Evidence of macroevolutionary selection of the mutation rate in sexual species

Gordon Irlam
gordoni@gordoni.com

Los Altos, California, United States

January 1, 2026

Abstract

For sexual species it has been hypothesized that, over macroevolutionary timescales, species-level processes such as differential extinction may bias the germline mutation rates of surviving species toward values that maximize long-term population mean fitness. If this hypothesis is correct, then the mutation rate is expected to lie near a population-optimal value that depends on several evolutionary and demographic parameters. Using previously published data, these parameters were estimated for several well-studied species pairs, enabling a quantitative evaluation of this prediction. Across the species examined, empirical estimates of mutation rates fall within the range that is predicted by the underlying model to yield appreciable population mean fitness, given substantial uncertainty in several parameter values. The underlying model is consistent with a previously reported inverse relationship between effective population sizes and mutation rates within broad clades. Although not intended to provide precise predictions for individual species, these results are compatible with the hypothesis that macroevolutionary processes contribute to shaping germline mutation rates toward population-optimal values.

Keywords: population-optimal mutation rate.

Introduction

This manuscript empirically assesses the determinants of germline spontaneous mutation rates in sexual eukaryotes.

Because most mutations are harmful rather than beneficial, adaptations that reduce the mutation rate are expected to be favored by microevolution. Despite the expectation that microevolution acts to minimize the mutation rate, spontaneous mutations cause a significant genetic load; the mutational load [1]. Why, then, is the mutation rate not lower than observed?

The drift-barrier hypothesis argues that genetic drift acts to create a barrier that impedes the fixation of mutations that contribute only small improvements in replication fidelity [2, 3].

What is beneficial at the level of genes or organisms is not necessarily optimal at the population or species level. If the mutation rate is driven too low, the species may be unable to adapt sufficiently and will face an elevated risk of extinction. Too high, and the species will bear a large deleterious mutational load. There therefore exists a population-optimal mutation rate between these two extremes that maximizes long-term geometric mean population mean fitness. For sexual species it is possible to estimate this population-optimal mutation rate [4]. Since microevolution might largely favor smaller and smaller mutation rates, one plausible mechanism for species to achieve the population-optimal mutation rate is through macroevolution. Species with mutation rates that are too high or too low are more likely to go extinct, leaving surviving species biased towards mutation rates near the population-optimal value. This is not a new argument; it was articulated by Kimura in 1967 [5]. For brevity, this hypothesis might be termed the mutation rate macroevolutionary selection hypothesis.

This manuscript seeks to evaluate the mutation rate macroevolutionary selection hypothesis. For the three animal species pairs considered, the predicted population-optimal mutation rates are found to be close to the observed mutation rates. That is, both are around 10^{-8} mutations per base pair per generation. Apart from one of the species pairs, where the a priori parameters might have been misestimated, the observed mutation rate is found to lie within the small range of values that are expected to yield an appreciable population mean fitness. Assuming species tend towards the population-optimal mutation rate, a quantitative relationship between effective population size and the mutation rate is predicted. This relationship is shown to largely agree with a previously observed empirical relationship between these two variables. These observations add evidence to the hypothesis that the mutation rates of surviving species are biased by macroevolutionary processes towards values that produce optimal population mean fitness.

Results

The rate of adaptive substitution

Let Γ_s denote the per sexual generation rate at which substituting sites become established in the population. For a species pair over macroevolutionary timescales, the temporal mean value of Γ_s equals the rate of positively selected adaptive substitution per generation and can be estimated by

	human-chimp	mouse-rat	*Drosophila*	Comments
α_n	0.25 [7]	0.41 [7]	0.54 [8]	
D_n	3.9×10^4 [9]	9.7×10^5	5.9×10^4 [8, 10]	mouse-rat: 3.5% [11] $\times$ 75% [12] $\times$ 3.7×10^7 [10]
$\alpha_{nc}D_{nc}$	0 [13]	≥ 0	1.3×10^5 [8, 10]	mouse-rat: not well known
g	26 [14, 15]	0.3 [16, 17]	0.07 [18, 19]	in years
T	6.4×10^6	1.2×10^7	4.6×10^6	in years [20]
$\mathrm{E}[\Gamma_s]$	2.0×10^{-2}	$\geq 5.0\times10^{-3}$	1.2×10^{-3}	
$\frac{\mathrm{E}[\Gamma_s]}{g}$	7.7×10^{-4}	$\geq 1.7\times10^{-2}$	1.8×10^{-2}	

Table 1: Estimation of Γ_s, the rate at which substituting sites become established per sexual generation. Analysis of three species pairs. *Drosophila* is *D. melanogaster* and *D. simulans*

[4],

$$\mathrm{E}[\Gamma_s] \approx \frac{(\alpha_n D_n + \alpha_{nc} D_{nc})g}{2T}$$

where α_n is the fraction of non-synonymous substitutions that are fixed by positive selection, D_n is the number of non-synonymous substitutions, α_{nc} is the fraction of non-coding substitutions that are fixed by positive selection, D_{nc} is the number of non-coding substitutions, g is the generation length, and T is the divergence time.

Table 1 shows the parameters used for, and estimates of, Γ_s for three species pairs. The *Drosophila* species studied were *D. melanogaster* and *D. simulans*. Although Γ_s is larger for humans and chimps than the two other species pairs, the much longer generation length results in a substantially lower per generation rate of genomic adaptation to environmental change. The value for *Drosophila* is broadly in line with an estimate of one coding adaptive substitution every 45 years for *D. simulans* and *D. yakuba* based on an analysis of just 35 genes [6].

The population-optimal mutation rate

Let μ denote the germline spontaneous mutation rate per base pair per sexual generation.

The population-optimal mutation rate $\hat{\mu}|g$ is expected to be given by [4],

$$\hat{\mu}|g \approx \sqrt{\mathrm{E}\left[\frac{3\pi_M \pi_e f_l \Gamma_s}{2P^2 \pi_f N_e L_n h}\right]} \tag{1}$$

where $\mathrm{E}[\cdot]$ denotes the expectation taken over generations, π_M is an adjustment to log fitness of the latent load (i.e., the genetic load from positively selected sites that are not yet established) due to macroevolutionary selection favoring fitter species, π_e is an adjustment to the log fitness

effect of selection due to epistasis, f_l is the fraction of new positively selected sites that are latent (i.e., not yet established) sites, P is the ploidy (haploids 1, diploids 2), π_f is an adjustment to the probability of fixation at new positively selected sites as a result of unaccounted for interference between positively selected sites, N_e is the neutral site variance effective population size, L_n is the effective number of conserved sites in a haploid genome, and h is the harmonic mean value of the beneficial alleles' dominance coefficients.

The given g qualifier in $\hat{\mu}|g$ signifies that this is the optimal mutation rate for a given generation length. This may be different from the optimal mutation rate if g is also free to vary.

Over evolutionary time frames of the order of hundreds of millions of years, most terms in equation 1 are likely to be approximately constant. The exceptions are Γ_s, N_e, and L_n. This means that for constant N_e and L_n or for constant Γ_s, the expectation operator can be dropped and $N_e L_n$ represented by its harmonic mean value or Γ_s represented by its arithmetic mean value.

Equation 1 is intended for broad macroevolutionary inference and is unlikely to yield precise predictions for individual species. This is because other stochastic factors also affect the survival of species.

N_e was approximated by the current neutral site coalescent effective population size, N_e^{Co}. Unfortunately N_e^{Co} is only approximately appropriate over the time to the coalescent, not over the entire time until the species merged. Ideally N_e might be estimated using the pairwise sequentially Markovian coalescent model [21], and the generationally weighted harmonic mean value taken. Since this was not done, it is assumed that the variance effective population size over the entire time interval is approximated by the estimated coalescent effective population size. This approximation biases estimates toward recent demographic history.

Table 2 shows actual mutation rate μ, the parameters used for, and estimates of, $\hat{\mu}|g$ assuming a constant N_e. In order to estimate the impact of macroevolutionary selection, π_M, it was necessary to estimate S_{niche}, the number of species occupying a given niche, λ, the rate of speciation along a lineage, and $\overline{s_p}$, the mean selection coefficient of new positively selected sites. $\overline{s_p}$ itself was determined from the mean fitness effects of new beneficial mutations, $\overline{s_b}$. Macroevolutionary selection was then determined to have little impact; π_M ranged from 0.90 to 1.0. L_n was computed as the product of f_n, the effective fraction of conserved sites in the genome, and L, the length of the genome. The final estimate of $\hat{\mu}|g$ for mice and rats is a lower bound since the number of non-coding adaptive substitutions, $\alpha_{nc} D_{nc}$, could not be reliably determined.

The values for the parameters in Table 2 were determined a priori, without reference to the resulting population-optimal mutation rate predictions. The predicted population-optimal mutation rates are reasonably close to the observed rates. I.e. both are around 10^{-8} mutations per base pair per sexual generation, and are at or within a factor of 4 of one another. Given uncertainties in several parameter values and the influence of additional stochastic factors, this correspondence provides supportive, though not definitive, evidence for the mutation rate macroevolutionary selection hypothesis.

	human-chimp	mouse-rat	*Drosophila*	Comments
S_{niche}	5 [22]	20 [23]	200 [24]	
λ	0.3×10^{-6}	0.3×10^{-6}	0.3×10^{-6}	in years^{-1} [25]
$\overline{s_b}$	$\geq 7\times10^{-4}$ [26]	4.5×10^{-5} [27]	1.3×10^{-5} [28]	inferred from coding region alone [29]
$\overline{s_p}$	$\geq 2.8\times10^{-3}$	1.5×10^{-4}	5.2×10^{-5}	[10]
f_n	5.5%	-	61%	human-chimp: median estimate [30] *Drosophila*: calculated [10, 31]
L	3.1×10^{9}	-	1.4×10^{8}	
π_M	1.0	0.98	0.90	all sites with same s_p [10]
π_e	0.8	0.8	0.8	[4]
f_l	100%	100%	99%	calculated using formula [4]
P	2	2	2	
π_f	1	1	1	[4]
N_e	16,000	160,000	880,000	harmonic mean of N_e^{Co} [32][EV1]
h	0.8	0.8	0.8	
L_n	1.7×10^{8}	1.7×10^{8}	8.5×10^{7}	mouse-rat: assume same as human
$\hat{\mu}\|g$	5.3×10^{-8}	$\geq 8.2\times10^{-9}$	2.3×10^{-9}	
μ	1.3×10^{-8}	5.6×10^{-9}	4.5×10^{-9}	mean [32][EV1]
$\frac{\mu}{\hat{\mu}\|g}$	25%	$\leq 68\%$	200%	

Table 2: Observed mutation rates are reasonably close to hypothesized population-optimal mutation rates. Parameters values and a priori estimates of the population-optimal mutation rate, $\hat{\mu}|g$, assuming a constant N_e, along with the observed mutation rate μ, for three species pairs.

Post hoc considerations of the difference between the population-optimal and actual mutation rates

The estimated value of $\hat{\mu}|g$ exceeds μ for humans and chimps, and for mice and rats, but is lower than μ for *Drosophila*.

Although exact concordance is not expected to occur, taken together, the factors listed below appear sufficient to allow $\hat{\mu}|g$ and μ to be in close agreement for each species pair. The uncertainty in the parameter values precludes determining whether μ systematically lags or exceeds $\hat{\mu}|g$.

Humans and chimps

For humans and chimps, the estimate $\alpha_n = 0.25$ appears high. Other estimates of α_n include 0.24 [33], 0.1–0.2 [34], 0.14 [35], 0.10–0.13 [36], 0.002 [37], 0.0 for human-macaque with a 95% confidence interval [-0.30, 0.24] [38], and < 0 [39]. Reducing α_n would reduce $\hat{\mu}|g$. For example, using $\alpha_n = 0.1$ would reduce $\hat{\mu}|g$ by 37%.

The ancestral human-chimp effective population size is estimated to have been 48,000 [40]. This is considerably larger than the estimate of 16,000 based on short-term coalescence that was used. A larger N_e would reduce $\hat{\mu}|g$. For example, using the ancestral estimate for $N_e$would reduce $\hat{\mu}|g$ by 42%.

Mice and rats

Mice and rats have a larger N_e than humans. This makes alleles that would be nearly neutral in humans, selected in mice and rats. This applies to regulatory elements in the non-coding region [41]. Mice and rats will thus have a greater number of effectively conserved sites than humans, L_n. L_n for mice and rats was assumed to be the same as for humans. Increasing it will reduce $\hat{\mu}|g$. For example, doubling L_n would reduce $\hat{\mu}|g$ for mice and rats by 29%.

Drosophila

The mutation rate in Drosophila was estimated as 4.5×10^{-9}. This is the mean of 6 studies ranging from 1.7×10^{-9} to 6.6×10^{-9} [32][EV1]. If the historical actual value is towards the low end of the estimates from these studies, then μ and $\hat{\mu}|g$ would be in agreement for Drosophila.

The time of divergence used for *Drosophila* was 4.6×10^{6} years ago. The 95% confidence interval is [2.5×10^{6}, 1.3×10^{7}] years ago [20], a considerable range. If the actual time was towards the lower end of this range, this would increase $\hat{\mu}|g$. For example, using the lower interval value would increase $\hat{\mu}|g$ by 36%.

The genus *Drosophila* contains around 2,000 species that are known to have diversified rapidly [42]. Hawaiian *Drosophila* are a clade of an estimated 1,000 species that arose through founder events and vicariance, and form an adaptive radiation that began 26 million years ago [43]. This suggests

that over much of the timespan of the genus *Drosophila*, N_e might have been influenced by founder events, and so be much lower than the estimate of 880,000 for *D. melanogaster* and *D. simulans*. Consequently, depending on the time frame over which macroevolutionary selection influences the mutation rate of surviving species, the observed mutation rate for *D. melanogaster* and *D. simulans* may be more appropriate to species with a lower N_e. Reducing N_e would increase $\hat{\mu}|g$.

Predicted fitness

Let $\overline{w}$ denote the population mean relative fitness of a sexual population. The geometric-mean value of $\overline{w}$ averaged over the number of generations as a function of μ is proposed to be given by [4],

$$\text{Geomean}[\overline{w}] \approx \exp\left(-\,\text{E}\left[\frac{3\pi_M \pi_e f_l \Gamma_s}{2P\pi_f N_e \mu h}\right] - PL_n\mu + K\right) \tag{2}$$

where K is either independent of, or only weakly dependent on, μ.

Using equation 2, Figure 1 illustrates Geomean$[\overline{w}]$ as a function of μ assuming constant N_e, along with the estimated actual mutation rate. Figure 1 shows that there is a reasonably narrow range of mutation rates, within a factor three, outside of which population mean fitness is substantially smaller than at the population-optimal mutation rate. Equation 2 for geometric-mean population mean fitness implies a mutation rate "filter" through which competing species must pass. Except for humans and chimps, where α_n and N_e might have been misestimated, the mutation rates occurred where the species were expected to experience appreciable fitness. For humans and chimps a post hoc analysis with $\alpha_n = 0.1$ and $N_e = 48{,}000$ showed the mutation rate occurred where it was expected to yield appreciable fitness. Such concordances would be more difficult to explain if the mutation rate macroevolutionary selection hypothesis were incorrect.

The relationship between effective population size and the mutation rate

π_M, π_e, π_f, f_l, P, and h are expected to have similar values within very broad clades (Linnaean kingdoms). Much of the time L_n might also be expected to have roughly similar values within slightly less broad clades (Linnaean phyla). If either N_e or Γ_s is constant over time for individual species, then equation 1 implies that within a less broad clade,

$$\log\mu \approx -0.5\log\text{Hm}[N_e] + 0.5\log\text{E}[\Gamma_s] + C$$

where Hm$[\cdot]$ denotes the harmonic mean taken over generations, and C is largely constant for the group of species.

Empirically, Lynch et al. report an approximately linear negative relationship between the log coalescent effective population size and the log mutation rate within broad clades of species (Linnaean phyla) [32]. The slopes of the log-log relationships ranged from -0.32 to -1.53 [32][Appendix S1].

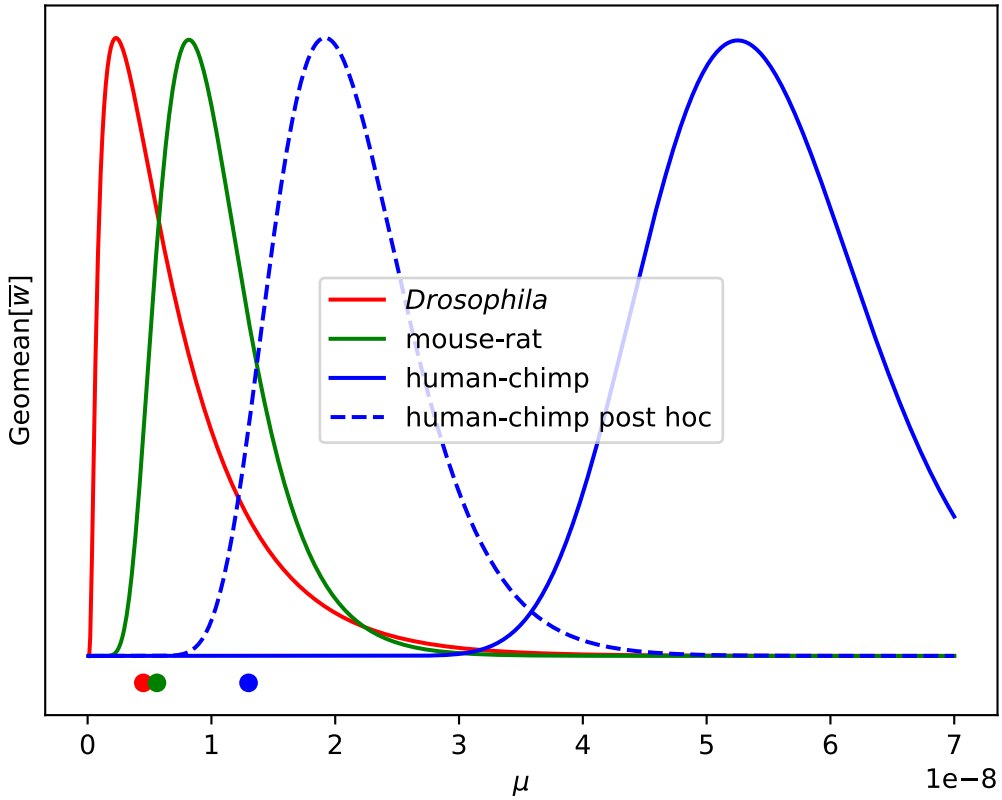

Figure 1: Population mean fitness implies reasonably narrow ranges for mutation rates. Generationally-averaged population mean fitness Geomean[$\overline{w}$] assuming a constant N_e, for a priori parameter values as a function of the mutation rate μ (solid lines), for post hoc parameter values (dashed line), and observed mutation rates (dots). The mouse-rat curve is the a priori lower bound.

Variability in Γ_s and possibly L_n between species presumably adds noise to the relationship. Lynch et al.'s finding of an inverse log-log relationship between N_e and μ counts as potential evidence for the mutation rate macroevolutionary selection hypothesis. It should be noted that the drift-barrier hypothesis also predicts an inverse relationship between N_e and μ, but does not uniquely predict the functional form or slope of the relationship [3].

Discussion

How might the mutation rate per sexual generation increase?

Proposing convergence toward a population-optimal mutation rate requires an accompanying biological mechanism for adjusting that rate. When the mutation rate is higher than optimal, microevolutionary forces can be invoked to bring it down. When the rate is lower than optimal it might appear that the only possibility is for a macroevolutionary force favoring extinction to take hold. However, this need not be the case.

One possibility is for a new species to simultaneously acquire a higher mutation rate and higher population-mean fitness through stochastic species founder effects. The scope for such a mechanism appears limited.

A more likely option appears to be through an increase in the generation length. Suppose substituting sites arise in response to changes in the external environment. Then Γ_s is given by the product of a rate per unit time, Γ'_s, which is determined by the external environment, and the generation length, g. From equation 1, the population-optimal mutation rate per sexual generation is then,

$$\mu = \sqrt{\mathrm{E}\left[\frac{3\pi_M \pi_e f_l \Gamma'_s g}{2P^2 \pi_f N_e L_n h}\right]} \tag{3}$$

For most species, microevolution favors an increase in the generation length; the longer the generation length of an organism, the greater the number of offspring. By equation 3, increasing the generation length will increase the population-optimal mutation rate. At first glance, this seems to be the opposite of what is required. The mutation rate is too low, and the population-optimal mutation rate is made larger. However, at least in some species, when the generation length increases, there is a biological mechanism that increases the germline spontaneous mutation rate. This mechanism is the stem cell cycle that leads to the production of new germ cells [44]. Thus the mutation rate increases in a more or less linear fashion as the generation length increases, whereas the population-optimal mutation rate increases at the square root of the generation length. Consequently, microevolutionary forces favoring increased generation lengths can elevate the mutation rate, and if initially low, potentially bring it closer to its population-optimal value.

The stem cell cycle occurs in men and male mice, and in both sexes in *Drosophila melanogaster* [44]. With each stem cell division there is the chance for mutations to occur as the DNA is replicated. It is hypothesized that something akin to the stem cell cycle in which mutation rates increase in proportion to the sexual generation time is widespread among sexual eukaryotes. Otherwise it

would appear that the only way for mutation rates to increase would be through stochastic species founder effects.

Conclusion

In this manuscript, the possibility that evolution drives the spontaneous mutation rate towards its population-optimal value was tested. When parameterized using independently estimated quantities, the mutation rate macroevolutionary selection hypothesis predicts the order of magnitude value of μ, an inverse relationship between μ and N_e, the log–log form of this relationship, and its approximate slope. Together, these results are compatible with the hypothesis that differential macroevolutionary selection biases the spontaneous mutation rate of surviving species toward the value that maximizes long-term population mean fitness.

Materials and methods

Assuming it exists, the variance effective population size, N_e, can be approximated by the eigenvalue effective population size, N_e^E [45]. N_e^E approaches the coalescent effective population size, N_e^{Co} as population size increases provided that N_e^{Co} exists [46]. Thus N_e can be estimated using N_e^{Co}, which is determined from the heterozygosity of neutral sites in the genome.

In order to determine π_M a simulation of the essential elements of macroevolutionary selection over time was performed. Fitness of a species was determined by the interplay between the stochastic arrival of latent sites, and their becoming substituting sites with fixed probability per unit time. At stochastic intervals a random species was cloned and replaced the then least fit species. All latent sites were assumed to share the same selection coefficient, equal to the mean value of the expected distribution of fitness effects for latent sites. This highly simplified model was intended only to estimate π_M.

The number of species occupying a niche, S_{niche}, was estimated as the average number of observed overlapping similar species (i.e., anthropoids, rodents, or drosophilids) in the wild. This value reflects both a reduction, since not all of the overlapping species are likely in competition, and an increase, on account of competition from more distantly related species.

The distribution of fitness effects for new positively selected sites was assumed to be exponential with mean $\overline{s_p}$ [47]. It differs from the distribution of fitness effects of beneficial mutations because beneficial mutations of small fitness effect are less likely to establish and fix than beneficial mutations of larger effect, and thus they will reoccur more frequently. The distribution of fitness effects of beneficial mutations is frequently studied. A mapping from the fitness effects of beneficial mutations to $\overline{s_p}$ was thus sought.

Let $\rho_b(s)$ denote the distribution of fitness effects for new beneficial mutations for selection coefficient s, and $\overline{s_b}$ be the mean selection coefficient of effective beneficial mutations. Consider a particular value for the selection coefficient of new positively selected sites, s. Since s is exponentially distributed with mean $\overline{s_p}$, and the probability of fixation is approximately $2\frac{N_e}{N}s$ where N is the census population size [48, 49],

$$\frac{1}{\overline{s_p}} \exp\left(\frac{-s}{\overline{s_p}}\right) \approx \frac{2\frac{N_e}{N} s\rho_b(s)}{\int_0^\infty 2\frac{N_e}{N} s\rho_b(s)ds}$$

which implies that,

$$\rho_b(s) \approx \frac{\exp\left(\frac{-s}{\overline{s_p}}\right)}{s}$$

For positively selected sites with beneficial alleles having selection coefficients less than $\frac{1}{PN_e}$ drift dominates over selection, and a beneficial mutation at the site is unlikely to contribute appreciably to the latent load, nor to the mutational load. Thus the mean selection coefficient for effective beneficial mutations is approximately given by,

$$\begin{aligned} \overline{s_b} &\approx \frac{\int_{\frac{1}{PN_e}}^\infty s\rho_b(s)ds}{\int_{\frac{1}{PN_e}}^\infty \rho_b(s)ds} \\ &= \frac{\exp\left(\frac{-1}{PN_e\overline{s_p}}\right)}{-\mathrm{Ei}\left(\frac{-1}{PN_e\overline{s_p}}\right)}\overline{s_p} \end{aligned} \tag{4}$$

where Ei is the exponential integral,

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

Equation 4 can then be solved for $\overline{s_p}$ using standard numerical root finding techniques.

h for deleterious alleles is typically around 0.2 or 0.3 [50], suggesting a value of h for beneficial alleles equal to 1 minus this value, or around 0.8.

Data availability

The code and data used in this study are publicly available as File S1.

File S1 - Population-optimal mutation rate software and results.
`https://doi.org/10.5281/zenodo.18111881`

Acknowledgments

I am grateful to Michael Lynch for the time he spent reviewing an early version of this manuscript and providing valuable feedback.

Conflict of interest statement

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

References

[1] 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.

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

[3] 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.

[4] Gordon Irlam. An integrated model of the effects on fitness of positively selected, conserved, and nearly neutral sites in sexual populations and the population-optimal mutation rate. *BioRxiv*, 2024. `doi:10.1101/2023.06.26.546616`.

[5] Motoo Kimura. On the evolutionary adjustment of spontaneous mutation rates. *Genetics Research*, 9(1):23–34, 1967.

[6] Nick GC Smith and Adam Eyre-Walker. Adaptive protein evolution in Drosophila. *Nature*, 415(6875):1022–1024, 2002.

[7] Ying Zhen, Christian D Huber, Robert W Davies, and Kirk E Lohmueller. Greater strength of selection and higher proportion of beneficial amino acid changing mutations in humans compared with mice and Drosophila melanogaster. *Genome Research*, 31(1):110–120, 2021.

[8] David J Begun, Alisha K Holloway, Kristian Stevens, LaDeana W Hillier, Yu-Ping Poh, Matthew W Hahn, Phillip M Nista, Corbin D Jones, Andrew D Kern, Colin N Dewey, et al. Population genomics: whole-genome analysis of polymorphism and divergence in Drosophila simulans. *PLOS Biology*, 5(11):e310, 2007.

[9] The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, 437(7055):69–87, 2005.

[10] Gordon Irlam. Evidence of macroevolutionary selection of the mutation rate: Supplement 1, 2025. `doi:10.5281/zenodo.18111881`.

[11] Robert Friedman and Austin L Hughes. The pattern of nucleotide difference at individual codons among mouse, rat, and human. *Molecular Biology and Evolution*, 22(5):1285–1289, 2005.

[12] Mauno Vihinen. When a synonymous variant is nonsynonymous. *Genes*, 13(8):1485, 2022.

[13] Peter D Keightley, Martin J Lercher, and Adam Eyre-Walker. Evidence for widespread degradation of gene control regions in hominid genomes. *PLOS Biology*, 3(2):e42, 2005.

[14] Richard J Wang, Samer I Al-Saffar, Jeffrey Rogers, and Matthew W Hahn. Human generation times across the past 250,000 years. *Science Advances*, 9(1):eabm7047, 2023.

[15] Kevin E Langergraber, Kay Prüfer, Carolyn Rowney, Christophe Boesch, Catherine Crockford, Katie Fawcett, Eiji Inoue, Miho Inoue-Muruyama, John C Mitani, Martin N Muller, et al. Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great

ape and human evolution. *Proceedings of the National Academy of Sciences*, 109(39):15716–15721, 2012.

[16] Michael W Nachman and Jeremy B Searle. Why is the house mouse karyotype so variable? *Trends in Ecology & Evolution*, 10(10):397–402, 1995.

[17] Eva E Deinum, Daniel L Halligan, Rob W Ness, Yao-Hua Zhang, Lin Cong, Jian-Xu Zhang, and Peter D Keightley. Recent evolution in Rattus norvegicus is shaped by declining effective population size. *Molecular Biology and Evolution*, 32(10):2547–2558, 2015.

[18] Jeremy D Lange, Héloïse Bastide, Justin B Lack, and John E Pool. A population genomic assessment of three decades of evolution in a natural Drosophila population. *Molecular Biology and Evolution*, 39(2):msab368, 2022.

[19] Peter Kriesner, Ary A Hoffmann, Siu F Lee, Michael Turelli, and Andrew R Weeks. Rapid sequential spread of two Wolbachia variants in Drosophila simulans. *PLOS Pathogens*, 9(9):e1003607, 2013.

[20] Sudhir Kumar, Glen Stecher, Michael Suleski, and S Blair Hedges. TimeTree: a resource for timelines, timetrees, and divergence times. *Molecular Biology and Evolution*, 34(7):1812–1819, 2017.

[21] Heng Li and Richard Durbin. Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357):493–496, 2011.

[22] John G Fleagle, Charles Janson, and Kaye Reed. *Primate Communities*. Cambridge University Press, 1999.

[23] Renan Maestri and Bruce D Patterson. Patterns of species richness and turnover for the South American rodent fauna. *PLOS One*, 11(3):e0151895, 2016.

[24] David A Grimaldi and Courtney Richenbacher. Exceptional species diversity of Drosophilidae (Diptera) in a Neotropical forest. *American Museum Novitates*, 2023(3997):1–28, 2023.

[25] J John Sepkoski Jr. Rates of speciation in the fossil record. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 353(1366):315–326, 1998.

[26] David Castellano, Moisès Coll Macià, Paula Tataru, Thomas Bataillon, and Kasper Munch. Comparison of the full distribution of fitness effects of new amino acid mutations across great apes. *Genetics*, 213(3):953–966, 2019.

[27] Tom R Booker and Peter D Keightley. Understanding the factors that shape patterns of nucleotide diversity in the house mouse genome. *Molecular Biology and Evolution*, 35(12):2971–2988, 2018.

[28] Peter D Keightley, José L Campos, Tom R Booker, and Brian Charlesworth. Inferring the frequency spectrum of derived variants to quantify adaptive molecular evolution in protein-coding genes of Drosophila melanogaster. *Genetics*, 203(2):975–984, 2016.

[29] Tom R Booker. Inferring parameters of the distribution of fitness effects of new mutations when beneficial mutations are strongly advantageous and rare. *G3: Genes, Genomes, Genetics*, 10(7):2317–2326, 2020.

[30] Chris P Ponting and Ross C Hardison. What fraction of the human genome is functional? *Genome Research*, 21(11):1769–1776, 2011.

[31] Guy Sella, Dmitri A Petrov, Molly Przeworski, and Peter Andolfatto. Pervasive natural selection in the Drosophila genome? *PLOS Genetics*, 5(6):e1000495, 2009.

[32] Michael Lynch, Farhan Ali, Tongtong Lin, Yaohai Wang, Jiahao Ni, and Hongan Long. The divergence of mutation rates and spectra across the Tree of Life. *EMBO Reports*, 24(10):e57561, 2023.

[33] Nicolas Galtier. Adaptive protein evolution in animals and the effective population size hypothesis. *PLOS Genetics*, 12(1):e1005774, 2016.

[34] Adam R Boyko, Scott H Williamson, Amit R Indap, Jeremiah D Degenhardt, Ryan D Hernandez, Kirk E Lohmueller, Mark D Adams, Steffen Schmidt, John J Sninsky, Shamil R Sunyaev, et al. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLOS Genetics*, 4(5):e1000083, 2008.

[35] Lawrence H Uricchio, Dmitri A Petrov, and David Enard. Exploiting selection at linked sites to infer the rate and strength of adaptation. *Nature Ecology & Evolution*, 3(6):977–984, 2019.

[36] Jun Gojobori, Hua Tang, Joshua M Akey, and Chung-I Wu. Adaptive evolution in humans revealed by the negative correlation between the polymorphism and fixation phases of evolution. *Proceedings of the National Academy of Sciences*, 104(10):3907–3912, 2007.

[37] Magdalena Gayà-Vidal and M Mar Albà. Uncovering adaptive evolution in the human lineage. *BMC Genomics*, 15(1):599, 2014.

[38] Adam Eyre-Walker and Peter D Keightley. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Molecular Biology and Evolution*, 26(9):2097–2108, 2009.

[39] Alexander Cagan, Christoph Theunert, Hafid Laayouni, Gabriel Santpere, Marc Pybus, Ferran Casals, Kay Prüfer, Arcadi Navarro, Tomas Marques-Bonet, Jaume Bertranpetit, et al. Natural selection in the great apes. *Molecular Biology and Evolution*, 33(12):3268–3283, 2016.

[40] Carlos G Schrago. The limiting distribution of the effective population size of the ancestor of humans and chimpanzees. *Journal of Theoretical Biology*, 357:55–61, 2014.

[41] Manas Joshi, Pablo Duchen, Adamandia Kapopoulou, and Stefan Laurent. Natural selection in transcription factor–DNA interaction motifs: a comparative and population genomics perspective. *Genome Biology and Evolution*, 17(11):evaf212, 2025.

[42] Patrick M O'Grady and Rob DeSalle. Phylogeny of the genus Drosophila. *Genetics*, 209(1):1–25, 2018.

[43] James Bonacum, Patrick M O'Grady, Michael Kambysellis, and Rob DeSalle. Phylogeny and age of diversification of the planitibia species group of the Hawaiian Drosophila. *Molecular Phylogenetics and Evolution*, 37(1):73–82, 2005.

[44] Joni B Drost and William R Lee. Biological basis of germline mutation: comparisons of spontaneous germline mutation rates among drosophila, mouse, and human. *Environmental and Molecular Mutagenesis*, 25(S2):48–64, 1995.

[45] WJ Ewens. On the concept of the effective population size. *Theoretical Population Biology*, 21(3):373–378, 1982.

[46] Ola Hössjer. On the eigenvalue effective size of structured populations. *Journal of Mathematical Biology*, 71(3):595–646, 2015.

[47] H Allen Orr. The distribution of fitness effects among beneficial mutations. *Genetics*, 163(4):1519–1526, 2003.

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

[49] Zaheerabbas Patwa and Lindi M Wahl. The fixation probability of beneficial mutations. *Journal of The Royal Society Interface*, 5(28):1279–1289, 2008.

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