

$$\mathbf{J} = \begin{pmatrix} \mathbf{L}_{in} & 0 \\ 0 & \mathbf{L}_{ex} \end{pmatrix}, \tag{27}$$

and $\mathbf{L}_{in}$ is the local stability matrix of the equilibrium $(\hat{v}, 0)$ in the space $\sum_{k=1}^{n} f_k = 1$. The zero block matrices are due to the complete linkage between the modifier and the types A_k and to the lack of transition (i.e., mutation) between the modifier alleles.

$\mathbf{L}_{ex}$ can be written as

$$\mathbf{L}_{ex} = (I - \tilde{C} + M\tilde{C})\tilde{D}, \tag{28}$$

where $\tilde{D} = D/\hat{\bar{w}}$ is the diagonal matrix with entries $w_k/\hat{\bar{w}}$ for all k, and $\hat{\bar{w}}$ is the stable population mean fitness in the absence of the modifier allele M. λ_1, the leading eigenvalue of $\mathbf{L}_{ex}$, coincides with $\tilde{\bar{w}}$, the maximal mean fitness associated with Eq. 28. Thus we can apply Result 1 to λ_1 and obtain the following result.

Result 2 (Evolution of increased genetic variation)**.** *Let λ_1 be the the leading eigenvalue of the external stability matrix $\mathbf{L}_{ex}$. If the transition rates induced by the modifier alleles m and M are equal, i.e., $\tilde{C}_k = C_k$ for all k, then*

$$\lambda_1 = 1, \tag{29}$$

and

$$sign \frac{\partial \lambda_1}{\partial \tilde{C}_k}\bigg|_{\tilde{C}_k = C_k} = sign(\hat{\bar{w}} - w_k). \tag{30}$$

Therefore, an initially rare modifier allele M with transition rates slightly different from the resident allele m can successfully invade the population ($\lambda_1 > 1$) if M increases the probability of transition from types with below-average fitness, thereby increasing the mean fitness $\tilde{\bar{w}}$.

Proof. Substituting $\tilde{C} = C$ in Eq. 28 and multiplying both sides by $\hat{\bar{w}}$,

$$\hat{\bar{w}}\mathbf{L}_{ex}\big|_{\tilde{C}_k = C_k} = (I - C + MC)D, \tag{31}$$

and since $\hat{\bar{w}}$ is the leading eigenvalue of the RHS (see Eq. 12), the leading eigenvalue of $\mathbf{L}_{ex}\big|_{\tilde{C}_k = C_k}$ is $\lambda_1 = 1$.

Now, applying Result 1 (eq. 14) to Eq. 28, the sign of the derivative of λ_1 with respect to $\tilde{C}_k$ is

$$sign \frac{\partial \lambda_1}{\partial \tilde{C}_k} = sign\left(\lambda_1 - \frac{w_k}{\hat{\bar{w}}}\right). \tag{32}$$

Thus

$$sign\ \frac{\partial \lambda_1}{\partial \tilde{C}_k}\bigg|_{\tilde{C}_k = C_k} = sign\left(1 - \frac{w_k}{\hat{\bar{w}}}\right) = sign(\hat{\bar{w}} - w_k), \tag{33}$$

since $\hat{\bar{w}} > 0$. This completes the proof.

Reduction principle and mutational loss. Note that if the modifier has the same effect on all types, then we can substitute $C = \mu I$ (with $\mu > 0$) in Eq. 12, and proceeding as in Eq. 22, we find a relationship previously described by Hermisson et al. (2002, eq. 24),

$$\frac{\partial \bar{w}}{\partial \mu} = -\frac{1}{\mu}\left(\sum_{k=1}^{n} u_k v_k w_k - \bar{w}\right) = -\frac{1}{\mu} G, \tag{34}$$

where u_k, v_k are computed at the equilibrium associated with μ and $\bar{w}$ is the mean fitness at that equilibrium. G is the difference between the *ancestral mean fitness* ($\sum_{k=1}^{n} u_k v_k w_k$) and the *stable population mean fitness* ($\bar{w}$) when $C_k = \mu$, is called the *mutational loss* (Hermisson et al., 2002).

If an invading allele M changes the transition probability from that of the resident allele m, i.e., $C_k = \mu$ and $\tilde{C}_k = \tilde{\mu}$, then $\mathbf{L}_{ex}$ becomes

$$\mathbf{L}_{ex} = \left[(1 - \tilde{\mu})I + \tilde{\mu}M\right]\tilde{D}, \tag{35}$$

where $\tilde{D} = D/\bar{w}$ is the diagonal matrix with entries $w_k/\bar{w}$ for all k. Using the *unified reduction principle* (Altenberg et al., 2017) the leading eigenvalue $\lambda_1 = \lambda(\tilde{\mu})$ of $\mathbf{L}_{ex}$ satisfies $\partial\lambda_1/\partial\tilde{\mu} < 0$, $\lambda_1(\tilde{\mu}) = 1$ when $\tilde{\mu} = \mu$, and $\lambda_1(\tilde{\mu}) = \bar{w}(\tilde{\mu})$. Thus we can conclude that the mutational loss G is positive, the reduction principle holds, and the mean fitness is a decreasing function of $\tilde{\mu}$.

Discussion

We have shown that under constant-viability selection and in an effectively infinite haploid population at mutation-selection or migration-selection equilibrium, the stable population mean fitness increases if individuals with below-average fitness increase the rate at which variation is generated. Furthermore, modifier alleles that increase generation of variation in such individuals are favored by natural selection. These results apply as long as there is a chance for the variation-generating process to transform an individual with below-average fitness into one with above-average fitness (e.g. M in Eq. 2 is primitive).

We have given several examples of variation-generating processes for which this principle applies – namely mutation, migration, and learning (see *Models* section) – but our model may apply to other processes as well. For example, the *reduction principle* applies to ecological models of dispersal, and Gueijman et al. (2013) have demonstrated that even in homogeneous environments, fitness-associated dispersal increases the mean fitness of diploid populations and is favored by selection over uniform dispersal. Similarly, if the transmission fidelity of culturally-transmitted traits depends on the type or fitness of the transmitting individual, we expect that our results will hold (see Learning model).

Eq. 13 is a generalization of a result of Ram and Hadany (2012, Eq. 4). Ram and Hadany modeled the accumulation of mutant alleles in a population (see Mutation model 2). Using Eq. 36 in Appendix A and a recursion on the ratios of the reproductive values (see Ram and Hadany, 2012, eqs. A5-6), they concluded that at the mutation-selection balance, if individuals with below-average fitness ($w_k < \hat{\bar{w}}$) increase their mutation rate, then the population mean fitness will increase – a result generalized by our *Mean fitness principle* in Eqs. 14 and 23.

Our analysis focuses on populations at equilibrium. Nevertheless, it has been demonstrated that during adaptive evolution

(i.e., in non-equilibrium populations), a modifier that increases the mutation rate of maladapted individuals can be favored by selection (Ram and Hadany, 2012; Lukačišinová et al., 2017) and increase the adaptation rate (Ram and Hadany, 2014), and empirical evidence suggests that *stress-induced mutagenesis* is common in bacteria and yeast, and may be prevalent in plants, flies, and human cancer cells (Rosenberg et al., 2012; Fitzgerald et al., 2017). Similar theoretical results have been demonstrated for a modifier that increases the recombination rate in maladapted individuals (Hadany and Beker, 2003*a*,*b*).

Conclusions. Departures from the *reduction principle* for mutation, recombination, and migration rates usually involve fluctuating selection, non-equilibrium dynamics, or departures from random mating (see Carja et al. (2014) and references therein). Here we have provided another general example, which suggests that a modifier allele that causes individuals with below-average fitness to increase the rate at which variation is generated, will be favored by selection and will lead to increased population mean fitness.

Appendices

Appendix A

Caswell (1978) gave a *formula for the sensitivity of the population growth rate to changes in life history parameters*. In this formula, the *population growth rate* is the leading eigenvalue of the population transformation matrix T, the *life history parameters* are entries of T, and the *sensitivity* is the derivative of the former with respect to the latter. This is a useful formula (Caswell, 1978; Hermisson et al., 2002; Ram and Hadany, 2012; Otto and Day, 2007, ch. 10), and therefore we reproduce it here.

Lemma 1. *T be a non-negative matrix with leading eigenvalue λ and left and right eigenvectors $\hat{u}$ and $\hat{v}$ such that $\sum \hat{v}_k = 1$ and $\hat{u}^\intercal \hat{v} = \sum \hat{u}_k \hat{v}_k = 1$. Then the sensitivity of λ to changes in any element t of the matrix T is*

$$\frac{\partial \lambda}{\partial t} = \hat{u}^\intercal \frac{\partial T}{\partial t} \hat{v} \tag{36}$$

Proof. Using the lemma assumptions, $\lambda = \lambda \hat{u}^\intercal \hat{v} = \hat{u}^\intercal \lambda \hat{v} = \hat{u}^\intercal T \hat{v}$ and differentiating both sides we get $\partial \lambda = \partial (\hat{u}^\intercal T \hat{v})$. Using the product rule (once in each direction),

$$\begin{aligned} \partial(\hat{u}^\intercal T \hat{v}) = \partial \hat{u}^\intercal T \hat{v} + \hat{u}^\intercal \partial T \hat{v} + \hat{u}^\intercal T \partial \hat{v} = \\ \hat{u}^\intercal \partial T \hat{v} + \partial \hat{u}^\intercal \lambda \hat{v} + \lambda \hat{u}^\intercal \partial \hat{v} = \\ \hat{u}^\intercal \partial T \hat{v} + \lambda(\partial \hat{u}^\intercal \hat{v} + \hat{u}^\intercal \partial \hat{v}) = \\ \hat{u}^\intercal \partial T \hat{v} + \lambda \partial(\hat{u}^\intercal \hat{v}). \end{aligned} \tag{37}$$

Because $\hat{u}^\intercal \hat{v} = 1$, we have $\partial(\hat{u}^\intercal \hat{v}) = 0$ and $\partial \lambda = \hat{u}^\intercal \partial T \hat{v}$.

Appendix B

Remark 1 (Fisher's reproductive value)**.** *Let M be an irreducible column-stochastic matrix and D be a positive diagonal matrix. The entries of the left* Perron *eigenvector $\hat{u}$ of the matrix MD can be regarded as* Fisher's reproductive values *(Fisher, 1930, pg. 27)*

Fisher's reproductive values can be understood as follows (Grafen, 2006; Otto and Day, 2007, ch. 10). Consider the dynamics not of frequencies but of absolute population sizes such that the vector of the number of individuals of each type at time t is $n(t)$ and the corresponding frequencies are $f_k(t) = n_k(t)/\sum_i n_i(t)$. The dynamics are

$$n(t) = (MD)^t n(0). \tag{38}$$

Let $n(k,t)$ be the vector when the initial population is a single individual of type k. The dynamics are

$$n(k,t) = (MD)^t e_k, \tag{39}$$

where e_k is a vector with 1 at position k and 0 elsewhere.

The total population size at time t starting with type k is then

$$N(k,t) = \sum_i n_i(k,t) = e^\intercal (MD)^t e_k. \tag{40}$$

Now we can compare the sizes of populations based on what type they started with:

$$\frac{N(j,t)}{N(k,t)} = \frac{e^\intercal (MD)^t e_j}{e^\intercal (MD)^t e_k}. \tag{41}$$

Now write MD in its Jordan canonical form

$$A = V \Lambda U^\intercal, \tag{42}$$

where V is the matrix of right (column) eigenvectors of MD, $U^\intercal$ is the transposed matrix of left (row) eigenvectors of MD, where we can take $VU^\intercal = U^\intercal V = I$, and Λ is the diagonal matrix of eigenvalues of A (for a non-generic set of matrices M, the geometric and algebraic multiplicities of the eigenvalues of MD differ, and Λ will not be a diagonal matrix, a case we can ignore).

Hence,

$$\begin{aligned} N(k,t) = e^\intercal (MD)^t e_k = e^\intercal (V \Lambda U^\intercal)^t e_k = \\ e^\intercal V \Lambda^t U^\intercal e_k. \end{aligned} \tag{43}$$

For the ratio, we can divide Λ by $\lambda_1 = \rho(MD)$, the spectral radius of MD:

$$\begin{aligned} \frac{N(j,t)}{N(k,t)} = \frac{e^\intercal V \Lambda^t U^\intercal e_j}{e^\intercal V \Lambda^t U^\intercal e_k} = \\ \frac{e^\intercal V \mathrm{diag}\left[1, \left(\frac{\lambda_2}{\lambda_1}\right)^t, \cdots\right] U^\intercal e_j}{e^\intercal V \mathrm{diag}\left[1, \left(\frac{\lambda_2}{\lambda_1}\right)^t, \cdots\right] U^\intercal e_k}. \end{aligned} \tag{44}$$

Now take the limit $t \to \infty$. By assumption, MD is irreducible, so $\lambda_i < \lambda_1$ for all $i > 1$. Therefore,

$$\lim_{t \to \infty} \left(\frac{\lambda_k}{\lambda_1}\right)^t = 0 \tag{45}$$

for all $k > 1$, and

$$\lim_{t \to \infty} \frac{N(j,t)}{N(k,t)} = \frac{e^\intercal V (e_1 e_1^\intercal) U^\intercal e_j}{e^\intercal V (e_1 e_1^\intercal) U^\intercal e_k} = \frac{e^\intercal \hat{v} \hat{u}_j}{e^\intercal \hat{v} \hat{u}_k} = \frac{\hat{u}_j}{\hat{u}_k}. \tag{46}$$

The vector $\hat{u}$ is the left *Perron* eigenvector of MD, and $\hat{u}_k$ is k-th element of $\hat{u}$. This is why the value $\hat{u}_k$ can be interpreted as the *reproductive value* of type k: it is a weighting for the size of the population generated by a single individual of type k.

If we begin with a population at the equilibrium distribution $\hat{v}$, and ask what fraction of long-term descendants descended from type k at that time, we weight the equilibrium frequency $\hat{v}_k$ by the reproductive value $\hat{u}_k$, to get $\hat{u}_k\hat{v}_k$. $\{\hat{u}_k\hat{v}_k\}_k$ is a probability distribution, since

$$\sum_k \hat{u}_k\hat{v}_k = \hat{u}^\intercal\hat{v} = 1. \tag{47}$$

Hermisson et al. (2002) called this distribution the *ancestor* or *ancestral distribution*.

Appendix C

Lemma 2 (Stability of the equilibrium $\hat{v}$)**.** *If C and D are positive diagonal matrices, and M is primitive, then the equilibrium $\hat{v}$ of the system in Eq. 12 is globally stable.*

Proof. Let $A = (I - C + MC)D$ and denote the leading eigenvalue of A as $\hat{\bar{w}}$.

According to the *Perron-Frobenius theorem*,

$$\left(\frac{A}{\hat{\bar{w}}}\right)^t \to \frac{\hat{v}\hat{u}^\intercal}{\hat{u}^\intercal\hat{v}} = \hat{v}\hat{u}^\intercal, \tag{48}$$

where $\hat{u}$ and $\hat{v}$ are the left and right Perron eigenvectors of A such that

$$\sum_{k=1}^n \hat{v}_k = 1, \quad \hat{u}^\intercal\hat{v} = \sum_{k=1}^n \hat{u}_k\hat{v}_k = 1, \tag{49}$$

and $\hat{v}\hat{u}^\intercal$ is the *Perron projection* into the eigenspace corresponding to $\hat{\bar{w}}$.

Therefore, for any positive frequency vector f

$$\left(\frac{A}{\hat{\bar{w}}}\right)^t \cdot f \to \hat{v}\hat{u}^\intercal \cdot f = \alpha\hat{v}, \tag{50}$$

where $\alpha \in \mathbb{R}$. Using eqs. 49 and 50

$$\alpha = \alpha\hat{u}^\intercal\hat{v} = \hat{u}^\intercal\alpha\hat{v} = \hat{u}^\intercal\hat{v}\hat{u}^\intercal f = \hat{u}^\intercal f = \sum_{k=1}^n \hat{u}_k f_k > 0, \tag{51}$$

because f and $\hat{u}$ are positive vectors.

Now, let $\|x\| = \sum_{k=1}^n |x_k|$, so that $x_t \to x$ implies

$$\|x_t\| \to \|x\|, \tag{52}$$

as $t \to \infty$ and rewrite Eq. 2 as

$$f' = \frac{Af}{\|Af\|}. \tag{53}$$

It is easily seen that $f^t = A^t f / \|A^t f\|$. Hence

$$\frac{A^t f}{\|A^t f\|} = \frac{\frac{1}{(\hat{\bar{w}})^t}A^t f}{\frac{1}{(\hat{\bar{w}})^t}\|A^t f\|} = \frac{\left(\frac{A}{\hat{\bar{w}}}\right)^t \cdot f}{\left\|\left(\frac{A}{\hat{\bar{w}}}\right)^t \cdot f\right\|} \to \frac{\alpha\hat{v}}{\|\alpha\hat{v}\|} = \hat{v}, \tag{54}$$

since $\|\hat{v}\| = 1$ (eq. 49) and $\alpha \neq 0$ (eq. 51). This completes the proof.

Acknowledgements

This work was supported in part by the Department of Information and Computer Sciences at the University of Hawai'i at Mānoa, the Konrad Lorenz Institute for Evolution and Cognition Research, the Mathematical Biosciences Institute through National Science Foundation Award #DMS 0931642, the Stanford Center for Computational, Evolutionary and Human Genomics, and the Morrison Institute for Population and Resources Studies, Stanford University.

References

Agrawal, A. F. (2002), 'Genetic loads under fitness-dependent mutation rates', *Journal of Evolutionary Biology* **15**(6), 1004–1010.

Altenberg, L. (1984), A generalization of theory on the evolution of modifier genes, PhD thesis, Stanford University.

Altenberg, L. (2009), 'The evolutionary reduction principle for linear variation in genetic transmission', *Bulletin of Mathematical Biology* **71**(5), 1264–1284.

Altenberg, L. (2012*a*), 'The evolution of dispersal in random environments and the principle of partial control', *Ecological Monographs* **82**(3), 297–333.

Altenberg, L. (2012*b*), 'Resolvent positive linear operators exhibit the reduction phenomenon.', *Proceedings of the National Academy of Sciences* **109**(10), 3705–3710.

Altenberg, L. and Feldman, M. W. (1987), 'Selection, generalized transmission and the evolution of modifier genes. i. the reduction principle.', *Genetics* **117**(3), 559–72.

Altenberg, L., Liberman, U. and Feldman, M. W. (2017), 'Unified reduction principle for the evolution of mutation, migration, and recombination', *Proceedings of the National Academy of Sciences* **114**(12), E2392–E2400.

Borenstein, E., Feldman, M. W. and Aoki, K. (2008), 'Evolution of learning in fluctuating environments: When selection favors both social and exploratory individual learning', *Evolution* **62**(3), 586–602.

Boyd, R. and Richerson, P. J. (1985), 'Culture and the evolutionary process', p. 331.

Carja, O., Liberman, U. and Feldman, M. W. (2014), 'Evolution in changing environments: Modifiers of mutation, recombination, and migration', *Proceedings of the National Academy of Sciences* p. 201417664.

Caswell, H. (1978), 'A general formula for the sensitivity of population growth rate to changes in life history parameters', *Theoretical population biology* **14**, 215–230.

Cavalli-Sforza, L. L. and Feldman, M. W. (1981), *Cultural Transmission and Evolution: A Quantitative Approach*, 1 edn, Princeton University Press, Princeton, New Jersey.

Deakin, M. A. B. (1966), 'Sufficent conditions for genetic polymorphism', *The American Naturalist* **100**(916), 690–692.

Eshel, I. and Feldman, M. W. (1982), 'On evolutionary genetic stability of the sex ratio', *Theoretical Population Biology* **21**(3), 430–439.

Feldman, M. W. (1972), 'Selection for linkage modification: I. random mating populations', *Theoretical Population Biology* **3**(3), 324–346.

Feldman, M. W., Christiansen, F. B. and Brooks, L. D. (1980), 'Evolution of recombination in a constant environment.', *Proceedings of the National Academy of Sciences of the United States of America* **77**(8), 4838–41.

Feldman, M. W. and Liberman, U. (1986), 'An evolutionary reduction principle for genetic modifiers.', *Proceedings of the National Academy of Sciences* **83**(13), 4824–7.

Fisher, R. A. (1930), *The Genetical Theory of Natural Selection*, Clarendon Press, Oxford.

Fitzgerald, D. M., Hastings, P. and Rosenberg, S. M. (2017), 'Stress-induced mutagenesis: Implications in cancer and drug resistance', *Annual Review of Cancer Biology* **1**(1), 119–140.

Foster, P. L. (2007), 'Stress-induced mutagenesis in bacteria.', *Critical reviews in biochemistry and molecular biology* **42**(5), 373–97.

Grafen, A. (2006), 'A theory of fisher's reproductive value', *Journal of Mathematical Biology* **53**(1), 15–60.

Gueijman, A., Ayali, A., Ram, Y. and Hadany, L. (2013), 'Dispersing away from bad genotypes: the evolution of fitness-associated dispersal (fad) in homogeneous environments', *BMC Evolutionary Biology* **13**(1), 125.

Hadany, L. and Beker, T. (2003*a*), 'Fitness-associated recombination on rugged adaptive landscapes', *Journal of evolutionary biology* **16**(5), 862–870.

Hadany, L. and Beker, T. (2003*b*), 'On the evolutionary advantage of fitness-associated recombination.', *Genetics* **165**(4), 2167–79.

Hadany, L. and Otto, S. P. (2007), 'The evolution of condition-dependent sex in the face of high costs.', *Genetics* **176**(3), 1713–27.

Haigh, J. (1978), 'The accumulation of deleterious genes in a population - muller's ratchet', *Theoretical Population Biology* **14**(2), 251–267.

Hermisson, J., Redner, O., Wagner, H. and Baake, E. (2002), 'Mutation-selection balance: ancestry, load, and maximum principle.', *Theoretical population biology* **62**(1), 9–46.

Holsinger, K. E., Feldman, M. W. and Altenberg, L. (1986), 'Selection for increased mutation rates with fertility differences between matings.', *Genetics* **112**(4), 909–22.

Karlin, S. (1982), 'Classifications of selection migration structures and conditions for a protected polymorphism', *Evolutionary Biology* **14**(1953), 61–204.

Kimura, M. and Maruyama, T. (1966), 'The mutational load with epistatic gene interactions in fitness.', *Genetics* **54**(6), 1337–51.

Lessard, S. (1990), 'Evolutionary stability: One concept, several meanings', *Theoretical Population Biology* **37**(1), 159–170.

Liberman, U. and Feldman, M. W. (1986), 'Modifiers of mutation rate: a general reduction principle.', *Theoretical population biology* **30**(1), 125–42.

Lukačišinová, M., Novak, S. and Paixão, T. (2017), 'Stress-induced mutagenesis: Stress diversity facilitates the persistence of mutator genes', *PLOS Computational Biology* **13**(7), e1005609.

Otto, S. P. and Day, T. (2007), *A biologist's guide to mathematical modeling in ecology and evolution*, Princeton University Press.

Ram, Y. and Hadany, L. (2012), 'The evolution of stress-induced hypermutation in asexual populations', *Evolution; international journal of organic evolution* **66**(7), 2315–2328.

Ram, Y. and Hadany, L. (2014), 'Stress-induced mutagenesis and complex adaptation', *Proceedings of the Royal Society B: Biological Sciences* **281**(1792), 20141025–20141025.

Ram, Y. and Hadany, L. (2016), 'Condition-dependent sex: who does it, when and why?', *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**(1706), 20150539.

Rosenberg, S. M., Shee, C., Frisch, R. L. and Hastings, P. J. (2012), 'Stress-induced mutation via dna breaks in *Escherichia coli*: A molecular mechanism with implications for evolution and medicine.', *BioEssays* pp. 1–8.

Wallace, B. (1975), 'Hard and soft selection revisited', *Evolution* **29**(3), 465–473.