The evolutionary scaling of cellular traits imposed by the drift barrier

Michael Lynch[a,1]

[a]Biodesign Center for Mechanisms of Evolution, Arizona State University, Tempe, AZ 85287

Contributed by Michael Lynch, March 12, 2020 (sent for review January 10, 2020; reviewed by Reinhard Burger and Alan M. Moses)

Owing to internal homeostatic mechanisms, cellular traits may experience long periods of stable selective pressures, during which the stochastic forces of drift and mutation conspire to generate variation. However, even in the face of invariant selection, the drift barrier defined by the genetic effective population size, which is negatively associated with organism size, can have a substantial influence on the location and dispersion of the long-term steady-state distribution of mean phenotypes. In addition, for multilocus traits, the multiplicity of alternative, functionally equivalent states can draw mean phenotypes away from selective optima, even in the absence of mutation bias. Using a framework for traits with an additive genetic basis, it is shown that 1) optimal phenotypic states may be only rarely achieved; 2) gradients of mean phenotypes with respect to organism size (i.e., allometric relationships) are likely to be molded by differences in the power of random genetic drift across the tree of life; and 3) for any particular set of population-genetic conditions, significant variation in mean phenotypes may exist among lineages exposed to identical selection pressures. These results provide a potentially useful framework for understanding numerous aspects of cellular diversification and illustrate the risks of interpreting such variation in a purely adaptive framework.

cellular evolution | evolutionary theory | optimal phenotype | mutation bias | random genetic drift

The field of theoretical population genetics typically considers the evolutionary behavior of alleles in a generic way, usually assigning general selection coefficients and mutation rates to alternative alleles. This platform has been enormously successful at elucidating the combined roles of selection, mutation, and drift in the evolutionary dynamics of single loci and quantitative traits (1, 2). Nonetheless, there remains a disconnect between theory, the underlying biology of specific traits, and the constraints on key population-genetic parameters that exist in different molecular, cellular, or phylogenetic contexts. Often, a continuous distribution of mutational effects is assumed such that any phenotype is viewed as being accessible, but with no explicit connections with actual biological features.

The details appear to matter in a number of contexts. For example, many aspects of cell biology are effectively manifest in a digital rather than an analog manner. Transcription-factor binding sites (typically no more than 12 bp in length) can be defined in terms of the number of matches to the preferred motifs of their cognate transcription factors (3, 4). MicroRNA-binding site interactions and protein–protein interfaces can be viewed in a similar manner (5). Posttranslational modifications such as phosphorylation and ubiquitinylation (involved in many signal-transduction and protein-disposal pathways) can be described in terms of a small number of modified sites per protein (6). The numbers of carbon atoms and double bonds in the fatty-acid tails of lipids (generally fewer than 20 and 6, respectively) influence membrane width and fluidity (7), and some have argued that specific amino acids or nucleotides may be selected for on the basis of numbers of carbon, nitrogen, or sulfur atoms (8, 9). Many other examples could be given.

The restriction of simple molecular traits to discontinuous values may have unique evolutionary consequences. For example, the optimum energy of molecular binding or transfer for a particular trait may be unattainable unless it coincides with an integer multiple of the underlying granularity. If this is not the case, two allelic states straddling the optimum may have nearly the same fitness, resulting in an essentially neutral process of molecular evolution combined with a permanent state of suboptimal fitness. In addition, if certain suboptimal allelic states are more accessible by mutation, this can compete with the ability of selection to promote higher-fitness states. As is discussed below, this is virtually always the case, even in the absence of mutation bias.

The theory developed below has two central goals. First, given the common view that all phenotypic variation can be explained in terms of adaptation, it is desirable to consider the degree to which phylogenetic lineages are free to diverge under a regime of invariant selection pressure. Here, results will be given on the extent to which expected mean phenotypes are subject to scaling relationships (e.g., with organism size and/or population size) associated with the power of random genetic drift (the drift barrier) and on the degree to which populations exposed to identical forces of mutation, selection, and drift are free to wander across mean-phenotype space. Second, because the primary forces of evolution, most notably the mutation rate and random genetic drift, are nonindependent across the tree of life (10), consideration will be given to how such covariation influences patterns of variation.

General Theory

We start with a simple model with L equivalent sites (factors), each with two alternative allelic states, $+$ and $-$, contributing

Significance

Owing to internal homeostatic mechanisms, cellular traits may experience long periods of stable selective pressures. Nonetheless, drift and mutation still conspire to generate significant variation in mean phenotypes among phylogenetic lineages. Provided there are classes of mutations with sufficiently small effects, even in the face of constant selection, variation in genetic effective population sizes will result in gradients of mean phenotypes with respect to organism size across the tree of life. Mutation is an important determinant of such patterns, even in the absence of directional bias. Thus, a substantial amount of variation in cellular features may be a simple consequence of lineage-specific differences in the power of drift rather than a reflection of adaptive divergence.

Author contributions: M.L. designed research, performed research, analyzed data, and wrote the paper.

Reviewers: R.B., University of Vienna; and A.M.M., University of Toronto.

The author declares no competing interest.

[1]Email: mlynch11@asu.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2000446117/-/DCSupplemental.

First published April 28, 2020.

positively and negatively to the trait (Fig. 1). Under this model, for all but the two most extreme genotypes, a multiplicity of functionally equivalent classes exists with respect to the number of positive alleles, m, defined by the binomial coefficients. As an example, for the case of $L=4$, there are five genotypic classes ($m=0$, 1, 2, 3, 4, and 5), with multiplicities 1, 4, 6, 4, and 1, respectively (Fig. 1). With equivalent fitness for all members (haplotypes) within a particular class, this variation in multiplicity of states plays an important role in determining the long-term evolutionary distribution of alternative classes. The site-specific per-generation mutation rates from the $-$ to the $+$ state, and vice versa, are defined as u_{01} and u_{10}, respectively. Unless stated otherwise, a haploid, nonrecombining population will be assumed. The absolute population size consists of N individuals, so that a de novo mutation has initial frequency $1/N$. However, the effective population size, N_e, which is generally no greater than N and often considerably smaller, governs the magnitude of random genetic drift and as discussed below is defined by the remaining features of the population-genetic environment (including the mutation bias and the strength of selection).

This type of biallelic model has been widely exploited in theoretical studies of the genetic structure of quantitative (multilocus) traits (2), including cell-biological features (4, 5). Most quantitative-genetic theory on such systems has focused on the equilibrium within-population variation under selection–mutation balance in infinite-sized populations. However, in finite populations, the stochasticity of mutation and drift ensures that genotype distributions are not fixed. Instead, the mean genotype is expected to wander over time within limits dictated by the strength of selection and the structure of the underlying genetic system. The goal here is to determine the long-term probability distribution of genotypic means defined by the joint forces of selection, mutation, and random genetic drift. Justification of this quasi–steady-state view derives from the fact that many internal cellular traits have functions (and cytoplasmic environments) that likely remain relatively stable for millions of years (even in the face of a changing external environment).

A Single Biallelic Locus. As a point of departure, the case of a single locus with reversible mutation will highlight several key points and provide a useful backdrop for the evaluation of more complex traits. Allele **A** (denoted as state 1 below) is taken to have a fractional selective advantage s over allele **a** (denoted as state 0). For this limiting case, previous work provides an exact solution for the equilibrium mean frequencies of alleles (*SI Appendix*). However, because the final expression is a bit cumbersome mathematically, it is common in population genetics to use one of two alternative approximations.

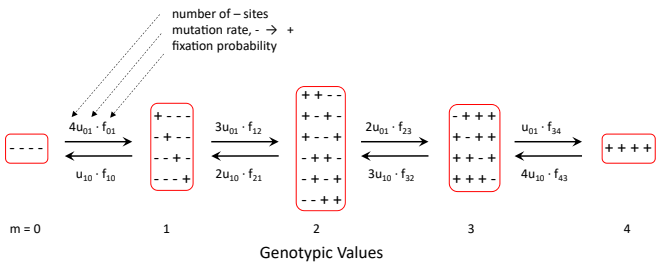

Fig. 1. Schematic for the transition rates between adjacent classes under the sequential-fixation model for the case of $L=4$ sites. This schematic readily generalizes to any value of L. u_{01} and u_{10} are the mutation rates from $-$ to $+$ allelic states, and vice versa, and f_{ij} is the probability of fixation of a newly arisen mutation to allele j from a background of i, defined as *SI Appendix*, Eq. **S5**. The number of $-$ alleles in a class is denoted by m, and except for the two extreme classes ($m=0$ and $m=4$), there are multiple equivalent genotypic states within each class of genotypic values.

One extreme view is that drift is weak enough relative to mutation and selection that the population can be treated as effectively infinite in size, yielding *SI Appendix*, Eqs. **S3a** and **S3b**. A second approach assumes a situation in which, relative to the power of drift, mutation rates are low enough and the strength of selection high enough that polymorphisms are almost never present, with fixation of alternative monomorphic states being the norm, and the long-term mean frequency of the beneficial allele being

$$p_1^* \simeq \frac{\beta e^S}{1+\beta e^S}, \qquad [1]$$

where $S=2N_e s$ is the strength of selection scaled to drift, and $\beta = u_{01}/u_{10}$ is the ratio of mutation rates. Under this weak-mutation/strong-selection scenario, referred to below as the sequential-fixation model, the equilibrium distribution depends only on the product of the two ratios of rates, not on the absolute values of their components. Although the concern here is with the long-term probabilities of alternative states at a particular locus, Eq. **1** was derived previously to describe the genome-wide expected frequencies of codons for particular amino acids (11, 12). Note that under neutrality ($s=0$), all three approaches yield identical results, with the mean frequency of the **A** allele being the neutral expectation $\eta = u_{01}/(u_{01}+u_{10})$.

Although the deterministic and sequential models have very different implications for the standing levels of variation within populations, often in comparative biology, just a single genome will be sampled per taxon, rendering the matter of within-population variation moot. Thus, it is desirable to know the domains in which these two approximations (if either) provide the most reasonable approximation of actual long-term mean allele frequencies. Comparison with the general analytical solution, *SI Appendix*, Eq. **S2**, allows several general conclusions. First, the sequential model begins to substantially overestimate the frequency of the beneficial allele once $N_e u_{01}$ exceeds 0.01 (Fig. 2). These deviations are particularly large when outside the realm of effective neutrality, i.e., for $N_e s > 0.1$. Second, for weak to moderately strong selection ($0.1 < N_e s < 1.0$), a critical point near $N_e u_{01} = 0.1$ separates domains in which the sequential vs. deterministic model fits the data more closely. At this crossover point, both of the simple models overestimate the actual mean frequency of the beneficial allele, often substantially so. Third, the predictions of the deterministic model converge on the results of the general model once the number of beneficial mutations ($N_e u_{01}$) entering the population exceeds 1.0 per generation. As the population-wide mutation rates ($N_e u_{01}$ and $N_e u_{10}$) increase beyond this point, mutation pressure begins to overwhelm both selection and drift, and the asymptotic frequency of the beneficial allele converges on the neutral expectation, η.

Two other points are illustrated in Fig. 2. First, in the presence of mutation bias, there can exist a wide range of population sizes in which the most common allelic state does not match the optimal state; e.g., $\widetilde{p}_1 < 0.5$. Second, situations may be common in which there is ample phylogenetic variation in allelic state despite the constancy of selection. For example, in the domain of the sequential model, when the directional forces of mutation and selection balance such that $\beta e^S = 1$, the mean frequencies of the alternative allelic states will be equal to 0.5. In the absence of knowledge on mutation bias, such cases could easily be misinterpreted as implying neutral genotypes, when in fact the population is under persistent selection in the opposite direction of mutation bias.

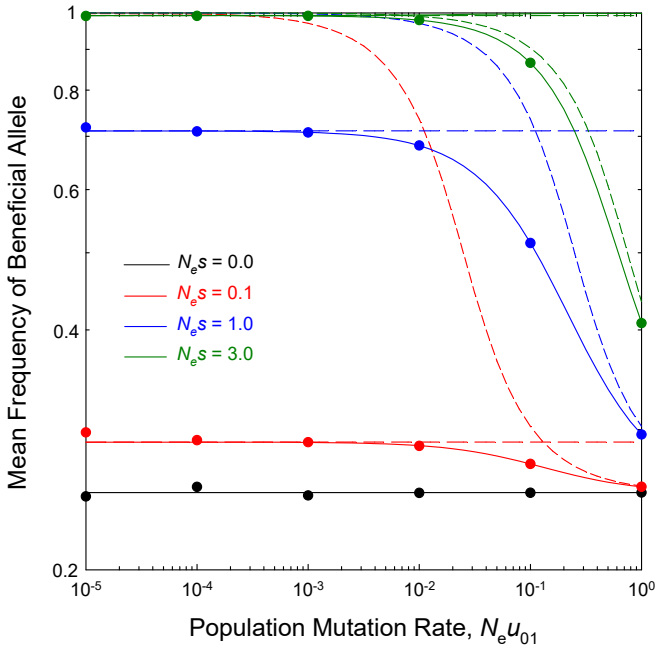

Fig. 2. Average beneficial allele frequency (long-term average) for a single-locus model, under the general model (*SI Appendix*, Eq. **S2**, solid lines), the infinite population size model (*SI Appendix*, Eqs. **S3a** and **S3b**, short-dashed lines), and the sequential-fixation model (Eq. **1**, long-dashed lines). Results from Wright–Fisher simulations are given as solid points for the case in which the actual and effective population sizes are equal ($N_e = N = 1{,}000$). The mutation rate of beneficial to deleterious alleles (u_{10}) is three times the reciprocal rate (as would be the approximate case for a single optimal nucleotide at a site). The infinite population size results use the selection coefficient inferred by the constant product *Ns* for each curve, so that *s* in this case scales inversely with *N*.

Two Sites. Proceeding to more complex traits, it becomes necessary to specify a function relating the genotypic states to fitness. Although the concepts developed hereafter are general with respect to the form of the fitness function, here we will assume a Gaussian (bell-shaped) function such that individuals in genotypic class m have fitness

$$W_m = e^{-(m-\theta)^2/(2\omega^2)}, \qquad [2]$$

where θ is the optimum phenotypic value, and ω is a measure of the width of the fitness function (analogous to the SD of a normal distribution). Selection is purely directional if $\theta = 0$ or L, and neutrality is approached as $\omega \to \infty$. Although m is confined to integer values, θ need not be, and if θ is outside of the $(0, 2)$ range, the optimum is unattainable.

As a central goal is to determine the relationship between the expected genotype distribution and the effective population size (N_e), it is desirable to perform analyses with realistic parameter values for N_e and the mutation rate. Across the tree of life, N_e generally falls in the range of 10^4 to 10^9, and the mutation rate per nucleotide site scales negatively with the $\sim$0.76 power of N_e (10, 13). Thus, the following analyses were performed under the assumption of a deleterious-mutation rate per site (which might be a cluster of adjacent nucleotides) of 10^{-7} at $N_e = 10^4$, such that $u_{10} = 0.00011 N_e^{-0.76}$, which is approximately 10 times the known rate per nucleotide site. As implied above and shown further below, provided the mutation rates are sufficiently low, it turns out that this scaling has no effect on the equilibrium distribution, which depends only on the ratio $\beta = u_{01}/u_{10}$. Unless stated otherwise, the beneficial rate (u_{01}) is set to 10% of the former one.

With two sites, there are three possible genotypic classes, $m = 0, 1, 2$, with the phenotypically equivalent $+-$ and $-+$ states being lumped into the $m = 1$ class. For reasons described in *Moderately Complex Traits*, the long-term equilibrium frequencies are given by

$$\widetilde{p}_m = C \cdot \binom{L}{m} \beta^m e^{2N_e s_m}, \qquad [3]$$

where the normalization constant C is equal to the reciprocal of the sum of the terms to the right of C for $m = 0$ to 2, and the binomial term takes on values of $1, 2, 1$ for $m = 0, 1, 2$. The indexed selection coefficients, s_m, are measures of the class-specific deviations of fitness from some reference genotype (e.g., the class with the highest fitness, in which case $s_m = 1 - W_m$, or class 0); the specific reference does not matter, as it cancels out through the normalization constant. The mean phenotype

$$\mu_m = \sum_{m=0}^{2} m \cdot \widetilde{p}_m \qquad [4]$$

reduces to 2η in the case of neutrality. (Throughout, $\widetilde{p}$ is used to indicate an equilibrium frequency and μ to denote a mean, in this case of m, the number of + alleles.)

This expansion to a second site introduces complexities not encountered with the one-site model, all of which can be understood by reference to Eq. **3**. First, for the case of $\theta = 1.5$, where the optimum is straddled by the class 1 and 2 genotypes, the genotypic mean (μ_m) never reaches the optimum, even at very large N_e, and instead remains much closer to $\mu_m = 1$ (Fig. 3). This bias results because although the class 1 and 2 genotypes have equivalent fitness, mutation pressure toward $-$ alleles weights the frequency of class 1 by a factor of 2β (the 2 being the multiplicity of this class), but class 2 by the smaller factor of β^2.

Second, for the case in which $\theta = 2$ (pure directional selection), there is a progressive succession of the prevailing genotype classes with increasing N_e. When N_e is sufficiently low to impose effective neutrality, class 0 predominates owing to the mutation bias toward $-$ alleles. With increasing N_e, selection becomes more effective at promoting class 1, but there remains effective mutation pressure against class 2. Finally, with very large N_e, selection becomes efficient enough to drive class 2 to near fixation, thereby decreasing the incidence of class 1. These results show that, in the face of a constant pattern and strength of selection, the genotypic mean can exhibit a considerable gradient with N_e owing entirely to changes in the power of drift and also that appreciable incidences of all three genotypic classes can be expected over time in lineages with intermediate N_e.

Moderately Complex Traits. We now turn to the general case of arbitrary L, starting with the sequential-fixation view that mutations are rare enough that evolution proceeds in an essentially stepwise manner, with incremental changes being restricted to adjacent states. The absolute flux rates between adjacent genotypic classes are then equal to the products of the expressions on the arrows in Fig. 1, where the numerical coefficients depend on the numbers of $-$ and $+$ sites within each class. Because the absolute population size N influences all mutational flux rates in the same way, it is omitted as a prefactor, but both N and N_e influence the equilibrium solution via the fixation probabilities, defined as *SI Appendix*, Eq. **S5**.

This linear sequential model has a relatively simple solution. Provided there are nonzero connections between all adjacent states, the steady-state frequencies are proportional to the products of all of the coefficients pointing upwardly and downwardly

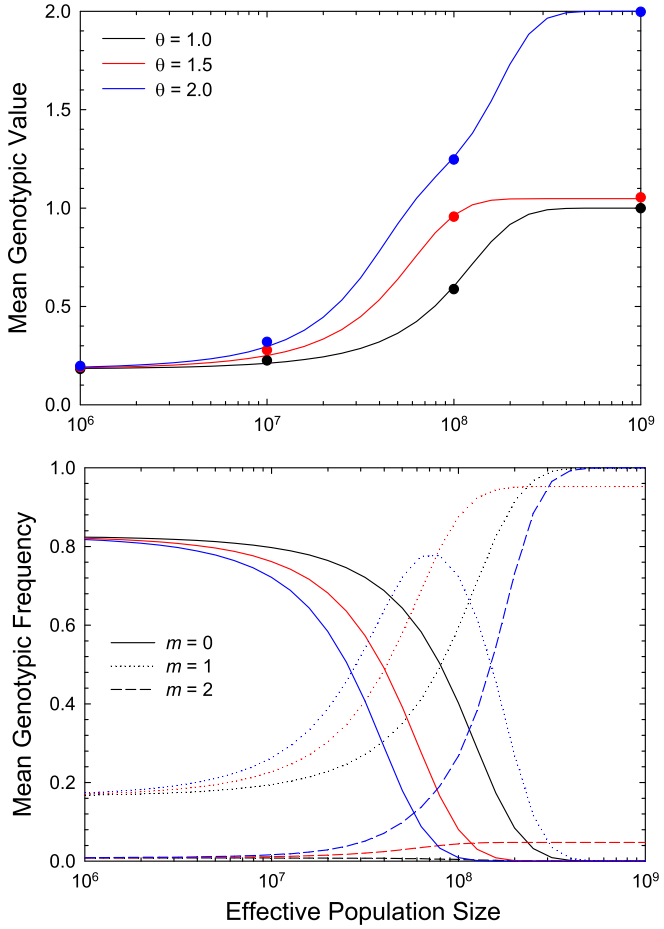

Fig. 3. The response of the long-term mean genotypic state (number of + alleles, *Upper*) and the underlying mean class frequencies (*Lower*) over a gradient of effective population sizes for a two-locus, two-allele model. The mutation rate to beneficial alleles is 10% of that to deleterious alleles. Results are given for three phenotypic optima, with the width of the fitness function $\omega = 5{,}000$. In *Lower* panel, for each color-coded optimum, the mean frequencies of the three genotypic classes (0, 1, 2) are given as solid, dotted, and dashed lines, and for each N_e sum to 1.0. Data points in *Upper* panel were obtained by computer simulations of a Wright–Fisher model, showing the near perfect agreement with the sequential-model approximation.

toward the state of interest. Prior results (4) lead to Eq. **3**, which shows that the equilibrium frequencies of the alternative genotypes are functions of three factors: 1) the multiplicity of configurations, as defined by the binomial coefficients; 2) the ratio of mutation rates; and 3) the strength of selection scaled by the power of random genetic drift. All other things being equal, the within-class multiplicity magnifies the likelihood of residing in such a state. However, for $\beta \neq 1$, the net effect of the multiplicity is further modified by the mutation bias term, β^m.

Under neutrality (denoted by n), $s_m = 0$ for all m, and Eq. **3** simplifies to a binomial distribution

$$\widetilde{p}_{n,m} = \binom{L}{m} \eta^m (1-\eta)^{L-m} \qquad \textbf{[5]}$$

with $\eta = u_{01}/(u_{01}+u_{10})$ being the expected frequency of + alleles at each site. In this case, the long-term mean and variance of the trait are $\mu_N = L\eta$ and $\sigma_N^2 = L\eta(1-\eta)$, respectively. Comparison of Eqs. **3** and **5** shows that the steady-state probabilities of genotypes under selection are simple transformations of the neutral expectations, with each class being weighted by the exponential function of the scaled strength of selection e^{S_m}, with $S_m = 2N_e s_m$,

$$\widetilde{p}_m = C \cdot \widetilde{p}_{n,m} \cdot e^{S_m}. \qquad \textbf{[6]}$$

Provided η is not overly close to 0.0 or 1.0, the neutral distribution is approximately normal, and if $0 \ll \theta \ll L$, the steady-state distribution of mean phenotypes with selection will also approach normality, with mean

$$\mu_m \simeq \frac{(\theta/\sigma_S^2) + (L\eta/\sigma_N^2)}{(1/\sigma_S^2) + (1/\sigma_N^2)} \qquad \textbf{[7a]}$$

and variance

$$\sigma_m^2 \simeq \frac{1}{(1/\sigma_S^2) + (1/\sigma_N^2)}, \qquad \textbf{[7b]}$$

where $\sigma_S^2 = \omega^2/(2N_e)$ (14). Under these conditions, the grand mean is the average of the expectations under mutation and selection alone, with each component weighted by the inverse of the variance under the relevant conditions.

As $\sigma_S^2 \to \infty$, which implies a flatter fitness function, the mean and variance converge on the expectations for a purely neutral process, $L\eta$ and σ_N^2. As $\sigma_N^2 \to \infty$, which implies a flattened distribution under mutation and drift alone, the mean and variance converge on the expectations for a purely selection-driven process, θ and σ_S^2. The pivot point separating these two domains is $N_e = \omega^2/[2L\eta(1-\eta)]$, showing that an x-fold increase in the width of the fitness function shifts the critical point to an x^2-fold larger N_e.

Eq. **6** predicts the steady-state distribution of the most recently fixed state, and one must acknowledge that some small amount of polymorphism exists in the time intervals between fixation. Thus, it is desirable to know how closely the preceding expressions represent the true sampling probabilities when polymorphism is allowed for. To evaluate this issue, extensive Wright–Fisher simulations were carried out, with recursive episodes of selection, mutation, and drift, for long enough periods to achieve highly precise estimates of the distributions of means to compare with the analytical expectations for $\widetilde{p}_m$ (*SI Appendix*). Because violations of the sequential model will be a function of the absolute number of mutations entering the population per generation, it is essential to perform such evaluations with realistic parameter estimates for N_e and the mutation rate, and the scaling relationship between the mutation rate and N_e noted above was adhered to. The consistency of the simulation results with the analytical approximations, even with L as high as 50 (Fig. 4, *Left*), justifies the use of Eq. **6** up to this level of granularity.

This expansion to larger numbers of sites again shows significant scaling of mean genotypic values with N_e for a wide range of conditions, upholding the conclusion that situations exist in which the optimal phenotype is rarely achieved even in very large populations (Fig. 4, *Left*). Moreover, for fixed selection and mutation functions, the direction of scaling of the mean genotype with N_e depends on the complexity of the underlying trait (L). If the genotypic mean in the absence of selection ($L\eta$) exceeds the genotypic optimum, there will be negative scaling of μ_m with N_e, regardless of the direction of mutation bias. In addition, there will always be an intermediate level of L, such that the mean under neutrality fortuitously equals the selective optimum, i.e., $L\eta = \theta$, at which point there is no response of the mean phenotype to N_e. Finally, for particular population-genetic

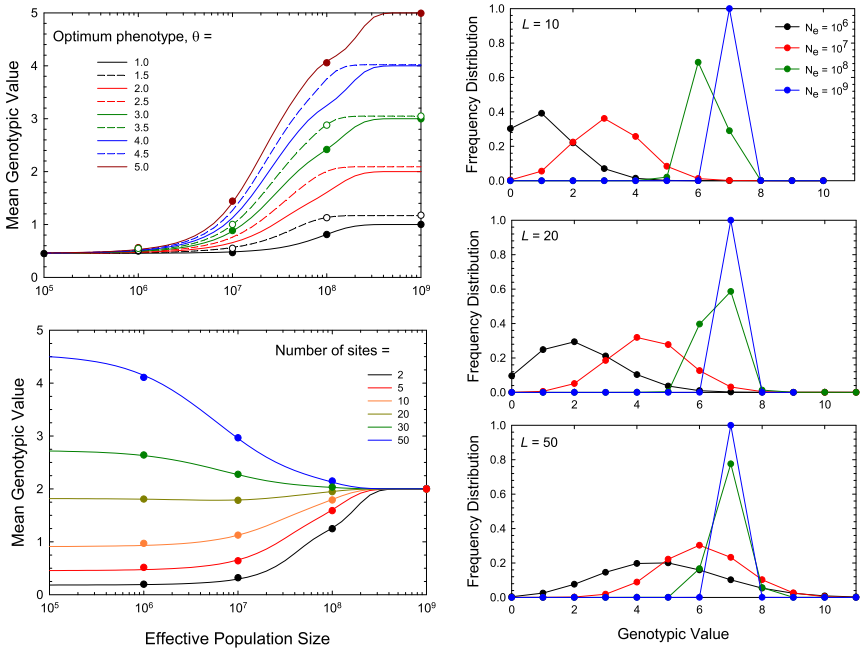

Fig. 4. (*Left*) Response of the mean genotypic value to change in the effective population size, for the case in which the fitness function has width $\omega = 5{,}000$, and the ratio of mutation rates is $u_{01}/u_{10} = 0.1$. (*Upper Left*) The effects of increasing the optimum phenotype (θ), for the case in which there are $L = 5$ sites. (*Lower Left*) The effects of increasing the number of sites from $L = 2$ to 50, for the case in which the optimum phenotype is $\theta = 2$. In both panels, the color-coded plotted points were obtained by computer simulations of a Wright–Fisher model, showing the near perfect agreement with the sequential-model approximation. (*Right*) Equilibrium genotype distributions (number of + alleles) for situations in which $L = 10$ (*Top*), 20 (*Middle*), and 50 (*Bottom*), with optimum genotypic value $\theta = 7.0$ and width of the fitness function $\omega = 5{,}000$. Results from Eq. **6** are given for four effective population sizes.

environments, multiple genotypic classes often have appreciable expected frequencies, and depending on the direction of mutation bias, these can be both above and below the optimum (Fig. 4, *Right*).

Large Numbers of Loci. When large numbers of loci contribute to a trait, the total number of mutations entering the population per generation will often exceed 0.01, violating the conditions necessary for the sequential model to closely approximate the steady-state distribution. One such example is the cellular growth rate, which likely depends on essentially every nucleotide in the genome in some way, e.g., the simple bioenergetic cost of a nucleotide pair (15). As a first-order generalization, one expects natural selection to relentlessly promote an organism's ability to convert energy and other limiting resources into biomass, and persistent directional selection may apply to many other traits such as catalytic rates and error minimization. To analyze the evolution of traits under such selective conditions, an exponential fitness function with independent mutational effects was evaluated,

$$W_m = (1-s)^{(L-m)}, \qquad [8]$$

where m is the number of beneficial alleles in the genome.

For random genetic drift to impose a significant barrier to the evolution of such a trait, there must be a substantial pool of deleterious mutations with small enough effects that they can drift to fixation in species with small N_e, and this process will be further facilitated if mutations are biased in the negative direction. Numerous lines of evidence are consistent with both conditions. First, studies of serially bottlenecked mutation-accumulation (MA) lines across diverse species consistently reveal a slow per-generation decline in growth rate and other fitness traits, in accordance with a strong bias of mutations toward deleterious effects (16, 17). Statistical inferences based on the distribution of MA-line performance imply highly skewed distributions of fitness effects, with the modes for both deleterious and beneficial mutations indistinguishable from zero and the bulk of the distributions having absolute effects $<1\%$ (18–20). Second, indirect inferences derived from studies on the site-frequency spectra of segregating alleles commonly suggest that 10 to 40% of mutations in diverse organisms have deleterious effects smaller than 10^{-5}, with the mode of this pool again being near (if not at) 0.0 (21–25). There are theoretical reasons for expecting this to be the case (26), and as many of these studies focus only on the nonsynonymous sites in protein-coding genes, the true distribution of effects can be expected to be even more skewed toward near-zero values. Third, bioenergetic considerations of the costs of small nucleotide insertions, which typically comprise $\sim 10\%$ of de novo mutations (27), imply fractional reductions in fitness typically far below 10^{-4} (15). Thus, the existence of large pools of mutations with deleterious effects small enough to allow fixation in some lineages but large enough to ensure removal by selection in others is not in doubt.

These observations allow some simple qualitative statements about the limits to selection. Sites with effects (s) smaller than $1/N_e$ will be highly susceptible to the vagaries of genetic drift and have expected deleterious-allele frequencies near $u_{10}/(u_{10}+u_{01})$. This quantity depends only on the mutation-rate ratio ($\beta = u_{01}/u_{10}$), not on the absolute mutation rate. In contrast, when $N_e \gg (1/s)$, selection overwhelms drift, and the expected frequency of deleterious alleles is approximately $u_{10}/(u_{10}+u_{01}+s)$, which is near zero when the strength of selection greatly exceeds that of mutation. The central issue is then the degree to which specific classes of deleterious alleles move from the domain of accumulation by mutation pressure to the domain of effective purging by selection as the population size increases. A complication that emerges with large L is that the effective population size (N_e) is often orders of magnitude below the actual

population size (N) owing to selective interference between simultaneously segregating mutations, which greatly elevates the level of N required for the effective purging of mildly deleterious alleles.

Although the ultimate goal is to derive analytical expressions for the distribution of mean phenotypes in this large-L domain, the system is sufficiently complex that it is necessary to perform Wright–Fisher simulations to determine the degree of validity of any equations that emerge. For these analyses, for each L, it was assumed initially that $s = 1/L$, which is roughly consistent with the empirical observation that the frequency of mutation types is a decreasing function of the selective effects. Each set of simulations involved a single fixed value of L: $L = 10^4$, $s = 10^{-4}$; $L = 10^5$, $s = 10^{-5}$; and $L = 10^6$, $s = 10^{-6}$. Example outcomes of the grand means are given in Fig. 5, along with the limiting expectations for single sites (from *SI Appendix*, Eq. **S2**, which assumes no background interference from linkage disequilibrium). These results highlight the expectations noted above—at sufficiently small N, the mean frequency of + alleles converges on the neutral expectation, whereas at large N, it converges on frequencies close to 1.0. Notably, however, the gradient in the mean frequency of + alleles is much shallower in the case of multiple linked loci than in the ideal single-locus case, particularly when s is small.

The key remaining challenge is to obtain an approximate analytical expression for the equilibrium behavior in Fig. 5. As outlined in *SI Appendix*, despite some intriguing relationships, a solution from first principles has not been obtained. Nonetheless, the general scaling behavior in Fig. 5 suggests a formula of the form

$$\widetilde{\mu}_m \simeq \frac{u_{01} L}{u_{01} + u_{10} f(s, N)}, \qquad \textbf{[9]}$$

where $f(s, N)$ is a function of s and N. For Eq. **9** to yield appropriate behavior, $f(s, N)$ must asymptotically approach 0 and 1 in the limits of large and small N. A simple expression that allows for such behavior, $f(s, N) = e^{-2Ns}$, is attractive because this yields the known result for the sequential model in which selection operates on a set of single sites without selective interference (Eq. **1**), and Charlesworth (ref. 28, equation A13) suggested this as a solution. However, whereas Eq. **1** yields correct results in the limits of $sN \gg 1$ and $\ll 1$, the predictions are too high at intermediate sN. The likely reason for the poor fit in this region of parameter space is that the genetic effective population size (N_e) can be substantially lower than N, owing to parallel selection operating on multiple polymorphic sites.

To gain insight into this matter, Eq. **1** can be rearranged to yield the effective population size necessary to yield the mean frequency of + alleles, $\widetilde{p}_1 = \widetilde{\mu}_m / L$, observed with computer simulations,

$$N_e = (1/2s) \ln \left(\frac{\widetilde{p}_1}{\beta(1 - \widetilde{p}_1)} \right). \qquad \textbf{[10]}$$

As can be seen in Fig. 6, there appear to be three central determinants of N_e using this formulation. First, provided the strength of selection is at least $0.1/N$, N_e/N scales negatively with the $\sim 3/4$ power of Ns, reflecting the increased involvement of selective interference. Second, N_e/N increases weakly as the incidence of beneficial mutations declines, scaling negatively with the $\sim 1/4$ power of β, presumably reflecting the reduced competition among simultaneously segregating beneficial mutations. Third, there is a further depression in N_e/N with increasing L, presumably reflecting more background variation in fitness resulting from elevated numbers of simultaneously segregating polymorphisms.

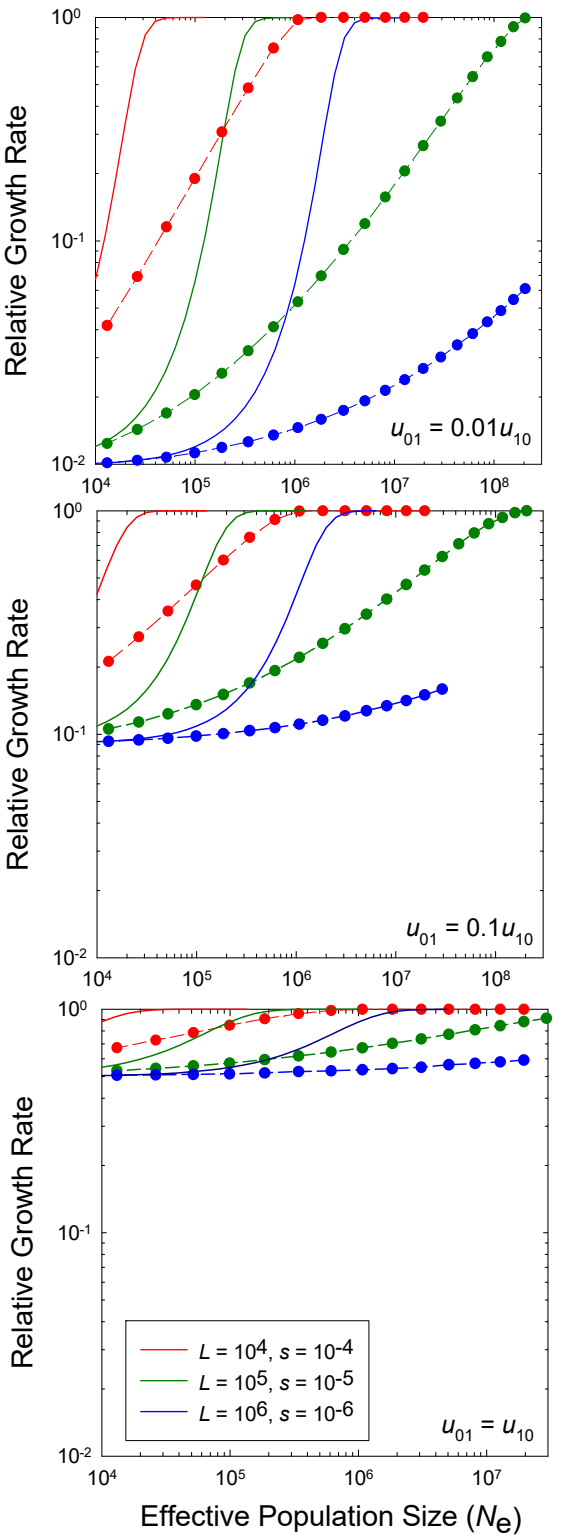

Fig. 5. The equilibrium mean frequency of + alleles (i.e., $\widetilde{\mu}_m$ relative to the maximum possible value, L) as a function of the absolute population size. Results are given for three different levels of mutation bias for three numbers of loci (L) with equivalent additive effects on the growth rate and multiplicative effects on fitness. Absolute mutation rates are defined by the empirical scaling relationship noted in the text. Analytical results (solid lines) are given for the case of free recombination, and simulation results (dashed lines and data points) for the case of complete linkage. Throughout, $s = 1/L$, the inverse of the number of factors contributing to the trait.

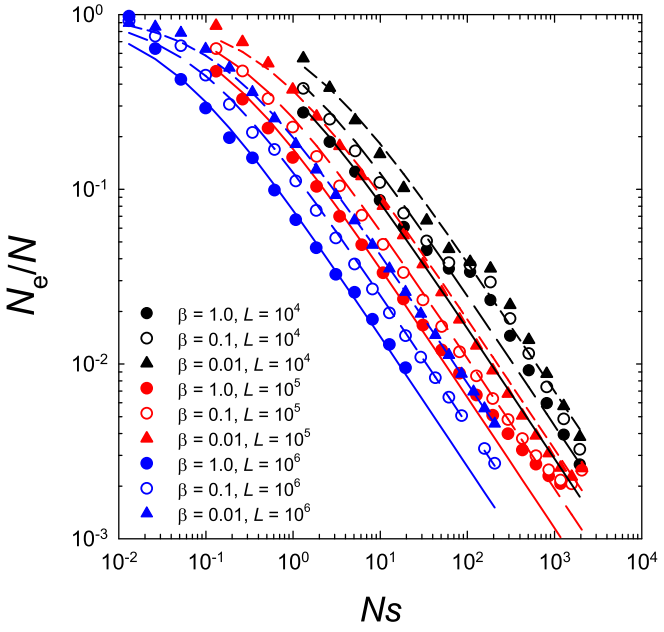

Fig. 6. Depression in the effective population size as a function of *Ns*, β, and $L = 1/s$. Points are estimates from the average behavior in simulations, whereas the solid lines map the expected behavior based on Eq. **11**. As noted in the text, β is varied by keeping the deleterious mutation rate u_{10} constant (although varying at each population size) and modifying u_{01}.

Regrettably, it has not been possible to derive an analytical expression encapsulating these three scaling relationships. However, by inspection, all three can be included in a single expression,

$$N_e \simeq \frac{N}{1 + [2.5^{\log_{10}(L)-1} - \ln \beta] \cdot \beta^{1/4} \cdot (Ns)^{3/4}}, \qquad [11]$$

which is consistent with observations for the full range of parameter space (Fig. 6), except for $Ns > 100$ and $s > 10^{-5}$, in which case the population behaves nearly deterministically and N_e/N is of minor significance. The key scaled selection parameter in Eq. **1**, $S = 2N_e s$, can then be approximated by multiplying Eq. **11** by $2s$, and provided $Ns > 1$, this reduces further to

$$S \simeq \frac{5(Ns/\beta)^{1/4}}{2.5^{\log_{10}(L)} - 2.5 \ln \beta}. \qquad [12]$$

Together, Eqs. **11** and **12** show that provided Ns is not too small, then S scales only slowly with Ns and the mutation bias β, with the key parameter being $(Ns/\beta)^{1/4}$. The effect of increasing the number of sites L is also nonlinear, with a 100-fold increase in L decreasing S only by a factor of $\simeq$6. Finally, decreasing β by a factor of 100 increases S by a factor $\simeq$3, showing that the bulk of the interference in promoting beneficial genotypes derives from simultaneously segregating beneficial alleles.

With Eq. **11** in hand, it then becomes possible to express the equilibrium expectation for the mean fraction of sites occupied by favorable alleles by substituting for N_e in Eq. **1**. This yields predictions that are generally within 10% of simulated values for the full range of parameter space and that are particularly accurate with large numbers of loci with small individual selective effects (*SI Appendix*, Fig. S2). Thus, although there must be additional second-order terms involved, Eq. **11** appears to capture the collective effects of selective interference, at least in the context of mutations with equal effects.

Notably, although the preceding analyses relied on a particular scaling of the mutation rate with N, the evolved genotypic mean appears to depend only on the mutational bias β. In addition, results for a wide range of L indicate that the approximate scalings indicated in Eqs. **11** and **12** are not a peculiarity of the use of $s = 1/L$ in the preceding analyses (*SI Appendix*, Fig. S3). Finally, although the exact scaling relationships are different, a similarly derived expression for the case of a half-Gaussian fitness function is given in *SI Appendix*.

These results also provide an interesting contrast to the "infinitesimal" model widely used in quantitative genetics (2, 29). Under this model, the genotypic values of traits are assumed to be influenced by alleles with infinitesimally small effects distributed over an infinite number of loci, such that the mean phenotypes of traits evolve with effectively no change in allele frequencies at individual loci. Despite its mathematical elegance, this model is inconsistent with the fact that all traits are encoded by finite numbers of genetic loci (bounded by the size of the genome) and more notably has the puzzling feature that directional selection is possible despite all allelic effects being miniscule. In contrast, the results in Fig. 5 show that with finite numbers of loci, subdivision of total fitness into increasingly fine contributions progressively diminishes the ability of directional selection to advance a mean phenotype, as expected under the principles of effective neutrality, illustrating the limitations of the mathematically convenient infinitesimal model.

Effects of Recombination. To this point, the chromosomal segments being modeled have been assumed to be nonrecombining. Given the population sizes employed here, the introduction of recombination into the simulations would be extremely demanding computationally, but an argument introduced by Good et al. (30) provides a simple way to address the matter qualitatively. They demonstrated that the overall behavior of a recombining system with multiple sites simultaneously segregating can be treated by subdividing the genome into blocks that are effectively nonrecombining on the timescale of the coalescent.

From a prior survey (31), we know that the average recombination rate between adjacent bases (c_0) is in the range of 10^{-7} to 10^{-5} for unicellular eukaryotes and 10^{-9} to 10^{-7} for multicellular species. The exact numbers depend on the overall genome size and numbers of chromosomes, as there is generally on the order of one crossover per chromosome arm. The mean time to coalescence of alleles within a population is on the order of $2N_e$ (haploids) to $4N_e$ (diploids) generations, so over this duration the expected number of recombination events within a span of length L_b is $2c_o L_b N_e$ to $4c_o L_b N_e$. Thus, because N_e is commonly in the range of 10^7 to 10^8 for unicellular eukaryotes (10, 13), a span as small as 2 bp will have a high probability of experiencing at least one recombination event in the coalescence time for individual alleles, suggesting that Eq. **1** might serve as a reasonable description of the expected allele frequency in many obligately sexual unicellular populations. However, most unicellular species primarily reproduce asexually, and allowing for recombination every x generations would expand the average linkage-block length by a factor of x. With N_e in multicellular species generally falling in the range of 10^4 to 10^6, nonrecombining block lengths may commonly fall in the range of 10^4 to 10^5 bp in such organisms.

Following these rough guidelines, results of simulations with a range of segment lengths (L) are given in Fig. 7. For the mutation rates employed herein, and the exponential fitness model, mean + allele frequencies are very close to those for the single-site expectations (Eq. **1**) when $L \leq 10$, and those for $L = 100$ are only moderately smaller. With increasing block lengths, however, the effects of linkage become increasingly substantial. With

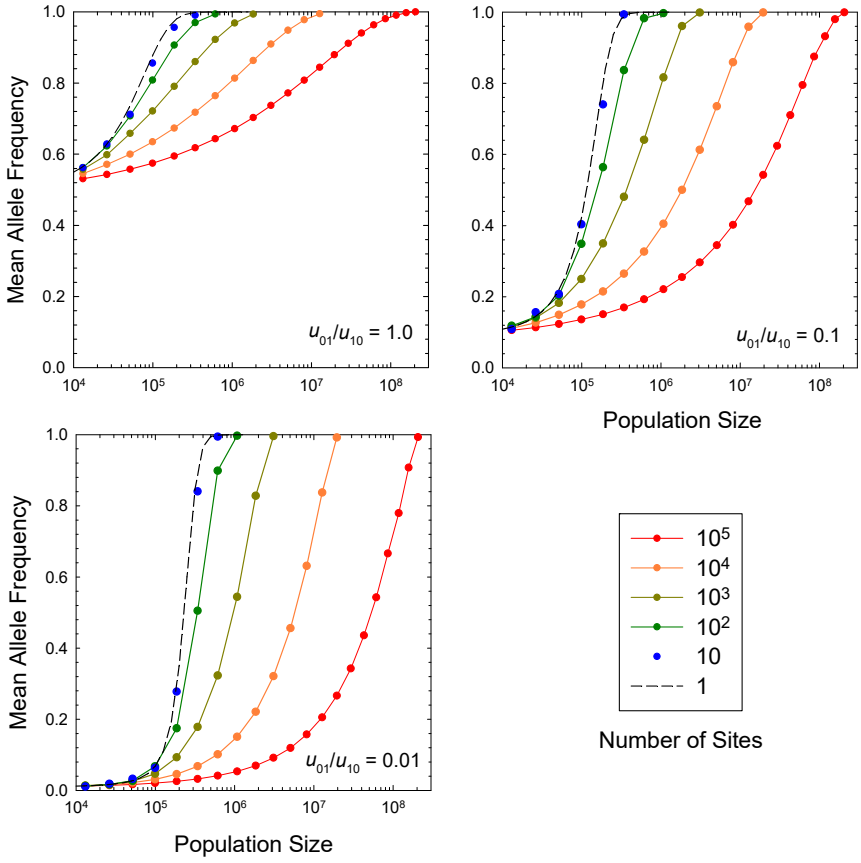

Fig. 7. Expected frequencies of beneficial alleles as a function of N_e and length of the chromosomal segment (in number of sites), for the case in which $s = 10^{-5}$. Dots are results from computer simulations, whereas the dashed line is the single-site expectation given by *SI Appendix*, Eq. **S2**.

$L = 1$, a transition from the neutral expectation to near fixation occurs over a 10-fold range of N, whereas with $L = 10^5$ this transition unfolds over a four orders-of-magnitude range of N_e. This is consistent with the gradual three orders-of-magnitude reduction in N_e/N with increasing N noted in Fig. 6. The key point here is that because linkage-block lengths are expected to decline with increasing N_e, in actual comparative analyses of different species, the scaling of the drift barrier with N_e is expected to be a hybrid of the kinds of functions shown in Fig. 7.

Discussion

Although much of population genetics is focused on standing levels of genetic variation and the impact on short-term responses to selection (2), comparative biology is generally concerned with the divergence of mean phenotypes among phylogenetic lineages. There is, therefore, a need for theory focused more on such long timescales (32, 33). The approach taken here starts with the assumption that the forces of mutation, drift, and selection remain relatively constant over long periods of time (tens to hundreds of millions of years), a feasible scenario for many intracellular traits. Even if the steady-state assumption is not fulfilled in every respect, the results still provide insight into the relative likelihoods of lineages wandering into alternative phenotypic states under a given set of population-genetic conditions (34), and the suggested approach can be readily modified to allow for additional factors such as fluctuating selection or alternative mutation functions. This framework of jointly accounting for the forces of selection, mutation, and drift is similar in spirit to the steady-state distributions of allele frequencies at single diallelic loci developed by Wright (35) and provides a conceptual contrast to the common approach among empiricists of simply assuming that the strength of selection is so overwhelming that observed mean phenotypes must precisely reflect optima dictated by the forces of natural selection.

The results here illustrate the riskiness of such an optimization approach. Mutation can cause mean phenotypes to deviate from the optimum in substantial and often unexpected ways that are not simply a function of the magnitude of mutation bias. Rather, when alternative, functionally equivalent underlying genotypes exist for a trait, the multiplicity of certain intermediate combinations can result in a mutational pull of the mean phenotype away from the optimum. This effect becomes especially significant when the phenotypic optimum is far from the expected mean under mutation alone and even more so if the level of multiplicity for the optimum is relatively small relative to other phenotypic states. Such effects can extend to populations with the largest known effective sizes, and, as noted previously (14), cases may even exist in which the selection gradient is sufficiently strong that the equilibrium mean-phenotype distribution can have two peaks, one driven by selection and the other by mutation.

With an increased understanding of the rates and molecular spectra of mutation, the approximate bounds on N_e, and the relationships between these two (10, 13), for cell biological features with well-understood genetic bases, this framework now provides a foundation for exploring the consequences of alternative selection functions for the scaling and diversification of cellular features with lineage-specific estimates of N_e. The challenge for future empirical work will be to obtain estimates of the key functional parameters, which minimally reduce to: 1) L, the number of sites (loci) encoding the trait; 2) the ratio of forward to reverse mutations rates, u_{01}/u_{10}; 3) N_e, the effective population

size; and 4) the mapping of genotypic values to a fitness function. With a focus on Gaussian stabilizing selection, there are just two fitness-function parameters: 1) θ, the phenotypic optimum, and 2) ω, the width of the fitness function; and with an exponential fitness function, there is just one parameter.

Notably, the direction and magnitude by which genotypic means scale with N_e are not simply functions of the pattern of selection and mutation bias, but are also influenced by the granularity of the system (i.e., the number of loci). The latter dictates the null distribution of genotypes in the absence of selection, with the selection function (scaled by the power of drift) then transforming the baseline distribution in a multiplicative manner. Expected gradients of genotypic means with respect to N_e can be positive or negative, depending on whether there is an excess or a deficit of mutational accessibility relative to the optimum state. One of the most striking examples consistent with this drift-barrier hypothesis is the negative scaling of the mutation rate per nucleotide site with N_e observed across the tree of life (10). Although the exact results will vary depending on the distribution of fitness effects associated with mutations for particular traits, the kinds of results illustrated in Fig. 6 suggest that for traits under persistent directional selection (such as growth rate), power-law scalings with exponents at least as extreme as 0.20 can be expected in large comparative studies. This raises challenges for fields such as evolutionary and physiological ecology that simply assume that such relationships are consequences of biophysical constraints.

Finally, it is worth noting that because of the degrees of freedom in systems like those illustrated above, isolated species can diverge to nonoverlapping underlying genetic constitutions despite residing in the same functional class. Consider, for example, a four-factor trait with initial state $++--$. If each of two descendant species experiences a $+\rightarrow-$ transition, there is a 50% chance that this will involve different factors, leading to $+---$ and $-+--$ configurations. Single independent reverse mutations in both lineages have a $2/9$ joint probability of converting these two states back to the class containing two $+$ sites but with nonoverlapping states $+--+$ and $-++-$, implying a $1/9$ chance of complete divergence. As L becomes larger, the probability of divergent states (in the face of constant selection) increases rapidly owing to the increased degrees of freedom in the overall system.

As one example of such a scenario, consider the phosphorylation of particular amino acid residues on proteins, which commonly plays a role in modifying activity level. Phosphorylation sites are typically clustered on the surface of a protein or in disordered regions, and the critical feature may simply be the acquisition of an adequate local charge. Comparative studies in yeasts and mammals indicate that many phosphorylated serines and threonines are under purifying selection to retain their phosphosite status (6, 36–38). Nonetheless, a large fraction of phosphosites appears free to vary among species in terms of status and location (39–42). For example, only $\sim$5% of all *Saccharomyces cerevisiae* phosphorylation sites appear to have been conserved across the entire yeast lineage (dating back $\sim$700 million y), and even when the same phosphorylatable residue is present in two moderately related species, their phosphorylation status may differ. In addition, Asp and Glu residues, which are naturally charged negatively, serve as potential replacements for their phosphorylatable counterparts; i.e., phosphosites often evolve from phosphomimetic Asp and Glu sites and vice versa (43–45). Taken together, these observations suggest a scenario whereby the degree of a protein's phosphorylation is under stabilizing selection for an appropriate charge, with the specific locations of many of the affected residues free to wander in an effectively neutral fashion (6, 46). That is, the number of phosphorylated residues on individual proteins appears to operate as a sort of quantitative trait under stabilizing selection, but with enough degrees of freedom that there can be considerable turnover of the specific phosphosites.

Aside from the need for a formal derivation of the expected distributions of mean phenotypes that fully account for the distinction between N_e and N for the case of large L, several things remain to be done to achieve a fuller understanding of the ways in which the drift barrier scales with population size and directional mutation bias. First, although the preceding expressions focus on the expected value of the mean phenotype, there can be considerable drift around the grand mean, so formal expressions for the temporal variance (or among-population variance) of the mean are desirable. This is a potentially significant issue for comparative studies based on small numbers of taxa, as the stochastic variance of means among taxa might be inappropriately assumed to reflect underlying adaptive differences. Second, there is a need to generalize the theory to allow for a distribution of mutational effects, in contrast to the fixed-effects approach taken herein. On the one hand, increasing the variance of mutational effects, while keeping the mean constant, should prolong the gradient of the response of the mean phenotype to population-size change as the fractions of mutations in the smallest populations that are subject to selection and in the largest populations that are subject to drift increase. On the other hand, greater variance in the fitness effects of alleles may increase the efficiency of selective discrimination among simultaneously segregating mutations, thereby reducing the depression of N_e relative to N. Finally, as noted above, there is a need for a more direct analysis of the role of recombination than that provided herein, particularly in combination with variable mutational effects.

ACKNOWLEDGMENTS. This research was supported by the Multidisciplinary University Research Initiative Awards W911NF-09-1-0444 and W911NF-14-1-0411 from the US Army Research Office, NIH Award R35-GM122566-01, and NSF Award MCB-1518060. I thank G. Bolstad, T. Hansen, K. Jain, and B. Trickovic for helpful comments.

1. B. Charlesworth, D. Charlesworth, *Elements of Evolutionary Genetics* (W. H. Freeman, 2010).
2. J. B. Walsh, M. Lynch, *Evolution and Selection of Quantitative Traits* (Oxford University Press, 2018).
3. J. Berg, S. Willmann, M. Lässig, Adaptive evolution of transcription factor binding sites. *BMC Evol. Biol.* **4**, 42 (2004).
4. M. Lynch, K. Hagner, Evolutionary meandering of intermolecular interactions along the drift barrier. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E30–E38 (2014).
5. M. Lynch, Evolutionary diversification of the multimeric states of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E2821–E2828 (2013).
6. C. R. Landry, L. Freschi, T. Zarin, A. M. Moses, Turnover of protein phosphorylation evolving under stabilizing selection. *Front. Genet.* **5**, 245 (2014).
7. W. Rawicz, K. C. Olbrich, T. McIntosh, D. Needham, E. Evans, Effect of chain length and unsaturation on elasticity of lipid bilayers. *Biophys. J.* **79**, 328–339 (2000).
8. P. Baudouin-Cornu, Y. Surdin-Kerjan, P. Marliére, D. Thomas, Molecular evolution of protein atomic composition. *Science* **293**, 297–300 (2001).
9. C. Acquisti, S. Kumar, J. J. Elser, Signatures of nitrogen limitation in the elemental composition of the proteins involved in the metabolic apparatus. *Proc. Biol. Sci.* **276**, 2605–2610 (2009).
10. M. Lynch *et al.*, Genetic drift, selection, and evolution of the mutation rate. *Nat. Rev. Genet.* **17**, 704–714 (2016).
11. W. H. Li, Models of nearly neutral mutations with particular implications for non-random usage of synonymous codons. *J. Mol. Evol.* **24**, 337–345 (1987).
12. M. Bulmer, The selection-mutation-drift theory of synonymous codon usage. *Genetics* **129**, 897–907 (1991).
13. H. Long *et al.*, Evolutionary determinants of genome-wide nucleotide composition. *Nature Ecol. Evol.* **2**, 237–240 (2017).
14. M. Lynch, Phylogenetic diversification of cell biological features. *Elife* **7**, e34820 (2018).
15. M. Lynch, G. K. Marinov, The bioenergetic costs of a gene. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15690–15695 (2015).
16. M. Lynch *et al.*, Spontaneous deleterious mutation. *Evolution* **53**, 645–663 (1999).

17. C. F. Baer, M. M. Miyamoto, D. R. Denver, Mutation rate variation in multicellular eukaryotes: Causes and consequences. *Nat. Rev. Genet.* **8**, 619–631 (2007).
18. P. D. Keightley, The distribution of mutation effects on viability in Drosophila melanogaster. *Genetics* **138**, 1315–1322 (1994).
19. L. Robert *et al.*, Mutation dynamics and fitness effects followed in single cells. *Science* **359**, 1283–1286 (2018).
20. K. B. Böndel *et al.*, Inferring the distribution of fitness effects of spontaneous mutations in Chlamydomonas reinhardtii. *PLoS Biol.* **17**, e3000192 (2019).
21. P. D. Keightley, A. Eyre-Walker, Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics* **177**, 2251–2261 (2007).
22. T. Bataillon, S. F. Bailey, Effects of new mutations on fitness: Insights from models and data. *Ann. N. Y. Acad. Sci.* **1320**, 76–92 (2014).
23. C. D. Huber, B. Y. Kim, C. D. Marsden, K. E. Lohmueller, Determining the factors driving selective effects of new nonsynonymous mutations. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4465–4470 (2015).
24. B. Y. Kim, C. D. Huber, K. E. Lohmueller, Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. *Genetics* **206**, 345–361 (2017).
25. M. Lynch, M. Ackerman, K. Spitze, Z. Ye, T. Maruki, Population genomics of Daphnia pulex. *Genetics* **206**, 315–332 (2017).
26. D. P. Rice, B. H. Good, M. M. Desai, The evolutionarily stable distribution of fitness effects. *Genetics* **200**, 321–329 (2015).
27. W. Sung *et al.*, Evolution of the insertion-deletion mutation rate across the tree of life. *G3* **6**, 2583–2591 (2016).
28. B. Charlesworth, Stabilizing selection, purifying selection, and mutational bias in finite populations. *Genetics* **194**, 955–971 (2013).
29. M. Bulmer, The effect of selection on genetic variability. *Am. Nat.* **105**, 201–211 (1971).
30. B. H. Good, A. M. Walczak, R. A. Neher, M. M. Desai, Genetic diversity in the interference selection limit. *PLoS Genet.* **10**, e1004222 (2014).
31. M. Lynch, L.-M. Bobay, F. Catania, J.-F. Gout, M. Rho, The repatterning of eukaryotic genomes by random genetic drift. *Annu. Rev. Genom. Hum. Genet.* **12**, 347–366 (2011).
32. R. Lande, Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30**, 314–334 (1976).
33. R. Bürger, R. Lande, On the distribution of the mean and variance of a quantitative trait under mutation-selection-drift balance. *Genetics* **138**, 901–912 (1994).
34. D. M. McCandlish, Long-term evolution on complex fitness landscapes when mutation is weak. *Heredity* **121**, 449–465 (2018).
35. S. Wright, *Evolution and the Genetics of Populations. The Theory of Gene Frequencies* (University of Chicago Press, 1969), vol. 2.
36. A. N. Nguyen Ba, A. M. Moses, Evolution of characterized phosphorylation sites in budding yeast. *Mol. Biol. Evol.* **27**, 2027–2037 (2010).
37. V. E. Gray, S. Kumar, Rampant purifying selection conserves positions with post-translational modifications in human proteins. *Mol. Biol. Evol.* **28**, 1565–1568 (2011).
38. E. D. Levy, S. W. Michnick, C. R. Landry, Protein abundance is key to distinguish promiscuous from functional phosphorylation based on evolutionary information. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 2594–2606 (2012).
39. A. M. Moses, M. E. Liku, J. J. Li, R. Durbin, Regulatory evolution in proteins by turnover and lineage-specific changes of cyclin-dependent kinase consensus sites. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17713–17718 (2007).
40. L. J. Holt *et al.*, Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science* **325**, 1682–1686 (2009).
41. L. Freschi, M. Osseni, C. R. Landry, Functional divergence and evolutionary turnover in mammalian phosphoproteomes. *PLoS Genet.* **10**, e1004062 (2014).
42. R. A. Studer *et al.*, Evolution of protein phosphorylation across 18 fungal species. *Science* **354**, 229–232 (2016).
43. Y. Z. Kurmangaliyev, A. Goland, M. S. Gelfand, Evolutionary patterns of phosphorylated serines. *Biol. Direct* **6**, 8 (2011).
44. S. M. Pearlman, Z. Serber, J. E. Ferrell, Jr., A mechanism for the evolution of phosphorylation sites. *Cell* **147**, 934–946 (2011).
45. G. Diss, L. Freschi, C. R. Landry, Where do phosphosites come from and where do they go after gene duplication? *Int. J. Evol. Biol.* **2012**, 843167 (2012).
46. G. E. Lienhard, Non-functional phosphorylations? *Trends Biochem. Sci.* **33**, 351–352 (2008).