An integrated model of the effects on fitness of beneficial, deleterious, and neutral mutations in sexual populations

Gordon Irlam
gordoni@gordoni.com

Los Altos, California, United States

October 21, 2024

Abstract

The fitness effects of beneficial, deleterious, and neutral mutations have historically been largely analyzed in isolation. Here, an integrated model of the fitness effects of mutations in sexual populations is assembled and analyzed. The model suggests the fitness effects associated with beneficial mutations can not be ignored simply because they are rare. The prospective load is defined as the genetic load associated with beneficial mutations that are feasible, but have yet to occur and begun the process of fixing. The optimal spontaneous mutation rate for a population is shown to occur when the prospective load is approximately equal to the mutational load. This population optimal mutation rate could be brought about through macroevolution. It differs from the as-small-as-possible rate that might be expected from microevolutionary considerations.

Keywords: beneficial mutation, adaptive mutation, positive selection, deleterious mutation, purifying selection, negative selection, neutral mutation, genetic load, prospective load, substitutional load, lag load, mutational load, neutral load, optimal mutation rate.

Introduction

The goals of this paper are twofold: to develop an equation that approximately describes the fitness of sexual populations incorporating the effects of beneficial, deleterious, and neutral mutation; and then to understand how the population fitness changes with changes in the spontaneous mutation rate. Much of the hard work has already been done. This manuscript is in part an integration of the works of Smith on lag load [1], Barton on the probability of fixation of a beneficial mutation [2], and Kimura on neutral mutations and the fixation of a beneficial mutation [3, 4, 5].

Before commencing it is worth having a clear understanding of what is meant here by "fitness". Fitness is traditionally defined as the growth rate of a genotype in a population. When organisms are haploid, and the genotype in question is that of the haploid organism, fitness takes on a particularly intuitive form. Fitness is simply the expected relative number of immediate offspring of a haploid organism in some environment. It is an expected value because random stochastic sampling effects may cause the actual observed number to vary from the expected value, but were it possible to repeat the same life cycle in the same environment fitness would equal the mean number of offspring observed. It is relative. An organism with twice the fitness value as some other organism would be expected to have twice as many offspring. In this manuscript the fitness of diploid organisms is defined in the same manner; as the expected relative number of immediate offspring. Most of the time it is this relative nature of fitness that concerns us, but so we can speak meaningfully of fitness values alone we need to anchor the fitness scale. We choose to discuss fitness here relative to a hypothetical organism that is perfectly adapted to its environment to which we assign a fitness value of 1. This means that log fitness, x, is negative, or zero in the one special case.

As used here, genetic load describes the difference between the fitness of an ideal organism at the current point in time and the average fitness of the current population. Four basic contributors to genetic load will be considered here:

- what we will term the "prospective load" which is the result of the population having the potential for beneficial mutations, but that those mutations have not yet occurred and begun the process of fixing;
- the substitutional load which results from beneficial mutations that are in the process of fixing;
- the mutational load which is the result of deleterious mutations;
- and what we will term the "neutral load" resulting from neutral and nearly neutral sites in the genome.

Beneficial mutations that die out rather than fix reduce the prospective load, but because such mutations fail to establish, they will be rare, and so only have a very small effect on the prospective load. The prospective and substitutional loads together make up what Smith calls the lag load [1]. They exist because the makeup of the population lags behind the ideal organism as dictated by the current environment, which is changing.

$\overline{x}$	Mean log fitness of the population.
N	Number of organisms in the population.
P	Ploidy: 1 haploid, or 2 diploid.
Γ_p	Population wide rate of occurrence of prospective beneficial sites.
μ_s	Germline mutation rate per base pair per haploid genome per sexual generation.
Π_0	Probability of fixation of a beneficial mutation, as determined by Kimura.
Π	Multiplicative adjustment to Π_0 to account for mutation and Γ_p.
s_p	Heterozygous selection coefficient associated with a beneficial (aka positive) mutation.
h	Dominance coefficient. Beneficial homozygotes have selection coefficient $\frac{s_p}{h}$.
γ	Euler's constant, $0.5772\ldots$.
Γ_n	Germline deleterious mutation rate per haploid genome per sexual generation.
s_n	Heterozygous selection coefficient associated with a deleterious (aka negative) mutation.
s_0	Mean selection coefficient of neutral sites.
L_0	Number of neutral or nearly neutral sites per haploid genome.
R	Genome length in Morgans.
L_n	Number of sites that are maintained by negative selection per haploid genome.

Table 1: Summary of key notation.

An important result to be derived here is that the spontaneous mutation rate per base pair that maximizes the population mean fitness is the one for which fitness losses associated with the prospective load and the mutational load are approximately equal. This mutation rate will be termed the population optimal mutation rate. This finding is significant because it is a mutation rate that is quite different from the as-small-as-possible rate that might be expected from microevolutionary considerations. The finding is also significant because, while mutational load has been considered to be a primary cause of genetic load [6], the possibility of a similarly sized impact for the prospective load appears to have rarely been considered.

Kimura makes related arguments to the arguments made here, but the prospective load is not considered [7]. It is argued in the present paper that the prospective load is a key component in determining the population optimal mutation rate.

Analysis

The population is modeled using a Wright-Fisher-like discrete time model consisting of repeating non-overlapping generations with random mating of the population in accordance with individual fitness.

A summary of the key notation used is provided by Table 1. Assume a fixed sexual population of organisms of size N. Let the ploidy, P, be 1 for a species that are counted and experience the force of selection during the haploid state, and 2 for organisms that are counted and experience selective forces while in the diploid state.

For simplicity, only single nucleotide point mutations are considered. Insertion and deletion (indel) mutations are not included in the analysis, although it would certainly be possible to extend the

analysis to incorporate them. A simple mutation fitness landscape is assumed. Double mutations in which either mutation alone is disadvantageous, but their combination is advantageous, are not considered. The analysis could be extended to incorporate such mutations occurring at a suitably low rate.

Epistasis is possible, but concurrent epistasis is assumed to be minimal. That is, a site that is either prospective or fixed may affect the fitness of another site, but there should be few interactions between the fitness effects of sites that are concurrently in the process of fixing.

Prospective load

The external environment may frequently or infrequently undergo change. This change reduces the fitness of an organism relative to the ideal organism, but it also creates opportunities for beneficial mutations to sites in the genome whose mutation would previously have been neutral or deleterious. In addition to beneficial mutational prospects arising from the external environment, the fixation of one beneficial mutation might lead to the creation of other mutational prospects on nearby or distant loci.

Let Γ_p be the long term rate of occurrence of prospective beneficial sites per generation. By notational convention, changes that render a previously fixed neutral site as beneficial without requiring any mutational change to the site will be excluded from Γ_p. This reduces the magnitude of Γ_p by approximately $\frac{1}{4}$.

Let the germline spontaneous mutation rate per base pair per haploid genome per sexual generation be μ_s. A base pair can take on 4 possible values. Typically only one of these will satisfy a particular prospective beneficial site. Consequently the rate of beneficial mutation per base pair per haploid genome is approximately $\frac{\mu_s}{3}$.

Not all beneficial mutations result in fixation. Most are lost. Let Π_0 be the probability of fixation of an individual mutant in the absence of other mutations and the emergence of beneficial sites, and Π be a multiplicative adjustment to Π_0 to take into account these two factors. Let the heterozygous selection coefficient associated with a beneficial mutation be s_p. Based on the work of Kimura [4],

$$\Pi_0 = \frac{1 - e^{-2s_p}}{1 - e^{-2PNs_p}}$$

For N large and s_p small, $\Pi_0 \approx 2s_p$. Let the dominance coefficient be h. By convention, let $h = 1$ for haploid eukaryotic species. Assuming s_p is small, the log fitness loss associated with a null homozygous site is approximately $\frac{s_p}{h}$. This is the log fitness to be gained should the site fix, and it is a contribution to the prospective genetic load.

The long term rate of occurrence of prospective beneficial sites must balance the long term rate at which such prospects are fixed. Otherwise the population would become less and less fit until it went extinct. Suppose there are a total of k prospective sites. Each site exists as a single copy in every haploid genome. The aggregate rate of beneficial mutation at these sites will be $kPN\frac{\mu_s}{3}$, and the probability of fixation is approximately $2s_p\Pi$. The product of these two terms must equal the rate at which beneficial mutational prospects occur,

$$\Gamma_p \approx \frac{2}{3} kPN\mu_s s_p \Pi$$

Let the mean log fitness associated with prospective sites be $\overline{x_p}$,

$$\begin{aligned} \overline{x_p} &\approx -\frac{ks_p}{h} \\ &\approx -\frac{3\Gamma_p}{2PN\mu_s h\Pi} \end{aligned} \tag{1}$$

In the event that h or Π is not constant, but a function of the prospective sites, $h\Pi$ in equation 1 would represent its harmonic mean value.

Substitutional load

It takes some time for a new beneficial mutation that is destined to fix, to actually fix. While a site is fixing it exerts a substitutional load.

Kimura has determined the number of generations until fixation for a single beneficial mutation with selective advantage s, $s \gg \frac{1}{N}$, $h = 0.5$, and proportion, $p = \frac{1}{PN}$, [5] as approximately,

$$\int_p^1 \frac{(e^{2PNs\xi} - 1)(e^{-2PNs\xi} - e^{-2PNs})}{s\xi(1-\xi)} d\xi$$

In order to properly use this result additional beneficial mutations at the same locus should not occur, as this would reduce the fixation time and hence the substitutional load.

The integrand denotes the time spent at each proportion. Multiplying the integrand within the integral by $1 - \xi$ gives the sum of the proportions of the population not containing the mutation multiplied by the time spent at each proportion. This is equivalent to the total of the fractional losses in log fitness associated with the fixation of a single site,

$$\begin{aligned} &\int_p^1 \frac{(e^{2PNs\xi} - 1)(e^{-2PNs\xi} - e^{-2PNs})}{s\xi} d\xi \\ &= \frac{-\mathrm{Ei}(-2PNs\xi) + \log(\xi) - e^{-2PNs}(\mathrm{Ei}(-2PNs\xi) - \log(\xi))}{s}\Bigg|_{\xi=p}^{1} \\ &\approx f(s,\xi)\Big|_{\xi=p}^{1} \text{ for } s \gg \frac{1}{2PN} \end{aligned}$$

in which,

$$f(s,\xi) = \frac{-\mathrm{Ei}(-2PNs\xi) + \log(\xi)}{s}$$

and Ei is the exponential integral,

$$\mathrm{Ei}(x) = \int_{-\infty}^{x} \frac{e^t}{t} dt$$

Let the increase in fitness associated with a double mutant relative to a single mutant be s'_p. The log fitness of a double mutant is approximately $\frac{s_p}{h}$, and that of a single mutant is approximately s_p, so,

$$s'_p \approx \frac{s_p}{h} - s_p = \frac{1-h}{h} s_p \tag{2}$$

Fixation can be approximately modeled as involving two adjacent components. One covering $\frac{1}{PN}$ to 50% mutational presence with $s = s_p$. And a second covering 50% to 100% mutational presence with $s = s'_p$. Ignoring rare double mutants in the first component, and null mutants in the second component, the mean log fitness, $\overline{x_s}$, associated with the substitutional load can be approximated in two parts. The rate of substitution multiplied by a cost of $\frac{s_p}{h}$ times the average fractional log fitness loss, $f(s_p,\xi)$, with ξ ranging from $\frac{1}{PN}$ to 0.5. And the rate of substitution multiplied by a cost of s'_p times the fractional log fitness loss for ξ ranging from 0.5 to 1,

$$\begin{aligned}
\overline{x_s} \approx &- \Gamma_p \frac{s_p}{h} f(s_p,\xi)\Big|_{\xi=\frac{1}{PN}}^{\frac{1}{2}} - \Gamma_p s'_p f(s'_p,\xi)\Big|_{\xi=\frac{1}{2}}^{1} \\
= &- \Gamma_p \Bigg(\frac{-\mathrm{Ei}(-PNs_p) + \log(0.5) + \mathrm{Ei}(-2s_p) + \log(PN)}{h} \\
&\qquad - \mathrm{Ei}(-2PNs'_p) + \mathrm{Ei}(-PNs'_p) - \log(0.5) \Bigg)
\end{aligned}$$

Now,

$$\mathrm{Ei}(x) \approx \begin{cases} 0 & \text{for } x \ll -1 \\ \gamma + \log(-x) & \text{for } -1 \ll x < 0 \end{cases}$$

where γ is Euler's constant, 0.5772.... Hence, assuming s_p, and $s'_p \gg \frac{1}{PN}$,

$$\overline{x_s} \approx -\Gamma_p \left(\frac{\log(PN) + \log(s_p) + \gamma}{h} + \log 2 \right) \tag{3}$$

Comparing equations 1 and 3, it can be seen that the effect of the substitutional load will be smaller than the effect of the prospective load whenever,

$$\log(PNs_p) + \gamma + h\log 2 < \frac{3}{2PN\mu_s\Pi}$$

This is likely to be true when $PN \ll \frac{1}{\mu_s}$. This will be the case most of the time; the substitutional load will typically be smaller than the prospective load.

The effect of concurrent beneficial mutations of a single site on the substitutional load

If two or more identical beneficial mutations to the same site establish and are concurrently fixing, this increases the speed of fixation at that site. A simple model can be used to assess whether concurrent mutation might be an issue. Fixation is roughly modeled as two logistic segments: from $\frac{1}{PN}$ to 50% presence, and from 50% to $1 - \frac{1}{PN}$. For a logistic curve, letting t be the time in generations, the proportion of mutated sites, p, is given by,

$$p = \frac{1}{1 + (1 + s_p)^{-t}}$$

Computing the time taken to go from $\frac{1}{PN}$ to 50% presence: $p \approx \frac{1}{PN}$ when $\log PN \approx -t\log(1 + s_p)$ assuming $PN \gg 1$, and $p = \frac{1}{2}$ at $t = 0$. Similar considerations apply to the second logistic segment. Thus the total fixation time is very roughly,

$$\frac{\log PN}{s_p} + \frac{\log PN}{s'_p}$$

Mutation of a particular site occurs at the rate $PN\frac{\mu_s}{3}$. The probability of fixation is approximately $2s_p\Pi$. This makes the expected time between consecutive fixation initiations at a particular locus, $\frac{3}{2PN\mu_s s_p\Pi}$.

Consequently, concurrent mutations at the same site will not be an issue when,

$$\frac{\log PN}{s_p} + \frac{\log PN}{s'_p} \ll \frac{3}{2PN\mu_s s_p\Pi}$$

Or, upon substituting equation 2, when,

$$PN\log PN \ll \frac{3(1-h)}{2\mu_s\Pi}$$

μ_s is around 10^{-8} mutations per base pair per sexual generation [8, 9, 10, 11, 12, 13]. h for deleterious mutations is around 0.2 or 0.3 [14], suggesting a value of h for beneficial mutations equal to 1 minus this value, or 0.7 to 0.8. Later on it is argued that Π will typically be reasonably close to 1. Hence, equation 3 will be accurate for $PN \ll 2\times10^6$. For values larger than this, $\overline{x_s}$ will

be smaller than that given by equation 3. The reduction will typically be small because beneficial mutations beyond the first to occur will typically only contribute a small amount to fixation.

Increasing the value of μ_s, increases the likelihood of concurrent mutations of the same site, and thus weakly reduces the magnitude of $\overline{x_s}$.

Mutational load

Many mutations are not beneficial, but deleterious, and will be removed from the population either by selection or by chance back mutation. Suppose deleterious mutations occur at random with a mean germline rate of Γ_n per organism per haploid genome per sexual generation. The symbol Γ_n was chosen as it is distinct from U, which commonly also includes neutral and beneficial mutations. Let the heterozygous selection coefficient associated with deleterious mutations be s_n. Also let the mean number of mutations present in the organisms be $\overline{k}$ per haploid genome. Assume the spontaneous mutation rate is very small, so that chance of back mutation is rare. And assume the genome is large so the chance of homozygotes occurring is small. Then mutation-selection balance dictates,

$$\Gamma_n PN = \overline{k} PN s_n$$

Or,

$$\overline{k} = \frac{\Gamma_n}{s_n}$$

As a result, if $\overline{x_m}$ is the log fitness cost of the mutational load,

$$\begin{aligned} \overline{x_m} &= -P\overline{k}\log(1+s_n) \\ &= -\frac{P\Gamma_n \log(1+s_n)}{s_n} \\ &= -P\Gamma_n \text{ provided } s_n \ll 1 \end{aligned} \tag{4}$$

This is Haldane's classical result for the mutational load [15, 16].

Neutral load

The analysis we have performed so far only applies when selection dominates over random drift. When $s \lesssim \frac{1}{N}$ this is not the case. The sites are neutral, or nearly neutral [3, 17, 18].

The motivation for considering the neutral load is not to provide a super accurate treatment of the region between neutrality and non-neutrality, but to provide an anchor point for future analysis

which may need to consider how fitness varies as a function of population size. When the population size varies, so does the number of neutral and nearly neutral sites.

The effect of neutral and nearly neutral sites will be determined in comparison to some hypothetical optimal reference genome. Neutrality means drift dominates over selection, so at any given time approximately $\frac{3}{4}$ of neutral sites will differ from their reference values. Let s_0 be the mean selection coefficient associated with such differing sites. Let L_0 be the number of neutral or nearly neutral sites per haploid genome.

The log fitness loss associated with neutral and nearly neutral sites, $\overline{x_0}$, is given by,

$$\overline{x_0} \approx -\frac{3PL_0s_0}{4} \tag{5}$$

Consequently, depending on the number of neutral and nearly neutral sites, and their selection coefficients, the log fitness loss associated with them could be substantial. They are neutral with respect to individual selection, not necessarily aggregate fitness.

Factors influencing the probability of fixation of beneficial mutations

The relative probability for the fixation of a beneficial mutation is given by Π. A number of factors contribute to Π.

Unlinked deleterious mutations

Unlinked deleterious mutations create background noise that decrease the efficiency of selection in the establishment of a beneficial mutation.

Letting Π_{un} be the decrease in the probability of fixation of a beneficial mutation due to uniformly distributed unlinked deleterious mutations, Barton derived the equation [2],

$$\Pi_{un} = \exp(-4\Gamma_n s_n) \tag{6}$$

The ploidy, P, should be inserted into this equation to reflect the fact that in diploids, not just the unlinked deleterious sites in one set of chromosomes have a probability of $\frac{1}{2}$ of being associated with the beneficial site in the next generation, but so do all the sites of the sister chromosomes. This results in,

$$\Pi_{un} = \exp(-4P\Gamma_n s_n) \tag{7}$$

Barton subsequently sidelined equation 6, favoring an alternative additive equation that more closely fit a numerical model. However, it seems likely the original approach was not wrong, but

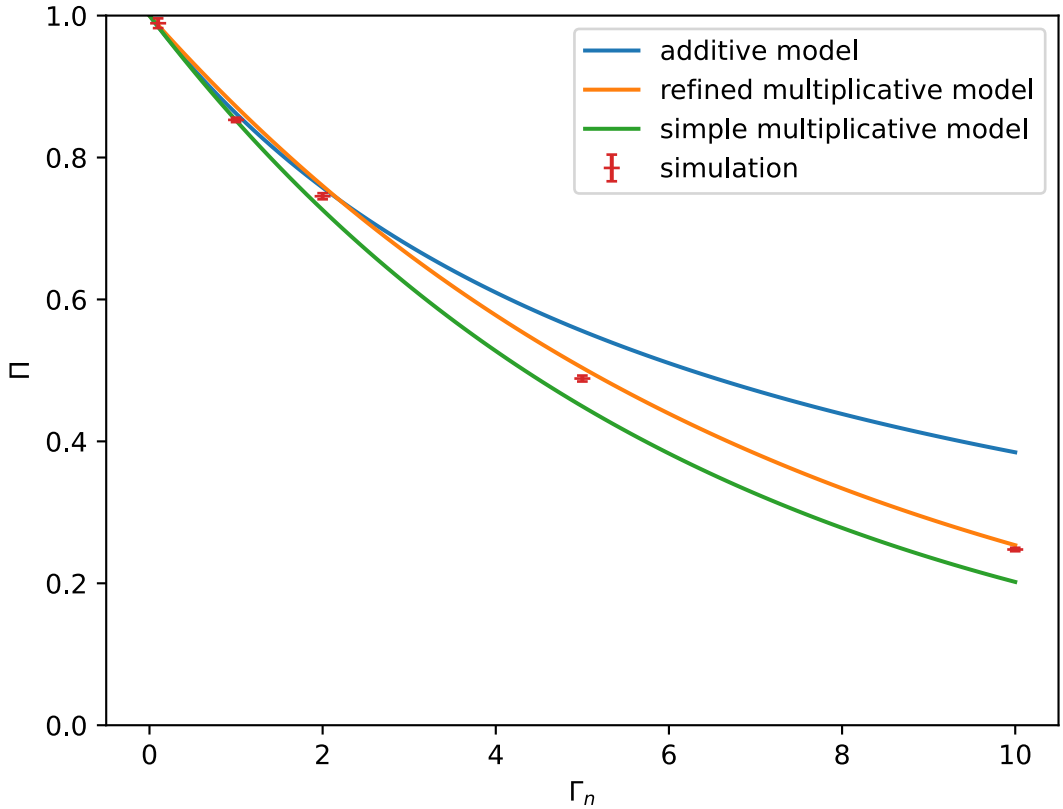

Figure 1: Relative fixation probability models and simulation results for beneficial mutations in the presence of unlinked deleterious mutations. Parameter values chosen to require minimal adjustment to Π_{un} by virtue of s_p and s_n being small, and to illustrate the divergence between the different models by virtue of Γ_n becoming large. Diploid, N=10,000, s_p=0.02, s_n=0.02, and Γ_n as indicated. Additive model: $\frac{1}{1+4P\Gamma_n s_n}$. Refined multiplicative model: equation 8. Simple multiplicative model: $\exp(-4P\Gamma_n s_n)$. Simulation data: mean of 10 runs of 100,000 fixation attempts per data point; error bars indicate the standard error of the mean.

merely had been simplified too much, making it less accurate. Going back to an earlier equation of Barton's leads to the form [2][equation 17c with $r = \frac{1}{2}$],

$$\Pi_{un} = \exp\left(\frac{-P\Gamma_n}{s_n\left(1+\frac{1}{2s_n}+\frac{s_p}{s_n}\right)^2}\right) \tag{8}$$

This more refined form for Π_{un} results in a good fit with simulation data. This is shown in Figure 1. See File S1 for the simulation code and data [19].

Empirically, U was found to be less than 1.1 mutations per generation across a range of species [16]. Γ_n is a fraction of U. Since the mean value of s_n is likely to be much less than 1, referring to

equation 7, Π_{un} will be reasonably close to 1.

Equation 8 assumes that s_p and $s_n \ll 1$. If this is not the case an adjustment to Π_{un} will be necessary.

Linkage between beneficial and nearby deleterious mutations

Linkage between a beneficial and nearby deleterious mutations will reduce the probability of fixation. Let Π_{ln} be the reduction in the probability of fixation of a beneficial site due to nearby linked deleterious mutations. Let R be the length of the genome in Morgans. Assume deleterious mutations are spread uniformly over the genome, crossovers are distributed randomly, the number of chromosomes is not excessively large, $R \gg 1$, and $s_n \ll 1$. Letting $\theta = \frac{s_p}{s_n}$, Barton has shown [2],

$$\Pi_{ln} \approx \begin{cases} \exp\left(\frac{-2\Gamma_n}{R}\left(1 - \frac{\theta}{3}\right)\right) & \text{for } \theta \leq 1 \\ \\ \exp\left(\frac{-2\Gamma_n}{R\theta}\left(1 - \frac{1}{3\theta}\right)\right) & \text{for } \theta \geq 1 \end{cases}$$

The effect of Π_{ln} is expected to be small. Were this not the case there would be strong evolutionary forces in favor of increasing fitness by increasing R by increasing the number of chromosomes. Crossovers are required to lock paired chromosomes together until anaphase 1 [20], and there is typically one crossover per chromosome, or chromosome arm [21]. Consequently, Π_{ln} could be increased by simply increasing the number of chromosomes.

Unlinked beneficial mutations

Analogous to deleterious sites, fixing beneficial sites create a log fitness variance that reduces the efficiency of selection in the establishment of a beneficial mutation.

For illustrative purposes, consider a eukaryotic species with a haploid life cycle (or a diploid with $h = 0.5$). The proportion of sites carrying a mutation that is in the process of fixing, p, can be modeled by the logistic function,

$$p = \frac{1}{1 + e^{-x}}$$

where,

$$x = \log(1 + s_p)t$$

and the time, t, is measured in generations.

The average variance in log fitness resulting from sites that are fixing is given by the products of the the rate at which sites fix, the ploidy (because the variance of a diploid is the sum of the variances

of two sets of chromosomes), and the aggregate variance over time from a single site fixing. The presence of the mutation on a particular haploid genome follows the Bernoulli p distribution, and if present it has log fitness effect of $\log(1+s_p)$. The variance of the Bernoulli distribution is $p(1-p)$. Consequently,

$$\begin{aligned}
\mathrm{Var}\left[x_s\right] &= \Gamma_p P \left(\log(1+s_p)\right)^2 \int_{-\infty}^{\infty} p(1-p)dt \\
&= \Gamma_p P \log(1+s_p) \frac{-1}{1+e^x}\bigg|_{x=-\infty}^{\infty} \\
&= \Gamma_p P \log(1+s_p)
\end{aligned}$$

Since Γ_p and s_p are typically both small, this variance will be very small. Based on the work of Barton [2], this produces a reduction in the probability of fixation of an unlinked beneficial mutation, Π_{up}, of approximate size,

$$\begin{aligned}
\Pi_{up} &\approx \frac{1}{1+4\,\mathrm{Var}\left[x_s\right]} \\
&\approx 1
\end{aligned}$$

Since the Π_{up} factor is very close to 1, it can typically be ignored. For this reason, it is not necessary to consider the effect in diploids when $h \neq 0.5$, nor the variance associated with beneficial mutations that fail to establish.

Linkage between beneficial mutations and nearby fixing mutations

Hitchhiking by non-mutated sites on a nearby mutation that is in the process of fixing reduces the probability that a new beneficial mutation establishes itself at the non-mutated site locus. Let the reduction in the probability of fixation be Π_{lp}.

I am not aware of any simple approximate expression for Π_{lp}. However, simulation suggests the effect of Π_{lp} will be small. Figure 2 shows population mean log fitness as a function of the mean number of crossings over per meiosis, R. Figure 2 also shows the results for a simulated infinite number of crossings over, for which Π_{lp} will be 1. As can be observed, the reduction in fitness when R is in the physiologically typical range of 20 to 50 Morgans is minimal. See File S2 for the simulation code and data [22].

Evolutionary considerations also suggest, Π_{lp} is likely to exert only a small effect. If this was not so, there would be a strong evolutionary force in favor of increasing R. An exception to this might occur if fixation of one site leads to mutational prospects in multiple very nearby sites. Clustering of mutations in this way is quite plausible, but is not captured by the model as developed here.

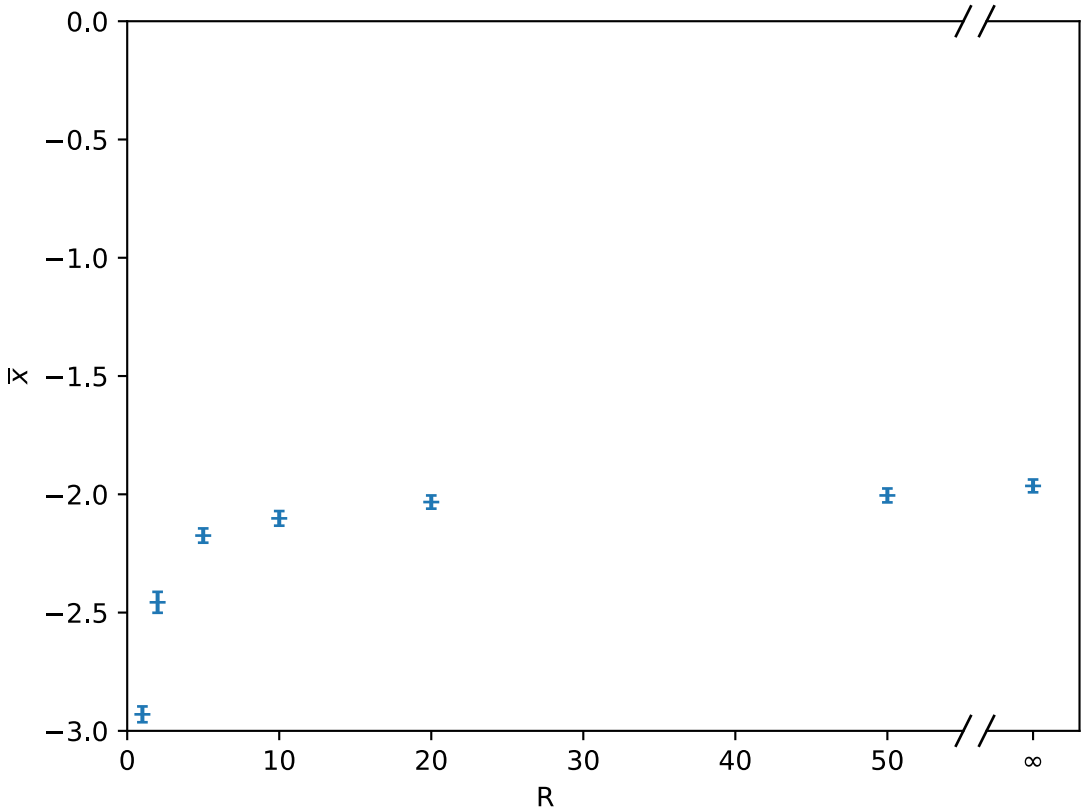

Figure 2: Simulated mean log fitness for linked beneficial mutations in the absence of deleterious mutations. Diploid, N=100,000, Γ_p=0.01, s_p=0.01, h=0.8, genome modeled as a single chromosome with a Poisson distributed number of crossovers having expectation R. Mean of 10 runs of 100,000 generations per data point; error bars indicate the standard error of the mean.

Interference between beneficial, deleterious, and neutral mutations

When combining multiple mutations of different types at different loci it is important to consider the possibility of interference between the fitness effects of different mutational types. Interference occurs when the presence of one mutational type affects the fitness of another.

No interference occurs between the prospective load and the other loads because, apart from rare beneficial mutations that are destined to die out, such mutations are not present in the population.

Interference by beneficial and deleterious mutations on the substitutional load is described by Π. At least to first order, Π is likely to be given by the product of its individual factors. Barton makes this assumption for the effects of linked and unlinked deleterious mutations [2]. In addition it has been argued that the effects of each of the Π factors will be small. Even if this were not the case, and there existed some unexpected interference between the Π factors, this would only warrant a change to Π not the overall fitness formula given below.

Interference by neutral mutations on beneficial and deleterious mutations is likely to be very small. Neutral loci are large in number, but only a small fraction of them show significant variability, and each is very small in individual effect. As a result the inter-organism variance in the aggregate fitness effects of such mutations will be very small.

As argued previously when discussing unlinked beneficial mutations, the variance associated with beneficial mutations is likely to be very small. Consequently, interference by beneficial mutations on the clearance of deleterious mutations is likely to be very small.

Finally, interference by beneficial and deleterious mutations on the neutral load is negligible. Neutral site frequencies are determined by drift rather than their selective effect. Consequently, while hitchhiking is possible, such hitchhiking is as likely to reduce the neutral load as it is to increase it.

Fitness of beneficial, deleterious, and neutral mutations

The fitness effects of beneficial mutations that have yet to occur, or are in the process of fixing, are captured by $\overline{x_p}$, and $\overline{x_s}$, respectively. The occurrence of deleterious mutations is captured by $\overline{x_m}$. And variability at neutral loci is captured by $\overline{x_0}$.

Let $\overline{x}$ be the mean log fitness of a population. Combining equations 1, 3, 4, and 5,

$$\begin{aligned}\overline{x} &= \overline{x_p} + \overline{x_s} + \overline{x_m} + \overline{x_0} + \cdots \\ &\approx -\frac{3\Gamma_p}{2PN\mu_s h\Pi} - \Gamma_p\left(\frac{\log(PN) + \log(s_p) + \gamma}{h} + \log 2\right) - P\Gamma_n - \frac{3PL_0 s_0}{4} + \cdots \end{aligned} \tag{9}$$

where,

$$\Pi \approx \Pi_{un}\Pi_{ln}\Pi_{up}\Pi_{lp}$$

and "$\cdots$" denotes the effect of other genetic loads, not considered here, such as loads related to inbreeding or heterozygote advantage.

For multiple loci with different dominance and selection coefficients, equation 9 is still applicable, provided $h\Pi$, h, and s_p are replaced by appropriately weighted average values.

The population optimal spontaneous mutation rate

The likelihood of a mutation being harmful is far greater than the likelihood of it being beneficial. Further, the magnitude of the fitness effect of harmful mutations is typically greater than that of beneficial mutations. Consequently, it has been hypothesized that natural selection will drive the mutation rate down to the lowest possible level taking into account the cost of producing such low mutation rates [23, 24], or that a lower bound is set by random genetic drift [25, 26].

Here we instead put forth the possibility that in the big picture evolution tunes the sexual mutation rate per generation so that it is as large as possible to maximize the ability to adapt and hence minimize the fitness cost associated with the prospective load, but not so large that negative, or purifying, selection incurs a heavy cost. That is, evolution tunes the mutation rate towards a species or population optimal value. This is decidedly different from the optimal mutation rate for the individual organism and its genes. This process may occur through macroevolution. Species that evolve mutation rates that prove to be too low, would fail to adapt, have a large prospective load, and become extinct. Species with mutation rates that are too high, would incur a heavy mutational load, fewer successful offspring, and again become extinct.

Let L_n be the number of sites under the control of negative selection per haploid genome. So that,

$$\Gamma_n = L_n \mu_s \tag{10}$$

Substituting this into equation 9, assuming any other loads are either small, or only weakly dependent on μ_s, produces,

$$\overline{x} \approx -\frac{3\Gamma_p}{2PN\mu_s h\Pi} - PL_n\mu_s + K \tag{11}$$

where K consists of terms that are either independent, or only weakly dependent on μ_s. Differentiating equation 11 with respect to μ_s, the population fitness takes on its maximal value when,

$$\frac{3\Gamma_p\left(\Pi + \mu_s \frac{d\Pi}{d\mu_s}\right)}{2PN\mu_s^2 h\Pi^2} - PL_n \approx 0$$

Or,

$$\frac{3\Gamma_p\left(\Pi + \Gamma_n \frac{d\Pi}{d\Gamma_n}\right)}{2PN\mu_s^2 h\Pi^2} \approx PL_n$$

Both Π_{un} and Π_{ln} depend on Γ_n and have a combined effect of the form $\exp(-\Gamma_n k)$ for some constant k. Previously it was argued that Π_{un} and Π_{ln} will be close to 1. Consequently, $\Gamma_n k \ll 1$. This makes,

$$\Pi + \Gamma_n \frac{d\Pi}{d\Gamma_n} \approx \Pi$$

So that the maximal population fitness occurs when,

$$\mu_s \approx \sqrt{\frac{3\Gamma_p}{2P^2 N L_n h \Pi}} \tag{12}$$

Or,

$$\mu_s \approx \frac{3\Gamma_p}{2P^2 N \Gamma_n h \Pi} \text{ by equation 10} \tag{13}$$

This maximal value is given by,

$$\overline{x} \approx -2P\Gamma_n + K$$

In other words, at the maximum population fitness value, the fitness losses coming from the prospective load and the mutational load are approximately equal. The lag load consists of the prospective and substitutional loads. If evolution favors the population optimal mutation rate, it follows that the lag load is likely greater than the mutational load.

Equation 12 suggests an inverse relationship between population sizes and mutation rates for sexual species. More precisely, assuming other parameters are held roughly constant, it predicts the log mutation rate should equal $-\frac{1}{2}$ times the log population size plus a constant. This appears to agree with empirical observations [26].

Discussion

This manuscript has considered sexual species. For the optimal spontaneous mutation rate of asexual populations similar considerations to those for sexual populations apply. The log fitness loss coming form the prospective load will be inversely proportional to the product of the population size and the mutation rate, while that of the mutational load will be proportional to the mutation rate. This will lead to a population optimal mutation rate in which the prospective load and the mutational load are approximately equal. And once again the population optimal mutation rate will be inversely proportional to the square root of the population size.

By assembling what is known of the fitness effects of beneficial, deleterious, and neutral mutations, it is possible to construct an overall model. A new load, the prospective load, was carved out

from the lag load and studied. The prospective load is typically likely to be greater than the other component of the lag load, the substitutional load. The constructed model suggests a spontaneous mutation rate per base pair that maximizes the mean fitness of a population. Too low a mutation rate, and prospective mutational sites take a long time to fix, increasing the prospective load. Too high a mutation rate, and negative selection incurs a large mutational load. The population optimal mutation rate is determined to be one for which the fitness losses from the prospective and mutational loads are equal. The extent to which evolution favors the population optimal mutation rate is a question for further study.

Data availability

The code and data used to produce Figures 1 and 2 are publicly available as File S1 and File S2 respectively.

File S1 - Fixation probability simulation software and results.
`https://doi.org/10.5281/zenodo.13901487`

File S2 - Fitness simulation software and results.
`https://doi.org/10.5281/zenodo.13927246`

Acknowledgments

I am very grateful for the time Steven Greidinger, Philip Gerrish, and Benjamin Good spent reviewing early versions of a related manuscript and providing me with feedback.

Conflict of interest statement

The author declares they have no conflicts of interest in relation to the content of this manuscript.

References

[1] J Maynard Smith. What determines the rate of evolution? *The American Naturalist*, 110(973):331–338, 1976.

[2] Nick H Barton. Linkage and the limits to natural selection. *Genetics*, 140(2):821–841, 1995.

[3] Motoo Kimura. *The neutral theory of molecular evolution*. Cambridge University Press, 1983.

[4] Motoo Kimura. On the probability of fixation of mutant genes in a population. *Genetics*, 47(6):713, 1962.

[5] Motoo Kimura and Tomoko Ohta. The average number of generations until fixation of a mutant gene in a finite population. *Genetics*, 61(3):763, 1969.

[6] Edward J Klekowski. Genetic load and its causes in long-lived plants. *Trees*, 2:195–203, 1988.

[7] Motoo Kimura. Optimum mutation rate and degree of dominance as determined by the principle of minimum genetic load. *Journal of Genetics*, 57(1):21–34, 1960.

[8] John W Drake. The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes. *Annals of the New York Academy of Sciences*, 870(1):100–107, 1999.

[9] Michael W Nachman and Susan L Crowell. Estimate of the mutation rate per nucleotide in humans. *Genetics*, 156(1):297–304, 2000.

[10] Adeolu B Adewoye, Sarah J Lindsay, Yuri E Dubrova, and Matthew E Hurles. The genome-wide effects of ionizing radiation on mutation induction in the mammalian germline. *Nature Communications*, 6(1):6684, 2015.

[11] Brandon Milholland, Xiao Dong, Lei Zhang, Xiaoxiao Hao, Yousin Suh, and Jan Vijg. Differences between germline and somatic mutation rates in humans and mice. *Nature Communications*, 8(1):1–8, 2017.

[12] Peter D Keightley, Rob W Ness, Daniel L Halligan, and Penelope R Haddrill. Estimation of the spontaneous mutation rate per nucleotide site in a Drosophila melanogaster full-sib family. *Genetics*, 196(1):313–320, 2014.

[13] Peter D Keightley, Urmi Trivedi, Marian Thomson, Fiona Oliver, Sujai Kumar, and Mark L Blaxter. Analysis of the genome sequences of three Drosophila melanogaster spontaneous mutation accumulation lines. *Genome Research*, 19(7):1195–1201, 2009.

[14] Chenlu Di and Kirk E Lohmueller. Revisiting dominance in population genetics. *Genome Biology and Evolution*, 16(8):evae147, 2024.

[15] JBS Haldane. The effect of variation of fitness. *The American Naturalist*, 71(735):337–349, 1937.

[16] Aneil F Agrawal and Michael C Whitlock. Mutation load: the fitness of individuals in populations where deleterious alleles are abundant. *Annual Review of Ecology, Evolution, and Systematics*, 43(1):115–135, 2012.

[17] Tomoko Ohta and Motoo Kimura. On the constancy of the evolutionary rate of cistrons. *Journal of Molecular Evolution*, 1:18–25, 1971.

[18] Wen-Hsiung Li. Maintenance of genetic variability under the joint effect of mutation, selection and random drift. *Genetics*, 90(2):349–382, 1978.

[19] Gordon Irlam. An integrated model of the effects on fitness of beneficial, detrimental, and neutral mutations in sexual populations: Supplement 1, 2024. doi:10.5281/zenodo.13901487.

[20] Stacie E Hughes and R Scott Hawley. Meiosis: location, location, location, how crossovers ensure segregation. *Current Biology*, 30(7):R311–R313, 2020.

[21] Joana Segura, Luca Ferretti, Sebastián Ramos-Onsins, Laia Capilla, Marta Farré, Fernanda Reis, Maria Oliver-Bonet, Hugo Fernández-Bellón, Francisca Garcia, Montserrat Garcia-Caldés, et al. Evolution of recombination in eutherian mammals: insights into mechanisms that affect recombination rates and crossover interference. *Proceedings of the Royal Society B: Biological Sciences*, 280(1771):20131945, 2013.

[22] Gordon Irlam. An integrated model of the effects on fitness of beneficial, detrimental, and neutral mutations in sexual populations: Supplement 2, 2024. doi:10.5281/zenodo.13927246.

[23] Egbert Giles Leigh Jr. Natural selection and mutability. *The American Naturalist*, 104(937):301–305, 1970.

[24] Paul D Sniegowski, Philip J Gerrish, Toby Johnson, and Aaron Shaver. The evolution of mutation rates: separating causes from consequences. *Bioessays*, 22(12):1057–1066, 2000.

[25] Michael Lynch. The lower bound to the evolution of mutation rates. *Genome Biology and Evolution*, 3:1107–1118, 2011.

[26] Michael Lynch, Matthew S Ackerman, Jean-Francois Gout, Hongan Long, Way Sung, W Kelley Thomas, and Patricia L Foster. Genetic drift, selection and the evolution of the mutation rate. *Nature Reviews Genetics*, 17(11):704–714, 2016.