æ

4. THE POPULATION-GENETIC ENVIRONMENT

8 October 2021

The paths open to evolutionary change depend on pre-existing conditions, as all evolutionary processes operate by modifying extant genotypic variation. However, to understand the kinds of modifications that are evolutionarily possible, one must start with an appreciation of population-genetic processes, as the three major dimensions of the population-genetic environment ultimately dictate what natural selection can and cannot accomplish. Mutation gives rise to the variation upon which evolution depends. Recombination reassorts variation between nucleotide sites among individuals in ways that can either accelerate or impede evolutionary progress. Finally, random genetic drift serves as a lens on the evolutionary process, modulating the level of noise in allele transmission across generations, hence dictating the efficiency of selection. Although it is tempting to stare at biodiversity and spin adaptive stories as to why things are so, doing this in an absence of an understanding of evolutionary genetic processes invites the pixies in.

The following three chapters introduce in a nontechnical way the minimum set of principles required to construct a logical evolutionary argument. In addition to introducing some of the most elementary aspects of evolutionary theory, this chapter will summarize the state of knowledge on the three dimensions of the population-genetic environment. All three factors will be shown to vary by several orders of magnitude across the Tree of Life, albeit in nonindependent ways. In particular, mutation and recombination rates will be shown to be strongly influenced by the power of random genetic drift.

Quantitative information on how these forces are distributed across species is critical to understanding inter-species divergence. For example, if the power of genetic drift exceeds the strength of selection operating on a particular variant, the latter will be essentially immune to selection and will evolve in the direction dictated by any prevailing mutation bias. This means that in species experiencing low levels of random genetic drift (small microbes), natural selection can take advantage of mutations of small effects that are unavailable to adaptive exploitation in species experiencing higher levels of noise.

We will start with a more formal presentation of the preceding ideas. First, random genetic drift will be shown to be influenced by both population size and chromosomal architecture, leading to the concept of a genetic effective population size (N_e), which can be orders of magnitude smaller than the actual census size. Second, it will be shown how drift competes with the selection process, with $1/N_e$ defining an approximate benchmark above and below which selection operating on a mutation is effective vs. ineffective. Third, patterns of variation in rates of mutation and recombination across the Tree of Life will be shown to be consistent with some relatively simple models based on random genetic drift. Further technical details on

these matters can be found in Walsh and Lynch (2018, Chapters 2-7). Drawing from the information presented here, the following two chapters will take things a step further by considering more specifically how the population-genetic environment modulates various processes involved in cellular evolution.

Demystifying Random Genetic Drift

Critical to understanding all aspects of evolution is the large role of chance in determining the fates of mutant alleles. Consider a newly arisen mutation in a haploid population of N individuals, each of which produces a large number of gametes. To produce the next generation, N newborns must be drawn from this pool. Because a new mutant allele is present in a single copy, it has a frequency of $1/N$, and the probability that a randomly drawn gamete is not of this type is $[1-(1/N)]$. There is then a $[1-(1/N)]^N \simeq e^{-1} \simeq 0.368$ chance that none of the newborns contain the mutation, in which case the new mutation is lost in the first generation. Although this result assumes a neutral mutation, the probability of immediate loss is not much different for an allele with fractional benefit s, as the previous expression generalizes to $\{1-[(1+s)/N]\}^N \simeq e^{-(1+s)}$, which is 0.333 when $s = 0.1$ (an enormous 10% selective advantage, far beyond what is typically observed in nature). If the population is diploid, a $2N$ is substituted for N in the preceding expressions, but the results are the same. Thus, natural selection is a wasteful process in that the vast majority of new mutations, no matter how beneficial, are lost by chance.

If the mutant allele is fortunate enough to survive the first generation, the process of stochastic sampling will occur again. Each generation, the allele frequency can wander up or down, but due to the cumulative effects of this sorting process over many generations, all mutant alleles eventually suffer the fate of either loss (returning to a frequency of 0.0) or fixation (progressing to a frequency of 1.0) (Figure 4.1). For a neutral mutation, there is no directional pressure on the frequency change across generations, and the probability of fixation is simply equal to the initial frequency ($1/N$ for a new mutation in a haploid population, and $1/(2N)$ in a diploid).

If over a series of generations, a beneficial mutation does wander to a sufficiently high density, so that the probability of chance loss in any particular generation becomes small relative to the strength of selection, then natural selection can propel it to eventual fixation in a nearly deterministic fashion. The critical frequency depends on a key feature known as the *effective population size,* generally abbreviated as N_e. Only in an ideal population, where each adult has an equal probability of contributing to each offspring and family sizes are completely random, does N_e equal the *absolute population size* N.

Nearly every conceivable property of natural populations conspires to cause a population to behave genetically as though it is much smaller than its actual size. For example, owing to some individuals acquiring more resources than others or attracting more predators or pathogens, nonrandom variation in family size causes gene transmission to the next generation to be dominated by the most successful individuals. In addition, because each individual in a sexual population has two parents, with an uneven sex ratio, the genetic effective population size will more

closely resemble the number of the rarer sex. Population subdivision can further reduce the species-wide N_e, as individual demes become increasingly inbred, and reflect the states of smaller numbers of gametes than expected under panmixia. And finally, fluctuations in population size have a major effect because reductions in numbers of individuals have a much more substantial influence on sampling noise than do increases.

All of the factors just noted can be viewed as being defined by ecological factors imposed on species, without regard to their genetic constitution. However, an even more important determinant of gene-transmission stochasticity is the physical structure of the genetic machinery itself, and this is particularly true in large populations. Because genes are physically connected on chromosomes, the fate of a new mutation is determined by the states of the nucleotide sites to which it is initially chromosomally linked (Figure 4.2). A beneficial mutation that arises adjacent to other segregating alleles whose collective effects are sufficiently deleterious will be removed from the population unless it can be rapidly freed from such a background by recombination. Because the vast majority of mutations have deleterious effects, this kind of background-selection process (Charlesworth 2012) is expected to be on-going in all populations, but will be more significant in large populations, which generally harbor more variation (below). In contrast, a sufficiently beneficial mutation that is propelled to fixation by natural selection will drag all other linked mutations to fixation as well (including those that are mildly deleterious). Such selective sweeps have the same effect as a bottleneck in population size, although the effects of each sweep are restricted to specific chromosomal regions.

No population can avoid stochastic effects associated with finite numbers of individuals and chromosomal linkage, but what is the expected magnitude of generation-to-generation fluctuations? We start with the simple situation in which the noise in the evolutionary process is entirely due to random sampling of parental gametes across generations, as this is identical to a coin-flipping problem. Imagine two alleles, A and a, with respective frequencies in the population equal to p and $(1-p)$. If N alleles are randomly drawn to produce the next generation, as in an ideal haploid population, the frequency of A in the next generation will almost certainly be slightly different than p, just as the summed fraction of heads drawn from throws of an unbiased coin will deviate slightly from 0.5. The variance in allele frequency among independent sets of N draws equals $p(1-p)/N$, showing that the smaller the number of draws (N), the larger the change in p across generations. (This becomes $p(1-p)/(2N)$ in a diploid population).

Although random genetic drift is a relentless process, these results imply that the consequences of drift unfold on longer time scales in larger populations. Consider the early generations of the drift process for a neutral allele with intermediate frequency. As the process has no memory, the variance in allele-frequency change is cumulative across generations, i.e., $p(1-p)/N$ after the first generation, and approximately $2p(1-p)/N$ after the second generation, and $tp(1-p)/N$ after the tth generation (provided $t \ll N$). Thus, if the population size is doubled, it will take twice the number of generations to achieve the same level of allele-frequency change as in a population of size N. The key issue is that the time scale of random genetic drift is inversely proportional to the population size (Figure 4.1). As discussed further below, this linear scaling eventually breaks down as all alleles become fixed or lost.

The deeper problem here is the one noted above – that the absolute number of individuals in the population, N, is generally not sufficient to define the stochasticity of allele-frequency change. To accommodate the substantial complexity of the problem, population geneticists rely on the effective-population size concept. The goal is to determine the size of an ideal population (like that envisioned in the preceding paragraph) that most closely mimics the between-generation dispersion of allele frequencies experienced by the actual population. In other words, given the myriad of usually undefined ecological and genome-architectural issues, we desire a composite measure of the effective population size (N_e) that yields a dispersion of allele frequencies across generations approximately equal to $p(1-p)/N_e$.

This concept of an effective measure of an assumed underlying parameter may be viewed as an oversimplification by some readers. However, although usually unstated, the use of surrogate measures underlies a myriad of model-fitting exercises in science – we attempt to reason out a theoretical framework to explain an expected set of observations based on assumed underlying mechanisms, and then obtain model parameter estimates that best fit the data. The limitations of the human mind demand such simplification, and in biology, we are often not bothered if an approximation leads us astray by no more than a few percent.

Numerous mathematical formulations have been developed to explicitly link N_e to N under various conditions involving demography and chromosomal architecture (Walsh and Lynch 2018, Chapter 3). Although these can be difficult to implement in a practical sense, the general view is that, owing to the totality of genetic-interference effects, for populations with absolute sizes in the range of unicellular species, N_e grows only logarithmically with N, increasing $\sim 2\times$ with each tenfold increase in the latter (Neher 2013; Lynch 2020). The following section shows that estimates of N_e indirectly derived from empirical data are in rough accord with this weak scaling.

The Genetic Effective Sizes of Populations

In Chapter 1, the case was made that the average number of individuals per species is on the order of 10^{21} for both bacteria and unicellular eukaryotes, and at least nine to ten orders of magnitude lower for multicellular eukaryotes (with a very large range of variation within each group). Absolute numbers like these are of relevance to the field of ecology, but they may also leave the false impression that there is little room for the role of chance in long-term evolutionary processes.

There are numerous ways to estimate genetic effective population sizes, the most logically compelling being an evaluation of the fluctuations in allele frequencies across generations. As noted above, the variance in allele-frequency change across generations for a neutral nucleotide site is simply $p(1-p)/N_e$ or $p(1-p)/(2N_e)$, for haploid and diploid populations respectively, where p is the initial allele frequency. However, observations of allele-frequency changes are only reliable with very small populations or very long periods of elapsed generations, as otherwise the bulk of such change is simply due to sampling error on the part of the investigator.

The most powerful alternative approach is to evaluate the standing level of variation in a population under the assumptions that the nucleotide sites being observed are neutral and have reached levels of variation expected under the balance

between recurrent input by mutation and loss by drift. As noted in Foundations 4.1, the expected average level of nucleotide variation (defined as heterozygosity) at neutral sites is $\theta \simeq 2N_e u$ or $4N_e u$, again for haploid vs. diploid populations, where u is the rate of base-substitution mutation per nucleotide site per generation. Note that the composite parameter θ has a simple interpretation. It is equivalent to the ratio of the power of mutation ($2u$ for two sequences being compared) and that of drift ($1/N_e$ or $1/2N_e$). Notably θ is also a function of the cumulative variation generated over multiple past generations.

Estimates of standing-levels of heterozygosity at neutral nucleotide sites are generally obtained by confining attention to synonymous positions within codons of protein-coding genes (e.g., third positions in codons that specify the same amino acid whether they are occupied by A, C, G, or T) or deep within introns or intergenic regions (where no functional sites are thought to reside). In practice, θ is estimated by obtaining average estimates of neutral-site heterozygosity over a large number of sites in a sample of individuals. Such estimates integrate information over approximately the past N_e generations, which is equivalent to the average number of generations separating two random alleles in haploid populations (Kimura and Ohta 1969; Ewens 2004). This point can be seen by noting that two alleles will accumulate mutational differences at a rate $2u$ per site, which after an average of N_e generations sums to $\theta = 2N_e u$ (for a diploid population, the average separation time is $2N_e$ generations, yielding $\theta = 4N_e u$).

Estimates of θ derived with this approach have been summarized for a wide range of species across the Tree of Life by Lynch (2007) and Leffler et al. (2012), and more specifically for metazoans and land plants by Romiguier et al. (2014), Corbett-Detig et al. (2015), and Chen et al. (2017). For this diverse assemblage of eukaryotic and prokaryotic species, there is a negative association between organism size and θ, with estimates for prokaryotes roughly averaging ~ 0.10, those for unicellular eukaryotes averaging ~ 0.05, invertebrates ~ 0.03, and land plants and vertebrates generally being < 0.01. These are very approximate average estimates, and there is considerable variation around the mean. The main point is that for diverse sets of organisms, the range in estimated θ (in part, a reflection of N_e) is only $\sim$ 10-fold. In contrast, as noted above, the absolute numbers of individuals per species differ by many orders of magnitude.

There are a number of sources of potential bias in these estimates. For example, for both bacteria and unicellular eukaryotes, most silent-site heterozygosity measures are derived from surveys of pathogens, whose N_e may be abnormally low because of the restricted distributions of their multicellular host species. In addition, observed levels of silent-site diversity will deviate from the neutral expectation if such sites experience some form of selection. The direction of bias depends on whether selection opposes or reinforces any prevailing mutation bias. If there is a conflict between selection and mutation bias, expected levels of heterozygosity can exceed the neutral expectation. Most results are consistent with this type of conflict, but the resultant levels of heterozygosity are inflated by no more than three to four-fold (Long et al. 2017). With these caveats in mind, the existing data lead to a compelling conclusion with respect to the relative power of mutation and random genetic drift – in essentially no species does the former exceed the latter (as this would cause $\theta = 4N_e u > 1$).

Taken at face value, the preceding results might suggest that average N_e varies by no more than an order of magnitude between organisms as diverse as bacteria and microbes. However, a problem remains in that θ is a function of the product of N_e and the mutation rate. The conclusion that N_e is relatively constant only follows if mutation rates are also relatively constant, which will be shown not to be the case in the following section. Factoring out known estimates of u from θ yields estimates of N_e ranging from 10^4 in some vertebrates to $> 10^8$ in many bacteria (Figure 4.3). Over a nearly 10^{20} range of variation in adult size, N_e exhibits a negative power-law relationship with organism size.

Why the structure of life has led to this particular range and scaling remains unclear. The fact that N_e estimates fall many orders of magnitude below the actual numbers of individuals per species, especially in the case of microbes, is consistent with the point made above that the predominant source of drift is the stochasticity in gene transmission resulting from jointly segregating polymorphisms linked on chromosomes. Recall from above the theoretical expectation that N_e increases logarithmically with N (Neher 2013; Lynch 2020). If we assume that species with the lowest N_e ($\simeq 10^4$) have $N \simeq 10^6$, then an average bacterial species with $N \simeq 10^{21}$ (from above) would be expected to have an $N_e \simeq e^{15} \times 10^4 \simeq 10^{10}$. Considering the crudeness of this estimate, it is remarkably close to the upper limit of observed $N_e \simeq 10^9$ (Figure 4.3; and Bobay and Ochman 2018).

As will be shown in the next section, the fact that no species appears to have $N_e > 10^9$ means that beneficial mutations with selective advantages smaller than 10^{-9} cannot be exploited by natural selection in any lineage. In contrast, mutations with advantages smaller than about 10^{-5} are unavailable to selection in multicellular species with $N_e \simeq 10^5$. Thus, the range of sensitivity to mutations with fitness effects expands by about four orders of magnitude with decreasing organism size, enabling selection to operate in a more fine-grained manner in small organisms.

Probability of Fixation of a Mutant Allele

We now consider in more quantitative detail the specific issue of the probability of fixation (i.e., of rising to frequency 1.0) of a newly arisen mutant allele, as this will clarify the different roles played by both absolute and effective population sizes, while also demonstrating more formally the way in which the efficiency of natural selection is dampened by the magnitude of random genetic drift. As noted above, fixation probabilities are virtually always $\ll 1.0$. Their magnitude depends on: 1) the initial frequency p_0, which for a new mutation is a function of the actual population size, with $p_0 = 1/N$ or $1/(2N)$ for haploid and diploid populations; 2) the strength of selection; and 3) the effective population size N_e.

Here, we focus on a diploid population, with the mutant allele having additive fitness effects (such that each copy of the allele changes fitness by an amount s, yielding genotypic fitnesses of 1, $1 + s$, and $1 + 2s$, with heterozygotes being intermediate to the two homozygotes); see Walsh and Lynch (2018, Chapter 7) for more complex situations. Taking into consideration the stochastic effects noted above, Malécot (1952) and Kimura (1957) found that the probability of fixation of an allele

starting at initial frequency p_0 is

$$\phi_f(p_0) \simeq \frac{1 - e^{-4N_e s p_0}}{1 - e^{-4N_e s}}. \tag{4.1a}$$

(The same formula applies to haploid populations if a 2 is substituted for each 4). For a newly arisen mutation, $p_0 = 1/(2N)$ for diploids, Equation 4.1a reduces to

$$\phi_f(1/2N) \simeq \frac{1 - e^{-2(N_e/N)s}}{1 - e^{-4N_e s}}. \tag{4.1b}$$

As discussed above, in almost all natural settings $N_e/N \ll 1$, and most mutations have only minor effects on overall fitness (Chapter 5), so that $|s| \ll 1$. Thus, noting that $e^{-x} \simeq 1 - x$ for $x \ll 1$, the numerator is closely approximated by $2s(N_e/N)$, simplifying things further to

$$\phi_f(1/2N) \simeq \frac{2s(N_e/N)}{1 - e^{-4N_e s}}. \tag{4.1c}$$

It is useful to note that $2N_e s = s/[1/(2N_e)]$ is equivalent to the ratio of the power of selection to that of drift.

Four limiting conditions are clear from Equations 4.1a-c. First, for strong selection relative to drift, $2N_e s \gg 1$, $e^{-4N_e s} \simeq 0$, and the denominator is essentially equal to 1.0, showing that the probability of fixation of a new beneficial mutation is just $2s\,(N_e/N)$. Thus, even in populations with very large N_e, the probability of fixation of a beneficial mutation is smaller than twice the selective advantage. These results formalize the point made earlier – owing to the high probability of stochastic loss in the earliest generations, even strongly beneficial mutations only rarely proceed to fixation.

Second, only after the frequency of a beneficial allele becomes sufficiently high, does the probability of fixation become almost certain. For example, from Equation 4.1a, if $N_e s p_0 > 0.5$, the probability of fixation exceeds 0.70, and if $N_e s p_0 > 1$, $\phi_f(p_0) > 0.93$. Thus, as a matter of convention, it is often argued that to be assured near-certain fixation, a beneficial mutation requires a starting frequency of $p_0 \gg 1/(4N_e s)$.

Third, as $|4N_e s| \to 0$, the probability of fixation converges to the initial frequency (p_0). This should be intuitive, as under these conditions drift is so dominant that there is effectively no directional pressure on allele-frequency change. For this reason, the domain in which $|s| \ll 1/(4N_e)$ is known as effective neutrality. The salient point is that natural selection is unable to purge deleterious mutations or promote beneficial mutations with effects $< 1/(4N_e)$. The central principle is that the domain of variation recognizable by natural selection is expanded in large populations.

Finally, it follows from the above that the probability of fixation of a newly arising, neutral mutation ($s = 0$) in a diploid population is always equal to its initial frequency, $1/(2N)$, regardless of N_e. This has interesting implications for some forms of molecular evolution. Letting the mutation rate per nucleotide site equal u, then $2Nu$ mutations arise in the population each generation, so that the long-term rate of evolution at a neutral site is simply equal to the product $2Nu \cdot [1/(2N)] = u$. Thus, the rate of molecular evolution at neutral sites is equal to the mutation rate,

independent of the mode of gene action, rates of recombination, and population size (Kimura 1983).

Evolution of the Mutation Rate

The fact that no organism has evolved to have 100% replication fidelity is consistent with basic thermodynamic principles, but the issues are much deeper than this. There is a 1000-fold range of variation in the mutation rate among species, raising the question as to why particular rates are associated with particular lineages. As first pointed out by Kimura (1967), selection operates on the mutation rate indirectly, via the effects of mutations linked to alleles associated with their production. Under this view, a newly arisen mutator allele progressively acquires an excess linked mutation load. Thus, given that the vast majority of mutations are deleterious (Lynch et al. 1999; Baer et al. 2007; Eyre-Walker and Keightley 2007; Katju and Bergthorsson 2019), it follows that natural selection generally strives to minimize the mutation rate.

The ability of natural selection to eradicate a mutator allele is a function of the magnitude of this associated mutation load, which is equal to the product of three terms: 1) the excess genome-wide rate of production of deleterious mutations relative to the pre-existing population mean, ΔU_D; 2) the reduction in fitness per mutation, s; and 3) the average number of generations that a mutation remains associated with the mutator, $\bar{t}$. The persistence time $\bar{t}$, in turn, is determined by two factors: 1) the selective disadvantage of mutations (s), with individuals harboring an additional mutation leaving a fraction s fewer progeny, and those with n mutations having fitness $(1-s)^n$; and 2) the rate of recombination (r), which physically dissociates the mutator from the load it creates.

The strength of selection against a mutator is greatest in the case of asexual reproduction, as the mutator is never separated from its mutation load by recombination. Provided $s \gg 1N_e$, s is equivalent to the rate of removal of a deleterious mutation from the population by selection, and the persistence time is simply equal to the reciprocal of the average rate of removal of a mutation (i.e., $\bar{t} = 1/s$). The selective disadvantage of a mutator allele is then the product of the three terms noted above: $s_m = \Delta U_D \cdot s \cdot (1/s) = \Delta U_D$, and hence is simply equal to the increased per-generation rate of production of deleterious mutations. The selective disadvantage is completely independent of the effects of mutations, as mutations with larger effects are selectively removed from a population at higher rates (along with the linked mutator allele).

Recombination weakens the selective disadvantage of a mutator allele by exporting the initially linked mutations to other members of the population. With free recombination ($r = 0.5$), mutant alleles are statistically uncoupled from their source in an average of just $\bar{t} = 1/r = 2$ sexual generations. The mutator-allele disadvantage then becomes $s_m \simeq 2s\Delta U_D$. Not every mutation will be freely recombining with respect to a mutator allele, but because most eukaryotic genomes contain multiple chromosomes, most new mutations will arise unlinked to the mutator, and the load is unlikely to be greater than twice the preceding value (Lynch 2008). The key point is that sexual reproduction weakens the magnitude of selection against a

mutator allele (relative to the case of complete linkage) by a factor of $\sim 2s$. As will be discussed in the following chapter, the average fitness effects of new mutations are generally < 0.1, implying that recombination reduces the strength of selection operating on the mutation rate by at least 80% relative to the case under asexuality.

Impeding the directional selection pressure to reduce the mutation rate is random genetic drift, which begins to prevail once the level of replication fidelity becomes so highly refined that the next incremental improvement is effectively neutral, i.e., $s_m \ll 1/N_e$. The general principle can be understood by reference to the logic outlined in the preceding section. Letting $\eta < 1$ be the fraction by which the mutation rate is reduced by an antimutator, so that $\Delta U_D = \eta U_D$, the drift-barrier hypothesis postulates that once the genome-wide deleterious mutation rate U_D in an asexual (haploid) population is reduced to the point that $\eta U_D < 1/N_e$, selection for further reduction in the mutation rate will be overwhelmed by stochastic noise. The exact location of such a barrier will also depend on the relative rates of production of mutator and antimutator alleles (Lynch et al. 2016), as a substantial rate of production of mutation-rate enhancers will counter downward selection on the mutation rate to some degree.

The key point is that owing to the stochastic consequences of finite population size, the degree of refinement that natural selection can achieve increases with N_e, which allows alleles with smaller incremental benefits to be promoted (Lynch 2011, 2020; James and Jain 2016; Lynch et al. 2016). This drift-barrier hypothesis leads to the prediction that the mutation rate will evolve to be negatively associated with N_e (Figure 4.4). If this hypothesis is correct, the very process necessary for producing adaptive mutations is selected against, with the fuel for evolution (the small fraction of beneficial mutations) simply being largely an inadvertent by-product of an imperfect process.

How do the data accord with the drift-barrier hypothesis? Prior to this century, almost all estimates of the mutation rate were derived indirectly using reporter constructs in microbes (Drake 1991; Drake et al. 1998), leaving considerable uncertainties with respect to efficiency of mutation detectability, marker position effects, etc. With the advent of whole-genome sequencing, it became possible to evaluate the genome-wide appearance of mutations in replicate lines maintained by single-progeny (or single full-sib mating) descent (i.e., $N_e \simeq 1$ or 2) for large numbers of generations (Lynch et al. 2016; Katju and Bergthorsson 2019). By maximizing the power of drift relative to selection, such treatment ensures that essentially all mutations (other than the small fraction causing lethality or sterility) will accumulate in an effectively neutral fashion.

The numerous results from such work demonstrate a 1000-fold range of interspecific variation in the mutation rate per nucleotide site per generation, from a low of $\sim 10^{-11}$ in some unicellular eukaryotes to a high of $\sim 10^{-8}$ in some mammals (including humans) (Lynch et al. 2016; Long et al. 2017; Lynch and Trickovic 2020). Consistent with expectations under the drift-barrier hypothesis, the rate of base-substitution mutation per nucleotide site (u) scales negatively with N_e (Figure 4.5, left). Although the three major groups (bacteria, unicellular eukaryotes, and multicellular eukaryotes) appear as fairly discrete clusters, recall that the theory outlined above implies that selection operates on the *genome-wide* deleterious rate, which is the product of the mutation rate per nucleotide site (per generation) and

the number of nucleotides under selection (the effective genome size, P_e). Multiplying u by the effective genome size (approximated by the amount protein-coding DNA) unifies the overall set of results, leading to an inverse scaling between uP_e and N_e (Figure 4.5, right).

Thus, mutation rates are lower in unicellular eukaryotes than in bacteria with similar effective population sizes because the genomes of the former contain many more genes than those of the latter, and hence are larger targets for deleterious mutations. The scaling patterns in Figure 4.5 also hold for insertion-deletion mutations, which are about 10% as common as the former across the Tree of Life (Sung et al. 2016).

High mutability of mutation rates. The preceding results make clear that the mutation rate is evolutionarily malleable, but what is the timescale of such change? The mutational target size for the mutation rate is likely to be very large, as it includes multiple DNA polymerases, DNA-repair proteins, and essentially all genes whose products alter the mutagenicity of the intracellular environment (including those influencing the production of free oxygen radicals via metabolic activity and those modulating the relative abundances of free nucleotides). Thus, with the expectation that both mutators and anti-mutators are recurrently introduced into all populations (Denamur and Matic 2006; Lynch 2008; Raynes and Sniegowski 2014), virtually all natural populations are expected to harbor polymorphisms for the mutation rate.

The idea that the mutation rate is capable of rapid change is supported by a diversity of observations. For example, Boe et al. (2000) estimate that *E. coli* cells with mutation rates elevated by 20 to 80× arise at rates of 5×10^{-6} per cell division, and one can imagine even higher rates of origin of milder (and less easily detected) mutators (as well as antimutators). Indeed, in an *E. coli* mutation-accumulation experiment initiated with a mutator strain that allowed accumulation of diversity over a period of 1250 generations, numerous lines evolved mutation rates $< 10\%$ of the baseline rate (antimutators), while a small fraction of them experienced up to 10× increases in the mutation rate (Singh et al. 2017).

Given this potential for rapid change in the level of replication fidelity, it is not surprising that microbes commonly evolve mutator genotypes when confronted with strong selective challenges (such as antibiotic treatment). Swings et al. (2017) found that in lethally stressful environments, laboratory *E. coli* populations rapidly evolve a mutator phenotype (on a time scale of ~ 100 generations), and then revert to background mutation-rate levels once adaptation has been achieved. In contrast, in constant environments, bacterial populations founded with a mutator genotype frequently evolve lower mutation rates on relatively short time scales via compensatory molecular changes at genomic sites not involved in the initial mutator construct (McDonald et al. 2012; Turrientes et al. 2013; Wielgoss et al. 2013; Williams et al. 2013). Taken together, these observations indicate that the mutation rate is among the most rapidly evolving traits known. In addition, the common appearance of antimutators implies the presence of substantial unexploited potential for improvement in replication fidelity.

As further evidence that despite their extremely low values, evolved mutation rates are not constrained by biophysical limitations, consider the fact that although

mammals harbor the highest known eukaryotic mutation rates per generation, the rate per germline cell division rivals the very low per-generation rates for unicellular species (Lynch 2010). The human germline mutation rate per nucleotide site is $\sim 6 \times 10^{-11}$ per cell division, approaching the lowest rates observed in unicellular eukaryotes, and 10 to 100× lower than rates in various human somatic tissues (Lynch 2010; Behjati et al. 2014; Milholland et al. 2017). The point here is that although selection operates on the per-generation mutation rate, this is accommodated by changes in replication fidelity at the cell-division level – an increased number of germline cell divisions is balanced by enhanced replication fidelity per division.

Error-prone polymerases. In all known organisms, almost all DNA replication is carried out by one or two major polymerases, each of which has a high baseline level of accuracy, with a substantial fraction of the few errors arising at this step being removed secondarily via a proof-reading step. However, the genomes of nearly all organisms also encode for one or more error-prone polymerases, whose usage is restricted mostly to times of stress or to dealing with bulky lesions in DNA. Stress-induced mutagenesis (SIM) has been found in virtually all organisms that have been examined, e.g., *E. coli* and many other bacteria (Kang et al. 2006; Foster 2007; Kivisaar 2010); yeast (Heidenreich 2007); *Chlamydomonas* (Goho and Bell 2000); *Caenorhabditis* (Matsuba et al. 2013); and *Drosophila* (Sharp and Agrawal 2012). The distinction between SIM and normal mutagenesis is often blurry, in that a variety of environmental stresses (e.g., nature of the limiting nutrient) alter the molecular spectrum of mutations without affecting the mutation rate itself (Maharjan and Ferenci 2017; Shewaramani et al. 2017).

An elevation in error rates under situations leading to physiological breakdown should not be too surprising, as this can be expected for virtually all traits. Nonetheless, some have argued that high mutation rates associated with error-prone polymerases have been promoted by selection as a means for generating adaptive responses to changing environments (Radman et al. 2000; Tenaillon et al. 2001; Earl and Deem 2004; Foster 2007; Galhardo et al. 2007; Rosenberg et al. 2012). Direct empirical support for such an argument is lacking, although special scenarios have been shown in theory to encourage selection for SIM, e.g., situations in which two mutations are required for an adaptation, with the first conferring reduced fitness and an elevated mutation rate when alone (Ram and Hadany 2014), or when stresses are sufficiently diverse and persistently fluctuating (Lukačišinová et al. 2017). However, establishing that an evolutionary outcome is theoretically possible is quite different than demonstrating a high likelihood of it actually occurring.

A simpler and more compelling explanation for the error-prone nature of some polymerases follows directly from the drift-barrier hypothesis – the net selection pressure to improve accuracy is expected to be proportional to the average number of nucleotide transactions that a DNA polymerase engages in per generation. Because error-prone polymerases generally replicate only small patches of DNA and do so quite infrequently, the strength of selection on accuracy will be correspondingly reduced (Lynch 2008, 2011; MacLean et al. 2013). This "use it or lose it" hypothesis is also consistent with the high error rates for polymerases deployed in the replacement of small RNA primers used in replication initiation (Lynch 2011). In addition, the secondary and tertiary fidelity mechanisms associated with replication

(proof-reading and mismatch repair), which necessarily involve far fewer nucleotide transactions than the earlier polymerization step, have greatly elevated error rates (Lynch 2008). None of these latter observations can be explained as specific adaptations to stress, as all of the factors involved are fundamental to normal replication cycles.

This view does not deny the critical importance of error-prone polymerases as mechanisms for dealing with bulky lesions or other forms of DNA damage, nor does it deny that induced mutagenesis can play a role in generating an appropriate adaptation in extreme times, sometimes being the only means for survival. It does, however, eliminate the need for an adaptive explanation for high error rates, implying instead that there is no way to avoid such an outcome.

Optimizing the mutation rate. Because mutational processes generate a large fraction of detrimental variants that must be removed by natural selection and a small fraction of beneficial mutations essential for adaptation in changing environments, considerable attention has been given to the idea that natural selection might fine-tune the mutation rate so as to maximize the long-term rate of adaptive evolution. In contrast to the drift-barrier hypothesis, under this view, selection does not constantly push the mutation rate to the lowest achievable level, but instead promotes higher levels of mutation via indirect effects associated with the small pool of beneficial mutations. This is a difficult area for theory development as the relative merits of increasing vs. decreasing the mutation rate depend on the distribution of mutational effects, the population size, the recombination rate, and the pattern of environmental change.

It is especially unclear how natural selection operating at the individual level can promote an elevated mutation rate in a sexual population by associated beneficial effects. The primary problem is that in a sexual population, hitch-hiking of a mutator with a linked beneficial mutation will generally be thwarted by their dissociation by recombination (on average in just two generations when the two loci are on different chromosome arms). Continuous reinforcement necessary for the promotion of a mutator allele requires a substantial rate of input of closely linked beneficial mutations. However, as noted above, because the vast majority of mutations are deleterious, there will be a steady-state background deleterious load associated with all mutator alleles, regardless of whether the mutator is involved in a transient (and most likely incomplete) beneficial sweep. An additional limitation in multicellular organisms is the direct negative effects that mutators impose via the production of somatic mutations, which have immediate detrimental effects on fitness (Lynch 2008, 2010).

Several attempts have been made to estimate the theoretically optimal mutation rates for maximizing long-term rates of adaptive evolution in nonrecombining, asexual populations, but the resultant models do not explain the most prominent pattern in the data – the inverse relationship between u and N_e (Figure 4.5). Nor do they explain why, if optimized, mutation rates are nearly $1000\times$ higher in large multicellular sexual species than in most microbes. Indeed, most models concerned with optimal mutation rates in persistently changing environments imply a positive association between N_e and u (Kimura 1967; Leigh 1970; Orr 2000; Desai and Fisher 2007; Johnson and Barton 2002; Good and Desai 2016). Thus, one could argue that

the utility of these models is not that they explain the data, but that they highlight how the kinds of imagined scenarios involving distributions of mutational effects are inconsistent with the data.

In summary, both data and theory are incompatible with the idea that mutation-rate evolution is guided by a population-level goal of maximizing the long-term rate of incorporation of beneficial mutations. Hypotheses based on numbers of cell-divisions per generation do not explain the data in Figure 4.5 either, as nearly the full range of variation in mutation rates per generation is encompassed by unicellular species alone. Nor does generation length explain the patterns, as unicellular eukaryotes have longer cell-division times but lower mutation rates than prokaryotes.

One lingering concern may be that, after accounting for effective proteome and population sizes, bacteria have no lower mutation rates than eukaryotes, whereas theory predicts that the efficiency of selection on the mutation rate is increased in the absence of recombination. As outlined below, however, a simple explanation for this is that although bacteria lack meiotic recombination, they nonetheless experience roughly the same amount of recombination per nucleotide site (via other mechanisms) at the population level as do eukaryotes.

The nonrandom nature of mutation. Decades of observations are consistent with the postulate that mutations arise randomly with respect to the demands for modifications imposed by natural selection. However, some have argued that selection is capable of modulating mutation rates on a gene-by gene basis by, for example, locating genes in regions with or without potentially mutagenic collisions between DNA and RNA polymerase or by somehow providing protection against mutagenic aspects of high rates of transcription (Martincorena et al. 2012; Paul et al. 2013). These claims have not held up to close scrutiny, and the theory outlined above explains why. The differences in mutation rates among genes associated with chromosomal locations and/or transcriptional activities are simply too small to be promoted by selection (Chen and Zhang 2013; Lynch et al. 2016).

Nonetheless, although mutations are random with respect to desirable gene targets, they are nonrandom in essentially every physical way. For example, in some bacteria there is a symmetrical wave-like pattern of the mutation rate around the circular chromosome (Foster et al. 2013; Long et al. 2015), although the amplitude of differences does not exceed 2.5× and the pattern differs among species. Up to two-fold differences have also been found among locations on eukaryotic chromosomes on spatial scales ranging from 200 bp to 100 kb (Stamatoyannopoulos et al. 2009; Chen et al. 2012; Lang and Murray 2011). The molecular mechanisms driving these large-scale patterns remain unclear, but may be associated with variation in the nucleotide pool composition during the cell cycle, regional variation in transcription rates and their influence on replication, protection by nucleosomes in eukaryotes, and/or alterations in the rates of processivity of DNA polymerase across different chromosomal regions.

On a more local scale, every genome that has been assayed reveals uneven frequencies of the six types of base-substitution mutations (Long et al. 2017). In most prokaryotes, and all eukaryotes so far observed, there is mutation bias in the direction of $G+C \rightarrow A+T$. In addition, the mutabilities of the individual nucleotides are context dependent, depending on the nature of the neighboring nucleotides (Sung

et al. 2015).

The appearance of mutations can also be temporally correlated within the same genomes. The naive view is that if mutations arise at an average rate u per nucleotide site, the rate of simultaneous origin of mutations at two specific sites would be u^2, and at three sites would be u^3. Given that average u is on the order of 10^{-9}, this would imply that double mutants would rarely ever occur except in large microbial populations. However, data from mutation-accumulation experiments suggest that on spatial scales of 100 bp or so, multinucleotide mutations commonly comprise 1 to 3% of mutational events in diverse lineages (Drake 2007; Schrider et al. 2011; Harris and Nielsen 2014; Terekhanova et al. 2013; Uphoff et al. 2016). Potential reasons for mutational clusters include local patches of DNA damage, the occasional use of a defective DNA polymerase molecule, accidental deployment of an innately error-prone polymerase, and mutagenic repair of double-strand breaks (Drake 2007; Hicks et al. 2010; Malkova and Haber 2012; Chan and Gordenin 2015).

The key point here is that transient, localized hypermutation is common enough that rates of occurrence of double mutations are often many orders of magnitude above the u^2 expectation under independent occurrence, with double-mutation rates commonly on the order of $u/1000$ to $u/100$, and perhaps triple-mutation rates only a few orders of magnitude lower. The occurrence of mutation clusters has major implications for the evolution of complex features, as modifications requiring multiple nucleotide changes need not await the sequential fixation of individual mutations, but can arise *de novo* and be promoted together. This change in view becomes particularly important with respect to complex adaptations in which first-step single-nucleotide variants are deleterious (Chapters 5 and 6).

Recombination

Recombination is a double-edged sword in evolution. On the one hand, by eliminating peculiarities associated with individual genetic backgrounds, the physical scrambling of linked loci increases the ability of natural selection to perceive mutations on the basis of their individual effects. In addition, recombination can create favorable genetic interactions by bringing together mutations that have arisen on independent backgrounds. On the other hand, high rates of recombination can inhibit the permanent establishment of pairs of mutations with favorable interactive effects if they are separated more rapidly than they are advanced as a unit by selection.

Before proceeding, it may be useful to review the mechanics of recombination, as this will help clarify how the recombination rate scales with distance L between nucleotide sites (Foundations 4.2; Figure 4.6). Contrary to common belief, the recombination rate between sites is not equal to the crossover rate, except in the case of distantly located sites (typically, > 10 kb). This is because much of recombination involves localized patches of gene conversion that do not, in themselves, cause exchange of flanking chromosomal regions.

Two general approaches provide insight into the level of recombination per physical distance along chromosomes. The first of these involves the construction of genetic maps, usually from observations on the frequency of meiotic crossovers between molecular markers in controlled crosses. Such maps have the power to yield

accurate estimates of average recombination rates over fairly long physical distances (usually with markers being separated by millions of nucleotide sites, which typically corresponds to $> 1\%$ recombination per generation). However, without enormous numbers of evaluated progeny, such exercises cannot reveal recombination rates at small spatial scales, simply because of the absence of observed recombination events over short intervals. The details of constructing genetic maps will not be elaborated on here, other than to note that they are assembled by use of mapping functions that convert observed recombination frequencies into the expected numbers of crossovers between pairs of markers (Chapter 14 in Lynch and Walsh 1998).

Despite the limitations, results from genetic-map construction allow a compelling general statement about average genome-wide levels of crossing-over. Although eukaryotic genome sizes (G, the total number of nucleotides per haploid genome) vary by four orders of magnitude, the range of variation in genetic-map lengths (in units of the total number of crossovers per genome per meiosis) among species is only about ten-fold (Lynch 2007; Lynch et al. 2011; Stapley et al. 2017). This behavior can be explained by a very simple physical constraint, which appears to be nearly invariant across the eukaryotic phylogeny. During meiosis, there are typically no more than two crossover events per chromosome arm, regardless of chromosome size. Because phylogenetic increases in genome size are generally associated with increases in chromosome size rather than chromosome number (Lynch 2007), the little variation in the total number of meiotic crossover events per genome that exists among eukaryotes is due to variation in chromosome number.

These observations lead to a simple structural model for the average crossover rate per physical distance across a genome, which is technically equal to the product of the rate of recombination initiation per nucleotide site (c_0) and the fraction of such events that lead to a crossover (x) (Foundations 4.2). Letting M be the haploid number of chromosomes per genome, G/M is the average physical length of chromosomes. Letting κ be the average number of crossovers per chromosome per meiosis, then the average amount of recombination per nucleotide site associated with crossing over is $\bar{c}_0 \simeq \kappa M/G$. If this model is correct, a regression of $\bar{c}_0$ on G on a log scale should have a slope of -1.0, with the vertical distribution (residual deviations) around the regression line being defined largely by variation in M (with species with the same genome size but more chromosomes having proportionally more crossing over per nucleotide site).

The data closely adhere to this predicted pattern, with the smallest genomes of microbial eukaryotes having recombination rates per physical distance $\sim 1000\times$ greater than those for the largest land plants (which have $\sim 1000\times$ larger genomes, but approximately the same numbers of chromosomes) (Figure 4.7). Thus, the smooth, overlapping decline in recombination intensity across unicellular species, invertebrates, vertebrates, and land plants, reflects the general increase in genome sizes among the latter eukaryotic domains, with the smaller level of vertical variation in the plot reflecting differences in chromosome numbers (Lynch et al. 2011).

Although these observations suggest that the vast majority of the variance in the average crossing-over rate among eukaryotic species is simply due to variation in genome size and chromosome number, even the highest density genetic maps are generally unable to reveal the features of short chromosomal regions. Finer-scale molecular analyses have shown that up to 100-fold differences in recombination

rates can exist on spatial scales of a few kb, although the locations and molecular mechanisms dictating their distributions are highly variable among species (Petes 2001; de Massy 2003, 2013; Jeffreys et al. 2004; Myers et al. 2005; Arnheim et al. 2007; Coop et al. 2008; Mancera et al. 2008; Kohl and Sekelsky 2013; Lam and Keeney 2015; Haenel et al. 2018). For example, recombination hotspots in budding yeast are enriched in the transcriptional promoters of genes, whereas in fission yeast, they are enriched in intergenic regions. Although in neither yeast species is the guiding mechanism known, in numerous mammals, a zinc-finger protein (PRDM9) marks nucleosomes with a distinct form of histone phosphorylation in the vicinity of specific DNA-sequence motifs, guiding the recruitment of Spo11 (Baudat et al. 2013; Capilla et al. 2016; Wells et al. 2020).

In the genomes of mice and great apes, $> 20,000$ hotspots are defined in this way, although the localization of sites varies dramatically even among closely related species. Such shifts are associated with the rapid sequence evolution of the zinc-finger used in DNA motif recognition, leading to recombination-hotspot variation among subspecies and even among individuals within species (Brick et al. 2012). Some mammals, such as dogs, do not even have such a system. A key issue with respect to such a system is that if the positions of double-strand breaks are closely associated with the recognition motif, the latter will tend to eventually be lost by gene conversion to the allele on the recipient chromosome, hence leading to the eventual loss of the hotspot. A shift in the zinc-finger recognition motif in PRDM9 could then lead to altered hotspot localization, unfolding a new series of events. There is, however, no evidence that recombination hotspots are promoted by natural selection, and as yet there is no evidence of their existence in unicellular species.

A second approach to estimating recombination rates uses measures of linkage disequilibrium (LD) in natural populations to quantify the statistical degree of association between allelic states at linked chromosomal sites. Just as theory reveals an equilibrium average level of molecular heterozygosity (variance) within neutral nucleotide sites at mutation-drift equilibrium (Foundations 4.1), the level of LD (covariance) between sites is expected to reach a balance between the forces of recombination, mutation, and drift. The equilibrium LD between two sites (i and j) is a function of the composite parameter $4N_e c_{ij}$, which is proportional to the ratio of the rate of recombination between sites (c_{ij}) and the power of random genetic drift ($1/2N_e$ for diploids) (Chapter 4 in Walsh and Lynch 2018).

The attractiveness of the LD approach is that the observations reflect the historical outcome of many thousands of generations (and equivalently, thousands meioses). This provides the power to obtain much more refined (kilobase scale) views of the recombinational landscape than is possible with short-term breeding experiments. To understand this benefit, note that for a mapping cross involving n gametes with recombination frequency c_{ij} between sites i and j, the expected number of recombinants is nc_{ij}, so for sufficiently close sites ($c_{ij} \ll 1/n$), the typical outcome of cross will be a complete absence of recombinants. On the other hand, if n random diploid individuals are sampled from a natural population, because the mean time to a common ancestor between random neutral alleles is $\simeq 2N_e$ generations (noted above), the expected number of recombination events is $4N_e nc_{ij}$.

Numerous population-level surveys have been made to estimate the normalized recombination parameter $4N_e c_0$, where c_0 is the rate of recombination per nucleotide

site. This is usually done by first estimating the population-level parameter at various distances between sites (L_{ij}), and then dividing by L_{ij} under the assumption that $c_{ij} = c_0 L_{ij}$, i.e., a linear relationship between the recombination rate and physical distance between sites. As noted in Foundations 4.2, this is a reasonable approximation provided the distance between sites is less than the average length of a gene-conversion tract, but will lead to an underestimate of $4N_e c_0$ when greater distances are relied upon (by a factor up to 20×). Using this procedure, all estimates of the per-site parameter $4N_e c_0$ are smaller than 0.1, with many falling below 0.01 (Walsh and Lynch 2018, Chapter 4). These observations provide support for the idea that, as with mutation, random genetic drift is generally a more powerful force than recombination at the level of individual nucleotide sites, with the caveat that most existing analyses involve animals and land plants.

By dividing estimates of $4N_e c_0$ by parallel estimates of $\theta = 4N_e u$, the effective population size cancels out, yielding an estimate of the ratio of recombination and mutation rates at the nucleotide level (c_0/u). All such estimates for eukaryotes are smaller than 5.0, and nearly half are smaller than 1.0 (Lynch 2007; Walsh and Lynch 2018). For example, the average estimate of c_0/u for *Drosophila* species is ~ 2.7, whereas that for humans is ~ 0.8, and the average for fourteen land plants is 1.1. These observations imply that the power of recombination between adjacent sites is often of the same order of magnitude as the power of mutation, or perhaps somewhat larger owing to the downward bias in c_0 estimates noted above.

Contrary to common belief, relative to the background rate of mutation, recombination at the nucleotide level is not exceptionally low in bacteria (Shapiro 2016; Bobay and Ochman 2017; Garud and Pollard 2019; Sakoparnig et al. 2021). Although bacteria do not engage in the kinds of organized meiotic activities of eukaryotes, they have several other pathways that can lead to homologous recombination, including transduction of sequences by bacteriophage, physical conjugation and DNA exchange between conspecifics, and even consumption and integration of free DNA from dead cells. Consistent with high levels of recombination induced by these alternative mechanisms, estimates of c_0/u for such species are often of the same order of magnitude as those for eukaryotes (Lynch 2007; Vos and Didelot 2009; Rosen et al. 2015).

Finally, as noted above, not all recombination events involve crossovers. For example, direct empirical observations suggest that the fraction of recombination events accompanied by crossing over is $x \simeq 0.30$ in the budding yeast *S. cerevisiae* (Malkova et al. 2004; Mancera et al. 2008), and $\simeq 0.15$ in the fly *D. melanogaster* (Hilliker et al. 1994). Indirect LD-based estimates suggest $x \simeq 0.14$ in humans (Frisse et al. 2001; Padhukasahasram and Rannala 2013), 0.09 on average in other vertebrates (Lynch et al. 2014), $\simeq 0.08$ in *D. melanogaster* (Langley et al. 2000; Yin et al. 2009), and $\simeq 0.10$ in plants (Morrell et al. 2006; Yang et al. 2012). Thus, the data universally point to $\sim 30\%$ to 90% of recombination events being simple local gene conversions unaccompanied by crossovers. This implies that recombination rates at short distances are typically 3 to 10× greater than expected based on crossing-over alone. Average conversion-tract lengths tend to be several hundred to a few thousand bp in diverse eukaryotes (Lynch et al. 2014; Liu et al. 2018).

Evolution of the recombination rate. Considerable attention has been devoted

to understanding how selection might favor recombination-rate modifiers in various contexts (e.g., Feldman et al. 1996; Barton and Otto 2005; Keightley and Otto 2006; Barton 2010; Hartfield et al. 2010). As in the case of mutation-rate evolution, selection on modifiers of the recombination rate will generally involve second-order effects, operating via fitness-altering recombination effects elsewhere in the genome. To this end, virtually all theoretical work on the evolution of the recombination rate is motivated by the idea that natural selection inadvertently encourages the build-up of linkage disequilibrium in ways that inhibit full evolutionary potential. Two general aspects of genetic systems can encourage the buildup of hidden genetic variance.

First, synergistic epistasis (with fitness declining at an increasing rate with increasing numbers of deleterious alleles) tends to encourage the maintenance of intermediate genotypes, thereby providing a selective advantage for recombinational production of the extreme genotypes and their more efficient promotion/elimination by selection (Eshel and Feldman 1970; Kondrashov 1988; Charlesworth 1990; Barton 1995). In contrast, diminishing-returns epistasis (with fitness declining at a decreasing rate with increasing numbers of deleterious alleles) has the opposite effect, thereby encouraging reduced recombination rates. As the evidence on the relative incidences of these two forms of epistasis is mixed, the role of epistasis in the evolution of recombination rates remains unclear.

Second, as noted above, linkage reduces the efficiency of selection on multilocus systems by inducing N_e-reducing background selection and selective sweeps. These general effects are expected to be more pronounced in larger populations, which generally harbor larger numbers of cosegregating polymorphic loci. Plausible arguments have been made that the power of this second effect in selecting for modifiers for increased recombination rates may substantially outweigh that resulting from synergistic epistasis (Felsenstein and Yokoyama 1976; Otto and Barton 2001; Pálsson 2002; Barton and Otto 2005; Keightley and Otto 2006; Roze and Barton 2006).

Nonetheless, despite all of the theory, the extent to which recombination-rate modifiers ever arise with substantial enough fitness consequences to be promoted by these kinds of associative effects remains unclear. Most modeling attempts have focused on rather extreme situations in which selection coefficients and/or the magnitude of the modifier's effect on the recombination rate are quite large, and yet even under these conditions the selective advantage of the modifier can be quite small (Barton and Otto 2005), perhaps too small to overcome the likelihood of being lost by drift in most cases.

This is not to dispute the adaptive utility of sexual reproduction, which promotes both independent segregation of chromosomes and recombination within chromosomes. Laboratory experiments with yeast populations do support the idea that fitness increases more rapidly in response to a selective challenge in the presence of outcrossing (Goddard et al. 2005; McDonald et al. 2016). The latter study is most notable in that it used pooled-population genome-wide sequencing to follow the fates of newly arising mutations, showing that in asexual populations many mutations go to fixation in a hitch-hiking fashion, with mildly deleterious mutations being dragged along by linked beneficials (Figure 4.8). In contrast, in sexual populations, fewer mutations arise to high frequency, but most of those that do are beneficial. In effect, sexual reproduction reduces the effects of selective interference (simultaneous

beneficial mutations competing with each other for fixation), while also reducing the "ruby in the rubbish" effect (beneficial mutations being permanently linked to a background containing deleterious mutations in an asexual; Chapters 5 and 6).

The key issue here is whether, conditional on sexual reproduction, natural selection further fine-tunes the level of within-chromosome recombination. As noted above, the fact remains that nearly all interspecific variation in the recombination rate per physical distance can be explained by a simple, phylogenetically invariant physical model of meiosis, leaving very little residual variation to be potentially assigned to adaptive fine-tuning. Variation in the recombination rate does exist among individuals and closely related species (Dapper and Payseur 2017; Ritz et al. 2017), so evolutionary modification is certainly possible, but the level of variation is generally less than two-fold, and may simply reflect the recurrent introduction of minor recombination-rate variants by mutation.

The most widely cited evidence in favor of the idea of adaptive modification of recombination rates involves examples of moderate increases in the crossover rate in populations exposed to strong directional selection (e.g., domestication, and insecticide resistance) (Ritz et al. 2017). In principle, such results might be examples of hitch-hiking of recombination-rate enhancers with strongly favored gene combinations, much like the situation with transient increases in mutator-allele frequencies. However, enough counterexamples have been presented (Muñoz-Fuentes et al. 2015; Stapley et al. 2017) that one must be concerned with reporting bias towards positive results.

An alternative view, consistent with the pattern in Figure 4.7, is that natural selection generally operates to minimize the amount of meiotic recombination (one crossover per arm being a minimum requirement for proper chromosome segregation), with phylogenetic divergence in recombination rates being largely an indirect and passive response to changes in the population-genetic environment. With reduced N_e in organisms with increasing cell/body size (Figure 4.3), genome sizes passively expand as selection becomes less capable of resisting the accumulation of intronic and mobile-element associated DNAs (Lynch 2007). With an increase in chromosome length but not in the number of crossovers per chromosome, the recombination rate per physical distance then naturally declines. Thus, as with the mutation rate, the bulk of the variation in the recombination rate among species may be largely governed by differences in the cumulative effects of random genetic drift. Notably, Dumont and Payseur (2007) find that variation in recombination rates across mammalian species evolves in a manner that cannot be discriminated from the expectations of a neutral model.

Summary

- Evolution is a population-genetic process governed by the joint forces of mutation, recombination, and random genetic drift, all of which vary by more than four orders of magnitude across the Tree of Life. As these three features define the population-genetic playing field upon which evolution operates, such quantitative knowledge is an essential resource for understanding the limits to all adaptive and nonadaptive evolutionary pathways.

- All newly arisen mutations experience stochastic fluctuations in allele frequencies, with only a small fraction of them being harvested by natural selection. The magnitude of noise in the evolutionary process (random genetic drift) is dictated by the genetic effective population size (N_e), which is influenced by the nonindependence of simultaneously interfering mutations at linked chromosomal sites, by the absolute numbers of individuals in the population (N), and by various ecological and behavioral factors. N_e is generally orders of magnitude smaller than N, scaling negatively with organism size, with no known species having $N_e > 10^9$ individuals.

- The mutation rate is generally under persistent selection in the downward direction, with the rate per nucleotide site per generation declining from $\sim 10^{-8}$ to $\sim 10^{-11}$ with increasing N_e. This gradient can be explained by the drift-barrier hypothesis, with the efficiency of selection on replication fidelity and DNA repair becoming stalled by random genetic drift as the room for improvement declines. This hypothesis also explains why microbial eukaryotes (with more functional DNA) have lower mutation rates than prokaryotes with the same N_e and why specialized polymerases that engage in relatively small numbers of nucleotide transactions have elevated error rates.

- Although mutations arise randomly with respect the selective demands operating on recipient genes, they are nonrandom in almost all other respects, including chromosomal location and nucleotide identity. Mutations are also commonly clustered, so that the incidence of double and triple mutants can be orders of magnitude greater than expected by chance.

- Average rates of recombination per physical distance decline in larger organisms with decreased N_e. This is due to the fact that eukaryotic meiosis almost always involves just one or two crossovers per chromosome arm. As average chromosome lengths increase via the passive accumulation of noncoding DNA in lineages with decreasing N_e, the recombination rate per nucleotide site then passively declines.

- A unifying view of these observations is that ecological and behavioral factors, combined with the influence of chromosomal linkage, conspire to define the effective sizes of populations, which in turn indirectly modify mutation and recombination rates. Small organisms with high N_e tend to have low mutation rates but high recombination rates per nucleotide site. In contrast, larger organisms have lower N_e, and owing to higher levels of random genetic drift, passively evolve higher mutation rates but lower recombination rates. These covarying aspects of the population genetic environment modify the ways in which evolution by natural selection can proceed in different phylogenetic lineages.

- Because mutations with selective effects $\ll 1/N_e$ are overwhelmed by drift, small

organisms with higher N_e are capable of utilizing a wider range of mutational effects in adaptive evolution. Larger organisms, with correspondingly smaller N_e, have a reduced capacity for evolutionary fine-tuning and hence are constrained to more coarse-grained evolution.

Foundations 4.1. The amount of neutral nucleotide variation maintained at mutation-drift equilibrium. Consider a population with a base-substitution mutation rate of u per genomic site per generation, with a long-term average effective population size of N_e. Here we assume that each nucleotide base mutates to each of the three other types at rate $u/3$. Each generation, new variation (defined as heterozygosity, which is the probability that randomly paired chromosomes differ at a nucleotide site) will arise by mutation. In addition, a fraction of pre-existing variation will be lost by drift. If u and N_e are kept approximately constant, a stable equilibrium level of heterozygosity per nucleotide site will eventually be reached, at which point the rates of gain and loss of heterozygosity will be equal. Such heterozygosity can be measured as either the fraction of all neutral sites that are heterozygous within single, random diploid individuals (assuming a randomly mating population) or as the average level of heterozygosity over a large number of neutral sites at the population level.

To obtain the expected equilibrium, we start with a formulation for the dynamics of neutral-site heterozygosity, and then seek the point at which the loss and gain rates are equal. As the focus is on the neutral situation, we will assume that A, C, G, and T have equivalent fitness effects, as might be the case for a four-fold redundant site in protein-coding sequence (e,g,, in the third positions of a number of amino-acid codons).

Letting H_t denote the expected level of heterozygosity in generation t, and assuming a diploid population with k allelic types per site, we wish to determine the dynamics of change in H_t and its eventual equilibrium value. It follows from basic theory (Walsh and Lynch 2018, Chapter 2) that drift causes a fractional loss of heterozygosity equal to $1/(2N_e)$ per generation. Letting $\lambda = 1 - (1/2N_e)$, the expected frequency of heterozygotes in generation $t+1$ in the absence of mutation is then λH_t, whereas the expected frequency of homozygotes is $1 - \lambda H_t$. Following mutation, the heterozygous state will be retained if: 1) neither allele mutates, the probability of which is $(1-2u)$, ignoring the very small probability of double mutations to the same state; or 2) one of the alleles mutates to a different state than the other, the probability of which is $[2u(k-2)/(k-1)]$ assuming that all allelic types are equally mutationally exchangeable. For nucleotide sites, there are $k = 4$ alternative states, and the preceding expression reduces to $4u/3$ – each of two sites mutates to two other possible nucleotides at rate $2u/3$. On the other hand, homozygotes will be mutationally converted to heterozygotes at rate $2u$.

Summing up, the expected dynamics of neutral-site heterozygosity under random mating can be expressed as

$$H_{t+1} = \lambda H_t \left((1-2u) + \frac{4u}{3} \right) + 2u(1 - \lambda H_t). \tag{4.1.1}$$

Setting $H_{t+1} = H_t = \widetilde{H}$, the expected level of heterozygosity under drift-mutation balance is found to be

$$\widetilde{H} = \frac{\theta}{1 + (4\theta/3)}, \tag{4.1.2a}$$

where $\theta = 4N_e u$ (Malécot 1948; Kimura 1968). The same expression applies to haploidy by setting $\theta = 2N_e u$.

There are two key points to note about Equation 4.1.2a. First, the final result is a function of one composite parameter, θ, which is equivalent to the ratio of the rates of mutational production of heterozygotes from homozygotes ($2u$) and the rate of loss of heterozygosity by drift ($1/2N_e$). Second, if θ is $\ll 3/4$, as is almost always the case in natural populations,

$$\widetilde{H} \simeq \theta = 4N_e u. \tag{4.1.2b}$$

Although Equation 4.1.1 needs to be modified if the different nucleotides mutate at different rates (Kimura 1983; Cockerham 1984), provided the product of the average mutation rate and $4N_e$ is $\ll 1$, the equilibrium approximation given by Equation 4.1.2b still holds.

Foundations 4.2. Relationship of the recombination rate to physical distance between sites. Meiotic recombination events involve heteroduplex formations between paired homologous chromosomes in diploid cells. Temporary physical annealing of homologous regions occurs as a single strand from one chromosome invades the double-stranded recipient homolog (Figure 4.6). Upon separation of recombining chromosomes, the heteroduplex DNA (containing one strand from each of the contributing chromosomes) remains. If heterozygous sites are contained within such a patch, the nonmatching sites have to be resolved by the mismatch-repair pathway. This leads to a process of gene conversion, as each mismatched pair of sites is restored to a Watson-Crick state, some of which are altered to the donor-strand state. Depending on how the heteroduplex is resolved, gene conversion may be accompanied by a crossover, which leads to a complete swapping of chromosomal material to one side of the conversion event. Gene conversion involves unidirectional exchange of information, whereas a crossover generates reciprocal exchange.

It is often assumed that the recombination rate is equivalent to the crossover rate between sites, but this is generally not true. Although all recombination events involve gene conversion, only a fraction lead to crossovers. If a gene-conversion tract unaccompanied by a crossover occurs within the span between two distantly located sites, there will be no recombination between the pair of sites, as they will retain their original status. Thus, when the sites under consideration are far apart, most recombination events involve crossing over because most conversion events are irrelevant. On the other hand, when sites are close together, recombination mostly results from the conversion of single sites (Figure 4.6).

To understand this behavior in a more quantitative way, let c_0 be the total rate of initiation of recombination events per nucleotide site (with or without crossing over), L be the number of sites separating the two focal positions (with $L = 1$ for adjacent sites), and x be the fraction of recombination events accompanied by crossing over. Using Haldane's (1919) mapping function, which assumes equal probabilities of recombination at all sites, the Poisson probability of no crossover between two homologous chromosomes during a meiotic event is e^{-2c_0xL}. The crossover rate can then be represented as $0.5(1 - e^{-2c_0xL})$, which is $\simeq c_0xL$ for $c_0xL \ll 1$, and asymptotically approaches 0.5 for large c_0xL. This asymptotic value of 0.5 follows from the fact that as the number of crossovers between markers increases, the recombination rate approaches 50% because even numbers of events restore the parental state.

How does gene conversion alter the recombination rate between sites? In the following, we assume distances between sites that are small enough that the crossover rate $\simeq c_0xL$; for $c_0xL > 1$, the total recombination rate is very close to 0.5. As noted by Andolfatto and Nordborg (1998), from the perspective of two sites, a gene-conversion event causes recombination between a pair of sites if the conversion tract encompasses just one of the sites. Under the assumption of an exponential distribution of tract lengths with mean length T (in bp), the total conversion rate per site is $(1-x)c_0T(1-e^{-L/T})$ (Langley et al. 2000; Frisse et al. 2001; Lynch et al. 2014). The total recombination rate between sites separated by distance L is then

$$c_L \simeq c_0[xL + (1-x)T(1-e^{-L/T})]. \tag{4.2.1a}$$

For sites that are much more closely spaced than the average conversion-tract length,

$L \ll T$,

$$c_L \simeq c_0 L, \tag{4.2.1b}$$

whereas for $L \gg T$,

$$c_L \simeq c_0 L x. \tag{4.2.1c}$$

These results show that unless all recombination events are accompanied by crossovers ($x = 1$), the use of recombination rates between distantly related sites to extrapolate to closely spaced sites will underestimate the true rate by a factor of $1/x$.

Literature Cited

Andolfatto, P., and M. Nordborg. 1998. The effect of gene conversion on intralocus associations. Genetics 148: 1397-1399.

Arnheim, N., P. Calabrese, and I. Tiemann-Boege. 2007. Mammalian meiotic recombination hot spots. Annu. Rev. Genet. 41: 369-399.

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

Barton, N. H. 1995. A general model for the evolution of recombination. Genet. Res. 65: 123-145.

Barton, N. H. 2010. Mutation and the evolution of recombination. Philos. Trans. R. Soc. Lond. B Biol. Sci. 365: 1281-1294.

Barton, N. H., and S. P. Otto. 2005. Evolution of recombination due to random drift. Genetics 169: 2353-2370.

Baudat, F., Y. Imai, and B. de Massy. 2013. Meiotic recombination in mammals: localization and regulation. Nat. Rev. Genet. 14: 794-806.

Behjati, S., M. Huch, R. van Boxtel, W. Karthaus, D. C. Wedge, A. U. Tamuri, I. Martincorena, M. Petljak, L. B. Alexandrov, G. Gundem, et al. 2014. Genome sequencing of normal cells reveals developmental lineages and mutational processes. Nature 513: 422-425.

Bobay, L. M., and H. Ochman. 2017. Biological species are universal across Life's domains. Genome Biol. Evol. 9: 491-501.

Bobay, L. M., and H. Ochman. 2018. Factors driving effective population size and pan-genome evolution in bacteria. BMC Evol. Biol. 18: 153.

Boe, L., M. Danielsen, S. Knudsen, J. B. Petersen, J. Maymann, and P. R. Jensen. 2000. The frequency of mutators in populations of *Escherichia coli*. Mutat. Res. 448: 47-55.

Brick, K., F. Smagulova, P. Khil, R. D. Camerini-Otero, and G. V. Petukhova. 2012. Genetic recombination is directed away from functional genomic elements in mice. Nature 485: 642-645.

Capilla, L., M. Garcia Caldés, and A. Ruiz-Herrera. 2016. Mammalian meiotic recombination: a toolbox for genome evolution. Cytogenet. Genome Res. 150: 1-16.

Chan, K., and D. A. Gordenin. 2015. Clusters of multiple mutations: incidence and molecular mechanisms. Annu. Rev. Genet. 49: 243-267.

Charlesworth, B. 1990. Mutation-selection balance and the evolutionary advantage of sex and recombination. Genet. Res. 55: 199-221.

Charlesworth, B. 2012. The effects of deleterious mutations on evolution at linked sites. Genetics 190: 5-22.

Chen, J., S. Glémin, and M. Lascoux. 2017. Genetic diversity and the efficacy of purifying selection across plant and animal species. Mol. Biol. Evol. 34: 1417-1428.

Chen, X., and J. Zhang. 2013. No gene-specific optimization of mutation rate in *Escherichia coli*. Mol. Biol. Evol. 30: 1559-1562.

Chen, X., Z. Chen, H. Chen, Z. Su, J. Yang, F. Lin, S. Shi, and X. He. 2012. Nucleosomes suppress spontaneous mutations base-specifically in eukaryotes. Science 335: 1235-1238.

Cockerham, C. C. 1984. Drift and mutation with a finite number of allelic states. Proc. Natl. Acad. Sci. USA 81: 530-534.

Coop, G., X. Wen, C. Ober, J. K. Pritchard, and M. Przeworski. 2008. High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. Science 319: 1395-1398.

Corbett-Detig, R. B., D. L. Hartl, and T. B. Sackton. 2015. Natural selection constrains neutral diversity across a wide range of species. PLoS Biol. 13: e1002112.

Dapper, A. L., and B. A. Payseur. 2017. Connecting theory and data to understand recombination rate evolution. Philos. Trans. R. Soc. Lond. B Biol. Sci. 372: 20160469.

de Massy, B. 2003. Distribution of meiotic recombination sites. Trends Genet. 19: 514-522.

de Massy, B. 2013. Initiation of meiotic recombination: how and where? Conservation and specificities among eukaryotes. Annu. Rev. Genet. 47: 563-599.

Denamur, E., and I. Matic. 2006. Evolution of mutation rates in bacteria. Mol. Microbiol. 60: 820-827.

Desai, M. M., and D. S. Fisher. 2007. Beneficial mutation selection balance and the effect of linkage on positive selection. Genetics 176: 1759-1798.

Drake, J. W. 1991. A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. USA 88: 7160-7164.

Drake, J. W. 2007. Too many mutants with multiple mutations. Crit. Rev. Biochem. Mol. Biol. 42: 247-258.

Drake, J. W., B. Charlesworth, D. Charlesworth, and J. F. Crow. 1998. Rates of spontaneous mutation. Genetics 148: 1667-1686.

Dumont, B. L., and B. A. Payseur. 2008. Evolution of the genomic rate of recombination in mammals. Evolution 62: 276-294.

Earl, D. J., and M. W. Deem. 2004. Evolvability is a selectable trait. Proc. Natl. Acad. Sci. USA 101: 11531-11536.

Eshel, I., and M. W. Feldman. 1970. On the evolutionary effect of recombination. Theor. Popul. Biol. 1: 88-100.

Ewens, W. J. 2004. Mathematical Population Genetics. 2nd Edition. Springer-Verlag, Berlin, Germany.

Eyre-Walker, A., and P. D. Keightley. 2007. The distribution of fitness effects of new mutations. Nature Rev. Genet. 8: 610-618.

Feldman, M. W., S. P. Otto, and F. B. Christiansen. 1996. Population genetic perspectives on the evolution of recombination. Annu. Rev. Genet. 30: 261-295.

Felsenstein, J., and S. Yokoyama. 1976. The evolutionary advantage of recombination. II. Individual selection for recombination. Genetics 83: 845-859.

Foster, P. L. 2007. Stress-induced mutagenesis in bacteria. Crit. Rev. Biochem. Mol. Biol. 42: 373-397.

Foster, P. L., A. J. Hanson, H. Lee, E. M. Popodi, and H. Tang. 2013. On the mutational topology of the bacterial genome. G3 (Bethesda) 3: 399-407.

Frisse, L., R. R. Hudson, A. Bartoszewicz, J. D. Wall, J. Donfack, and A. Di Rienzo. 2001. Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. Amer. J. Hum. Genet. 69: 831-843.

Galhardo, R. S., P. J. Hastings, and S. M. Rosenberg. 2007. Mutation as a stress response and the regulation of evolvability. Crit. Rev. Biochem. Mol. Biol. 42: 399-435.

Garud, N. R., and K. S. Pollard. 2020. Population genetics in the human microbiome. Trends Genet. 36: 53-67.

Goddard, M. R., H. C. Godfray, and A. Burt. 2005. Sex increases the efficacy of natural selection in experimental yeast populations. Nature 434: 636-640.

Goho, S., and G. Bell. 2000. Mild environmental stress elicits mutations affecting fitness in *Chlamydomonas.* Proc. Biol. Sci. 267: 123-129.

Good, B. H., and M. M. Desai. 2016. Evolution of mutation rates in rapidly adapting asexual populations. Genetics 204: 1249-1266.

Haenel, Q., T. G. Laurentino, M. Roesti, and D. Berner. 2018. Meta-analysis of chromosome-scale crossover rate variation in eukaryotes and its significance to evolutionary genomics. Mol. Ecol. 27: 2477-2497.

Haldane, J. B. S. 1919. The combination of linkage values, and the calculation of distance between the loci of linked factors. J. Genetics 8: 299-309.

Harris, K., and R. Nielsen. 2014. Error-prone polymerase activity causes multinucleotide mutations in humans. Genome Res. 24: 1445-1454.

Hartfield, M., S. P. Otto, and P. D. Keightley. 2010. The role of advantageous mutations in enhancing the evolution of a recombination modifier. Genetics 184: 1153-1164.

Heidenreich, E. 2007. Adaptive mutation in *Saccharomyces cerevisiae.* Crit. Rev. Biochem. Mol. Biol. 42: 285-311.

Hicks, W. M., M. Kim, and J. E. Haber. 2010. Increased mutagenesis and unique mutation signature associated with mitotic gene conversion. Science 329: 82-85.

Hilliker, A. J., G. Harauz, A. G. Reaume, M. Gray, S. H. Clark, and A. Chovnick. 1994. Meiotic gene conversion tract length distribution within the *rosy* locus of *Drosophila melanogaster.* Genetics 137: 1019-1026.

James, A., and K. Jain. 2016. Fixation probability of rare nonmutator and evolution of mutation rates. Ecol. Evol. 6: 755-764.

Jeffreys, A. J., J. K. Holloway, L. Kauppi, C. A. May, R. Neumann, M. T. Slingsby, and A. J. Webb. 2004. Meiotic recombination hot spots and human DNA diversity. Phil. Trans. Roy. Soc. Lond. B Biol. Sci. 359: 141-152.

Johnson, T., and N. H. Barton. 2002. The effect of deleterious alleles on adaptation in asexual populations. Genetics 162: 395-411.

Kang, J. M., N. M. Iovine, and M. J. Blaser. 2006. A paradigm for direct stress-induced mutation in prokaryotes. FASEB J. 20: 2476-2485.

Katju, V., and U. Bergthorsson. 2019. Old trade, new tricks: insights into the spontaneous mutation process from the partnering of classical mutation accumulation experiments with high-

throughput genomic approaches. Genome Biol. Evol. 11: 136-165.

Keightley, P. D., and S. P. Otto. 2006. Interference among deleterious mutations favours sex and recombination in finite populations. Nature 443: 89-92.

Kimura, M. 1957. Some problems of stochastic processes in genetics. Ann. Math. Stat. 28: 882-901.

Kimura, M. 1967. On the evolutionary adjustment of spontaneous mutation rates. Genet. Res. 9: 23-34.

Kimura, M. 1968. Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. Genet. Res. 11: 247-269.

Kimura, M. 1983. The Neutral Theory of Molecular Evolution. Cambridge Univ. Press, Cambridge, UK.

Kimura, M., and T. Ohta. 1969. The average number of generations until fixation of a mutant gene in a finite population. Genetics 61: 763-771.

Kivisaar, M. 2010. Mechanisms of stationary-phase mutagenesis in bacteria: mutational processes in pseudomonads. FEMS Microbiol. Lett. 312: 1-14.

Kohl, K. P., and J. Sekelsky. 2013. Meiotic and mitotic recombination in meiosis. Genetics 194: 327-334.

Kondrashov, A. S. 1988. Deleterious mutations and the evolution of sexual reproduction. Nature 336: 435-440.

Lam, I., and S. Keeney. 2015. Nonparadoxical evolutionary stability of the recombination initiation landscape in yeast. Science 350: 932-937.

Lang, G. I., and A. W. Murray. 2011. Mutation rates across budding yeast chromosome VI are correlated with replication timing. Genome Biol. Evol. 3: 799-811.

Langley, C. H., B. P. Lazzaro, W. Phillips, E. Heikkinen, and J. M. Braverman. 2000. Linkage disequilibria and the site frequency spectra in the *su(s)* and *su(w^a)* regions of the *Drosophila melanogaster* X chromosome. Genetics 156: 1837-1852.

Leffler, E. M., K. Bullaughey, D. R. Matute, W. K. Meyer, L. Ségurel, A. Venkat, P. Andolfatto, and M. Przeworski. 2012. Revisiting an old riddle: what determines genetic diversity levels within species? PLoS Biol. 10: e1001388.

Leigh, E. G., Jr. 1970. Natural selection and mutability. Amer. Natur. 104: 301-305.

Liu, H., J. Huang, X. Sun, J. Li, Y. Hu, L. Yu, G. Liti, D. Tian, L. D. Hurst, and S. Yang. 2018. Tetrad analysis in plants and fungi finds large differences in gene conversion rates but no GC bias. Nat. Ecol. Evol. 2: 164-173.

Long, H., W. Sung, S. F. Miller, M. S. Ackerman, T. G. Doak, and M. Lynch. 2015. Mutation rate, spectrum, topology, and context-dependency in the DNA mismatch-repair deficient *Pseudomonas fluorescens* ATCC948. Genome Biol. Evol. 7: 262-271.

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

Lukačišinová, M., S. Novak, and T. Paixão. 2017. Stress-induced mutagenesis: stress diversity facilitates the persistence of mutator genes. PLoS Comput. Biol. 13: e1005609.

Lynch, M. 2006. The origins of eukaryotic gene structure. Mol. Biol. Evol. 23: 450-468.

Lynch, M. 2007. The Origins of Genome Complexity. Sinauer Assocs., Inc. Sunderland, MA.

Lynch, M. 2008. The cellular, developmental, and population-genetic determinants of mutation-rate evolution. Genetics 180: 933-943.

Lynch, M. 2010. Evolution of the mutation rate. Trends Genetics 26: 345-352.

Lynch, M. 2011. The lower bound to the evolution of mutation rates. Genome Biol. Evol. 3: 1107-1118.

Lynch, M. 2020. The evolutionary scaling of cellular traits imposed by the drift barrier. Proc. Natl. Acad. Sci. USA 117: 10435-10444.

Lynch, M., M. Ackerman, J.-F. Gout, H. Long, W. Sung, W. K. Thomas, and P. L. Foster. 2016. Genetic drift, selection, and evolution of the mutation rate. Nature Rev. Genetics 17: 704-714.

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., L. M. Bobay, F. Catania, J.-F. Gout, and M. Rho. 2011. The repatterning of eukaryotic genomes by random genetic drift. Annu. Rev. Genomics Hum. Genet. 12: 347-366.

Lynch, M., and B. Trickovic. 2020. A theoretical framework for evolutionary cell biology. J. Mol. Biol. 432: 1861-1879.

Lynch, M., and J. B. Walsh. 1998. Genetics and Analysis of Quantitative Traits. Sinauer Assocs., Inc., Sunderland, MA.

Lynch, M., S. Xu, T. Maruki, P. Pfaffelhuber, and B. Haubold. 2014. Genome-wide linkage-disequilibrium profiles from single individuals. Genetics 198: 269-281.

MacLean, R. C., C. Torres-Barceló, and R. Moxon. 2013. Evaluating evolutionary models of stress-induced mutagenesis in bacteria. Nat. Rev. Genet. 14: 221-227.

Maharjan, R. P., and T. Ferenci. 2017. A shifting mutational landscape in 6 nutritional states: stress-induced mutagenesis as a series of distinct stress input-mutation output relationships. PLoS Biol. 15: e2001477.

Malécot, G. 1948. Les Mathmatiques de lHérédité. Masson, Paris.

Malécot, G. 1952. Les processes stochastiques et la méthode des fonctions génératrices ou caracteréstiques. Publ. Inst. Stat. Univ. Paris 1: Fasc. 3: 1-16.

Malkova, A., and J. E. Haber. 2012. Mutations arising during repair of chromosome breaks. Annu. Rev. Genet. 46: 455-473.

Malkova, A., J. Swanson, M. German, J. H. McCusker, E. A. Housworth, F. W. Stahl, abnd J. E. Haber. 2004. Gene conversion and crossing over along the 405-kb left arm of *Saccharomyces cerevisiae* chromosome VII. Genetics 168: 49-63.

Mancera, E., R. Bourgon, A. Brozzi, W. Huber, and L. M. Steinmetz. 2008. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. Nature 454: 479-485.

Martincorena, I., A. S. Seshasayee, and N. M. Luscombe. 2012. Evidence of non-random mutation rates suggests an evolutionary risk management strategy. Nature 485: 95-98.

Matsuba, C., D. G. Ostrow, M. P. Salomon, A. Tolani, and C. F. Baer. 2013. Temperature, stress

and spontaneous mutation in *Caenorhabditis briggsae* and *Caenorhabditis elegans*. Biol. Lett. 9: 20120334.

McDonald, M. J., Y. Y. Hsieh, Y. H. Yu, S. L. Chang, and J. Y. Leu. 2012. The evolution of low mutation rates in experimental mutator populations of *Saccharomyces cerevisiae*. Curr. Biol. 22: 1235-1240.

McDonald, M. J., D. P. Rice, and M. M. Desai. 2016. Sex speeds adaptation by altering the dynamics of molecular evolution. Nature 531: 233-236.

Milholland, B., X. Dong, L. Zhang, X. Hao, Y. Suh, and J. Vijg. 2017. Differences between germline and somatic mutation rates in humans and mice. Nat. Commun. 8: 15183.

Morrell, P. L., D. M. Toleno, K. E. Lundy, and M. T. Clegg. 2006. Estimating the contribution of mutation, recombination and gene conversion in the generation of haplotypic diversity. Genetics 173: 1705-1723.

Muñoz-Fuentes, V., M. Marcet-Ortega, G. Alkorta-Aranburu, C. Linde Forsberg, J. M. Morrell, E. Manzano-Piedras, A. Söderberg, K. Daniel, A. Villalba, A. Toth, A. Di Rienzo, et al. 2015. Strong artificial selection in domestic mammals did not result in an increased recombination rate. Mol. Biol. Evol. 32: 510-523.

Myers, S., L. Bottolo, C. Freeman, G. McVean, and P. Donnelly. 2005. A fine-scale map of recombination rates and hotspots across the human genome. Science 310: 321-324.

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

Orr, H. A. 2000. The rate of adaptation in asexuals. Genetics 155: 961-968.

Otto, S. P., and N. H. Barton. 2001. Selection for recombination in small populations. Evolution 55: 1921-1931.

Padhukasahasram, B., and B. Rannala. 2013. Meiotic gene-conversion rate and tract length variation in the human genome. Eur. J. Hum. Genet. 2013: 1-8.

Pálsson, S. 2002. Selection on a modifier of recombination rate due to linked deleterious mutations. J. Hered. 93: 22-26.

Paul, S., S. Million-Weaver, S. Chattopadhyay, E. Sokurenko, and H. Merrikh. 2013. Accelerated gene evolution through replication-transcription conflicts. Nature 495: 512-515.

Petes, T. D. 2001. Meiotic recombination hot spots and cold spots. Nat. Rev. Genetics 2: 360-369.

Radman, M., F. Taddei, and I. Matic. 2000. Evolution-driving genes. Res. Microbiol. 151: 91-95.

Ram, Y., and L. Hadany. 2014. Stress-induced mutagenesis and complex adaptation. Proc. Biol. Sci. 281: 20141025.

Raynes, Y., and P. D. Sniegowski. 2014. Experimental evolution and the dynamics of genomic mutation rate modifiers. Heredity 113: 375-380.

Ritz, K. R., M. A. F. Noor, and N. D. Singh. 2017. Variation in recombination rate: adaptive or not? Trends Genet. 33: 364-374.

Romiguier, J., P. Gayral, M. Ballenghien, A. Bernard, V. Cahais, A. Chenuil, Y. Chiari, R. Dernat, L. Duret, N. Faivre, et al. 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature 515: 261-263.

Rosen, M. J., M. Davison, D. Bhaya, and D. S. Fisher. 2015. Fine-scale diversity and extensive recombination in a quasisexual bacterial population occupying a broad niche. Science 348: 1019-1023.

Rosenberg, S. M. 2001. Evolving responsively: adaptive mutation. Nat. Rev. Genet. 2: 504-515.

Rosenberg, S. M., C. Shee, R. L. Frisch, and P. J. Hastings. 2012. Stress-induced mutation via DNA breaks in *Escherichia coli:* a molecular mechanism with implications for evolution and medicine. Bioessays 34: 885-892.

Roze, D., and N. H. Barton. 2006. The Hill-Robertson effect and the evolution of recombination. Genetics 173: 1793-1811.

Sakoparnig, T., C. Field, and E. van Nimwegen. 2021. Whole genome phylogenies reflect the distributions of recombination rates for many bacterial species. eLife 10: e65366.

Schrider, D. R., J. N. Hourmozdi, and M. W. Hahn. 2011. Pervasive multinucleotide mutational events in eukaryotes. Curr. Biol. 21: 1051-1054.

Shapiro, B. J. 2016. How clonal are bacteria over time? Curr. Opin. Microbiol. 31: 116-123.

Sharp, N. P., and A. F. Agrawal. 2012. Evidence for elevated mutation rates in low-quality genotypes. Proc. Natl. Acad. Sci. USA 109: 6142-6146.

Shewaramani, S., T. J. Finn, S. C. Leahy, R. Kassen, P. B. Rainey, and C. D. Moon. 2017. Anaerobically grown *Escherichia coli* has an enhanced mutation rate and distinct mutational spectra. PLoS Genet. 13: e1006570.

Singh, T., M. Hyun, and P. Sniegowski. 2017. Evolution of mutation rates in hypermutable populations of *Escherichia coli* propagated at very small effective population size. Biol. Lett. 13: 20160849.

Stamatoyannopoulos, J. A., I. Adzhubei, R. E. Thurman, G. V. Kryukov, S. M. Mirkin, and S. R. Sunyaev. 2009. Human mutation rate associated with DNA replication timing. Nat. Genet. 41: 393-395.

Stapley, J., P. G. D. Feulner, S. E. Johnston, A. W. Santure, and C. M. Smadja. 2017. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. Philos. Trans. R. Soc. Lond. B Biol. Sci. 372: 20160455.

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

Sung, W., M. S. Ackerman, J. F. Gout, S. F. Miller, P. Foster, and M. Lynch. 2015. Asymmetric context-dependent mutation patterns revealed through mutation-accumulation experiments. Mol. Biol. Evol. 32: 1672-1683.

Swings, T., B. Weytjens, T. Schalck, C. Bonte, N. Verstraeten, J. Michiels, and K. Marchal. 2017. Network-based identification of adaptive pathways in evolved ethanol-tolerant bacterial populations. Mol. Biol. Evol. 34: 2927-2943.

Tenaillon, O., F. Taddei, M. Radman, and I. Matic. 2001. Second-order selection in bacterial evolution: selection acting on mutation and recombination rates in the course of adaptation. Res. Microbiol. 152: 11-16.

Terekhanova, N. V., G. A. Bazykin, A. Neverov, A. S. Kondrashov, and V. B. Seplyarskiy. 2013.

Prevalence of multinucleotide replacements in evolution of primates and *Drosophila.* Mol. Biol. Evol. 30: 1315-1325.

Turrientes, M. C., F. Baquero, B. R Levin, J.-L. Martínez, A. Ripoll, J.-M. González-Alba, R. Tobes, M. Manrique, M.-R. Baquero, M.-J. Rodríguez-Domínguez, et al. 2013. Normal mutation rate variants arise in a mutator (Mut S) *Escherichia coli* population. PLoS One 8: e72963.

Uphoff, S., N. D. Lord, B. Okumus, L. Potvin-Trottier, D. J. Sherratt, and J. Paulsson. 2016. Stochastic activation of a DNA damage response causes cell-to-cell mutation rate variation. Science 351: 1094-1097.

Vos, M., and X. Didelot. 2009. A comparison of homologous recombination rates in bacteria and archaea. ISME J. 3: 199-208.

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

Wells, D., E. Bitoun, D. Moralli, G. Zhang, A. Hinch, J. Jankowska, P. Donnelly, C. Green, and S. R. Myers. 2020. ZCWPW1 is recruited to recombination hotspots by PRDM9 and is essential for meiotic double strand break repair. eLife 9: e53392.

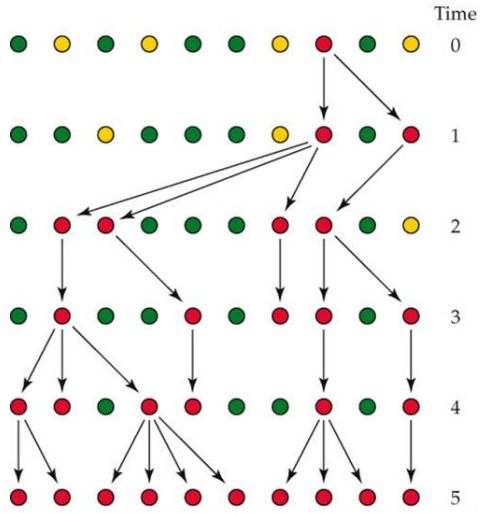
Wielgoss, S., J. E. Barrick, O. Tenaillon, M. J. Wiser, W. J. Dittmar, S. Cruveiller, B. Chane-Woon-Ming, C. Médigue, R. E. Lenski, and D. Schneider. 2013. Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. Proc. Natl. Acad. Sci. USA 110: 222-227.

Williams, L. N., A. J. Herr, and B. D. Preston. 2013. Emergence of DNA polymerase ϵ antimutators that escape error-induced extinction in yeast. Genetics 193: 751-770.

Yang, S., Y. Yuan, L. Wang, J. Li, W. Wang, H. Liu, J. Q. Chen, L. D. Hurst, and D. Tian. 2012. Great majority of recombination events in *Arabidopsis* are gene conversion events. Proc. Natl. Acad. Sci. USA 109: 20992-20997.

Yin, J., M. I. Jordan, and Y. S. Song. 2009. Joint estimation of gene conversion rates and mean conversion tract lengths from population SNP data. Bioinformatics 25: i231-i239.

Figure 4.1. Left) Random genetic drift of a single mutation to fixation (denoted in red), with the population initially containing two additional alleles (denoted in yellow and green). Each ball can be viewed as a single haploid individual. Each generation, population replacement is achieved by randomly drawing offspring from the members of the parental population. After three generations, the yellow allele has been lost by chance, and the green allele is lost soon thereafter. The arrows trace the full gene genealogy of the population at generation 5 back to the single mutation at time 0. Different numbers of descendants of each gene, in this case between 0 and 4 per generation, are purely a matter of chance. Ultimately, the red allele would be replaced by a novel mutation. **Right)** Long-term sample trajectories for three populations with different sizes (N), all starting at allele frequency 0.5. Each trajectory is unique, and the magnitude of fluctuations increases with decreasing population size.



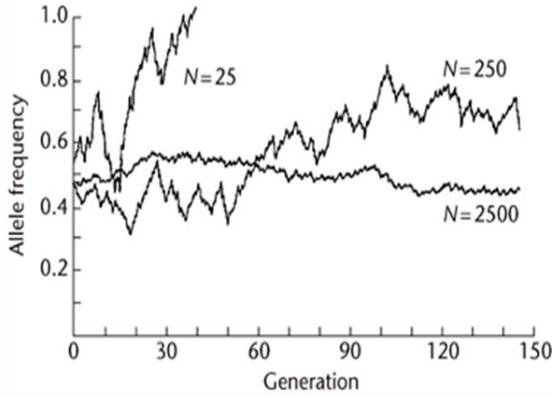


Figure 4.2. Effects of chromosomal linkage on the efficiency of natural selection. **Left**) A selective sweep involving a strongly beneficial mutation (green circle) drags the chromosomal segment upon which this mutation arose to fixation, and in doing so jointly fixes a weakly deleterious mutation (blue circle) to which it is linked, while also causing the loss of another more weakly beneficial mutation (red circles). The tick marks denote additional nucleotide polymorphisms at other sites in the ancestral population, which are also lost in the region of the sweep. With free recombination, both the green and red mutations would eventually become fixed with high probability, and the deleterious mutations would be purged. **Right**) Background selection against recurrent deleterious mutations (blue) causes their associated chromosomal segments to be purged from the population by negative selection, removing variation.

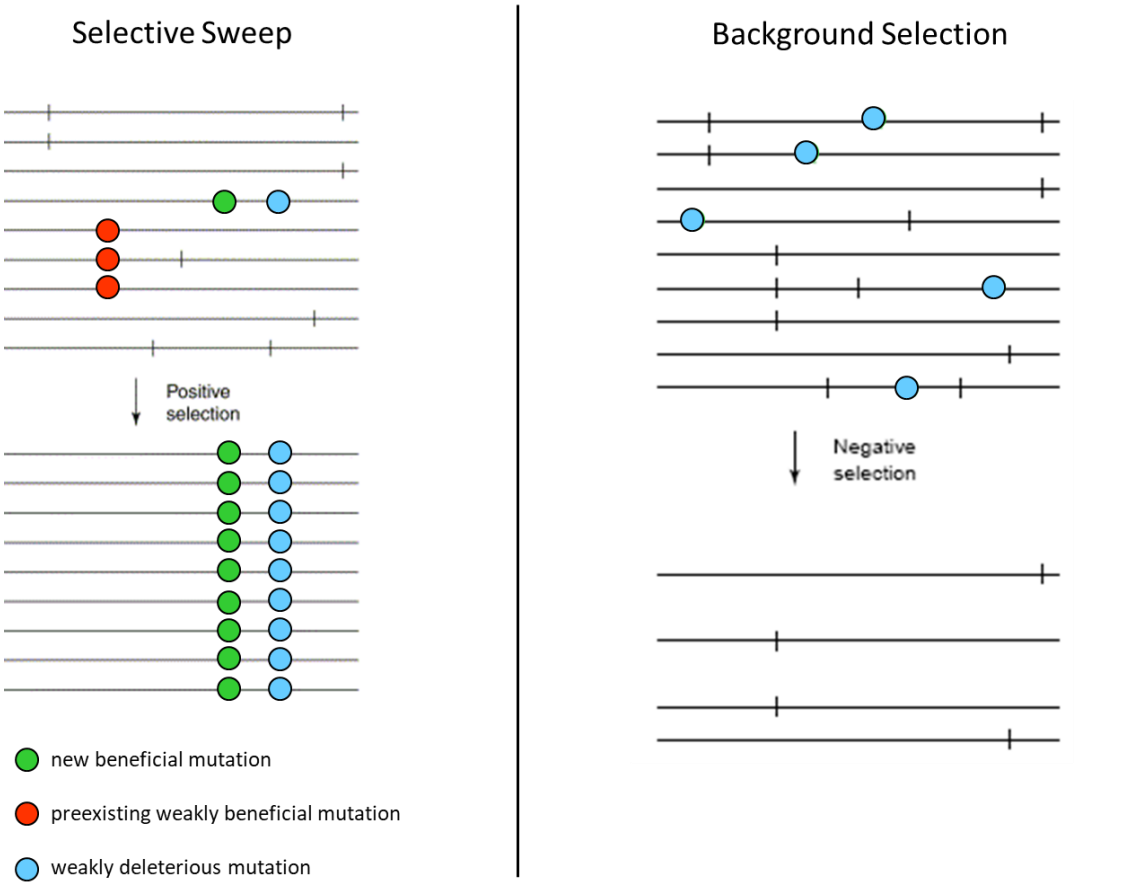


Figure 4.3. The negative scaling of effective population size (N_e) with organism size across the Tree of Life. The shaded squares envelope the full range of estimated variation for each major group. The line is the least-squares fit to the overall data $N_e = (4.5 \times 10^6)B^{-0.20}$, where B is adult dry weight (in μg). As indicated on the right, mutations with selective effects much smaller than $1/N_e$ fall in the range of effective neutrality and are impervious to natural selection. From Lynch and Trickovic (2020).

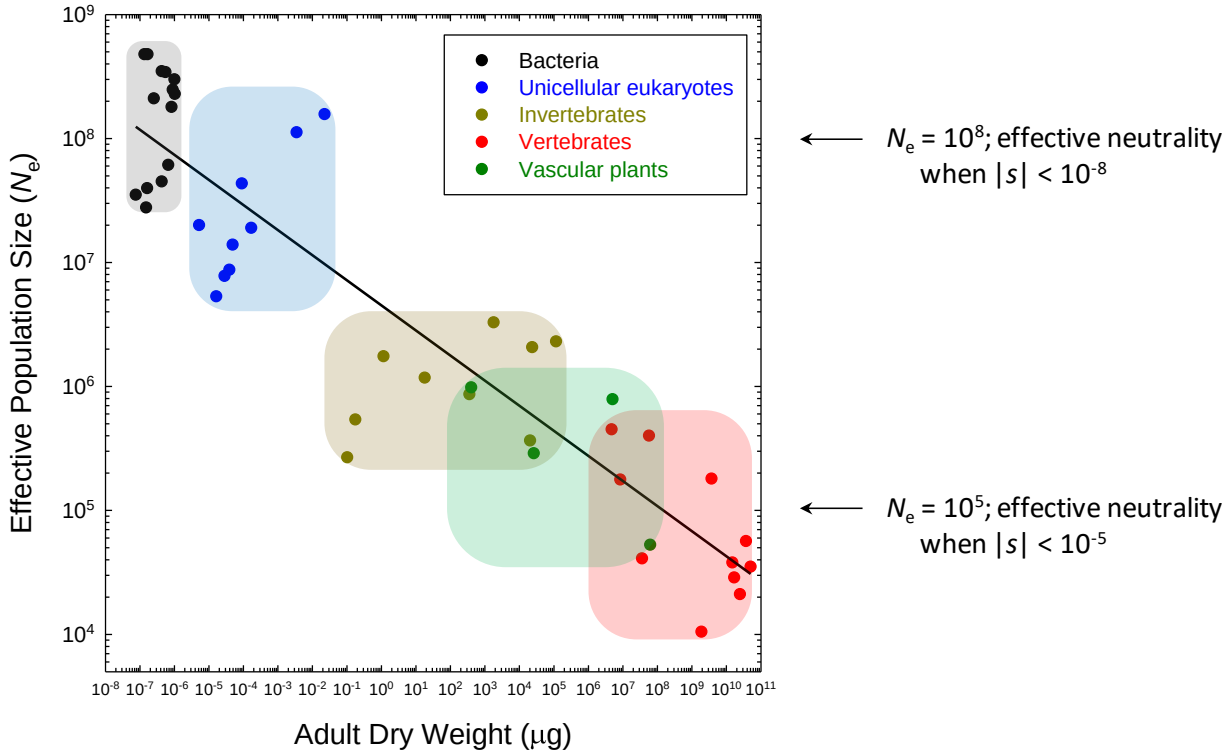


Figure 4.4. The evolution of quasi-equilibrium distributions of mutation rates under the drift-barrier hypothesis. Mutator and antimutator alleles are recurrently introduced, and these acquire associated loads of linked deleterious mutations, imposing indirect selection on the mutation rate. Regardless of the starting point, the mutation rate eventually evolves to reside within the confines of the gray area (the upper and lower bounds to the drift barrier). Above the drift barrier, the mutation rate is sufficiently high that antimutators have a large enough fitness advantage to be advanced by selection; below the drift barrier, the selective disadvantage of mutators is sufficiently small that they are able to drift to fixation. Within the grey domain, the mutation rate fluctuates stochastically through time owing to the sporadic introduction of mutator and antimutator alleles. From Lynch (2011).

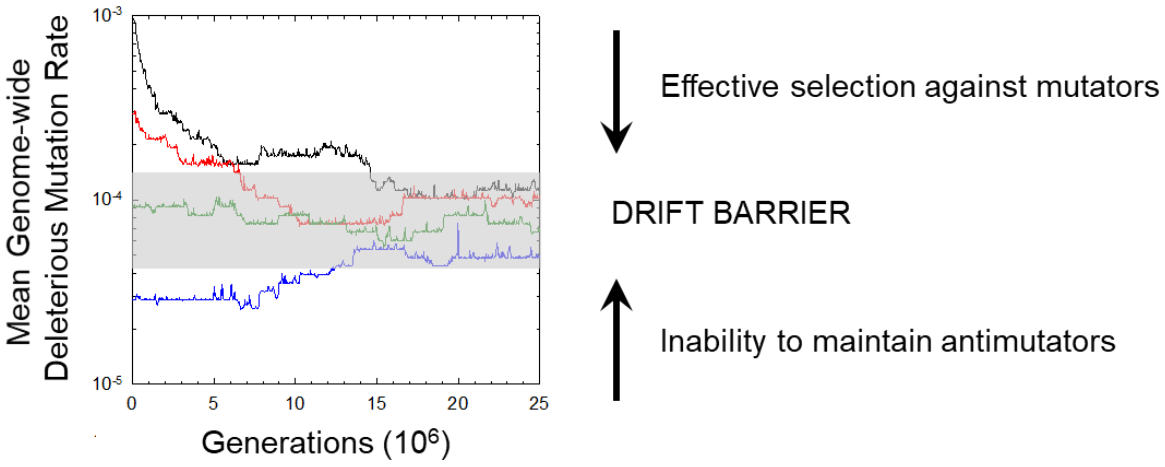


Figure 4.5. Negative scaling of per-generation mutation rates (per nucleotide site) with effective population sizes across the Tree of Life. **Left**) Regressions are given for three major phylogenetic groups: bacteria, $0.00021N_e^{-0.72}$; unicellular eukaryotes, $0.00039N_e^{-0.90}$; and multicellular eukaryotes, $0.00000078N_e^{-0.41}$. **Right**) Rates are given as the total haploid genomic rate for sites likely to be under selection (the sum over all nucleotide sites in protein-coding genes), with the overall regression being $1410N_e^{-0.76}$. Coding nucleotides comprise almost the entire genomes of bacteria, so there is little room for underestimation; the upper left arrow denotes the likely magnitude of the elevation of the slope due to additional regulatory sites in eukaryotes (see Lynch et al. (2016) for further details). Dashed lines denote slopes of -1. The plotted data are from Lynch et al. (2016) and Long et al. (2018), with updates (Supplemental Table 4.1).

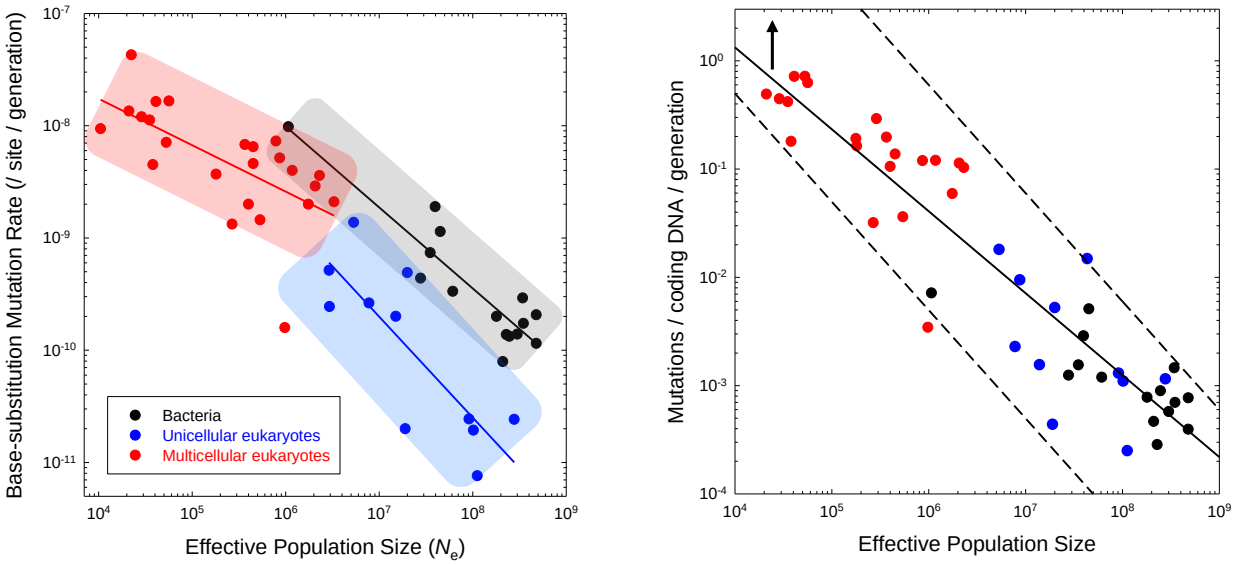


Figure 4.6. **Left)** The physical mechanics of meiotic recombination. After the first stage of meiosis in a diploid organism, chromosomes exist as sister chromatids, fused at the centromere, and these undergo pairing with their homologous partners (one from the paternal and the other from the maternal chromosome that formed the individual). Here, double-stranded sisters (each a DNA molecule) are shown for each of the homologous parental chromosomes (one yellow, and one blue). Single-strand breaks are created enzymatically during meiosis, and then resected, with strand invasion resulting in the two homologous chromosomes becoming intercalated at the breakpoint. This creates patches of heteroduplex DNA consisting of complementary molecules derived from each parental chromosome. To complete meiosis, this complex must be restored to the two-chromosome state, and depending on how the intermediates are separated, this can lead to a complete exchange of material distal to the breakpoint, a process known as crossing-over. However, regardless of whether crossing-over occurs, there is potential for a gene conversion in the heteroduplex (yellow-blue) patches – if such patches contain base mismatches (which will be the case if the individual is heterozygous in the region), these must be restored to Watson:Crick (A:T or G:C) states by the mismatch-repair pathway. If, for example, one parental chromosome had a A:T double-strand state, and the other C:G, invasion would produce an A:G or a C:T mismatch, which would then be restored to one of the parental states. Depending on the directionality of change, this can lead to a short patch of exchange between chromosomes, independent of whether crossing-over occurred. **Right)** The relative importance of gene conversion and crossing over in recombination between sites separated by different distances. The black diamonds represent nucleotide sites at which a crossover can occur, and horizontal black lines denote conversion tracts. Recombination between sites represented by the black balls occurs if a conversion tract covers a single site or a crossover event falls between sites. At the top, the sites are close together, and relative to the number of relevant gene-conversion patches, only a few recombination events can lead to potential crossovers. At the bottom, sites are farther apart, and whereas the potential for crossing over increases linearly with distance, the potential for single-site gene conversion does not, as many conversion tracts reside between the markers of interest.

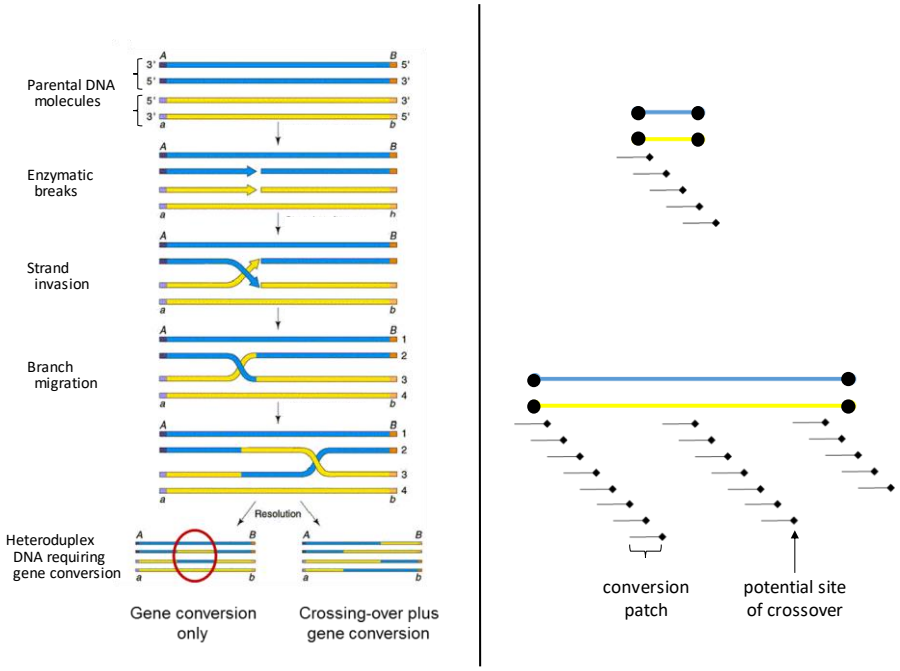


Figure 4.7. Average rates of recombination per physical distance for four major groupings of eukaryotes, determined from information on total physical and genetic map sizes. Each data point is based on the genetic map of a different species. The two dashed lines have slopes of -1.0 in accordance with the theory discussed in the text. Letting κ be the average number of crossovers per chromosome, and M be the number of haploid chromosomes, the upper line assumes $\kappa M = 50$, i.e., 50 chromosomes with an average of one crossover, 25 with averages of two crossovers, etc. The lower line assumes $\kappa M = 3$. For the plotted species, κ is in the range of 0.3 to 3.1 (with one exception) and M is in the range of 3 to 44. From Lynch et al. (2011).

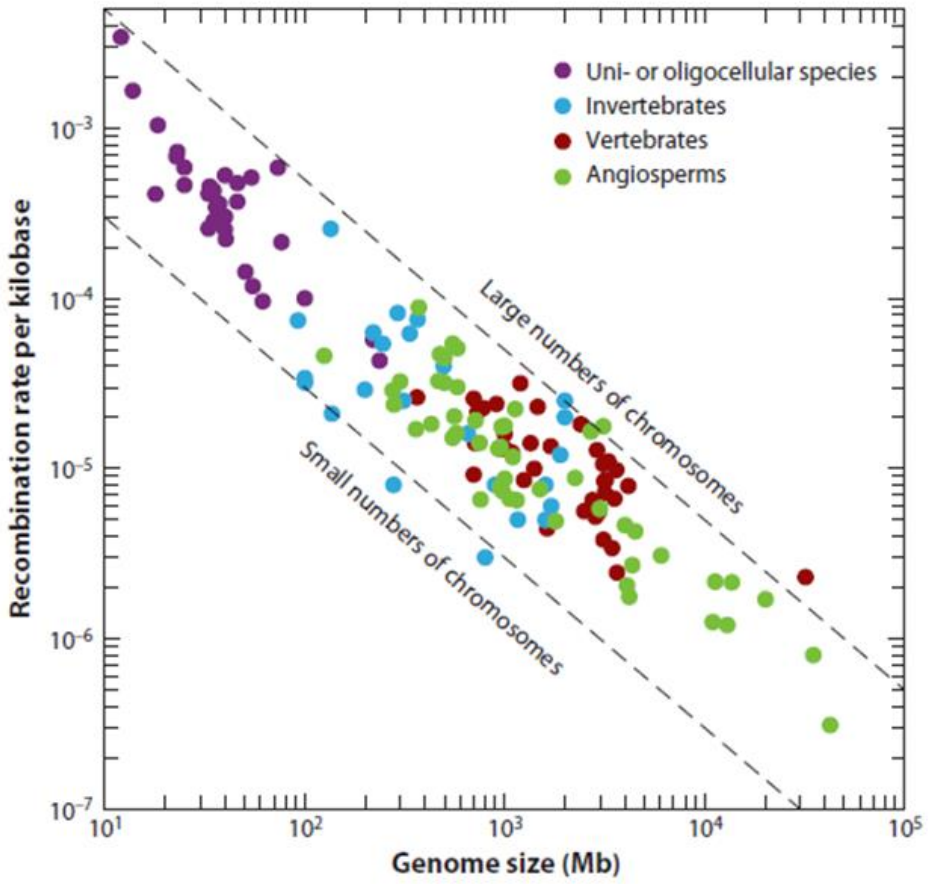


Figure 4.8. Temporal changes in frequencies of single-nucelotide variants in eight experimental populations (four asexual and four sexual) of yeast (*S. cerevisiae*) over a period of 1000 generations. All populations were initiated with the same single clone, and the frequencies of spontaneously arising mutations were estimated by pooling entire populations and sequencing the entire genome at high coverage. Solid lines denote amino-acid altering mutations; dashed lines denote silent-changes in coding regions that leave the encoded amino acid unaltered; and dotted lines denote mutations in intergenic regions. Note that in the asexual populations, bundles of trajectories of individual mutations follow the same pattern of frequency change; these represent the history of individual clones within which all relevant variants are permanently linked and therefore dragged together to fixation or loss). In sexual populations, segregation and/or recombination places individual mutations on multiple genetic backgrounds, rendering the trajectories of individual variants more independent. From McDonald et al. (2016).