æ

5. EVOLUTION AS A POPULATION-GENETIC PROCESS

5 April 2020

With knowledge on rates of mutation, recombination, and random genetic drift in hand, we now consider how the magnitudes of these population-genetic features dictate the paths that are open vs. closed to evolutionary exploitation in various phylogenetic lineages. Because historical contingencies exist throughout the Tree of Life, we cannot expect to derive from first principles the source of every molecular detail of cellular diversification. We can, however, use established theory to address more general issues, such as the degree of attainable molecular refinement, rates of transition from one state to another, and the degree to which nonadaptive processes (mutation and random genetic drift) contribute to phylogenetic diversification.

Substantial reviews of the field of evolutionary theory appear in Charlesworth and Charlesworth (2010) and Walsh and Lynch (2018). Much of the field is concerned with the mechanisms maintaining genetic variation within populations, as this ultimately dictates various aspects of the short-term response to selection. Here, however, we are primarily concerned with long-term patterns of phylogenetic diversification, so the focus is on the divergence of mean phenotypes. This still requires some knowledge of the principles of population genetics, as evolutionary divergence is ultimately a consequence of the accrual of genetic modifications at the population level. All evolutionary change initiates as a transient phase of genetic polymorphism, during which mutant alleles navigate the rough sea of random genetic drift, often being evaluated on various genetic backgrounds, with some paths being more accessible to natural selection than others.

The goal of this most technical of chapters is to summarize some of the more general themes and challenges to understanding how evolutionary change is accomplished, ideally endowing the reader with an appreciation for why the population-genetic details matter. With a specific focus on the ways in which selection acts to promote novel adaptive changes, emphasis will be placed on how the efficiency of selection is compromised or enhanced in different population-genetic environments, sometimes in counter-intuitive ways. Special attention will be given to the ways in which evolutionary rates and outcomes are expected to vary with the effective sizes of populations (N_e).

Most of the theory presented here will be discussed in a generic way, focusing for example on a mutation with selective advantage or disadvantage s, with no connection to the actual underlying trait(s). Such an approach allows for the development of general statements as to what natural selection can and cannot accomplish, and is a necessary prelude to more explicit exploration of particular traits where genotypes can be directly connected to phenotypes and then to fitness. Specific examples to be presented in later chapters include the evolution of protein-protein interfaces, the coevolution of transcription factors and their binding sites, and the evolution of

maximum growth-rate potential. Although the technical level of presentation here may be disappointing to high-level theoreticians, the goal is not to overwhelm the reader with a litany of equations and formal derivations, but to facilitate understanding as to how population-genetic theory transforms comparative cell biology into evolutionary biology.

The Perils of the Adaptive Paradigm

Ever since Darwin, most discussions with any connection to evolutionary thought generally start with the implicit assumption that all organismal traits are products of the promotion of favorable mutations by natural selection. Such logic underlies, for example, virtually every study in the field of evolutionary ecology. Closer to the subject material herein, a massive number of papers in cell biology end with a paragraph on why the trait being studied (and its sometimes arcane structure) must have been refined by selective forces.

An appreciation for the power of natural selection is one of the great advances of the life sciences over the past century. However, problems arise when the wand of natural selection is deemed to be the only mechanism relevant to evolutionary change, as this eliminates any hope for broader understanding of evolutionary processes, and often leads to false narratives. Starting with the conclusion that the phenotype under investigation is a necessary product of natural selection, the only remaining challenge is to identify the actual agent of selection. If one hypothesis fails, one moves on to another possibility, but always with unwavering certainty that selection must somehow be involved. Many biologists have spent entire careers wandering down such paths in search of an adaptive explanation for a particular biological feature.

This is not to say that optimization thinking has completely mislead us with respect to the evolution of alternative behavioral and/or life-history strategies. However, as touched upon in Chapter 4, the evolutionary outcomes that are actually achievable by natural selection depend critically on levels of mutation, drift, and recombination. Moreover, owing to the stochastic nature of these genetic features, even under constant selective pressures, the phenotypic states of populations are expected to wander over time. Indeed, as outlined below, the most common phenotype need not even be the optimum for a given environment, depending on the bias and granularity of mutational effects.

To begin to explore these ideas, this chapter will close with an overview of the concept of long-term steady-state distributions of mean phenotypes. This will also provide a more formal analysis of the drift-barrier hypothesis introduced in the previous chapter in the context of mutation-rate evolution, demonstrating the conditions under which traits are expected to exhibit gradients in performance scaling with N_e. The points being made here are not just arcane technical nuances. As will unfold in subsequent chapters focused on particular cellular traits, owing to the population-genetic and molecular features of biology, many aspects of evolution at the cellular level are best understood not by invoking the all-powerful guiding hand of natural selection, but through an appreciation of the factors that limit the reach of selection.

The Fitness Effects of New Mutations

Before proceeding with the theory, an overview of the fitness effects of mutations is necessary, as this defines the landscape and potential granularity of evolutionary change accessible to natural selection. As pointed out in Chapter 4, the vast majority of mutations are stochastically lost early in their life histories, regardless of their fitness effects, and if the absolute selective advantage/disadvantage (s) of a mutation is much smaller than the reciprocal of the effective population size ($1/N_e$), natural selection is incapacitated by drift. The key message below is that in all but the largest-N_e species, a substantial fraction of mutations fall within the domain of effective neutrality, $4N_es < 1$. This means that the molecular resources for natural selection differ among lineages with different N_e.

Although much remains to be learned, multiple lines of evidence point to the vast majority of mutations having very small, detrimental fitness effects. Organisms and their genomes are structured in such a way that most mutations have effects on fitness orders of magnitude smaller than 0.1. Mutations with larger detrimental effects do occur, but these are rare and rapidly eliminated from populations unless rescued by compensatory mutations elsewhere in the genome (below). Part of the reason for the rarity of mutations with large effects is that, within the prevailing bounds of random genetic drift, natural selection is often able to bring most phenotypes to the vicinity of optimum achievable values (Fisher 1930; Rice et al. 2015).

For random genetic drift to impose a significant barrier to the evolution of a trait, there must be a substantial pool of mutations with small enough effects that they can drift to fixation in species with small but not large 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 (Keightley and Eyre-Walker 1999; Lynch et al. 1999; Baer et al. 2007). Such experiments start with a set of isogenic lines, which are then maintained for large numbers of generations by propagation of just one (clone or selfer) or two (full-sib) individuals per generation. With an N_e this small, natural selection is incapable of promoting or removing mutations with fitness effects $< 25\%$ in these experiments, so the data are in full accord with a strong bias of mutations towards deleterious effects. Furthermore, statistical inferences based on the distribution of MA-line performances imply highly skewed distributions of fitness effects. The modes for such distributions are often indistinguishable from zero, with the bulk of mutations having absolute effects $< 1\%$ (almost all negative), although mean deleterious effects can sometimes be as high as 1 to 10% owing to the presence of rare mutations with large negative effects (Keightley 1994; Robert et al. 2018; Böndel et al. 2019).

Second, indirect inferences derived from allele-frequency distributions in natural populations of diverse multicellular species commonly suggest that 10 to 50% of mutations have deleterious effects smaller than 10^{-5}, with the inferred distribution sometimes being bimodal, but always with one mode being near (if not at) 0.0 (Keightley and Eyre-Walker 2007; Bataillon and Bailey 2014; Huber et al. 2015; Kim et al. 2017; Lynch et al. 2017; Booker and Keightley 2018; Johri et al. 2020). As

many of these studies focus only on the nonsynonomous sites in protein-coding genes, the full distributions of effects (which would include synonymous sites, introns, and intergenic DNA) can be expected to be even more skewed towards near-zero values. Substantial evidence also supports the idea that there is near universal selection for G/C (relative to A/T) composition across the Tree of Life, and estimates of the scaled strength of selection favoring G/C content at silent (synonymous) sites in protein-coding genes are almost all in the range of $N_e s = 0.1$ to 1.0, i.e., on the edge of the domain of effective neutrality (Long et al. 2017).

Third, although the preceding inferences are based on indirect extrapolations from statistical distributions, the costs of some kinds of mutations can be derived from first principles. For example, from a knowledge of the total energy budget of a cell and the biosynthetic costs of its building blocks, it is possible to estimate the fractional reduction in cell growth rates resulting from various kinds of mutations (Chapter 17). Bioenergetic considerations of the costs of small nucleotide insertions, which typically comprise $\sim 10\%$ of *de novo* mutations (Sung et al. 2016), imply fractional reductions in fitness far below 10^{-5} (Lynch and Marinov 2015). Likewise, the costs of using alternative amino acids or nucleotides from the standpoint of elemental (e.g., C, N, or S) composition imply that s associated with such substitutions is generally $\ll 10^{-5}$ (Chapter 18). Single residue changes in protein-protein interfaces or DNA binding sites are expected to have similarly small effects (Chapters 13 and 21). Finally, broad surveys of single amino-acid substitutions in a range of proteins generally imply that most such changes influence protein performance by $< 1\%$, with a large class indistinguishable from zero, and a second peak with very substantial effects (Chapter 12).

Taking all of these observations into consideration, 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. This being said, however, a number of caveats remain. First, statistical limitations prevent us from knowing with certainty the full distribution of mutations of small effects, e.g., the fraction of mutations with effects $< 10^{-5}$, $< 10^{-6}$, $< 10^{-7}$, etc. This is a significant concern – as we know that N_e ranges from $\sim 10^4$ to 10^9 (Chapter 4), these are the mutations that can be utilized vs. purged in some lineages but not in others.

Second, the data are inadequate to tell us whether the distribution of fitness effects is substantially different among lineages, e.g., in bacteria vs. unicellular eukaryotes vs. multicellular species. Such differences might be expected purely on physical grounds – the larger cells in lower-N_e species have higher total energy budgets, and hence single nucleotide and/or amino-acid insertion/deletions will have smaller fractional energy costs (Chapter 17). Whether the functional consequences are also altered is unknown.

Finally, and most importantly, although most of the observations noted above address the general fitness properties of random mutations, they are disconnected from the actual cellular traits that we will wish to eventually explore, i.e., they do not inform us as to the precise molecular targets/phenotypic effects of fitness-altering mutations.

The Classical Model of Sequential Fixation

The simplest entrée for considering the temporal dynamics of evolution under natural selection invokes the situation in which a trait is under persistent directional selection, with the pace of evolution being slow enough that each consecutive adaptive mutation is fixed before the next beneficial mutation destined to fix arises. In principle, such a scenario can exist if the supply of adaptive mutations is quite limited owing to either a relatively small population size, a low mutation rate to beneficial variants, or both. In this limiting situation, recombination is irrelevant because no two loci are ever simultaneously segregating polymorphisms at meaningful frequencies.

This sequential model of molecular evolution (sometimes also called the strong selection/weak mutation model or the origin-fixation model; McCandlish and Stoltzfus 2014) may only rarely represent reality as it assumes a constant march towards higher fitness. However, it serves as a useful heuristic for thinking about several issues concerning the limits to rates of adaptive evolution, and how they might scale with population size.

The long-term rate of adaptation under this model is equal to the product of the rate of introduction of beneficial mutations, the fixation probability of such mutations, and their fitness effects: 1) The population-wide rate of origin of new beneficial mutations is NU_b for haploids and $2NU_b$ for diploids, where U_b is the mutation rate (per generation) to beneficial alleles. Depending on the focus, U_b can represent a single gene or an entire haploid genome. 2) The probability of fixation of a new beneficial mutation is $\simeq 2s(N_e/N)$, where N and N_e are the actual and effective population sizes, and s is the selective advantage relative to the ancestral allele (Chapter 4). As N_e/N is almost always less than one, it can be thought of as the efficiency with which selection promotes new beneficial mutations, with the maximum rate being $2s$. (Note that for diploids, it is assumed here that mutational effects are additive, with heterozygotes having a fitness advantage $1+s$ intermediate to the two homozygotes, 1 and $1+2s$). 3) The increase in fitness per fixation is s for haploids, but $2s$ for diploids.

The resultant expected rate of evolution in terms of fixations is the product of the first two terms,

$$r_e = NU_b \cdot 2s(N_e/N) = 2N_eU_bs, \tag{5.1a}$$

for haploids, and twice this for diploids. The increase in fitness is then equal to

$$\Delta W = r_e \cdot s = 2N_eU_bs^2, \tag{5.1b}$$

for haploids, and $4\times$ this for diploids. (Note that this difference between haploids and diploids is merely a function of how the fitness effects are annotated in diploids. If it is assumed that mutant heterozygotes and homozygotes have fitnesses $1+(s/2)$ and $1+s$, respectively, then Equations 5.1a,b hold for diploids).

This model is idealized in many ways, as it assumes long-term persistent selection in one direction. Nonetheless, this simple approach highlights the key roles that the individual population-genetic parameters play in dictating the potential for evolutionary change. For example, all other things being equal, Equation 5.1 suggests that the rate of adaptive evolution should scale linearly with the effective population size (not with the absolute population size) and with the genome-wide beneficial mutation rate.

However, there is room for caution in interpreting this expression as the speed limit to the rate of adaptation. First, the conditions under which mutations are likely to fix sequentially are limited. Sequential fixation requires that the average time between fixations (the inverse of $4N_eU_bs$ for diploids) be greater than the mean time required for each mutation to fix, which is $\simeq (2/s)\ln(2N)$ generations for a diploid population (Walsh and Lynch 2018, Equation 8.4c). It follows then that for sequential fixation to be the rule, $4N_eU_b$ must be smaller than $1/[2\ln(2N)]$. Because $\ln(2N)$ falls in the narrow range of 10 to 58 over a range of $N = 10^4$ to 10^{20}, as a first-order approximation, the sequential model will hold if the effective number of beneficial mutations arising per generation is < 0.01, i.e., if no more than one beneficial mutation for the trait arises per 100 generations at the population level.

How likely is this condition to be met? Recall from Chapter 4 that the product of $2N_e$ and the mutation rate per nucleotide site per generation (μ) generally falls in the range of $2N_e\mu = 10^{-3}$ to 10^{-2}. Multiplying this number by the number of selected sites in a chromosomal region and the fraction of mutations that are beneficial converts this quantity to $2N_eU_b$. With a moderate-sized region of 10^5 fitness relevant sites and just 0.01% of mutations being beneficial, then $2U_bN_e$ for such a region would be in the range of 10^{-2} to 10^{-1}, bracketing the strict cutoff for the sequential model. This crude calculation demonstrates that the sequential model cannot be assumed to be generally valid. These issues are evaluated in great detail in (Weissman and Barton (2012) and Weissman and Hallatschek 2014).

One reason for concern here is that simultaneously segregating mutations (both beneficial and deleterious) interfere with each other in the selection process, diminishing their individual probabilities of fixation (Frenkel et al. 2014). As new beneficial mutations arise, their fates will be determined by the fitness of the linked backgrounds in which they appear (Figure 4.2). Theoretical work suggests that selective interference between linked beneficial mutations reduces the scaling of the overall rate of adaptation from linear to as little as logarithmic in N_e beyond the domain of the sequential model (Neher 2013; Weissman and Barton 2012; Weissman and Hallatschek 2014). Cosegregating deleterious mutations may play a more prominent role in reducing the rate of adaptive evolution, as the majority of *de novo* mutations are deleterious. If tightly linked to a segregating deleterious mutation, a beneficial mutation will not experience its full intrinsic advantage, and in some cases will be completely overshadowed by the linked background load (Good and Desai 2014). The central point is that although the effects of background variation do not alter the expectation that the rate of adaptive evolution will scale positively with N_e (all other things being equal), the gradient of scaling will likely decline with increasing N_e.

All of these interference effects can be observed in long-term laboratory evolution experiments involving microbes. For example, Figure 5.2 illustrates the trajectories of genome-wide mutant-allele frequencies in three replicate populations of *E. coli* grown in just 10 ml of medium. Over a period of 60,000 generations, these populations experienced an average 30% increase in fitness in the culture conditions, albeit at a diminishing rate. The norm was for many mutations to be simultaneously polymorphic, and although many fixations of individual mutations can be seen, more often than not, groups of mutations increase (and sometimes decline) in a coordinated manner. This is a simple consequence of the clonal nature of the

experimental populations, as a single, positively selected mutation driving to fixation sweeps along all linked "passenger" mutations (some of which themselves have beneficial or deleterious fitness effects).

Examples can also be seen of mutations reaching very high frequencies in a short period of time (which can only be due to positive selection), followed by a subsequent decline to 0.0 as other more fit mutant clones take over. In one population, two major clones, each containing multiple mutations, appear to reach equilibrium frequencies, with neither going to fixation (middle panel); this may be a result of some form of frequency-dependent selection, with each clone providing a metabolic product beneficial to the other. In another case, there is a massive accumulation of mutations near the midpoint of the experiment (lower panel), owing to the appearance of a mutator strain, which may have hitchhiked to fixation in linkage with a beneficial mutation that it promoted.

Compensating for these constraining effects from selective interference is the fact that a larger fraction of beneficial mutations is exploitable in larger populations. Owing to the fact that efficient selection requires $|N_e s| > 1$ (Chapter 4), larger populations have access to mutations with smaller s. Although this expansion in the pool of available beneficial mutations will further tip the balance in favor of higher rates of evolution in larger populations, from Equation 5.1 it can be seen that the contribution of beneficial mutations to increases in fitness scales with the square of the selective advantage, s^2. Thus, because average s is typically $\ll 1$, broadening the window of mutational availability will have a less than linear effect on the rate of adaptation unless the pool of mutations is strongly skewed towards those with small fitness benefits. Moreover, recall from Chapter 4 that there is a nearly inverse relationship between N_e and the mutation rate (μ). From Equation 5.1, such a compensatory effect would yield near independence of the rate of adaptation and N_e, other than the potential expansion of the window of exploitable mutations.

A final consideration of the factors influencing rates of adaptation involves the matter of time scale. The previous derivations consider the rate of adaptive evolution on a per-generation basis. However, smaller organisms typically have shorter generation times, which will elevate the rate of evolution on an absolute time scale. For example, if the generation length scaled inversely with N_e, this would render the expected rate of evolution (Equation 5.1) on an absolute time scale proportional to N_e^2. The following simple argument supports the idea of such a generation-time effect. Recall from Figure 4.3 that across the Tree of Life, N_e varies by about five orders of magnitude, from $\sim 10^4$ for some vertebrates to nearly 10^9 for some bacteria. Generation lengths in bacteria tend to be on the order of 0.1 to 1.0 days, whereas those in vertebrates and land plants are generally of order 10^2 to 10^4 days thus spanning nearly five orders of magnitude in the opposite direction of N_e. There can, of course, be considerable variation in generation lengths among organisms with the same N_e, but a general negative relationship between N_e and generation length is not in question.

Taken together, for simple adaptations involving mutations with additive effects, the above observations point toward the potential rate of evolutionary change (per absolute time unit) being greater for organisms with small size, short generation times, and large N_e. However, whether the scaling of evolutionary potential with N_e is sublinear, linear, or superlinear remains uncertain. Moreover, as will be discussed

below, for evolutionary changes involving interactions among loci, the scaling of evolutionary rates with N_e can deviate from that described above.

Before proceeding, two other points merit discussion. First, although the quantity $2N_eU_bs$ is a measure of the expected number of long-term fixations per unit time (for haploids), because mutation and fixation events are stochastic, considerable variation is expected around this expectation. For a Poisson process, where each rare event is independent of the others, the variance in the amount of long-term change among replicate populations is equal to the expectation. If, for example, the time interval under consideration is long enough that one beneficial mutation is expected per lineage, the probability that one event actually accrues is just 0.368, but the probability of no fixations is also 0.368, of two is 0.182, of three is 0.061, and of four or more is 0.021. The central point is that considerable variation is expected among lineages exposed to identical selection pressures, and that such dispersion should not be taken as evidence of adaptive differentiation or of intrinsic differences in evolutionary potential. Foundations 5.1 provides an even more dramatic example of how divergence among populations exposed to identical selection pressures can exceed that expected under neutral drift.

Vaulting Barriers to More Complex Adaptations

To this point, we have been assuming that the fitness effect of an allele is independent of the genetic background on which it resides. Under this view, Equation 5.1 provides the simplest possible model for the rate of adaptation by new mutations, as prior fixations have no bearing on subsequent events. However, this assumption can be violated for at least two reasons. First, for the case of stabilizing selection for an intermediate optimum phenotype or directional selection up the edge of a fitness plateau, the fixation of a mutation will alter the selection coefficients of future mutations by moving the mean phenotype closer to the optimal state, reducing the capacity for further improvement.

Second, when mutations have epistatic effects (i.e., interact in a nonadditive fashion), the possibility exists for neutral or even deleterious mutations to become beneficial in certain contexts. Multilocus traits exhibiting the latter types of genetic behavior will be referred to here as complex adaptations, as the paths for their evolution and the rapidity with which they are acquired are much less obvious than under conditions of additive fitness effects.

One broad category of complex-trait evolution involves compensatory mutations, wherein specific single mutations at either of two loci cause a fitness reduction, while their joint appearance can restore or even elevate fitness beyond the ancestral state. Such epistatic interactions play a prominent role in Wright's (1931, 1932) shifting-balance theory of evolution, which postulates that adaptive valleys between fitness peaks are typically traversed by random genetic drift in a small subpopulation, with the locally fixed advantageous genotype then being exported to surrounding demes by migration. Compensatory mutations appear to play important roles in protein-sequence evolution (Chapter 12) and in the composition of nucleotides in the stems of RNA molecules (Stephan and Kirby 1993; Kondrashov et al. 2002; Kulathinal et al. 2004; Azevedo et al. 2006; DePristo et al. 2007; Breen

et al. 2012; Wu et al. 2016). Situations likely also exist in which mutations that are effectively neutral in isolation yield an increase in fitness when combined, and many other scenarios are possible, including the involvement of intermediate pathways that allow the bypass of shorter but deleterious pathways (Figure 5.3).

As complex adaptations are expected to evolve over relatively long periods, and to be accompanied by numerous mutations (some drivers and some passengers), establishing the molecular paths of establishment from patterns in comparative data is challenging. However, clear examples observed in real time do exist. For example, in a long-term ($> 40,000$ generation) evolution experiment with *E. coli* simply selected for growth in flasks on a defined medium, the ability to utilize citrate as a carbon source emerged in one of twelve cultures (Blount et al. 2008; Quandt et al. 2014). Drawing from the historical record of evolution by resurrecting frozen samples, it was found that a weak variant for citrate utilization with a promoter-region mutation provided a potential mutational target for further refinement of the trait. While this initial mutation was still infrequent in the population (and possibly effectively neutral), a linked mutation appearing at a second locus conveyed a much greater ability to uptake citrate, conferring a substantial increase in fitness that drove the double mutant to fixation.

Sequential fixation vs. stochastic tunneling. Ascertaining the population-genetic conditions under which complex adaptations are likely to occur is challenging because unlike the situation in which a single mutation fixes at a rate depending only on its own initial frequency, the success of a mutation involved in an interlocus interaction depends on the frequency of alleles at the interacting locus, on the fitnesses associated with all possible multi-locus genotypes, and on the recombination rate between the two loci.

The focus here will be on the rate of establishment of a complex adaptation, defined to be the inverse of the expected arrival time of the ultimate multi-mutation configuration destined to be fixed in the population. Although this excludes the additional time required for fixation, the latter will generally be considerably smaller than the time to establishment.

Population size alone can dictate the kinds of evolutionary pathways that are open to the establishment of complex traits (Figure 5.4). For populations of sufficiently small size, the path toward adaptation almost always involves sequential fixations of the contributing mutations, owing to the extreme rarity of occasions in which multiple mutations are simultaneously segregating at key sites. Consider, for example, an adaptation involving two mutations, the first of which is neutral (Walsh 1995; Lynch and Abegg 2010). The mean time for a new neutral mutation to fix is $4N_e$ generations in a diploid population, with an average frequency during this period of 0.5 (which implies an average $0.5 \cdot 2N = N$ copies), and from the theory outlined above, the rate of appearance of second-step mutations destined to fix is then $\mu_2 \cdot 2s(N_e/N)$ per mutational target per generation, where μ_2 is the rate of second-step mutation. The product of these three quantities gives the approximate probability of a secondary mutation arising on a segregating first-step mutation background (and also being destined to fix) of $4N_e \cdot N \cdot u_2 \cdot 2s(N_e/N) = 8N_e^2\mu_2 s$. Setting this equal to 1 and rearranging shows that there is a negligible chance of arrival of a successful secondary mutation before fixation of the first if $N_e \ll (8\mu_2 s; -1/2$.

With $\mu_2 = 10^{-9}$ and $s = 0.01$, for example, the critical effective population size is $\simeq 10^5$. For N_e below this threshold value, selection is restricted to exploring the fitness landscape by single mutational steps (sequential fixations).

In contrast, key secondary (and even tertiary) mutations can arise in large populations prior to the fixation of earlier-step contributors (Figure 5.3). This raises the possibility of the joint, simultaneous fixation of combinations of mutations as a single haplotype without any mutation having been common in isolation. For example, a conditionally beneficial secondary mutation may arise in linkage with a low-frequency deleterious first-step mutation, with the joint fixation of the double-mutation haplotype in effect rescuing the first-step mutation otherwise destined to be lost. Such a process, often referred to as stochastic tunneling (Komarova et al. 2003; Iwasa et al. 2004), provides a smooth route for the establishment of complex adaptations, allowing large populations to explore the fitness surface more broadly than possible by single-step mutations. Most notably, the process makes possible progression through intermediate deleterious alleles without the population ever experiencing the transient decline in fitness that would necessarily occur with sequential fixation (Gillespie 1984; Weinreich and Chao 2005; Gokhale et al. 2009; Weissman et al. 2009, 2010; Lynch and Abegg 2010; Lynch 2010). The following analyses will focus on the domain in which stochastic tunneling dominates, i.e., populations of moderate to very large size.

Before proceeding, it should be emphasized that the theory explored in the following paragraphs starts with the premise that each mutation contributing to a final adaptation arises independently of all others. Recall, however, from Chapter 4 that mutations often arise in clusters, such that if the mutation rate per site is μ, the rate of simultaneous appearance of pairs of mutations on the same chromosome is often orders of magnitude greater than μ^2. This means that adaptations involving two or three, and perhaps even more, site-specific mutations will often arise spontaneously in a single individual on realistic time scales. In such cases, assuming negligible recombination between these sites, the rate of fixation of the mutant haplotype follows directly from the single-mutation theory noted in the preceding chapter, Equations 4.1a,b.

Two-locus transitions. We start with a simple selection scenario, first explored by Kimura (1985), in which haplotypes Ab and aB have reduced but equivalent fitness $(1 - s_d)$ relative to AB and ab, both with fitnesses of 1.0 (Figure 5.3, upper left). In this case, two-step transitions between pure population states of AB and ab render no gain in fitness, but do involve an intermediate deleterious genotype. Initially, the two sites will be assumed to be completely linked, and μ_d and μ_b will denote the rate of mutation to first (potentially deleterious) and second (potentially beneficial) step variants. (If $s_d = 0$, the intermediate states are neutral).

As an explicit example, AB and ab might represent beneficial pairs of amino acids involved in protein folding or stability, with Ab and aB representing nonmatching combinations. The Watson-Crick pairs in the stems of ribosomal RNAs provide another compelling example. Although the overall stem/loop structure of rRNAs is highly conserved across species, orthologous complementary nucleotide pairs in stems often have different states in different species. Barring a rare double mutation, such shifts require passage through an intermediate deleterious state, e.g., A:T →

A:C → G:C. Provided the overall secondary structure is maintained, which is presumably essential for proper protein binding in the ribosome, rRNAs from different bacterial species with up to 20% divergence can substitute for each other with only small effects on fitness (Kitahara et al. 2012).

Starting with a population in state AB, we wish to determine the mean time for the population to reach an alternative state of fixation at both loci, with respective alleles a and b. Assuming that selection against the intermediate haplotypes is sufficiently strong that fixation of such states is extremely unlikely, in the stochastic tunneling domain mutation will recurrently introduce new Ab/aB alleles despite their being selected against. Provided the strength of selection exceeds that of mutation, both the aB and Ab haplotypes will then be expected to have steady-state frequencies of μ_d/s_d, resulting from the balance between the rates of mutational input μ_d and selective removal s_d (Walsh and Lynch 2018). These low-frequency haplotypes then serve as reliable substrates for secondary mutations to the ab type, as mutant ab gametes arise at rate μ_b from each of the $4N\mu_d/s_d$ intermediate types (there being two loci in N diploid target individuals). However, even though the ab type has an advantage over its parental haplotype, Ab or $aB \rightarrow ab$ mutations fix in an essentially neutral fashion with probability $1/(2N)$. This follows because $\mu_d/s_d \ll 1$, so almost all resident haplotypes are of type AB, with equivalent fitness to ab. Thus, the rate of establishment of the ab type by stochastic tunneling from AB is

$$r_e \simeq \frac{4N\mu_d}{s_d} \cdot \mu_b \cdot \frac{1}{2N} = \frac{2\mu_d\mu_b}{s_d} \tag{5.2}$$

(Gillespie 1984; Stephan 1996). This rate is independent of population size (provided the conditions for selection-mutation balance, $4N_e s_d \gg 1$, are met, and ignoring for the moment any population-size dependence of the mutation rate).

Next, suppose that the secondary mutation has advantage s_b, such that the fitness of the AB and ab haplotypes are 1 and $1 + s_b$, respectively (Figure 5.3, upper right). Equation 5.2 must then be modified to account for the fact that the fixation probability of the double mutant is no longer $1/(2N)$ but $\simeq 2s_b(N_e/N)$. This leads to

$$r_e \simeq \frac{4N\mu_d}{s_d} \cdot \mu_b \cdot 2s_b(N_e/N) = \frac{8N_e\mu_d\mu_b s_b}{s_d}. \tag{5.3}$$

As in the case of selectively equivalent end states, the rate of establishment scales positively with the square of the mutation rate, but now also with the factor $4N_{eb}$, which as noted above is the strength of selection scaled to that of drift. A key feature of these stochastic-tunneling events is that they result in the fixation of mutations that considered alone would be viewed as deleterious, but in doing so the population never experiences a reduction in fitness because the first-step mutation experiences enhanced fitness in linkage with the second. In this sense, deleterious mutations can play a central role in evolutionary diversification.

Finally, consider the special situation in which first-step mutations are effectively neutral, arising at rate μ_n per site, and fixing with probability $1/(2N)$ (Figure 5.3, upper middle). To obtain the expected rate of tunneling for the case of neutral intermediates, we require the probability that tunneling occurs within a descendant lineage of a first-step mutation before it becomes lost from the population. Under the assumption that first-step mutations are absent at the outset of the process,

again assuming complete linkage, this probability has been found to be approximately $\sqrt{2\mu_b s_b(N_e/N)}$ in large populations (Komarova et al. 2003; Iwasa et al. 2004; Weissman et al. 2009, 2010; Lynch and Abegg 2010). With two sites in a diploid population, there are $4N\mu_n$ neutral first-step mutations arising per generation, so the rate of establishment via tunneling is then

$$r_e \simeq 4N\mu_n\sqrt{2\mu_b s_b N_e/N} = 4\mu_n\sqrt{2\mu_b s_b N_e N} \tag{5.4}$$

The key observation here is that when the intermediate steps are neutral, the probability of tunneling scales positively with the absolute population size, contrary to the situation with deleterious intermediates.

Note, however, that Equation 5.4 assumes the extreme situation in which both first- and second-step mutations must arise anew. Given that first-step mutations are actually neutral, there should be a nonzero probability of their being present at some low frequency at the outset. Denoting this initial frequency as p_0, and simply substituting this for μ_d/s_d in Equation 5.3 yields

$$r_e = 4Np_0 \cdot \mu_b \cdot 2s_b(N_e/N) = 8N_e p_0 \mu_b s_b. \tag{5.5}$$

This is a potentially much higher rate of tunneling than implied by Equations 5.3 and 5.4, owing to the expectation that mutations at neutral sites are expected to have much higher equilibrium frequencies than deleterious mutations. If, for example, the genetic substrate here is a nucleotide site, assuming no mutational bias, the long-term average frequency of each of the four nucleotide types is 0.25, yielding $r_e = 2N_e\mu_b s_b$, which is just half the expectation for the single-site model (because one of the two potential starting nucleotides, out of four, is present at the outset). This kind of scenario would apply to the situation in which a codon requires two changes for a transition to a more beneficial amino-acid.

What can be inferred about the likely scaling of two-site adaptations from these results? A key issue is that the algebraic scaling implied by the preceding expressions is confounded by the nonindependent behavior of the biological components. Most notably, as outlined in Chapter 4, there is a roughly inverse scaling between the mutation rate per nucleotide site per generation and N_e across the Tree of Life. Thus, treating the product of N_e and the mutation rate as an approximate constant provides a more realistic view of how the per-generation rate of evolution scales with population size.

Consider, for example, Equation 5.2 for the rate of transition between two equivalent fitness states via deleterious intermediates. Although this expression suggests that the evolutionary rate scales with $\mu_d\mu_b$, independent of N_e, because both mutation rates scale inversely with N_e, the expectation is that r_e will actually scale inversely with N_e^2. Extending this logic to Equation 5.3 for the rate of transition to a higher fitness state through deleterious intermediates implies an inverse scaling of r_e with N_e. On the other hand, Equation 5.5 for the rate of transition to a higher fitness state through neutral intermediates implies a scaling that is independent of N_e, as in the case of the single-site model.

The key point here is that the rate of exploitation of various kinds of evolutionary pathways can depend critically on the nature of the adaptive change. Relatively speaking, complex adaptations involving two sites and neutral intermediates are

more likely in large populations, whereas those involving deleterious intermediates are more likely in small populations. Recall, however, that simultaneous mutations arise at two sites much more rapidly than expected by chance, and in this case, the rate of origin of two-site mutants in Equation 5.3 becomes $2N\mu_2$, yielding a rate of evolution equal to $4N_e\mu_2 s_b$. Provided $u_2 \gg \mu_b\mu_d$, this rate will be much higher than that expected by tunneling, and also implies a scaling independent of N_e.

More complex scenarios. While the above analyses assume an evolutionary path to a final adaptation through just a single intermediate step, the routes to many molecular/cellular modifications are more complex, with a variety of potential pathways through any number of mutations, e.g., Figure 5.3 (bottom). The rates of establishment under these alternative scenarios have been examined by Gokhale et al. (2009), Weissman et al. (2009), and Lynch and Abegg (2010), again with the primary focus being on situations in which the intermediate states are neutral or deleterious. Simple analytical approximations have been found in only a few cases, two of which are considered below. The primary focus again is on how the establishment rate r_e scales with the underlying features of population size, mutation rate, and selection intensity.

For the case of deleterious intermediates, suppose that all haplotypes involving 1 to $d-1$ mutations are equally deleterious (with fitness $1-s_d$), with the final mutation conferring an advantage s_b. First-step mutations then arise at rate $2Nd\mu_d$, but owing to selection have an expected survivorship time of $1/s_d$ generations, during which period $d-2$ additional intermediate step mutations must be acquired, followed by the appearance of a final-step mutation destined to fixation. This leads to a rate of establishment via tunneling of

$$r_e \simeq 4N_e d!(\mu_d/s_d)^{d-1}\mu_b s_b \tag{5.6}$$

which reduces to Equation 5.3 when $d = 2$. Here we see that r_e scales with the dth power of the mutation rate, owing to the limited opportunities for mutation during the short sojourn times of deleterious mutations. Thus, the acquisition of a novel adaptation involving multiple, deleterious intermediate steps is a very low probability event, diminishingly so for populations with large N_e, as the expected scaling is now with N_e^{1-d}. Again, however, multinucleotide mutations are likely to dramatically accelerate this process.

For the case of neutral intermediates with d mutations required for the final adaptation (and the order of events assumed to be irrelevant), there is again the conceptual issue of the starting conditions. The worst-case scenario is the one in which all contributing mutations are absent at time zero, with establishment then requiring a series of nested tunneling events. For example, for the case of $d = 3$ (with two neutral mutations required before the final adaptation is assembled with a third mutation), a secondary mutation must arise on a haplotype lineage containing the first mutation, and before being lost by drift, the still smaller two-mutation lineage must acquire a third mutation destined to fixation. Equation 5.4 then expands to

$$r_e = 6N\mu_n\sqrt{2\mu_n\sqrt{2\mu_b s_b(N_e/N)}} \tag{5.7a}$$

Note that the first term is now $6N\mu_n$ because first-step mutations can arise at three diploid sites. The next step then initiates at either of the two remaining sites,

bringing in the $2\mu_n$ term, with the final stage being initiated at the one remaining site and fixing at the usual rate for a single beneficial mutation. With $d = 4$, this expression expands one step further to

$$r_e = 8N\mu_n\sqrt{3\mu_n\sqrt{2\mu_n\sqrt{2\mu_b s_b(N_e/N)}}}, \tag{5.7b}$$

and so on.

These results show that with neutral intermediates, the rate of establishment of complex adaptations can be much more rapid than expected under the naive assumption that independently arising mutations would lead to a scaling with the dth power of the mutation rate. Regardless of the number of sites involved in this case, the rate of establishment by tunneling scales with no more than the square of the mutation rate. Adhering to the empirical observation of an approximately inverse relationship between mutation rates and population size, and it it is assumed that $N \simeq N_e^2$ (which is very roughly in accord with the data in Chapter 4), the rate of establishment is entirely independent of N_e regardless of the number of required neutral intermediates. If it can be assumed that $(d-1)$-stage mutants are present at some low frequency p_0 in the base population, Equation 5.5 applies, which again implies N-independent scaling.

Finally, it should be noted that the above examples are just a small sample of the kinds of evolutionary pathways that can exist between two complex genotypes. In principle, multiple pathway types may connect two presumed endpoints, including those with mixtures of neutral and deleterious intermediates, different numbers of links, and so on. The kind of theory just outlined can be used to evaluate the relative probabilities of alternative routes, as well as the possibility of becoming transiently trapped at points with suboptimal fitness, and of back-tracking and exploring alternative paths (McCandlish 2018). Experimental evolution studies with microbes are being increasingly exploited to evaluate these issues (e.g., Lind et al. 2019; Rodrigues and Shakhnovich 2019; Zheng et al. 2019).

Effects of recombination. Finally, we note that all of the above analyses assume an absence of recombination. In the sequential-fixation regime, recombination can be ignored simply because multiple polymorphic sites are never present simultaneously. However, in the stochastic-tunneling domain, opportunities will exist for both the creation and breakdown of optimal haplotypes. Higgs (1998), Lynch (2010), and Weissman et al. (2010) have examined this problem with a broad class of models, reaching the conclusion that recombination is most likely to have either a minor or an inhibitory effect on the *de novo* establishment of a complex adaptation.

Consider, for example, the case of a two-site adaptation with a deleterious intermediate, starting with a population fixed for the suboptimal ab haplotype. The overall influence of recombination on the rate of establishment of the AB haplotype will then be a function of two opposing effects. On the one hand, the rate of origin of AB gametes by recombination between the two single-mutation haplotypes (aB and Ab) will be proportional to the rate of recombination between the sites (c). On the other hand, recombinational breakdown discounts the net selective advantage of resultant AB haplotypes from s_b to $s_b - c$. This occurs because in the early stages of establishment, ab haplotypes still predominate, and hence are the primary partners

in recombination events with AB, generating the maladaptive Ab and aB products. Thus, because the product $c(s_b - c)$ is maximized at $c = s_b/2$, two-site adaptations are expected to emerge most rapidly in chromosomal settings where the recombination rate is half the selective advantage of the final adaptation, a rather specific requirement. Moreover, the domain for any improvement by recombination is quite narrow. For example, when first-step mutations are deleterious, if the rate of recombination exceeds the selective advantage of the AB haplotype (i.e., $s_b - c < 0$), recombination presents an extremely strong barrier to establishment of the AB haplotype. Even in the case of neutral intermediates and at the optimal recombination rate, the rate of establishment is generally enhanced by much less than an order of magnitude relative to the situation with complete linkage (Lynch 2010).

Taken together, these results suggest that: 1) only a narrow range of recombination rates (in the neighborhood of $s_b/2$) can enhance the rate of establishment of a complex adaptation from *de novo* mutations; and 2) overly high rates of recombination in large populations can be prohibitive. Thus, because the role that recombination plays in the origin of specific adaptations depends on both the selective advantage of the final product and the physical distance between the genomic sites of the underlying mutations, as discussed in Chapter 4, recombination is far from universally advantageous.

The Phylogenetic Dispersion of Mean Phenotypes

The theory discussed above provides insight into the rapidity with which populations can respond to novel and/or persistent directional selective challenges. Such scenarios might be encountered in a continuous coevolutionary arms race between a host cell and a pathogen, or in situations involving a sudden environmental shift. However, numerous cellular traits may have been under very similar selective pressures across phylogenetic lineages since their origin. This is likely to be true, for example, for enzymes whose sole function has always been to convert a specific substrate into a specific product, membrane channels specialized to admitting and/or excluding specific ions, or polymerases responsible for generating complementary base-pair matches. Homeostatic mechanisms further buffer cells from external environmental changes.

In such cases, the dynamical response to changing selection pressures is no longer the key issue. The more appropriate evolutionary perspective is the long-term steady-state probability distribution of alternative genotypes. Although natural selection may relentlessly strive to improve trait performance, there are numerous reasons why perfection will seldom, if ever, be obtained for more than transient periods. First, absolute limits to the refinements of chemical and physical processes can be dictated by diffusion limitation and by effectively discrete processes such as the energy associated with individual hydrogen bonds. Second, the stochastic processes of mutation and drift can result in the dispersion of mean phenotypes around an expected value, to a degree that depends on the range of effectively neutral parameter space. Third, mutation pressure will almost never be perfectly aligned with the goals of selection, and this will cause the average phenotype to deviate from an optimum, even in the absence of mutation bias.

The latter two points imply that: 1) populations under identical selection pressures will not have identical mean phenotypes, but instead will exhibit a dispersion of such measures; 2) the most common phenotype need not be the optimum phenotype; and 3) gradients of mean phenotypes with respect to N_e are likely to be molded by differences in the power of random genetic drift across the Tree of Life. To reduce the likelihood of evolutionary cell biology succumbing to the common practice of interpreting all phylogenetic variation in phenotypes as necessary reflections of differences in selective environments, these basic principles will first be sketched out for the case of a very simple two-state trait, and then further explored for traits encoded by multiple genetic loci.

Two-state traits. Consider a single-locus situation in which one allele (denoted as state 1) has a fractional selective advantage s over another allele (denoted as state 0). Although allele 1 has the highest achievable fitness, this need not mean that it is a perfect fit to the environment. Nor does it ensure that once achieved, allele 1 will be immune to replacement by the suboptimal type. The mutation rate from allele 0 to 1 is denoted μ_{01}, with μ_{10} denoting the reciprocal mutation rate, and the population will be assumed to be haploid.

The simplest situation involves a population with a small enough effective size that the waiting times between mutations destined to fixation are large enough that there is effectively no polymorphism and the population is nearly always fixed for one allele or the other, with probabilities $\widetilde{p}_0$ and $\widetilde{p}_1$. The conditions necessary for such a situation are equivalent to those noted above for the sequential-fixation model, with the more general model being outlined in Foundations 5.2.

Under these weak-mutation, strong-selection conditions, a lineage spends a long period of time in one particular monomorphic state before making a stochastic shift to another. The intervening intervals (waiting times for transitions) are functions of the relative strengths of selection, mutation, and random genetic drift, but over a very long time period, the rate of movement from state 0 to 1 must equilibrate to equal that in the opposite direction (a principle known as detailed balance in the statistical physics literature). This implies

$$\widetilde{p}_0 \cdot (N\mu_{01}) \cdot \phi_{01} = \widetilde{p}_1 \cdot (N\mu_{10}) \cdot \phi_{10}, \tag{5.8}$$

where ϕ_{01} and ϕ_{10} denote the probabilities of fixation of newly arisen beneficial and deleterious alleles, defined by Equation 4.1b. Using the useful identity $\phi_{01}/\phi_{10} = e^S$, where $S = 2N_e s$, and the fact that $\widetilde{p}_1 \simeq 1 - \widetilde{p}_0$,

$$\widetilde{p}_0 \simeq \frac{1}{1 + \beta e^S}, \tag{5.9}$$

where $\beta = \mu_{01}/\mu_{10}$ is the ratio of the mutation rates in both directions (mutation bias being implied when $\beta \neq 1$).

This simple model illustrates three key points. First, unless completely lethal, the low-fitness state has a non-zero probability of occurrence. Thus, despite constant selection pressure, a lineage (or series of parallel lineages) is not expected to remain in a stable fixed state. In fact, at equilibrium, the rate of $0 \rightarrow 1$ transitions is the same as the reverse. This is because, whether favorable or encouraged by mutation

pressure, the most abundant state provides more opportunities for transitions, which are individually less likely to proceed to fixation; the rarer state provides fewer transition opportunities, but when these arise they are more likely fix.

Second, the two alleles approach equal probabilities as $\beta e^S \rightarrow 1$. This composite parameter is simply equal to the product of the mutation and selection biases in favor of state 1, so that even if state 1 is selectively favored ($S > 0$), state 0 will be more common if there is mutation bias in the opposite direction of sufficient strength ($\beta < e^{-S}$). The relevance of this point is that maximum divergence among fixed lineages occurs if $\widetilde{p}_0 = \widetilde{p}_1 = 0.5$, again demonstrating the potential for substantial variation in the face of uniform selection.

Third, if the effective population size and/or strength of selection is sufficiently small that $S \ll 1$ (the domain of effective neutrality), the equilibrium frequency of the disfavored allele will be entirely driven by mutation pressure. In this case, because $e^S \simeq 1$, $\widetilde{p}_0 \simeq 1/(1+\beta)$, which is entirely a function of the relative (but not absolute) mutation rates.

These expectations are altered when the population-level mutation rate exceeds the limits of the domain of the sequential model (Foundations 5.2). Most notably, Equation 5.9 defines the lower bound to the expected frequency of the low-fitness allele. The expected frequency of the beneficial allele declines once $N_e\mu_{01}$ exceeds ~ 0.01 (i.e., a new mutation enters the population at least every 100 generations), asymptotically approaching the neutral expectation $\widetilde{p}_1 \simeq \beta/(1+\beta)$ as $N_e\mu_{01}$ exceeds 1.0 (Figure 5.5). The latter condition arises when mutation brings in allelic variants faster than natural selection can promote beneficial over detrimental alleles, and the population is almost always represented by a polymorphic collection of both alleles.

Multistate-traits and the drift-barrier hypothesis. Extension of the preceding single-locus model to an arbitrary number L sites (factors) yields additional insights into the limits to what natural selection can accomplish. To appreciate the fundamental points in a relatively simple manner, it will be adequate to assume that all genetic factors are equivalent with two alternative (+ and −) allelic states contributing positively and negatively to the trait. Depending on the context, these factors may be viewed as single nucleotides, amino-acid codons, or entire genes. Although there are L^2 potential genotypic states, there is redundancy among members of genetic classes containing the same numbers of +/− alleles.

For all but the two most extreme genotypes, a multiplicity of functionally equivalent classes exists with respect to the number of + alleles, i, defined by the binomial coefficients. As an example, for the case of $L = 4$, there are $L^4 = 16$ possible genotypes, but just five genotypic classes ($i = 0$, 1, 2, 3, 4, and 5), with multiplicities 1, 4, 6, 4, and 1, respectively (Figure 5.6). 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, as classes with higher multiplicities are more mutationally accessible. This type of biallelic model has been widely exploited in theoretical studies of the genetic structure of quantitative (multilocus) traits (Walsh and Lynch 2018), and will be encountered in a number of different contexts in subsequent chapters, including the evolution of protein-protein interfaces, transcription-factor binding sites, and growth rate. Here, we will assume a haploid, nonrecombining population of N individuals.

The site-specific per-generation mutation rates from the − to the + state, and vice versa, are defined as μ_{01} and μ_{10}, respectively.

As with the single-factor model, the multiple-factor model has a long-term equilibrium distribution of population residence in the $L+1$ alternative states. Again, starting with the assumption that population-level mutation rates are low enough that transitions only occur between adjacent classes, the relative flux rates between classes are equal to the expressions on the arrows in Figure 5.6, which are proportional to the products of rates of mutational production and probabilities of subsequent fixation, with the numerical coefficients being defined by the numbers of − and + sites within each class. The absolute population size N defines the number of mutational targets per generation, but because N influences all mutational flux rates in the same way, it is omitted as a prefactor, although both N and N_e do influence the equilibrium solution via the fixation probabilities (Equation 4.1b).

This linear sequential model has a relatively simple solution (Berg et al. 2004; Sella and Hirsh 2005; Lynch and Hagner 2014; Lynch 2020). As a reference point, consider first the extreme case of effective neutrality. In this situation, $\eta = u_{01}/(u_{01}+u_{10}) = \beta/(1+\beta)$ is the expected equilibrium frequency of + alleles at each site, which arises when the net flux of $+ \rightarrow -$ and $- \rightarrow +$ mutations is balanced. With no selection for particular combinations of alleles, each site evolves in an independent fashion, so the steady-state distribution of phenotypes under neutrality is simply equal to the binomial probability distribution,

$$\begin{aligned} \widetilde{p}_{n,i} &= \binom{L}{i} \eta^i (1-\eta)^{L-i} \\ &= C \cdot \binom{L}{i} \beta^i \end{aligned} \tag{5.10}$$

where $C = (1+\beta)^{-L}$. Denoting the overall genotypic state as the sum of + alleles, the long-term mean and variance of the trait are $L\eta$ and $L\eta(1-\eta)$, respectively. Equation 5.10 defines the long-term probability of a population residing in each of the $L+1$ possible genotypic classes, i.e., the fractional time wandering over the evolutionary landscape that is spent in each class. Note that the neutral steady-state distribution depends only on the ratio of mutation rates, not on their absolute values.

Selection transforms the neutral distribution in a remarkably simple way, with each class being weighted by the exponential function of the scaled strength of selection e^{S_i}, with $S_i = 2N_e s_i$,

$$\widetilde{p}_i = C \cdot \widetilde{p}_{n,i} \cdot e^{S_i}, \tag{5.11}$$

where C is a new normalization constant that ensures that the $\widetilde{p}_i$ sum to 1.0. The selection coefficients associated with each class are generally defined as deviations from some benchmark in the population (say the optimum type), but this does not matter, as the reference is a constant that simply enters the normalization constant. The utility of this approach is that, provided there are mutational connections between all adjacent states, Equation 5.11 can be applied to any fitness function describing the relationship between s and i.

Taken together, Equations 5.10 and 5.11 show that the equilibrium frequencies of the genotypic classes 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. This demonstrates that mutation need not be directionally biased to have an impact on the overall distribution. All that is required is that neutral distribution not coincide with the expectations under selection alone, which will almost always be the case.

Two examples will now be explored to illustrate how these three factors jointly define the distribution of phenotypes within and among alternative population-genetic environments. First, consider the simple case in which a trait determined by just $L = 2$ factors is under stabilizing selection, such that there is an optimum phenotype θ, with fitness dropping off at a rate determined by the width of the fitness function ω (analogous to the standard deviation of a normal distribution). This Gaussian (bell-shaped) function is defined by

$$W_i = e^{-(i-\theta)^2/(2\omega^2)}. \tag{5.12}$$

With two sites, there are three possible genotypic classes, $i = 0, 1, 2$, with the phenotypically equivalent $+-$ and $-+$ states being lumped into the $i = 1$ class. Selection is purely directional if the optimum is at or beyond an end state, i.e., $\theta \leq 0$ or $\geq L$, and neutrality is approached as the fitness function becomes flatter, i.e., $\omega \to \infty$. Although i is confined to integer values, θ need not be, and if θ is outside of the $(0, 2)$ range, the optimum is unattainable. The selection coefficients can be simply defined as deviations of fitness from the maximum value of 1, $s_i = 1 - W_i$. The mean phenotype is $\widetilde{p}_1 + 2\widetilde{p}_2$, which reduces to 2η in the case of neutrality.

This expansion to a second site introduces complexities not encountered with the one-site model (Figure 5.7a). For example, for the case of $\theta = 1.5$, where the optimum is straddled by the class 1 and 2 genotypes, assuming mutation bias towards $-$ alleles, the long-term genotypic mean never reaches the optimum, even at very large N_e, and instead remains much closer to 1. This bias results because although the class 1 and 2 genotypes have equivalent fitness, mutation pressure towards $-$ alleles weights the frequency of class 1 by a factor of 2β (the two being the multiplicity of this class), but class 2 by the smaller factor of β^2 (from Equation 5.10)

For the case in which $\theta = 2$ (pure directional selection), there is a progressive succession of the prevailing genotype classes with increasing N_e (Figure 5.7b). When N_e is sufficiently low to impose effective neutrality, class 0 predominates owing to the mutation bias towards $-$ 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. They also show that appreciable incidences of all three genotypic classes can be expected over time in lineages with intermediate N_e.

As a second example, consider the case in which a large number of loci contribute in an additive fashion to the expression of a trait such as growth rate, with fitness declining with the number of $-$ alleles in a multiplicative fashion,

$$W_i = (1 - s)^{(L-i)}, \tag{5.13}$$

where i is the number of beneficial (+) alleles in the genome. Three examples are shown in Figure 5.8 for different numbers of contributing loci. In each case, the overall performance is equally subdivided across all L factors, such that $s = 1/L$, with $L = 10^4$, 10^5, and 10^6. With such large numbers of loci, the analytical solution noted above is not reliable at large N_e, as the number of mutations arising per generation may exceed the limit of the domain of the sequential model, although results have been obtained by computer simulations (Lynch 2020).

Three general features are clear (Figure 5.8). First, as already noted, at N_e sufficiently small that the mutational effects are rendered effectively neutral, the mean fraction of + alleles is defined by the neutral expectation, with the probability of a + allele being η at each locus.

Second, with increasing N_e, the mean fraction of + alleles progressively increases, converging on 1.0 as N_e becomes much greater than $2s$. This gradient in trait means with N_e is a result of the drift barrier, which increasingly compromises the ability of natural selection to alter the frequencies of mutations as the population size declines. The exact location of the drift barrier is defined by the relative power of drift and selection, becoming shallower with smaller s. It is also influenced by the mutation bias towards alleles with deleterious effects and by the multiplicity effect.

Third, the observed gradients are much shallower than the expectations under free recombination. This illustrates the point made above that owing to selective interference among linked loci, populations behave genetically as though they are much smaller than implied by their actual size N. For example, in the absence of selective interference, sites with selection coefficients equal to 10^{-5} are expected to be nearly fixed for + alleles once the absolute population size exceeds 500,000, but with linked loci at the same absolute population size, the vast majority of alleles are of the − type owing to the combination of mutation pressure and random genetic drift (Figure 5.8).

These results highlight the riskiness of an evolutionary biology that simply assumes that all phenotypes simply reflect optimal outcomes dictated by natural selection, with any deviations from optimality being consequences of unavoidable biophysical constraints. In addition to the pervasive influence of drift, mutation can cause mean phenotypes to deviate from the optimum in substantial and often unexpected ways that are not simply functions 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, especially if the level of multiplicity for the optimum is relatively small relative to other phenotypic states. Moreover, cases may even exist in which the opposing pressures of selection and mutation are sufficiently strong that the equilibrium mean-phenotype distribution can have two peaks, one driven by selection and the other by mutation (Lynch and Hagner 2014; Lynch 2018).

Notably, the direction and magnitude by which genotypic means scale with N_e is not simply a function of the pattern of selection and mutation bias, but is also influenced by the granularity of the system, i.e., the magnitude of mutational effects (Figure 5.8). As noted above, substantial evidence suggests that a large fraction of mutations have effects small enough that these issues do matter. One of the most

striking examples consistent with the drift-barrier hypothesis is the negative scaling of the mutation rate per nucleotide site with N_e observed across the Tree of Life presented in Chapter 4.

Summary

- Critical to understanding the evolutionary potential and limitations of phylogenetic lineages is information on the distribution of fitness and phenotypic effects of new mutations. Multiple lines of evidence indicate that the vast majority of mutations have deleterious effects, most being very mild, and the mode effect being near zero. There are statistical limitations to discerning the fine-scale features of the distribution of mutation with small effects, but there is strong evidence for a substantial pool of mutations that are only available for evolutionary exploitation in species with large effective population sizes (N_e). These mutations play a key role in defining the limits to adaptation in different phylogenetic lineages.

- Despite the common view that populations under identical selection pressures will tend to be highly similarly phenotypically, many plausible situations exist in which uniform selection combined with random genetic drift can lead to substantial interspecies divergence, sometimes more than expected under drift alone.

- Many molecular adaptations require the co-occurrence of two or more mutations to elicit a phenotype with elevated fitness. Theory suggests that the rapidity (per generation) of acquiring such adaptations is roughly independent of N_e if the intermediate states are neutral, but scales negatively with N_e if the intermediate states are deleterious, and more rapidly so with more intermediate steps. This is one example of how the likelihood of alternative paths of evolution are modulated by changes in N_e. Such scaling, however, may be radically changed with sufficiently high rates of multinucleotide mutation, as this will instantaneously embark a complex genotype on a more rapid path to establishment.

- Many cellular traits have retained the same function for hundreds of millions of years, and may have been under nearly invariant selection pressures for this same amount of time. This shifts the evolutionary focus away from dynamical changes in allele frequency under directional selection to the long-term steady-state probability distribution of alternative phenotypic states in constant population-genetic environments. Motivated by the evidence for a large pool of mutations with small effects, the drift-barrier hypothesis predicts that mean phenotypes of traits will commonly exhibit gradients with respect to N_e, with the level of functional refinement increasing with the latter.

- Mutation has a fundamental influence on the expected distribution of mean phe-

notypes because genotypic states differ in the multiplicity of ways in which they can be constructed from the underlying set of genetic loci. Mutation bias further influences the evolutionary attraction towards a particular region of phenotypic space, in ways that may conflict with or reinforce prevailing selection pressures.

- Taken together, these results from evolutionary theory call into question the common habit of assuming an essentially perfect match between mean phenotypes and prevailing selection pressures. Often, the most common phenotype is an unreliable indicator of the optimum defined by natural selection.

Foundations 5.1. Divergence under uniform selection. Although it is generally thought that selection will increase the evolutionary determinism of a system, causing pairs of populations under identical selection pressures to be more similar than expected on the basis of random assortment of variation, this is not necessarily the case (Cohan 1984; Lynch 1986). Consider a pair of populations exposed to identical conditions and starting with identical frequencies of two alleles, A and a, p and $(1-p)$, respectively. Letting $\phi(p)$ be the probability of fixation of allele A, the probability that a pair of populations will ultimately experience fixation for different alleles is $\Delta = 2\phi(p)[1-\phi(p)]$, which reaches a maximum value when $\phi(p) = 0.5$. Under the naive view that the beneficial allele always fixes, one expects $\phi(p) = 1$ and $\Delta = 0$.

That populations can sometimes diverge to a greater extent under uniform selection than under pure neutral drift can be seen as follows. In the absence of selection, the probability of fixation of allele A is simply p, and the probability of alternative outcomes is $\Delta = 2p(1-p)$. The probability of divergence is increased by selection if $\phi(p)$ under selection is closer to 0.5 than the initial frequency p. Thus, because $\phi(p) > p$ for a selectively favored allele, a minimum requirement for increased divergence under pan-selection is that the starting frequency of the advantageous allele be < 0.5.

The conditions for excess divergence under drift plus selection to exceed that under drift alone are not very restrictive. Consider two replicate populations with identical initial frequencies of the A allele, $p = 0.25$. Under pure drift, the probability that one replicate becomes fixed for A and the other for a is $2 \cdot 0.25 \cdot (1 - 0.25) = 0.375$. Now suppose that A is weakly favored by selection, with $N_e s = 0.5$. Again assuming $p_0 = 0.25$, Equation 4.1a gives the fixation probability of A as 0.46, implying a probability of fixing alternative alleles of $2 \cdot 0.46 \cdot 0.54 = 0.496$. Divergence is increased by the action of selection, which in this case moves the fixation probability very close to 0.5. Thus, even when under identical directional selection pressures, populations that initiate with low-frequency, advantageous alleles can exhibit levels of divergence conventionally interpreted as being associated with diversifying selection.

Foundations 5.2. Mean probabilities of alternative alleles at steady state. A concern with the sequential model outlined in the text is that populations of moderate size are expected to reside in both fixed (monomorphic) and polymorphic states for significant amounts of time. This would not be a problem if the frequencies of alleles when monomorphic were the same as those while polymorphic, but such a condition is only met in the special case of neutrality. This is because deleterious mutations that are strongly inhibited from going to fixation can nonetheless maintain measurable frequencies owing to recurrent mutational input. As population sizes increase, the likelihood of residing in a polymorphic state necessarily increases, owing to the greater total influx of variation per generation. Under such conditions, one can still inquire as to the average frequency of a sampled allele over a long-term steady-state equilibrium, but this average must also factor in all possible polymorphic states, ranging from allele frequency $(1/N)$ to $[1-(1/N)]$ (for haploids).

Let P_1, P_0, and P_p denote the steady-state probabilities of a population being monomorphic for the optimal allele (1), monomorphic for the suboptimal allele (0), or polymorphic (p). Under the sequential model, $P_1 + P_0 \simeq 1$. Here, we make use of a result from diffusion theory that describes the steady-state probability distribution of allele frequency x for the deleterious state 0 (which is equivalent to the beneficial allele 1 being present at frequency $1-x$), described more fully by Kimura et al. (1963), Wright (1969), and Charlesworth and Jain (2014). Although actual allele-

frequency distributions are discrete, with large N, the overall pattern can be accurately approximated by the continuous distribution

$$\phi(x) = Cx^{U_{10}-1}(1-x)^{U_{01}-1}e^{-Sx}, \tag{5.2.1a}$$

where $U_{01} = 2N_e\mu_{01}$, $U_{10} = 2N_e\mu_{10}$, and the normalization constant

$$C = \frac{\Gamma(U_{01}+U_{10})}{\Gamma(U_{01})\cdot\Gamma(U_{10})\cdot {}_1F_1(U_{10};(U_{01}+U_{10});-S)}, \tag{5.2.1b}$$

ensures that integration of the distribution over the full range of allele frequencies sums to 1.0. Γ denotes the gamma function, and ${}_1F_1$ is the confluent hypergeometric function. These two functions can be calculated numerically using series expansions defined respectively as Equations 6.1.2 and 13.1.2 in Abramowitz and Stegun (1964).

The probability of being monomorphic for state 1 can be approximated by integration of the end class

$$P_1 = \int_0^{1/N} \phi(x)\cdot dx.$$

Because x is very small in this region, both $(1-x)$ and e^{-Sx} can be approximated as 1, leading to

$$P_1 \simeq \left(\frac{C}{U_{10}}\right)\left(\frac{1}{N}\right)^{U_{10}}. \tag{5.2.2a}$$

At the opposite end of the spectrum, using $x \simeq 1$ and $e^{-Sx} \simeq e^{-S}$ yields the probability of being monomorphic for state 0,

$$P_0 = \int_{1-(1/N)}^{1} \phi(x)\cdot dx \simeq \left(\frac{C}{U_{01}}\right)\left(\frac{1}{N}\right)^{U_{01}} e^{-S}. \tag{5.2.2b}$$

Here, it can be seen that the ratio P_1/P_0 obtained with this approach deviates from the prediction of the sequential model, $\widetilde{p}_1/\widetilde{p}_0 = (\mu_{01}/\mu_{10})e^S$ (inferred from Equation 5.9), unless $\mu_{01} = \mu_{10}$. Although the details will not be covered here, it can be shown that the probability of polymorphism, $P_p = 1 - P_0 - P_1$, is only weakly dependent on the magnitude of selection, and generally does not exceed 0.1 until $N_e\mu_{01} > 0.01$.

The average frequencies of the two alleles over the stationary distribution can be obtained by weighting the frequency classes by their densities, Equation 5.2.1a, and integrating over (0,1), which yields

$$\overline{p}_0 = \frac{\mu_{10}\cdot {}_1F_1[(U_{10}+1);(U_{01}+U_{10}+1);-S]}{(\mu_{01}+\mu_{10})\cdot {}_1F_1[U_{10};(U_{01}+U_{10});-S]}. \tag{5.2.3}$$

Foundations 5.3. The detailed-balance solution for the evolutionary distribution of alternative molecular states. Here we assume a linear array of alternative molecular states, with population-level transitions only occurring between adjacent states (Figure 5.6). For the latter condition to be met, each transition rate must be sufficiently small that a population generally resides in one state for an extended period of time before fixation of a subsequent mutation leads to a switch between states. Under these conditions, a relatively simple model defines the probability of residing in each class after a sufficient amount of time has elapsed to ensure

occupancy over the entire distribution of states. At this equilibrium, for any particular state, the rates of entry and exit must be equal, a condition is known detailed balance. The overall form of the steady-state distribution depends on the full set of transition rates and is reached regardless of the starting conditions.

Letting $m_{i,j}$ denote the rate of evolutionary transition from state i to state j, we have a system of $L+1$ simultaneous equations (where L denotes the final state in the series, which starts with index 0),

$$\begin{aligned} p_0(t+1) &= (1-m_{0,1})p_0(t) + m_{1,0}p_1(t), \\ p_i(t+1) &= m_{i-1,i}p_{i-1}(t) + (1-m_{i,i-1}-m_{i,i+1})p_i(t) + m_{i+1,i}p_{i+1}(t), \\ p_n(t+1) &= m_{L-1,L}p_{L-1}(t) + (1-m_{L,L-1})p_n(t). \end{aligned}$$

Assuming nonzero transition rates between all adjacent classes, the equilibrium solution (the steady-state probability of being in state i) takes on a simple, intuitive form (Lynch 2013),

$$\widetilde{p}_i = \frac{\left(\prod_{j=0}^{i-1} m_{j,j+1}\right)\left(\prod_{k=i+1}^{L} m_{k,k-1}\right)}{C} \tag{5.3.1}$$

where the first term in the numerator is equal to the product of all transition rates pointing up toward the class, the second term is the product of all transition rates pointing down toward the class, and C is simply a normalization constant that ensures that all of the $\widetilde{p}_i$ sum to one (it is equal to the sum of numerators for all i, and is generally called the partition function).

For example, with four alternative states (indexed 0, 1, 2, 3), the equilibrium probabilities become

$$\begin{aligned} \widetilde{p}_0 &= m_{1,0}m_{2,1}m_{3,2}/C, \\ \widetilde{p}_1 &= m_{0,1}m_{2,1}m_{3,2}/C, \\ \widetilde{p}_2 &= m_{0,1}m_{1,2}m_{3,2}/C, \\ \widetilde{p}_3 &= m_{0,1}m_{1,2}m_{2,3}/C, \end{aligned}$$

where C is the sum of the numerators all four expressions. The steady-state probabilities, $\widetilde{p}_i$, can be equivalently viewed as the proportion of time a specific lineage spends in state i over a long evolutionary period, or as the fraction of populations experiencing identical population-genetic environments that are expected to reside in class i at any specific time.

In the context of the model introduced in the text, each coefficient can be viewed as the product of the number of mutations arising in the population per generation and the probability of fixation. Letting N be the population size, $u_{i,j}$ be the mutation rate from allelic state i to j (where j can be only $i-1$ or $i+1$), and $\phi_{i,j}$ be the probability of fixation of a newly arisen j allele in a population predominated by allele i, the transition rates are equal to the products of the relevant numbers of new mutant alleles arising per generation and their probabilities of fixation, i.e., $m_{i,j} = 2Nu_{i,j}\phi_{i,j}$ assuming diploidy (or one half that assuming haploidy). Because every coefficient has the same prefactor, $2N$ or N, this can be ignored, reducing the coefficients to $m_{i,j} = u_{i,j}\phi_{i,j}$. (In the text, the $u_{i,j}$ are functions of per-site mutation rates and the number of sites relevant to the particular transition).

A second key simplification arises from the behavior of the probability of fixation in opposite directions between adjacent states. Letting s_i denote the selective disadvantage of allele i, measured relative to a perfect fitness of 1.0, then $s_{i+1} < s_i$ implies that allele $i+1$ is beneficial compared to allele i. Assuming mutations with additive effects on fitness, application of the formula for the fixation probability of new mutations, Equation 4.1b, yields the convenient result that $\phi_{i,i+1}/\phi_{i+1,i} = e^{4N_e(s_i - s_{i+1})}$

for diploids (with a 2 substituted for the 4 in haploids) (Sella and Hirsh 2005; Lynch 2013; Lynch and Hagner 2014).

As a simple application of the preceding methods, consider the situation in which there are just two alternative states, A and B, with the mutation rate from A to B being u, from B to A being v, and s being the selective advantage of B (negative if B is disadvantageous). The combined mutation/selection pressure towards B is then $u\phi_{A,B}$, while that towards A is $v\phi_{B,A}$, implying that

$$\widetilde{p}_B = \frac{u\phi_{A,B}}{u\phi_{A,B} + v\phi_{B,A}}, \tag{5.3.2a}$$

where the denominator is the partition function. Dividing all terms by $v\phi_{B,A}$, and using the relationship just noted for ratios of opposite fixation probabilities for diploids, leads to the simplification

$$\widetilde{p}_B = \frac{(u/v)e^{4N_e s}}{(u/v)e^{4N_e s} + 1}. \tag{5.3.2b}$$

Literature Cited

Abramowitz, M., and I. A. Stegun (eds.) 1964. Handbook of Mathematical Functions. Dover Publ., Inc., New York, NY.

Azevedo, L., G. Suriano, B. van Asch, R. M. Harding, and A. Amorim. 2006. Epistatic interactions: how strong in disease and evolution? Trends Genet. 22: 581-585.

Baer, C. F., M. M. Miyamoto, and D. R. Denver. 2007. Mutation rate variation in multicellular eukaryotes: causes and consequences. Nat. Rev. Genet. 8: 619-631.

Bataillon, T., and S. F. Bailey. 2014. Effects of new mutations on fitness: insights from models and data. Ann. N. Y. Acad. Sci. 1320: 76-92.

Berg, J., S. Willmann, and M. Lässig. 2004. Adaptive evolution of transcription factor binding sites. BMC Evol. Biol. 4: 42.

Blount, Z. D., C. Z. Borland, and R. E. Lenski. 2008. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli.* Proc. Natl. Acad. Sci. USA 105: 7899-7906.

Böndel, K. B., S. A. Kraemer, T. Samuels, D. McClean, J. Lachapelle, R. W. Ness, N. Colegrave, and P. D. Keightley. 2019. Inferring the distribution of fitness effects of spontaneous mutations in *Chlamydomonas reinhardtii.* PLoS Biol. 17: e3000192.

Booker, T. R., and P. D. Keightley. 2018. Understanding the factors that shape patterns of nucleotide diversity in the house mouse genome. Mol. Biol. Evol. 35: 2971-2988.

Breen, M. S., C. Kemena, P. K. Vlasov, C. Notredame, and F. A. Kondrashov. 2012. Epistasis as the primary factor in molecular evolution. Nature 490: 535-538.

Charlesworth, B., and D. Charlesworth. 2010. Elements of Evolutionary Genetics. Roberts and Co. Publ., Greenwood Village, CO.

Charlesworth, B., and K. Jain. 2014. Purifying selection, drift, and reversible mutation with arbitrarily high mutation rates. Genetics 198: 1587-1602.

Cohan, F. M. 1984. Can uniform selection retard random genetic divergence between isolated populations? Evolution 38: 495-504.

DePristo, M. A., D. L. Hartl, and D. M. Weinreich. 2007. Mutational reversions during adaptive protein evolution. Mol. Biol. Evol. 24: 1608-1610.

Fisher, R. A. 1930. The Genetical Theory of Natural Selection. Oxford Univ. Press, Oxford, UK.

Frenkel, E. M., B. H. Good, and M. M. Desai. 2014. The fates of mutant lineages and the distribution of fitness effects of beneficial mutations in laboratory budding yeast populations. Genetics 196: 1217-1226.

Gillespie, J. H. 1984. Molecular evolution over the mutational landscape. Evolution 38: 1116-1129.

Gokhale, C. S., Y. Iwasa, M. A. Nowak, and A. Traulsen. 2009. The pace of evolution across fitness valleys. J. Theor. Biol. 259: 613-620.

Good, B. H., and M. M. Desai. 2014. Deleterious passengers in adapting populations. Genetics 198: 1183-1208.

Good, B. H., M. J. McDonald, J. E. Barrick, R. E. Lenski, and M. M. Desai. 2017. The dynamics of molecular evolution over 60,000 generations. Nature 551: 45-50.

Higgs, P. G. 1998. Compensatory neutral mutations and the evolution of RNA. Genetica 102/103: 91-101.

Huber, C. D., B. Y. Kim, C. D. Marsden, and K. E. Lohmueller. 2015. Determining the factors driving selective effects of new nonsynonymous mutations. Proc. Natl. Acad. Sci. USA 114: 4465-4470.

Kim, B. Y., C. D. Huber, and K. E. Lohmueller. 2017. Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. Genetics 206: 345-361.

Iwasa, Y., F. Michor, and M. A. Nowak. 2004. Stochastic tunnels in evolutionary dynamics. Genetics 166: 1571-1579.

Johri, P., B. Charlesworth, and J. D. Jensen. 2020. Towards an evolutionarily appropriate null model: jointly inferring demography and purifying selection. Genetics (in press).

Keightley, P. D. 1994. The distribution of mutation effects on viability in *Drosophila melanogaster*. Genetics 138: 1315-1322.

Keightley, P. D., and A. Eyre-Walker. 1999. Terumi Mukai and the riddle of deleterious mutation rates. Genetics 153: 515-523.

Keightley, P. D., and A. Eyre-Walker. 2007. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. Genetics 177: 2251-2261.

Kimura, M. 1985. The role of compensatory neutral mutations in molecular evolution. J. Genetics 64: 7-19.

Kimura, M., T. Maruyama, and J. F. Crow. 1963. The mutation load in small populations. Genetics 48: 1303-1312.

Kitahara, K., Y. Yasutake, and K. Miyazaki. 2012. Mutational robustness of 16S ribosomal RNA, shown by experimental horizontal gene transfer in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 109: 19220-19225.

Komarova, N. L., A. Sengupta, and M. A. Nowak. 2003. Mutation-selection networks of cancer initiation: tumor suppresser genes and chromosomal instability. J. Theor. Biol. 223: 433-450.

Kondrashov, A. S., S. Sunyaev, and F. A. Kondrashov. 2002. Dobzhansky-Muller incompatibilities in protein evolution. Proc. Natl. Acad. Sci. USA 99: 14878-14883.

Kulathinal, R. J., B. R. Bettencourt, and D. L. Hartl. 2004. Compensated deleterious mutations in insect genomes. Science 306: 1553-1554.

Lind, P. A., E. Libby, J. Herzog, and P. B. Rainey. 2019. Predicting mutational routes to new adaptive phenotypes. eLife 8: e38822.

Long, H., W. Sung, S. Kucukyildirim, E. Williams, S. W. Guo, C. Patterson, C. Gregory, C. Strauss, C. Stone, C. Berne, et al. 2017. Evolutionary determinants of genome-wide nucleotide composition. Nature Ecol. Evol. 2: 237-240.

Lynch, M. 1986. Random drift, uniform selection, and the degree of population differentiation. Evolution 40: 640-643.

Lynch, M. 2010. Scaling expectations for the time to establishment of complex adaptations. Proc. Natl. Acad. Sci. USA 107: 16577-16582.

Lynch, M. 2013. Evolutionary diversification of the multimeric states of proteins. Proc. Natl. Acad. Sci. USA 110: E2821-E2828.

Lynch, M. 2018. Phylogenetic diversification of cell biological features. eLife 7: e34820.

Lynch, M., and A. Abegg. 2010. The rate of establishment of complex adaptations. Mol. Biol. Evol. 27: 1404-1414.

Lynch, M., M. Ackerman, K. Spitze, Z. Ye, and T. Maruki. 2017. Population genomics of *Daphnia pulex.* Genetics 206: 315-332.

Lynch, M., J. Blanchard, D. Houle, T. Kibota, S. Schultz, L. Vassilieva, and J. Willis. 1999. Spontaneous deleterious mutation. Evolution 53: 645-663.

Lynch, M., and K. Hagner. 2014. Evolutionary meandering of intermolecular interactions along the drift barrier. Proc. Natl. Acad. Sci. USA 112: E30-E38.

Lynch, M., and G. K. Marinov. 2015. The bioenergetic costs of a gene. Proc. Natl. Acad. Sci. USA 112: 15690-15695.

McCandlish, D. M. 2018. Long-term evolution on complex fitness landscapes when mutation is weak. Heredity 121: 449-465.

McCandlish, D. M., and A. Stoltzfus. 2014. Modeling evolution using the probability of fixation: history and implications. Quart. Rev. Biol. 89: 225-252.

Neher, R. A. 2013. Genetic draft, selective interference, and population genetics of rapid adaptation. Ann. Rev. Ecol. Evol. Syst. 44: 195-215.

Quandt, E. M., D. E. Deatherage, A. D. Ellington, G. Georgiou, and J. E. Barrick. 2014. Recursive genomewide recombination and sequencing reveals a key refinement step in the evolution of a metabolic innovation in *Escherichia coli.* Proc. Natl. Acad. Sci. USA 111: 2217-2222.

Rice, D. P., B. H. Good, and M. M. Desai. 2015. The evolutionarily stable distribution of fitness effects. Genetics 200: 321-329.

Robert, L., J. Ollion, J. Robert, X. Song, I. Matic, and M. Elez. 2018. Mutation dynamics and fitness effects followed in single cells. Science 359: 1283-1286.

Rodrigues, J. V., and E. I. Shakhnovich. 1019. Adaptation to mutational inactivation of an essential gene converges to an accessible suboptimal fitness peak. eLife 8: e50509.

Sella, G., and A. E. Hirsh. 2005. The application of statistical physics to evolutionary biology. Proc. Natl. Acad. Sci. USA 102: 9541-9546.

Stephan, W. 1996. The rate of compensatory evolution. Genetics 144: 419-426.

Stephan, W., and D. A. Kirby. 1993. RNA folding in *Drosophila* shows a distance effect for compensatory fitness interactions. Genetics 135: 97-103.

Sung, W., M. S. Ackerman, M. Dillon, T. Platt, C. Fuqua, V. Cooper, and M. Lynch. 2016. Evolution of the insertion-deletion mutation rate across the tree of life. G3 (Bethesda) 6: 2583-2591.

Walsh, J. B. 1995. How often do duplicated genes evolve new functions? Genetics 139: 421-428.

Walsh, J. B., and M. Lynch. 2018. Evolution of Quantitative Traits. Sinauer Assocs., Inc., Sunderland, MA.

Weinreich, D. W., and L. Chao. 2005. Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. Evolution 59: 1175-1182.

Weissman, D. B., and N. H. Barton. 2012. Limits to the rate of adaptive substitution in sexual populations. PLoS Genet. 8: e1002740.

Weissman, D. B., M. M. Desai, D. S. Fisher, and M. W. Feldman. 2009. The rate at which asexual populations cross fitness valleys. Theor. Pop. Biol. 75: 286-300.

Weissman, D. B., D. S. Fisher, and M. W. Feldman. 2010. The rate of fitness-valley crossing in sexual populations. Genetics 186: 1389-1410.

Weissman, D. B., and O. Hallatschek. 2014. The rate of adaptation in large sexual populations with linear chromosomes. Genetics 196: 1167-1183.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16: 97-159.

Wright, S. 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proc. 6th Internat. Cong. Genetics 1: 356-366.

Wright, S. 1969. Evolution and Genetics of Populations. The Theory of Gene Frequencies. Univ. Chicago Press, Chicago, IL.

Wu, N. C., L. Dai, C. A. Olson, J. O. Lloyd-Smith, and R. Sun. 2016. Adaptation in protein fitness landscapes is facilitated by indirect paths. eLife 5: e16965.

Zheng, J., J. L. Payne, and A. Wagner. 2019. Cryptic genetic variation accelerates evolution by opening access to diverse adaptive peaks. Science 365: 347-353.

Figure 5.1. An idealized representation of the distribution of deleterious fitness effects of new mutations, based on arguments presented in the text. Although it is clear that most mutations have relatively small and deleterious effects, there remain substantial uncertainties about the detailed form of the distribution in the neighborhood of 0.0, as well as about the degree to which the distribution varies among phylogenetic lineages. The secondary peak to the right is meant to represent a class of mutations of relatively large effects, as expected with amino-acid substitutions in catalytic sites of proteins, premature stop codons, and gene deletions. The distribution of beneficial effects is not shown, but in total this is expected to be $< 1\%$ of that for deleterious mutations, again with the mode being near zero.

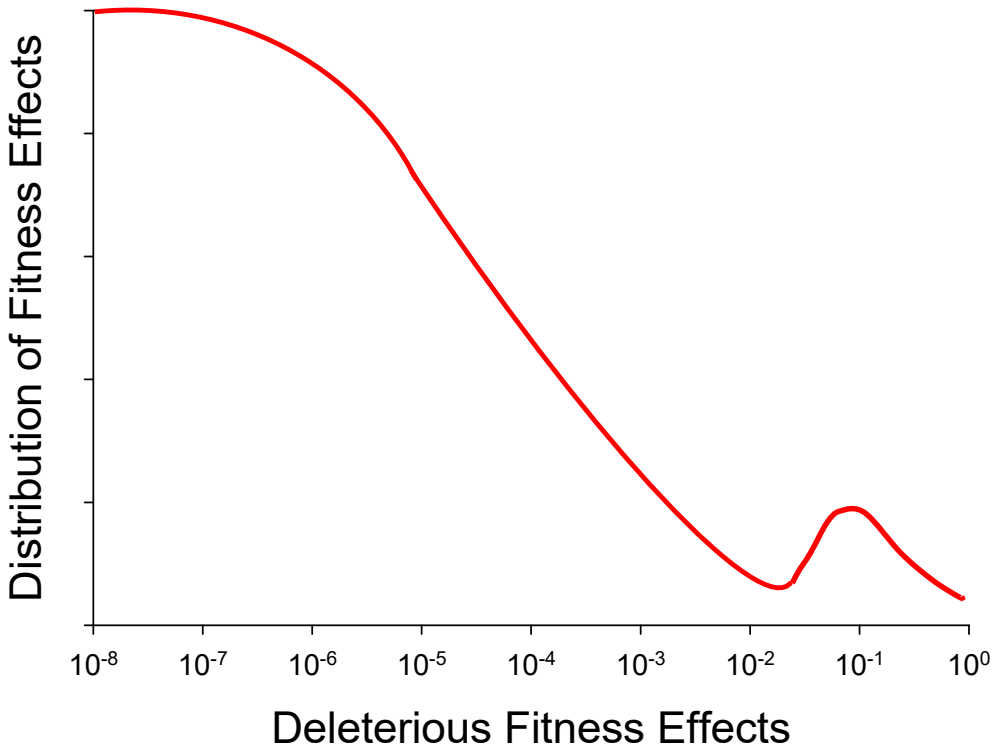


Figure 5.2. Allele-frequency trajectories in three replicate cultures of *E. coli* grown in 10-mL cultures serially diluted 100-fold on a daily basis. To estimate allele frequencies, the complete genomes of each mixed culture were subject to pooled-population sequencing, to an average of 50× depth of coverage, every 500 generations over a 60,000-generation period. Each individual line denotes a mutation that arose to frequency 0.1 on at least one occasion. All cultures were genetically identical and monomorphic at time zero. From Good et al. (2017).

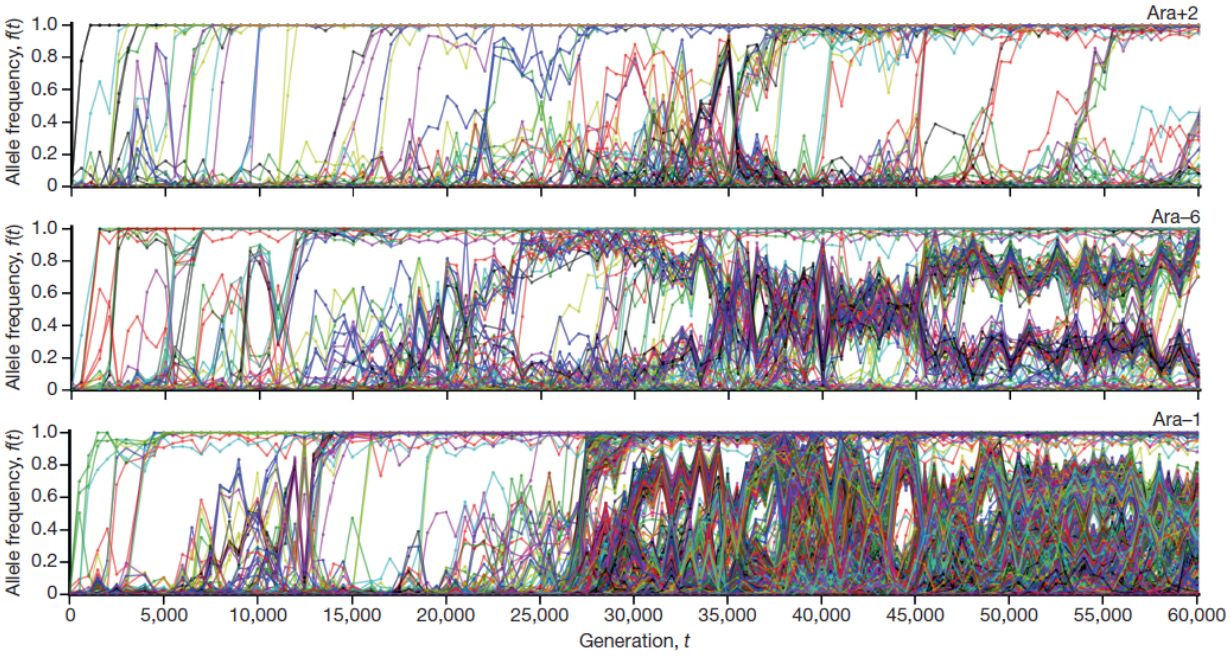


Figure 5.3. Some possible routes to the establishment of adaptations involving two or more mutations at different genetic loci (or nucleotide sites within single loci). In each case, the starting genotype is on the left. Mutations that are deleterious with respect to their ancestral allele are denoted in blue, whereas neutral changes are in white, and beneficial changes are denoted in red. Intermediate states can also be beneficial, although such instances are not shown. The lower left denotes a case involving three loci, with all intermediate states being deleterious, and the final three-mutation genotype being beneficial. The lower right illustrates a situation in which the first site has three alternative allelic states, with one derived allele being deleterious and the other providing a three-step neutral path to the final adaptation.

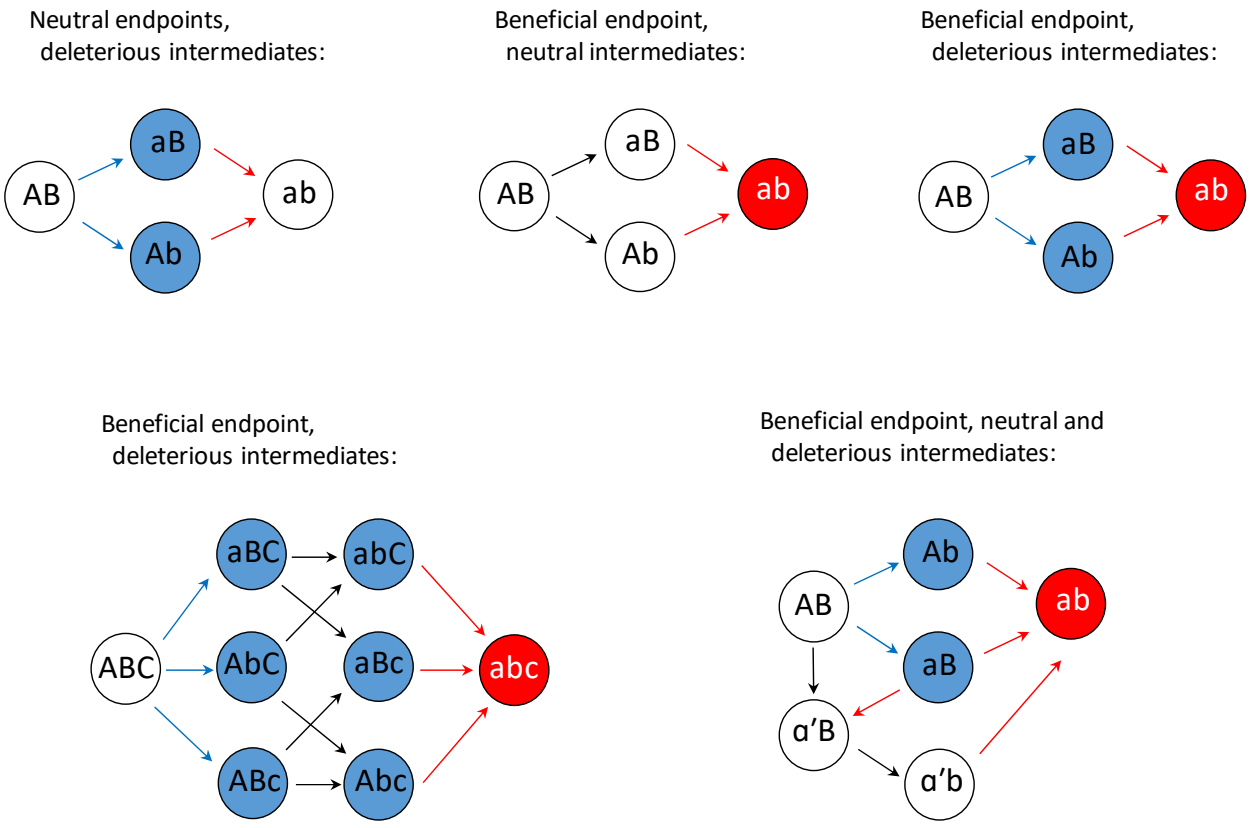
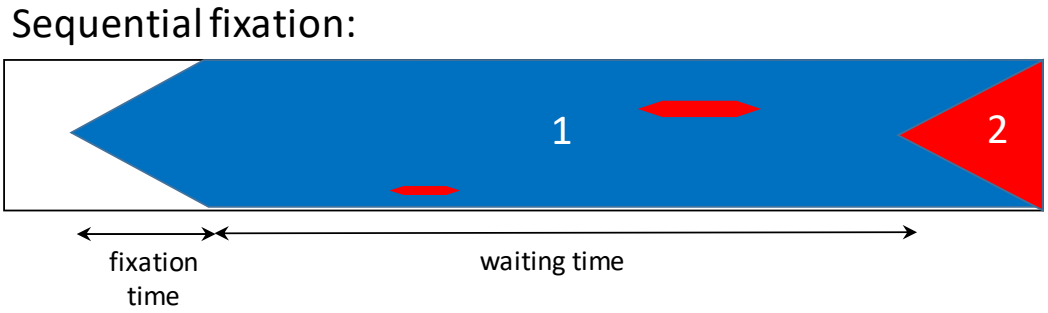


Figure 5.4. Origin of a complex adaptation involving three mutations (blue (1), red (2), and green (3)) in small and large populations. Fixation is implied when a vertical line through the temporal profile is all of one color. **Top)** Here, the population size is small enough and mutation sufficiently weak that each contributing mutation goes to fixation prior to the appearance of the next mutation destined to fix; this leads to sequential fixation of single-step changes, with potentially long waiting times between fixation events. A few secondary mutations may arise before one reaches a high enough frequency to be susceptible to fixation. In the illustrated case, the final (green) mutation has not arisen yet. **Bottom)** In large populations, allelic variants will typically be segregating at low frequencies at all time, even if deleterious, owing to the recurrent input by mutation. Secondary and tertiary mutations can then arise on such backgrounds, occasionally generating a beneficial combination that leads to fixation of the entire linked haplotype by positive selection.



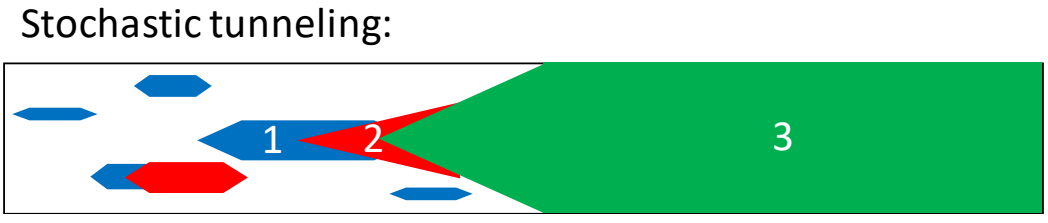


Figure 5.5. Expected frequency of the beneficial allele at a biallelic locus under the joint forces of drift, mutation, and selection. The mutation rate to deleterious alleles μ_{10} is assumed to be $3\times$ that to beneficial alleles (μ_{01}), as would be approximately the case for complementary nucleotides in an RNA stem. Results are given for four intensities of selection (with s denoting the selective advantage of the beneficial allele) relative to drift. Solid lines give the exact results from Foundations 5.2, whereas the dashed lines are the results under the sequential model, which assumes the population to be nearly always in a monomorphic state, Equation 5.9.

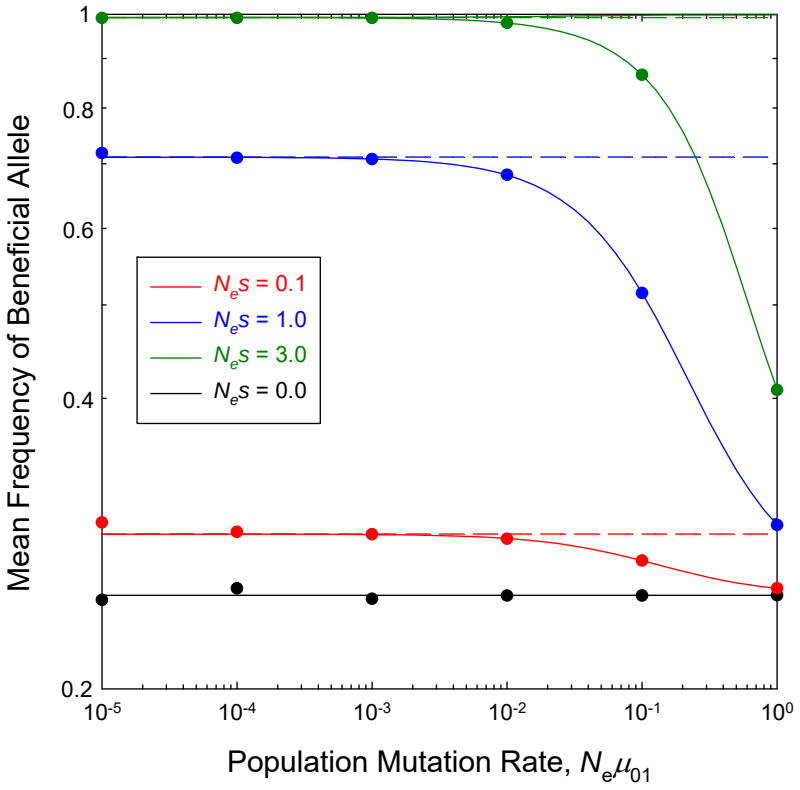


Figure 5.6. Schematic for the transition rates between adjacent classes under the sequential-fixation model for the case of $L = 4$ sites. This layout readily generalizes to any value of L. μ_{01} and μ_{10} are the mutation rates from $-$ to $+$ allelic states, and vice versa, and ϕ_{ij} is the probability of fixation of a newly arisen mutation to allele j from a background of i, defined by Equation 4.1b. The number of $-$ alleles in a class is denoted by i, and except for the two extreme classes ($i = 0$ and $i = 4$), there are multiple equivalent genotypic states within each class of genotypic values.

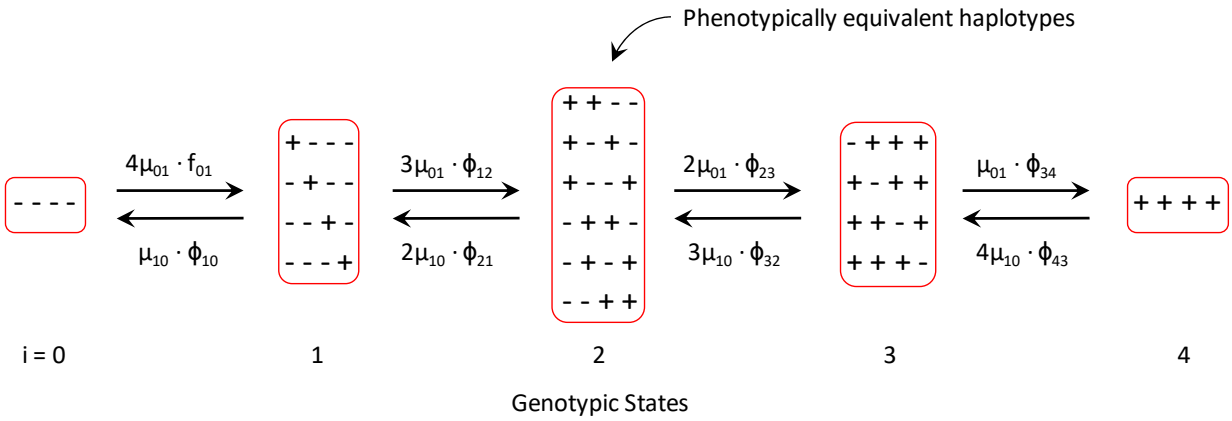


Figure 5.7. a) The response of the long-term mean genotypic state (number of + alleles; upper panel) and the underlying mean class frequencies (lower panel) over a gradient of effective population sizes for a two-locus, two-allele model. The mutation rate to beneficial alleles is 10% of that to deleterious alleles ($\beta = 0.01$). Results are given for three phenotypic optima, with the width of the fitness function $\omega = 5000$. b) For each color-coded optimum, the mean frequencies of the three genotypic classes (0, 1, 2) are given as solid, short-dashed, and long-dashed lines, which for each N_e sum to 1.0. Results are derived from Equations 5.10 and 5.11. From Lynch (2020).

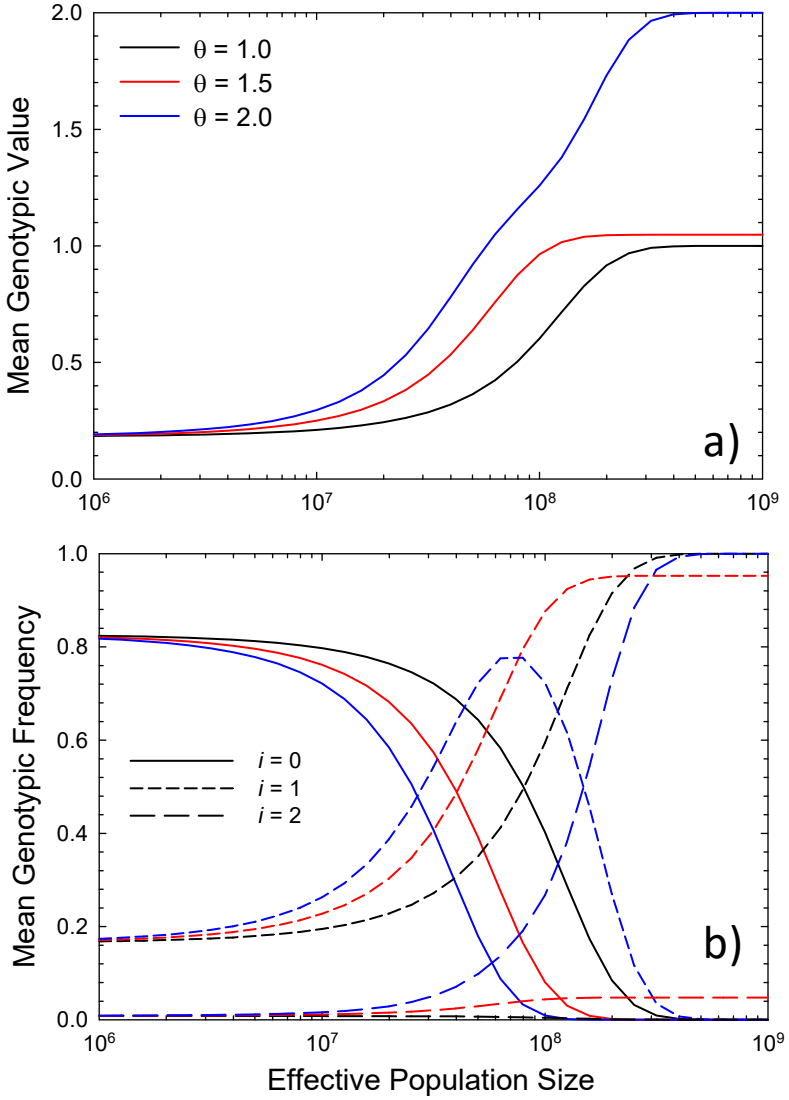


Figure 5.8. The equilibrium mean frequency of + alleles as a function of the absolute population size for the case in which the mutation bias to + alleles is 0.01. Results are given for three numbers of loci (L) with equivalent additive effects on the growth rate and multiplicative effects on fitness. 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, the selective disadvantage of a $-$ allele is $s = 1/L$, the inverse of the number of factors contributing to the trait: red, $L = 10^4, s = 10^{-4}$; green, $L = 10^5, s = 10^{-5}$; blue, $L = 10^6, s = 10^{-6}$. From Lynch (2020).