The Rate of Observable Molecular Evolution When Mutation May Not Be Weak

A.P. Jason de Koning[1,*] and Bianca D. De Sanctis[1,2]

[1]*Department of Biochemistry and Molecular Biology, Department of Medical Genetics, University of Calgary, Cumming School of Medicine and Alberta Children's Hospital Research Institute. Calgary, Alberta. Canada.*

[2]*Current address: Graduate Program in Computational Biology, Department of Applied Mathematics and Theoretical Physics, University of Cambridge. Centre for Mathematical Sciences, Wilberforce Rd, Cambridge, UK, CB3 0WA*

[*] *Corresponding author: jason.dekoning@ucalgary.ca*

1 Abstract

One of the most fundamental rules of molecular evolution is that the rate of neutral evolution equals the mutation rate and is independent of effective population size[1–4]. This result lies at the heart of the Neutral Theory, and is the basis for numerous analytic approaches that are widely applied to infer the action of natural selection across the genome and through time[5–17], and for dating divergence events using the molecular clock[18,19]. However, this result was derived under the assumption that evolution is strongly mutation-limited[3,4,20], and it has not been known whether it generalizes across the range of mutation pressures or the spectrum of mutation types observed in natural populations. Validated by both simulations and exact computational analyses, we present a direct and transparent theoretical analysis of the Wright-Fisher model of population genetics, which shows that some of the most important rules of molecular evolution are fundamentally changed by considering recurrent mutation's full effect. Surprisingly, the rate of the neutral molecular clock is found to have population-size dependence and to not be equal to the mutation rate in general. This is because, for increasing population mutation rates (θ), the time spent waiting for mutations quickly becomes smaller than the cumulative time mutants spend segregating before a fixation, resulting in a net deceleration compared to classical theory that depends on the population mutation rate. Furthermore, selection exacerbates this effect such that more adaptive alleles experience a greater deceleration than less adaptive alleles, introducing systematic bias in a wide variety

of methods for inferring the strength and direction of natural selection from across-species sequence comparisons. Critically, the classical weak mutation approximation performs well only when $\theta < 0.1$, a threshold that many biological populations seem to exceed.

Classical population genetic theory was largely based on the assumption that mutation is "weak" and can therefore be conveniently ignored. Although it has been fairly clear that this assumption is often reasonable, estimates of the population mutation rate (θ; see Materials and Methods) across many organisms have only become available recently in the genomics era, and it has been unclear what values of θ might violate weak mutation. A variety of examples are now known of populations, and mutation types, where mutation might very well not be weak. These include hyperdiverse eukaryotes[21], many prokaryotes[22,23], and a variety of rapidly evolving viruses (including HIV[24–26]). Similarly, mutation types with fast natural rates such as some context-dependent nucleotide mutations in the nuclear genomes of mammals[?], mutations in the mitochondrial genomes of some vertebrates[28], microsatellite and simple sequence repeat polymorphisms[29], somatic mutations[30], and heritable epigenetic changes[31] all occur at rates fast enough that it is reasonable to question the correctness of the weak mutation assumptions upon which most analytic approaches rely. Furthermore, as has been recently pointed out[32], the apparent ubiquity of "soft sweeps" in nature[33–35], where adaptive mutations appear to have multiple origins by recurrent mutation or immigration, has been interpreted as supporting the idea that θ in some populations may be significantly larger than is widely believed. It is therefore critical to determine how the implications of classical population genetic theory might change under the degrees of mutation pressure observed in natural populations[31–34,36].

Here we explore the impact of general, recurrent mutation processes on the rate of molecular evolution. The rate of evolution is among the most fundamental and useful quantities in all of evolutionary genetics, and is the basis for analytic approaches used widely in the study of molecular evolution, genome evolution, population genetics, phylogenetics, and related fields. Despite its central importance to understanding patterns of genomic sequence variations and their causes, remarkably, no explicit and complete derivation has appeared in the literature. Kimura[1], like Wright[37], seems to have simply written down the standard equation from intuition, and it has since become second nature to population geneticists. However, this rate of evolution depends upon several important implicit assumptions that appear to not be widely appreciated. Here we make those assumptions explicit, and after doing so, show that it is surprisingly easy and valuable to generalize the *rate of observable evolution* at individual genomic positions with respect to mutation.

Sequence evolution is often considered from two rather different perspectives. When making across-species sequence comparisons, molecular evolution is typically modelled at individual genomic positions following a phylogenetic *substitution process*, wherein recurrent mutations occur over long timescales and evolution proceeds by successive substitution events, at the same position, along diverging lineages. At the population genetic level, each substitution corresponds to the turnover of the entire population for a new allelic state. It has long been considered essential to account for the possibility of serial substitutions at the same positions[38,39]. Indeed, standard phylogenetic likelihood computations allow for an infinity of unobservable serial substitutions, sometimes represented as a distribution of substitution histories (reviewed in Ref.[41]). This is the canonical approach that is implicitly applied when using continuous-time Markov chain models of sequence evolution.

Working from a rather different perspective, Kimura was among the first to consider how the rate of molecular evolution could be approximated in terms of the underlying population genetic processes that generate substitutions[1,2,37] (also see Bustamante Ref.[40] for an excellent review). In particular, he considered the rate of long-term evolution by fixation under free recombination, no epistasis, and weak mutation. We will refer to this quantity as the *weak-mutation rate of evolution* (defined and discussed in detail in the next section). Importantly, although Kimura justified his approach by referring to the assumptions of the infinite sites model, he never appears to have claimed that his approach requires it (Ref.[4], p. 46). Although the infinite sites assumption should generally preclude the application of Kimura's model to substitution processes at individual positions, in practice his model and its insights are widely applied to individual sites. For example, they are applied implicitly in d_N/d_S approaches for inferring the strength and direction of natural selection[7–12], and explicitly in methods for inferring population-genetic parameters from across-species comparisons[13–17]. As we discuss in detail below, this apparent contradiction is explained by noting that Kimura's derivation actually does not require infinite sites *per se*, but rather it requires some specific, related assumptions about the weakness of mutation.

Before we derive the rate of evolution under general mutation, some potentially counter-intuitive ideas must first be introduced. Owing to variation in definitions of a fixation, substitution and fixation may or may not correspond. Fixation is most often defined as the takeover of the entire population by a single mutant lineage. This is a convenient definition because it has an inverse correspondence to coalescence. However, in standard models of population genetics, such as the Wright-Fisher model, fixation is usually defined as simply reaching 100% frequency for the mutant *state*. These definitions coincide solely when mutation is unnaturally disallowed in the instantaneous generator of the underlying model, so that only a single lineage of mutants is permitted to exist in the population at a time. However, even over short timescales, this can be highly unrealistic. For example, neutral mutations persist on average for $4N_e$ generations in diploid populations, where N_e is the effective population size and may differ from the census population size, N. Since approximately $2Nv$ mutations are expected to arise in each generation, for forward mutation rate v, the number of additional mutations expected in the population during an average neutral fixation trajectory exceeds 1 when the mutation rate is as small as $v = 1/(8NN_e)$, or $1/(8N^2)$ when $N_e = N$. Thus, even for exceedingly small mutation rates, it is plausible that multiple lineages of the same variant commonly arise simultaneously. When population mutation rates are relatively large, it is substantially more likely that this will occur (e.g., by soft selective sweeps[33]). This is important because the standard, weak-mutation rate of evolution considers only the fixation of lineages with single mutational origins. However, especially when making across-species sequence comparisons, we generally assume that so many generations have elapsed that positions evolved independently via effectively free recombination. We therefore can not tell, and should not necessarily care, whether an apparent substitution had single or multiple mutational origins when measuring the rate of long-term evolution. Therefore, the *rate of observable evolution* in sequence comparisons must correspond to the rate of substitution by either single *or* multiple mutational origins.

The rate of evolution under weak mutation Kimura[1,4], and King and Jukes[2], building on earlier work by Wright[37], first showed that the rate of neutral evolution is expected to equal the mutation rate. To obtain this result, they started with an expression for the diploid *weak-mutation rate of evolution*,

$$k_{\text{weak}} = 2N\mu \cdot P_{\text{Fix}} \tag{1}$$

where N is the number of reproducing individuals, μ the mutation rate per locus per generation, and P_{Fix} the probability that a mutation will eventually go to fixation. Consistent with common practices described above, we define a locus as an individual genomic position or site. Since the probability of fixation for neutral mutations is $1/(2N)$ under weak mutation assumptions (discussed below), the *weak-mutation rate of neutral evolution* is

$$k_{\text{weak}} = 2N\mu \cdot \frac{1}{2N} = \mu$$

This result is a cornerstone principle of the Neutral Theory of Molecular Evolution[4] and is deeply embedded in our thinking about the relationship between population genetics and molecular evolution. Indeed, it has been called "one of the most elegant and widely applied results in population genetics"[42]. Although Kimura described this equation in the context of the infinite sites model (Ref.[4], p. 46), it is just as consistent when interpreted in a finite-sites context where mutation is assumed to be weak. Indeed, as we argued above, it is this context in which equation 1 is usually applied. However, as we will show, relaxing the weak mutation assumptions used to derive this result leads to a different rate of evolution, which can have strikingly different characteristics.

Several assumptions about weak mutation are implied by equation 1. These are: 1) that mutations arise and go to their fates one by one, so that only one segregating lineage of mutations may exist in a population at a given time; 2) that evolution is fundamentally mutation-limited, so that the timescale of mutation dominates over segregation times[43]; and 3) that mutations originate in a single individual at a time (i.e., the initial number of mutant alleles is $p = 1$). Recent work on the effects of arbitrarily fast mutation[31] has retained these assumptions, perhaps due to the perception that they are needed for analytic tractability. Contrariwise, we will first show how the first two assumptions can be easily relaxed without approximation, and will turn our attention to the third in the supplementary methods (SI Methods 1.1). For generality, we assume a biallelic locus undergoing recurrent bidirectional mutation. Because we are interested in when the mutant *state* takes over the population (e.g., by either a hard or a soft sweep), we do not distinguish between individuals who are identical by state or identical by descent. Importantly, this means that the usual inverse correspondence between fixation and coalescence is lost (see Materials and Methods for full details).

2 Results

The rate of observable evolution. Following Kimura[4], we define the rate of evolution as one over the mean time between substitutions. The rate is thus measured in expected substitutions per generation. In the finite sites context, this refers to the rate of substitution of different allelic states at the same position. Guess and Ewens[44] referred to this as the rate of “quasifixation”, however, we will avoid this term since others have used it to mean something different. From this point forward, we will use the term fixation to refer to attaining a population frequency of 100% for the mutant state (by either single or multiple origins). Because fixations are rare even for advantageous mutations[43], for every mutation that arises and becomes fixed, we expect many more mutations to have arisen and gone extinct. We call these mutation-fixation (MF) and mutation-extinction (ME) cycles respectively (or mutation-absorption cycles, in the general case). The mean time between fixations can then be written as a function of the expected number of cycles and their respective lengths,

$$\begin{aligned} k &= \frac{1}{N_{\text{ME}} \cdot T_{\text{ME}} + N_{\text{MF}} \cdot T_{\text{MF}}} \\ &= \frac{1}{N_{\text{ME}} \cdot (T_\mu + T^*_{\text{Ext}}) + 1 \cdot (T_\mu + T^*_{\text{Fix}})} \end{aligned} \tag{2}$$

where $N_{M\cdot}$ and $T_{M\cdot}$ denote the mean number of cycles and the mean length of cycles, respectively. T_μ, T^*_{Ext}, and T^*_{Fix} are defined as the mean time in numbers of generations to get a mutation (or mutations), the mean time to extinction (calculated including the effect of bidirectional mutation, as indicated by the '*'), and the mean time to fixation (also including mutation).

Since P^*_{Fix} represents the probability that an absorption is a fixation, $1/P^*_{\text{Fix}}$ is the expected number of absorptions to get a fixation. Since one of these will be a mutation-fixation cycle, $(1/P^*_{\text{Fix}}) - 1$ of these are expected to be mutation-extinction cycles. We can therefore write

$$
\begin{aligned}
k &= \frac{1}{(\frac{1}{P^*_{\text{Fix}}} - 1) \cdot (T_\mu + T^*_{\text{Ext}}) + (T_\mu + T^*_{\text{Fix}})} \\
&= \frac{P^*_{\text{Fix}}}{T_\mu + \big(T^*_{\text{Ext}}(1 - P^*_{\text{Fix}}) + T^*_{\text{Fix}} P^*_{\text{Fix}}\big)} \\
&= \frac{P^*_{\text{Fix}}}{T_\mu + T^*_{\text{Abs}}} \qquad (3)
\end{aligned}
$$

where T^*_{Abs} is the unconditional time to absorption allowing for bidirectional mutation. We will refer to equation 3 as the *rate of observable evolution*. Notably, by reintroducing Kimura's assumptions into equation 3, this expression becomes equal to the weak-mutation rate of evolution (equation 1; SI Methods 1.2). We also provide a more formal derivation of equation 3 in the supplement (SI Methods 1.3), and show how it can be integrated over an initial distribution, $f(p)$ (Supp. Meth. 1.1), thus relaxing the third weak mutation assumption that $p = 1$. All subsequent results are integrated over $f(p)$.

Unlike the weak-mutation rate of evolution, equation 3 makes no assumptions about the strength of mutation and indeed introduces no additional assumptions beyond those of the model of population genetics used to calculate its component quantities. Dominance, selection, and other forces may thus be easily considered by including their effects in the underlying model. Importantly, simply incorporating mutation into the probability of fixation in equation 1 does not work without also including the absorption times (Fig. S1). Although simple closed-form expressions for the component quantities are not available in general, they can be easily calculated using efficient computational techniques we recently described[45,46].

For validation, we developed a direct way to compute the time between fixations, without requiring any of the above theory. This direct approach uses a modified Wright-Fisher model, where the extinction state is treated as transient rather than absorbing, and the population is initialized with $p = 0$ mutants. This allows the time between fixations to be directly calculated as the expected time to absorption, using standard absorbing Markov chain theory[45] (see Methods). A similar approach can be used to directly calculate the variance of the time between fixations, which is useful for testing hypotheses about the dispersion of the molecular clock (Fig. S3). When equation 3 is integrated over $f(p)$, it numerically agrees with the direct approach (see Materials and Methods; Fig. S2).

Implications and observation of non-classical phenomena. Based on equation 3, when mutation is weak ($T_\mu \gg T^*_{\text{Abs}}$), the time between fixations should be dominated by the time spent waiting for mutations (Fig. 1A; as first recognized by Kimura[4]). However, when mutation is not weak (Fig. 1B), the cumulative time mutants spend segregating in the population before a fixation can become significant, causing the weak mutation model to underestimate the time between fixations and therefore to overestimate the substitution rate. This effect can be directly observed by numerically comparing the weak-mutation sub-

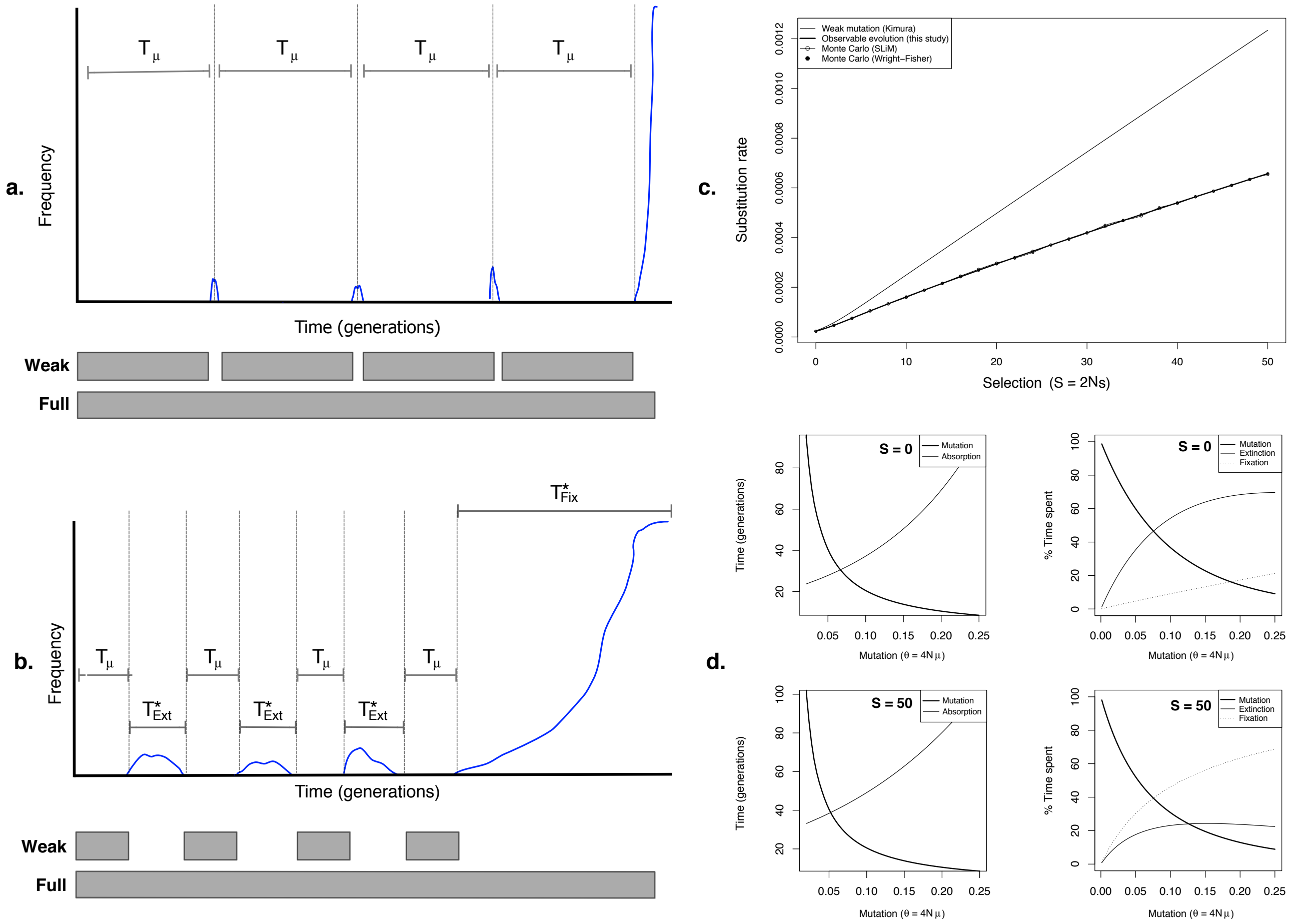

Figure 1: **The rate of evolution when mutation may not be weak. A.** When mutation is weak, the time spent waiting for mutations dominates the time between fixations. **B.** When mutation is not weak, the segregation times also become important. **C.** Demonstration and validation of a deceleration in the rate of evolution under the Wright-Fisher model ($\theta = 0.1$, $h = 0.5$, $u = 0$; where u is the backward mutation rate). Monte Carlo simulations measured the average number of fixations per generation for a large number of absorptions (SLiM[47]: 5M generations, averaged over 5 runs; Wright-Fisher: 10M absorptions, averaged over 100 runs). Note that both simulations and the rate of observable evolution produced nearly identical results, which are fully overlapping. At the origin ($S = 0$), the weak-mutation and observable rates of evolution differ by about 10%. D. Left: Mean time to mutation (T_μ) and absorption (T^*_{Abs}). Right: Fraction of the time between fixations spent waiting for mutations, extinctions, and fixations. As explained in the main text, extinction and fixation do not necessarily refer to the fate of a particular lineage of segregating mutants, but to the frequency of the mutant state.

stitution rate and the observable substitution rate under Wright-Fisher assumptions (Fig. 1C), where the overestimation by the weak mutation model is found to be exaggerated by increasing positive selection. This exaggeration is especially concerning because it implies that methods comparing the rates of substitution for neutral and non-neutral changes (e.g., Refs.[5,8,12,15]) are subject to systematic error without correcting for the deceleration effect. These results were replicated and validated in two types of simulations (Fig. 1C; also see Fig. S4, and Materials and Methods).

To provide some intuition about the deceleration effect, the mean time spent waiting for a mutation or an absorption is shown in Fig. 1D (left) for increasing population mutation rates. Remarkably, the mean time per absorption for neutral variants overtakes the mean time to a mutation when θ is as small as 0.07, strongly violating Kimura's second weak mutation assumption (see Fig. S5 for a larger parameter range). When θ exceeds this value, the rate of neutral evolution is dominated by the time mutants spend segregating in the population prior to a mutation-fixation cycle (Fig. 1D, upper right, "Extinctions"; also see Fig. S6). For selected variants, T_μ is overtaken by T^*_{Abs} at even smaller values of θ (e.g., $\theta = 0.05$, Fig. 1D, bottom left). Interestingly, when positive selection is strong, the rate of evolution is dominated by the time it takes for a mutant, once arisen, to go to fixation (Fig. 1D lower right, "fixation"; also see Fig. S6). These observations conspicuously contradict Kimura's oft-repeated claim that the time it takes for variants to reach their fates does not

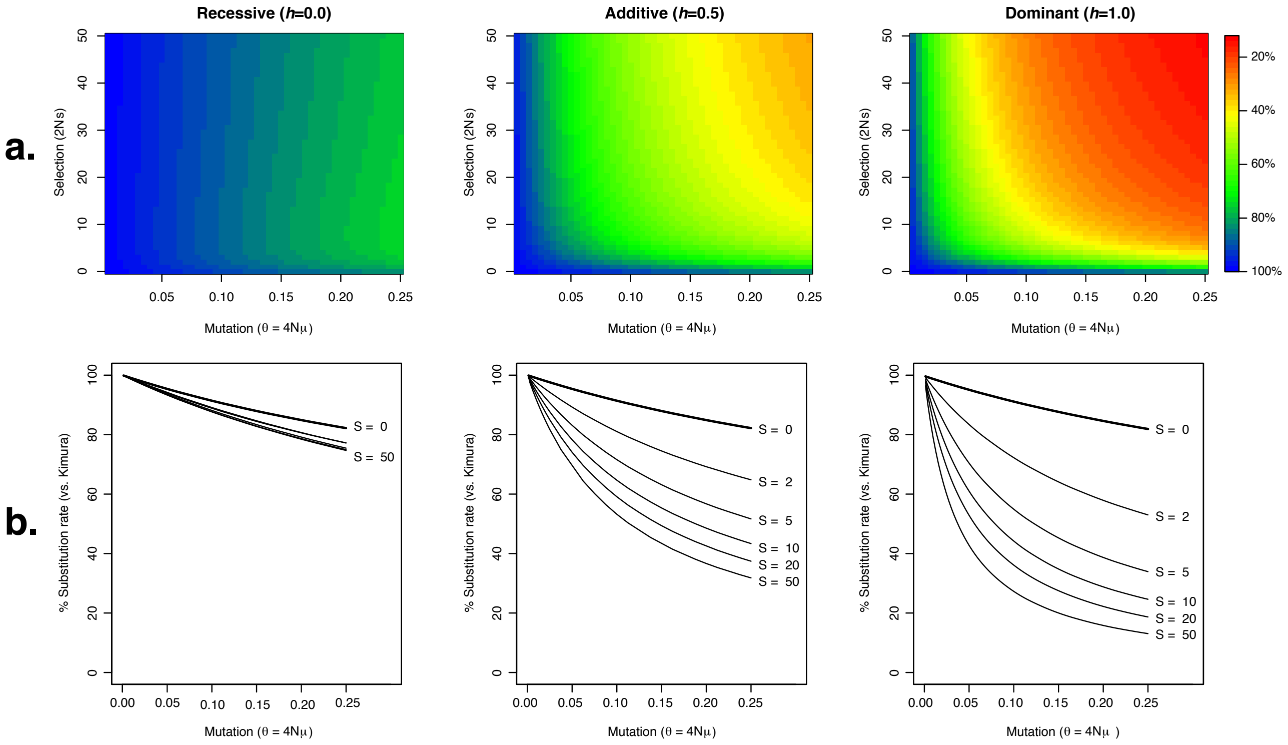

Figure 2: **Effect of mutation and selection on the rate of evolution.** The rate of observable evolution is displayed as a fraction of Kimura's weak-mutation rate of evolution. **A.** Overall effect of mutation, selection, and dominance across a fine grid. **B.** Specific effect of mutation, selection, and dominance across a range of biologically relevant values.

affect the rate of evolution[4,43]. They also clarify that this claim was justified only by the assumption that mutation is always weak.

Examination of the joint effect of mutation, selection, and dominance on the two rates of evolution shows that the deceleration effect in Fig. 1C becomes exaggerated as mutation, selection and dominance are increased (Fig. 2A). Color corresponds to the rate of observable evolution as a percentage of the weak-mutation rate of evolution, including dominance where appropriate. As can be more clearly observed in Fig. 2B, for modestly high but biologically realistic values of θ (e.g., $\theta \geq 0.1$), a deceleration in the neutral molecular clock ($S = 0$) is predicted when compared to the weak mutation model. In fact, the rate of neutral evolution becomes dependent on the population size such that greater decelerations are observed as population mutation rates are increased. This surprising result means that even for strictly neutral mutations, the rate of the molecular clock is expected to be erratic over time if any lineages grow large enough in terms of their population mutation rates. It should also be noticed that nearly neutral variants ($S = 2$) experience a significantly larger deceleration than do neutral variants, which may be consequential for the molecular clock when functional or constrained sequences are used for dating divergence events. These results thus have the potential to help resolve some persistent paradoxes, for instance, where mutation rates from pedigrees and phylogenetic substitution rates unexpectedly differ[18].

The relationship between d_N/d_S (ω) and selection unexpectedly depends on the population mutation rate. The ratio of non-synonymous to synonymous substitution rates, d_N/d_S, is widely used to measure the strength and direction of selection in protein-coding genes. Evidence of heterogeneity in d_N/d_S across the lineages of a phylogeny is usually interpreted as evidence of fluctuating or episodically varying selective constraints over time. Contrariwise, we find that when mutation is not weak, adaptive substitutions are

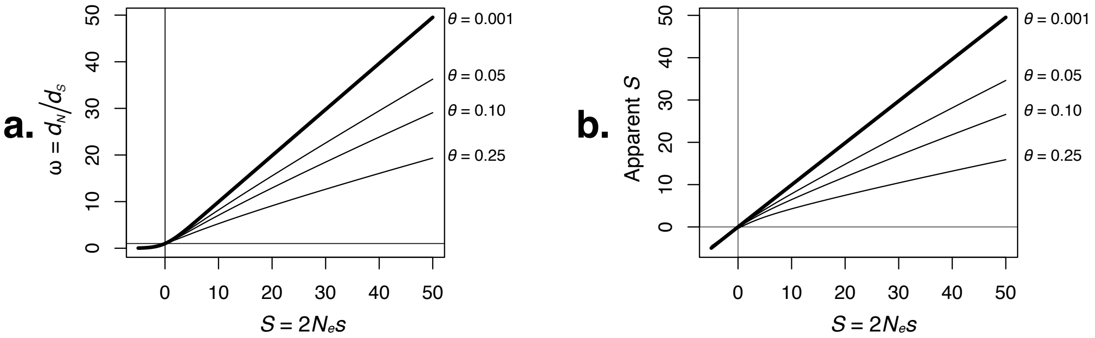

Figure 3: **Effect of the weak mutation approximation on inferences of the strength and direction of natural selection. A.** The relationship between d_N/d_S and selection is modified by the population mutation rate. The true substitution rate was calculated using equation 3 for both d_N and d_S (given the true value of S). **B.** S inferred under weak mutation assumptions systematically underestimates the true strength of adaptive evolution, particularly for large values of θ. Similarly, neutral evolution increasingly appears as weak negative selection for large θ.

expected to have different true d_N/d_S values under different population mutation rates, even while the population-scaled selection coefficient is held constant (Figure 3A). When θ is small, d_N/d_S increases approximately linearly with population-scaled selection coefficients, with a slope close to 1 for adaptive substitutions (Figure 3A). However, for larger values of θ, the slope of this relationship decreases substantially. Consequently, d_N/d_S, has different meanings with respect to the actual strength of selection in populations with different population mutation rates. This phenomenon has the potential to make comparison of d_N/d_S between species problematic whenever θ is large and varies between lineages. It should be noted that while these results may appear superficially similar to previously published results on the population genetics of d_N/d_S when unfixed polymorphisms are used to approximate divergence[42], they are in fact unrelated.

Inferred selection coefficients from phylogenetic data become increasingly biased for larger population mutation rates. Across-species population genetics approaches have recently become an important way to make inferences about the relative fitness of different sequence states at the same position[13–17]. All such approaches to date use equation 1 and thus assume weak mutation. To determine how the weak mutation approximation might effect inferences about selection when population mutation rates are not small, we numerically solved for the selection coefficient in equation 1 that best approximates the true rate of evolution (using equation 3), across a range of true selection coefficients and population mutation rates. Weak mutation approximations lead to systematic underestimation of the strength of selection, increasingly for larger values of the population mutation rate (Figure 3B). Importantly, when $S = 0$, inferences of S, $\hat{S}$ (or apparent S), suggest increasingly strong negative selection for increasing population mutation rates. For example, for neutral evolution ($S = 0$) and $\theta = 0.1$, $\hat{S}$ under weak mutation assumptions is -0.18. Similarly for $\theta = 0.25$, $\hat{S} = -0.38$ under weak mutation assumptions. Such findings have the potential to erroneously imply weak constraint in its absence, or saturation of substitutions. A similar effect is observed for adaptive substitutions, where weak mutation approaches lead to systematically underestimated selection coefficients. For example, at $\theta = 0.1$, the strength of strong positive selection ($S = 50$) is inferred using weak mutation calculations as $\hat{S} = 26.6$ (53.2% of actual). When theta is larger, $\theta = 0.25$, the underestimation is even more extreme, with $\hat{S} = 15.9$ (32% of actual). To overcome these problems, the rate of observable evolution could simply be used instead of the weak-mutation rate of evolution. More development, however, would be required to properly account for mutation to a finite number of alternate alleles in a fully multi-allelic framework (e.g., Ref.[48])

3 Discussion

Classical population genetic theory is a source of deep insight into a variety of critical problems in the post-genomic era. However, classical results were built upon assumptions that may now be questioned in the light of recently accumulated knowledge. Many species seem to have population mutation rates substantially below $\theta = 0.1$, and in these cases, the use of the classical weak mutation assumption appears unproblematic. Nevertheless, it is not uncommon for estimates of θ to exceed 0.1, as is true in some hyperdiverse eukaryotes[21] ($\hat{\theta} \approx 0.1 - 0.15$), many prokaryotes ($\hat{\theta} \approx 0.15 - 0.53$; e.g., *Helicobacter pylori*[23], *Salmonella enterica*[23], *Pseudomonas syringae*[22]), pathogens including Plasmodium[49] (a protist; $\hat{\theta} \approx 0.02 - 0.122$), and HIV-1[24–26] ($\theta > 1$). Furthermore, since most such estimates are made using the infinite-sites assumption, which forbids recurrent mutation, they are expected to be conservative and to thus reflect an underestimation of population mutation rates in general. The population mutation rate can be similarly large, or larger, for mutation types with faster natural rates, such as sequence transitions in some organellular genomes, microsatellite mutations, simple repeat polymorphisms, and epigenetic variations. Even in vertebrate mitochondrial DNA, which is not a particularly extreme example, there is evidence of θ falling in the range of $0.1 - 0.3$ in many species[50]. Our findings suggest that in such cases, ignoring the full effect of mutation is unwise.

Our general approach is not without precedent. During the preparation of this manuscript, we became aware of the largely overlooked study by Guess and Ewens [44], which aimed to determine the effect of mutation on the rate of evolution using an infinite alleles approach. That approach was criticized by Kimura (p. 47 of Ref.[4]), owing to the inappropriateness of modelling substitution processes with infinite alleles. Nevertheless, the spirit of their approach is captured by including mutation in the transition matrix of the underlying model as we have done, and by then approximating the time between fixations as $1/P^*_{\text{Fix}} \cdot T_\mu + T^*_{\text{Fix}}$, which follows from their equation 22. Note that this approach differs from the rate of observable evolution by ignoring the expected time it takes for each mutation that is destined for extinction to go extinct. Even though extinctions generally happen quickly, cumulatively, this time can be very large per expected fixation (e.g., Fig. 1D, right). As shown in Fig. S1, ignoring this fact leads to wild overestimation of the rate of evolution (when the Guess and Ewens model is implemented as described above, in a biallelic context with forward mutation). It is somewhat remarkable that Guess and Ewens also concluded that the weak-mutation rate of evolution overestimates the true rate when $\theta > 0.1$. They reached this conclusion, however, for a totally different reason than we did. In their calculations, they assumed a piecewise biallelic model to approximate infinite alleles. In the biallelic context, however, this model had no forward mutation and included only back mutation (their equation 1). The deceleration in the rate of evolution they inferred is therefore likely a trivial consequence of the slowdown expected by back mutation making it harder for an initial allele to escape extinction. We therefore believe that both studies identified the same critical threshold ($\theta = 0.1$) by coincidence.

Our model, like others, may be criticized for being overly simplistic, as it does not allow for the effect of clonal interference among different allelic types, nor account for the effects of linkage. However, there are good reasons to believe that both of these forces should exaggerate the effects we identify, further slowing the rate of evolution compared to the predictions of classical theory. Therefore the rate of observable evolution (equation 3) should be considered as an upper bound on the expected rate of evolution when applied to real sequence data. An obvious future direction will be to extend this work to account for the effects of non-equilibrium demography on the rate of observable evolution. This may be particularly important in species having average population mutation rates that are low, but that experience periods of high population size (e.g., Drosophila[34]), or in those populations that experience frequent bottlenecks.

Based on recent direct estimates of mutation rates in mice[51], we predict that the rate of neutral evolution of CpG transitions in large rodent populations should experience a detectable deceleration compared to slower, non-CpG substitutions (c.f., Figs. S7A and S7B). This prediction is particularly interesting, as reports have suggested that CpG transitions are "saturated" in rodents[52]. We find this to be unlikely, as saturation (which is caused by repeated serial fixations of different states at the same position) would require cycles of losing and gaining CpG dinucleotides at individual positions, when only the forward mutation (loss of CpG) is unusually fast. We hypothesize that it is more likely that the reduced rate of evolution at CpGs is caused by the deceleration identified above and predicted by the unadulterated Wright-Fisher model.

Conclusions Our results suggest that several of the most fundamental rules of molecular evolution fail to generalize when population mutation rates grow beyond a critical threshold. Through the lens of standard theory, these effects would be interpreted as either weak negative selection when there is none, as weak positive selection when it is actually strong, or even worse, simply as saturation–a phenomenon where branch lengths are supposedly underestimated due to information loss following many recurrent substitutions happening in serial over long evolutionary distances. While our findings are not wholly unexpected when population mutation rates are very large, we have identified significant deviations from the predictions of classical theory for modest, biologically relevant population mutation rates. Taken together, our results suggest that the regime of “weak mutation” is substantially narrower than is widely believed. Great caution should thus be applied in the use of classical population genetics approaches in organisms with large population-scaled mutation rates such as HIV, hyperdiverse eukaryotes, and many prokaryotes.

4 Materials and Methods

Definitions. Throughout this work, and following convention, we refer to the backward mutation rate as u, the forward mutation rate as v, the dominance coefficient as h, and the selection coefficient as s. The population mutation rate, θ, is defined for diploids as $\theta = 4Nv$, and the population scaled selection coefficient is defined as $S = 2Ns$. Since backward mutation will invariably reduce the rate of evolution, to be conservative, we assumed a backward mutation rate of $u = 0$ throughout this study. As expected, when $u > 0$, the slowdowns we observe and predict increase in magnitude (not shown).

Direct computation of the rate of evolution. Let x be the current number of mutants in a Wright-Fisher population of size N, and p the initial number of mutants to arise on a background of $x = 0$. To directly calculate the substitution rate without invoking the theory presented above, we modified our program WFES[45] by making $x = 0$ a transient state so that computation of the mean time to fixation from a starting count of zero ($p = 0$)

will represent the mean time it takes to go from being 100% wildtype to 100% mutant (i.e., the time between fixations). Similarly, we calculated the variance of the time between fixations as the variance of the time to fixation under these same conditions.

For numerical computations, transition probabilities, $P(i, j)$, were calculated under a Wright-Fisher model including bidirectional mutation, selection, and dominance[53],

$$P_{i,j} = \binom{2N}{j} (\psi_i)^j (1 - \psi_i)^{2N-j}, \tag{4}$$

with

$$\psi_i = \frac{\left[(1+s)f_i^2 + (1+sh)f_i(1-f_i)\right](1-u) + \left[(1+sh)f_i(1-f_i) + (1-f_i)^2\right]v}{(1+s)f_i^2 + 2(1+sh)f_i(1-f_i) + (1-f_i)^2} \tag{5}$$

where $f_i = i/(2N)$.

Computation of mean times and probabilities. Mean times and probabilities in equation 3 were computed exactly from a computationally efficient analysis of the appropriate Wright-Fisher Markov model, using absorbing Markov chain methods we previously described[45]. To measure properties of expected allele frequency trajectories demarcated by visits to either of the extinction or fixation boundaries, the boundary states were treated as absorbing (except when directly calculating the mean and variance of the time between fixations; see above), even though they may be escaped by mutation. This implies that the population evolves until reaching one of the two boundaries, after which it is instantaneously "restarted" by a return process. This return process is equivalent to simply permuting the wildtype and mutant state labels and does therefore not disrupt the behaviour of the model when computing functions of the expected trajectories.

Simulations. Simulations were performed by two methods so that the effect of variation in the underlying model assumptions or implementations could be examined. First, we simulated directly from the same Wright-Fisher model used to make calculations throughout the manuscript (equations 4 and 5). Populations were assumed to begin as 100% wildtype. The time to the origination of the next founder mutation, or mutations, was drawn from a geometric distribution with success probability, ρ, equal to the probability of leaving $x = 0$ under the Wright-Fisher model (with $x = 0$ treated as a transient state).

$$\begin{aligned} t &\sim \text{Geom}(\rho = 1 - P(0,0)) \\ &= \text{Geom}(\rho = 1 - (1 - v)^{2N}) \end{aligned}$$

Next, the number of initial mutations, p, was drawn from the Wright-Fisher model such that

$$p \sim \frac{P(0,i)}{1-P(0,0)}, i \in \{1,2,...,2N-1\}$$

where $P(0,i)$ is the probability of going from zero copies to i copies. The sampling distribution was truncated once the probability of transition became smaller than 10^{-8}.

The frequencies of the mutant state were then updated iteratively using $P(i,j)$ until either $x = 0$ or $x = 2N$ were reached. After the population hit either boundary, the simulation was restarted from $x = 0$. Numbers of fixations over many replicates, and the total time spent in generations were recorded. Reported Monte Carlo estimates of the substitution rate were taken to be the number of fixations that occurred divided by the number of generations spent over all simulated absorption cycles. Simulations were repeated 100 times and averaged, where each simulation consisted of 10,000,000 absorptions.

Second, we performed individual-based simulations using SLiM[47] under conditions as close as possible to those employed in the first set of simulations. Estimating the Monte Carlo substitution rate as above, estimates of the substitution rate agreed well between the two simulation types. However, for extreme parameter ranges, we found that SLiM slightly over-predicted the substitution rate compared to the Wright-Fisher simulations and the rate of observable evolution (Fig. S4). This variation is likely due to minor differences in assumptions or implementation details. Simulation error bars were all very small and appeared as points in Fig. 1C. They were thus omitted from the display item. The Erdos code used to conduct SLiM simulations is included in SI Methods 1.4. Simulations were repeated 5 times and consisted of 5,000,000 generations, where the simulation was restarted after each absorption.

Acknowledgements. The authors gratefully acknowledge Ivan Krukov, Sam Yeaman, Caro-Beth Stewart, David Liberles, Quan Long, and Edwin Wang for comments on the manuscript. This work was supported by a Discovery Grant to APJdK from the Natural Sciences and Engineering Research Council of Canada (NSERC) and by an NSERC undergraduate research award (BDS). The authors gratefully acknowledge infrastructure support from the Canada Foundation for Innovation, Alberta Enterprise and Advanced Education, and the Alberta Children's Hospital Research Institute.

Author contributions. APJdK designed the research, developed and implemented the model, performed the analyses, made the figures, and wrote the manuscript. BDS contributed ideas, developed the proof, performed simulations, and edited the manuscript.

The authors declare no conflict of interest.

References

1. Kimura, M. Evolutionary rate at the molecular level. *Nature* **217,** 624-626 (1968)

2. King, J.L., Jukes, T.H. Non-Darwinian evolution. *Science* **164,** 788-798 (1969)

3. Crow, J.F., Kimura, M. Introduction to population genetics theory. Harper and Row publishers (1970)

4. Kimura, M. The neutral theory of molecular evolution. Cambridge University Press (1983)

5. McDonald, J.H., Kreitman, M. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* **351,** 652-654 (1991)

6. Zhai, W., Nielsen, R., Slatkin, M. An investigation of the statistical power of neutrality tests based on comparative and population genetic data. *Mol Biol Evol* **26,** 273-283 (2008)

7. Muse, S.V., Gaut, B.S. A likelihood approach for comparing synonymous and non-synonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol. Biol. Evol.* **11,** 715-724 (1994)

8. Goldman, N., Yang, Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11,** 726-736 (1994)

9. Yang, Z., Nielsen, R., Goldman, N., Pedersen, A.M. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155,** 431-449 (2000)

10. Messier, W., Stewart, C.B. Episodic adaptive evolution of primate lysozymes. *Nature* **385,** 151-154 (1997)

11. Yang, Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **15,** 568-573 (1998)

12. Zhang, J., Nielsen, R., Yang, Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* **22,** 2472-2479 (2005)

13. Halpern, A.L., Bruno, W.J. Evolutionary distances for protein-coding sequences: Modeling site-specific residue frequencies. *Mol. Biol. Evol.* **15,** 910-917 (1998)

14. Dimmic, M.W., Mindell, D.P., Goldstein, R.A. Modeling evolution at the protein level using an adjustable amino acid fitness model. *Pac. Symp. Biocomput.* **5,** 18-29 (2000)

15. Nielsen, R., Yang, Z. Estimating the distribution of selection coefficients from phylogenetic data with applications to mitochondrial and viral DNA. *Mol. Biol. Evol.* **20,** 1231-1239 (2003)

16. Thorne, J.L., Chul Choi, S., Yu, J., Higgs, P.G., Kishino, H. Population genetics without intraspecific data. *Mol. Biol. Evol.* **24,** 1667-1677 (2007)

17. Rodrigue, N., Philippe, H., Lartillot, N. Mutation-selection models of coding sequence evolution with site-heterogeneous amino acid fitness profiles. *Proc. Nat. Acad. Sci. USA* **107,** 4629-4634 (2010)

18. Moorjani, P., Gao, Z., Przeworski, M. Human germline mutation and the erratic evolutionary clock. *PLoS Genetics* **14,** e2000744 (2016)

19. Zuckerkandl, E., Pauling, L.B. ”Molecular disease, evolution, and genic heterogeneity” in Kasha, M. and Pullman, B (editors). *Horizons in Biochemistry*. Academic Press, New York. pp. 189-225 (1962)

20. Kimura, M., and Ohta, T. Mutation and evolution at the molecular level. *Genetics Suppl.* **73,** 19-35 (1973)

21. Cutter, A.D., Jovelin, R., Dey, A. Molecular hyperdiversity and evolution in very large populations. *Mol. Ecol.* **22,** 2074-2095 (2013)

22. Hughes, A.L., Friedman, R., Rivailler, P., French, J.O. Synonymous and nonsynonymous polymorphisms versus divergences in bacterial genomes. *Mol. Biol. Evol.* **25,** 2199-2209 (2008)

23. Sung, W., Ackerman, M.S., Miller, S.F., Doak, T.G., Lynch, M. Drift-barrier hypothesis and mutation-rate evolution. *Proc. Natl. Acad. Sci. U.S.A.* **109 (45)** 18488-18492 (2012)

24. Maldarelli, F., et al. HIV populations are large and accumulate high genetic diversity in a nonlinear fashion. *J. Virol.* **87** 10313-10323 (2013)

25. Rouzine, I.M., Coffin, J.M., Weinberger, L.S. Fifteen years later: hard and soft selection sweeps confirm a large population number for HIV in vivo. *PLoS Genet.* **10 (2)** e1004179 (2014)

26. Pennings, P., Kryazhimskiy, S., Wakeley, J. Loss and recovery of genetic diversity in adapting populations of HIV. *PLoS Genet.* **10 (1)** e1004000 (2014)

27. Erlich, M., Wang, R.Y. 5-Methylcytosine in eukaryotic DNA. *Science* **212,** 1350-1357 (1981)

28. Brown, W., George Jr, M., Wilson, A.C. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* **76,** 1967-1971 (1979)

29. Richard, G.F., Kerrest, A., Dujon, R. Comparative genomics and molecular dynamics of DNA repeats in Eukaryotes *Microbiol. Mol. Biol. Rev.* **72,** 1686-727 (2008)

30. Lynch, M. Evolution of the mutation rate. *Trends in Genetics* **26,** 345-352 (2010)

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

32. Messer, P.W., Ellner, S.P., Hairston, N.G. Jr. Can Population Genetics Adapt to Rapid Evolution? *Trends in Genetics* **32,** 408-418 (2016)

33. Pennings, P.S., Hermisson, J. Soft sweeps II–molecular population genetics of adaptation from recurrent mutation or migration. *Mol. Biol. Evol.* **23** 1076-1084 (2006)

34. Karasov, T., Messer, P., Petrov, D. Evidence that adaptation in drosophila is not limited by mutation at single sites. *PLoS Genet.* **6,** e1000924 (2010)

35. Messer, P.M., Petrov, D. Population genomics of rapid adaptation by soft selective sweeps. *TREE* **28,** 659-669 (2013)

36. Kryazhimskiy, S., Plotkin, J.B. The population genetics of dN/dS. *PLoS Genetics* **187** 1139-1152 (2008)

37. Wright, S. The distribution of gene frequencies under irreversible mutation. *Proc. Natl. Acad. Sci. U.S.A.* **23,** 307-320 (1938)

38. Yang, Z. Molecular evolution: a statistical approach. Oxford University Press (2014)

39. Felsenstein, J. Inferring Phylogenies. Sinauer (2003)

40. Bustamante, C. ”Population genetics of molecular evolution” in Nielsen, R. (editor). *Statistical Methods in Molecular Evolution*. Springer, New York. pp. 63-99 (2005)

41. de Koning, A.P.J., Gu, W., Pollock, D.D. Rapid likelihood analysis on large phylogenies using partial sampling of substitution histories. *Mol. Biol. Evol.* **27,** 249-265 (2010)

42. Draghi, J.A., Parsons, T.L., Plotkin, J.B. Epistasis increases the rate of conditionally neutral substitution in an adapting population. *Genetics* **4**(12): e1000304 (2011)

43. Kimura, M., Ohta, T. On the rate of molecular evolution. *J. Molec. Evolution* **1** 1-17 (1971)

44. Guess, H.A., Ewens, W.J. Theoretical and simulation results relating to the neutral allele theory. *Theor. Pop. Biol.* **3,** 434-447 (1972)

45. Krukov, I., De Sanctis, B.D., de Koning, A.P.J. Wright-Fisher exact solver (WFES): scalable analysis of population genetic models without simulation or diffusion theory. *Bioinformatics* **33,** 1416-1417 (2017)

46. De Sanctis, B.D., Krukov, I., de Koning, A.P.J. Allele age under non-classical assumptions is clarified by an exact computational Markov chain approach. *Scientific Reports* **7,** 11869 (2017)

47. Haller, B.C., Messer, P.W. SLiM 2: Flexible, interactive forward genetic simulations. *Molecular Biology and Evolution* **34** 230-240 (2016)

48. Wilson, D.J., Hernandez, R.D., Andolfatto, Przeworksi, M. A population genetics-phylogenetics approach to inferring natural selection in coding sequences. *PLoS Genetics* **7,** e1002395 (2011)

49. Anderson, T.J.C., et al. Population parameters underlying an ongoing soft sweep in southeast Asian malaria parasites. *Mol. Biol. Evol.* **34** 131-144 (2017)

50. Piganeau, G., Eyre-Walker, A. Evidence for variation in the effective population size of animal mitochondrial DNA. *PLoS One* **4,** e4396 (2009)

51. Uchimura, A., Higuchi, M., Minakuchi, Y., Ohno, M., Toyoda, A., Fujiyama, A., Miura, I., Wakana, S., Nishino, J., Yagi, T. Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice. *Genome Research* **25** 1-10 (2015)

52. Keightley, P.D., Lercher, M.J., Eyre-Walker, A. Evidence for widespread degradation of gene control regions in hominid genomes. *PLoS Biology* **3** e42 (2005)

53. Ewens, W.J. *Mathematical Population Genetics 1: Theoretical Introduction* Edn. 2 (New York: Springer-Verlag, USA, 2004).