Review

EMBO reports

The divergence of mutation rates and spectra across the Tree of Life

Michael Lynch[1,*], Farhan Ali[1], Tongtong Lin[2], Yaohai Wang[2], Jiahao Ni[2] & Hongan Long[2]

Abstract

Owing to advances in genome sequencing, genome stability has become one of the most scrutinized cellular traits across the Tree of Life. Despite its centrality to all things biological, the mutation rate (per nucleotide site per generation) ranges over three orders of magnitude among species and several-fold within individual phylogenetic lineages. Within all major organismal groups, mutation rates scale negatively with the effective population size of a species and with the amount of functional DNA in the genome. This relationship is most parsimoniously explained by the drift-barrier hypothesis, which postulates that natural selection typically operates to reduce mutation rates until further improvement is thwarted by the power of random genetic drift. Despite this constraint, the molecular mechanisms underlying DNA replication fidelity and repair are free to wander, provided the performance of the entire system is maintained at the prevailing level. The evolutionary flexibility of the mutation rate bears on the resolution of several prior conundrums in phylogenetic and population-genetic analysis and raises challenges for future applications in these areas.

Keywords drift-barrier hypothesis; effective population size; mutation rate; mutation spectrum; nucleotide composition

Subject Categories Chromatin, Transcription & Genomics; Evolution & Ecology

DOI 10.15252/embr.202357561 | Received 29 May 2023 | Revised 1 August 2023 | Accepted 2 August 2023

EMBO Reports (2023) e57561

Introduction

Motivated by interest in the emergence of genetic disorders and in the rate of origin of genetic fuel for evolutionary change, biologists have long sought to estimate the rate at which *de novo* mutations arise in various organisms. Early approaches ranged from indirect inferences from the incidence of monogenic disorders of debilitating effects in humans (Haldane, 1935; Kondrashov, 2003; Lynch, 2009) to single-locus reporter constructs in microbes (Rosche & Foster, 2000). Three decades ago, Drake (1991) summarized everything that had been learned about microbial mutation rates using the latter approach. Extrapolating from observations in three bacteriophages and three cellular organisms, he inferred an inverse relationship between the per-generation base-substitution mutation rate per nucleotide site (u_{bs}) and genome size in microbial species, concluding that on average ~ 0.003 mutations arise per genome per cell division in such species. Advances in genome sequencing over the past two decades have dramatically altered the state of knowledge in this area, to the point where no molecular trait has been more finely resolved across the Tree of Life than the mutation rate.

For short-lived organisms, mutation-rate estimates are now commonly obtained from mutation-accumulation experiments that rigidly enforce the propagation of replicate lines by single-progeny descent, often for hundreds to thousands of generations (Lynch *et al*, 2016). With bottlenecks of single individuals (or two in the case of species with separate sexes), all mutations other than those with extremely strong effects accumulate in an effectively neutral fashion, and in the end the full molecular spectrum of mutations is enumerated by whole-genome sequencing. For longer-lived organisms with known pedigrees, including humans, trio studies are now routinely performed in which the genomes of progeny are compared with those of the parents (e.g., Wu *et al*, 2020; Bergeron *et al*, 2023). These analyses also enable quantification of mutations in a nearly unbiased way, as selection is unable to eradicate mutations in a single generation except in the case of lethals.

Drawing from these kinds of results, we provide an overview on the phylogenetic distribution of the mutation rate across the Tree of Life, along with an attempt to explain the patterns in a unifying evolutionary framework. While this paper was under review, Wang & Obbard (2023) presented a meta-analysis of eukaryotic mutation rates. Our conclusions are essentially concordant with theirs, but we extend the analyses to prokaryotes and cover numerous additional topics.

Based on his inferred constant number of mutations arising per genome per cell division, Drake (1991) argued that because these patterns are consistent across diverse microbial organisms, they are "likely to be determined by deep general forces, perhaps by a balance between the usually deleterious effects of mutation and the physiological costs of further reducing mutation rates." Invoking more specific biophysical arguments, others have suggested that

1 Biodesign Center for Mechanisms of Evolution, Arizona State University, Tempe, AZ, USA
2 Institute of Evolution and Marine Biodiversity, KLMME, Ocean University of China, Qingdao, China
*Corresponding author. Tel: +1 812 606 0681; E-mail: mlynch11@asu.edu

evolved cellular error rates reflect a balance between speed and efficiency (Hopfield, 1974; Ninio, 1975; Qian, 2006; Banerjee *et al*, 2017). As will be seen below, however, these arguments are inconsistent with the fact that the lowest known mutation rates are generally found in the fastest-growing organisms, for example, bacteria and ciliates (Sung *et al*, 2012).

An alternative to such optimization arguments is the drift-barrier hypothesis (DBH), which postulates that because the majority of mutations are deleterious, natural selection predominantly operates to reduce the mutation rate per functional base pair down to the point at which the small advantage of any further reduction is thwarted by the power of random genetic drift (Lynch, 2008, 2010, 2011). Although the idea that selection strives to reduce the mutation rate dates to Sturtevant (1937), he did not consider the role played by drift. Because selection on a mutator allele is an indirect effect of the deleterious effects inflicted on linked sites (Kimura, 1967; Kondrashov, 1995; Dawson, 1999; Lynch, 2008, 2011), genomes with larger amounts of functional DNA are also expected to evolve lower mutation rates, and as most microbes have genomes consisting of 75 to 95% coding DNA, this in part explains the negative associations between u_{bs} and genome size in these groups recognized by Drake (1991) and Drake *et al* (1998).

Mutation-rate variation in light of the drift-barrier hypothesis

Drake's (1991) inference of a constant number of mutations per cell division in microbes can now be evaluated in a more rigorous fashion with direct observations from mutation-accumulation experiments followed by whole-genome sequencing. Based on observations from 38 prokaryotes, the slope for the regression of the base-substitution mutation rate (u_{bs}) on genome size ($\simeq -2.0$) is significantly more negative than Drake's expectation of -1.0 (on a log–log plot), and the $\sim$ 10-fold range of variation around the regression is much greater than Drake inferred (Fig 1A; see Dataset EV1 for all of the data used in this and the following analyses). As a consequence, the average number of mutations per genome per cell division ranges from ~ 0.01 for the smallest to ~ 0.001 for the largest prokaryotic genomes. The two estimates for archaeal species are consistent with those for bacteria with similar genome sizes.

On the other hand, the regression coefficient for 20 unicellular eukaryotes (excluding *Chlamydomonas* and *Emiliania*, both of which have genomes with substantial fractions of noncoding DNA) is not significantly different from -1.0 (Fig 1A). Estimated mutation rates for three filamentous fungi have a broad range, roughly consistent with unicellular eukaryotes, but there is substantial uncertainty in these estimates owing to challenges in estimating cell division number. In striking contrast, for multicellular organisms, there is a positive association between u_{bs} and genome size.

To obtain a broader understanding of the phylogenetic dispersion of mutation rates, we require information on genetic effective population sizes (N_e), as the inverse of this number defines the power of random genetic drift. The only feasible way of estimating N_e is to rely on measures of within-population nucleotide diversity at neutral genomic sites for natural populations (π_s), as the expected level of such variation under mutation-drift equilibrium $\simeq 4N_e u_{bs}$ for diploid species (and half that for haploids). For species for which such measures are available (generally from silent sites in protein-coding genes) along with an estimated u_{bs}, N_e can then be extracted by dividing by $4u_{bs}$ (or $2u_{bs}$ for haploids). For each of the five major organismal groupings for which this kind of analysis can be performed, there is a negative association between u_{bs} and N_e, as anticipated by the drift-barrier hypothesis; the average of the slopes being -0.82 (SE = 0.21) (Fig 1B).

There are two potential concerns with the statistical analyses in Fig 1B. Although sampling error of u_{bs} will not directly bias the estimated regression coefficient, there are two indirect effects associated with the fact that u_{bs} is used to estimate N_e from silent-site variation. First, the independent variable N_e is estimated with error, owing to the fact that both π_s and u_{bs} are themselves estimates. Errors associated with an x-axis variable will cause overdispersion on this axis, leading to a downward (in this case less negative) bias in the slope. Although based on whole-genome analyses in most cases, estimates of π_s still have uncertain variances associated with the sampling of individuals and subpopulations. However, the average coefficients of variation (standard errors divided by the means) of individual u_{bs} estimates are 0.09 for bacteria, 0.11 for unicellular eukaryotes, 0.15 for invertebrates, 0.24 for vertebrates, and 0.27 for land plants. Thus, given that the range in N_e estimates in Fig 1B extends over orders of magnitude, this effect is unlikely to cause much more than 10% bias in the overall pattern across the Tree of Life, although the regressions for individual phylogenetic groups may be biased toward 0 by up to 25% in the multicellular groups. Second, in estimating N_e by dividing π_s by the mutation rate, an intrinsic negative sampling correlation is created between N_e and u_{bs}, owing to the fact that an upward error in the estimate of the latter will translate to a downward error in the estimate of N_e. If u_{bs} were estimated with 100% accuracy, this also would not be an issue. These two opposing sources of bias might roughly cancel each other out within organismal groupings, and prior work shows that the magnitude of sampling variation in u_{bs} is insufficient to alter the conclusion that there is a negative association between u_{bs} and N_e across the Tree of Life (Sung *et al*, 2012).

There is, however, a more formal way to infer the form of the relationship between u_{bs} and N_e. If N_e and u_{bs} are independent, a log–log regression of π_s on u_{bs} should yield a slope of +1, with the intercept simply being related to the average N_e. If, on the other hand, u_{bs} and N_e are negatively correlated with allometric scaling coefficient $-x$, the expected slope will equal $1-x$, for example, if u_{bs} is inversely related to N_e (with $x = 1$), the expected slope would equal 0. When such regressions are performed for the five major organismal groups in Fig 1B, in each case the slope is substantially less than 1.0, and in only one case it is significantly different from 0.0 (Appendix Fig S1; Dataset EV1). Thus, the negative association of mutation rates with N_e, both within and among groups, is a true evolutionary feature and not a statistical artifact. Wang & Obbard (2023) applied a similar analysis to eukaryotic mutation rates and reached the same conclusion.

The question remains as to what determines the effective population size of a species. N_e is governed not just by the absolute number of individuals within a species but also by various aspects of the breeding system, temporal patterns of demographic variation, and the amount of background selection operating on linked chromosomal sites. Linkage of nucleotide sites on chromosomes results in a

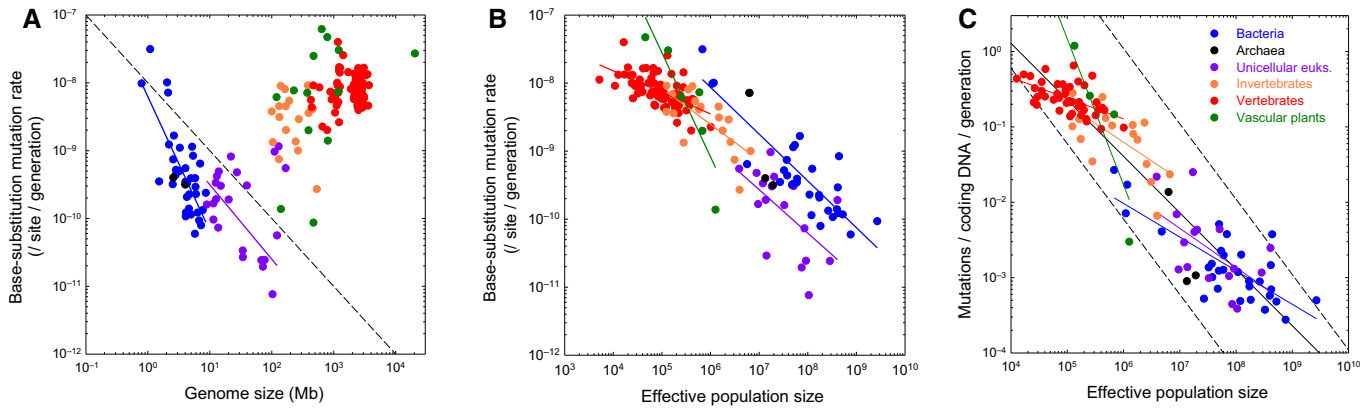

Figure 1. Scaling of base-substitution mutation rates with genome size and effective population size.

(A) Base-substitution mutation rates per generation (u_{bs}) as a function of genome size (G, in Mb). Fitted log–log regression for prokaryotes: $y = -8.21-1.98x$, with standard errors (SEs) of 0.18 and 0.30 for the intercept and slope, respectively; $r^2 = 0.553$ and sample size $n = 38$. Fitted regression for unicellular eukaryotes (excluding *Chlamydomonas* and *Emiliana*: $y = -8.40-1.10x$) with SEs of 0.40 and 0.27 for the intercept and slope, respectively; $r^2 = 0.442$ and $n = 20$. The dashed line is a reference with slope = −1.0. (B) Group-specific regressions relating u_{bs} to the effective population size (N_e). (C) Relationship between U_{bs} (the product of u_{bs} and the total number of nucleotides contained within protein-coding sequence) and N_e; $y = 3.10-0.75x$, with SEs of 0.18 and 0.03 for the intercept and slope, respectively; $r^2 = 0.861$ and $n = 117$. Here, the diagonal dashed lines are references with slopes =−1 All raw data used in these regressions are in Dataset EV1. Regression statistics are summarized in Appendix Tables S1–S3.

reduced amount of variation retained within populations, owing to selective sweeps of beneficial mutations and the relentless removal of recurrent deleterious mutations, and these factors may almost completely override the influence of absolute population size in most cases (Charlesworth, 2009; Walsh & Lynch, 2018; Charlesworth & Jensen, 2022; Lynch, 2023). Over the scale of the entire Tree of Life, the average size of an adult individual serves as a good predictor of the expected value of N_e, explaining about 85% of the range of variation in the latter, with a 10-fold increase in organismal size leading to an expected 37% reduction in N_e (Fig 2). However, for any particular organism size, there is also an order-of-magnitude range of variation in N_e, and the upper limit to the latter is of order 10^9 even in the smallest microbes. Given this range of residual variation and the narrower ranges in body size within organismal groups, the scaling of N_e and size is less obvious within groups.

Theory suggests that it is not simply the per-site mutation rate that should decrease with increasing N_e, but more specifically the genome-wide mutation rate summed over functionally significant sites with influences on fitness. This may help explain, for example, why unicellular eukaryotes tend to have lower mutation rates than bacterial species with similar N_e (Fig 1B), as the former tend to have genomes harboring larger numbers of genes (Lynch *et al*, 2016). It remains difficult to enumerate the full pool of genomic sites of functional significance except in a few model species, and even then there are uncertainties involved with respect to intergenic DNA. As a surrogate, we use the total number of nucleotides employed in protein-coding genes (including alternatively spliced exons in species in which this occurs), P, noting that this will proportionately underestimate the actual functional genome size somewhat more for multicellular species with complex regulatory regions than for bacterial species for which > 90% of the genome is generally within protein-coding sequences.

Across the entire Tree of Life, $U_{bs} = u_{bs} \cdot P$, scales negatively with N_e with an allometric slope very close to −0.75 (Fig 1C). The regression coefficients are significantly negative for each of

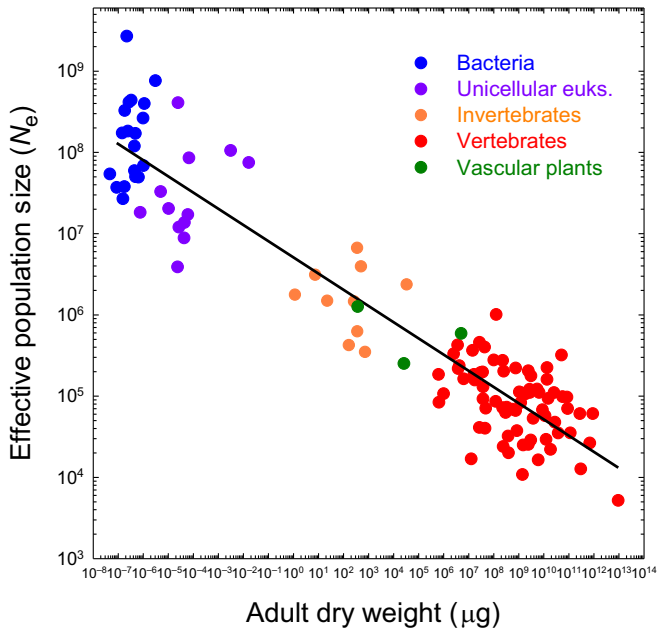

Figure 2. Scaling of effective population size and organism size.

Relationship between the effective population size (N_e) and adult dry weight (in μg). Fitted log–log regression for the entire pool of data (Dataset EV1): $y = 6.717-0.199x$, with standard errors (SEs) of 0.054 and 0.007 for the intercept and slope, respectively; $r^2 = 0.875$, sample size $n = 113$, with $P < 10^{-53}$. The data for four pathogenic bacteria and two self-fertilizing nematodes, all of which have unusually low N_e, are excluded from this analysis.

the narrower organismal groups, although less strongly so, except in the case of vascular plants (where there are only five data points). This is expected for reasons given above—because the range of observed N_e is reduced within groups, a larger fraction of

its total variance will be a consequence of sampling error (rather than true phylogenetic variation). Again, within each phylogenetic group, there is a ~ 5- to 10-fold range of variation in U_{bs} conditional on a particular N_e, and given the small error in inference of individual mutation-rate estimates noted above, much of this dispersion will be true evolutionary variance.

Less extensive data exist on rates of insertion/deletion (indel) mutation, and some studies do not report them at all, but two generalizations seem to hold across the Tree of Life (Fig 3). First, the indel mutation rate, u_{id}, is strongly correlated with the base-substitution mutation rate. Given this correlation, it follows that u_{id} is also strongly negatively correlated with estimates of N_e, and that this correlation cannot be driven by circularities in estimating N_e from u_{bs} (Sung *et al*, 2016). Second, u_{id} is generally less than u_{bs}, typically on the order of 10 to 30% of the latter; three species with u_{id} slightly elevated over u_{id} have genomes with highly repetitive DNAs subject to replication slippage. On average, for the few species with sufficient data, there is approximate parity between the numbers of small insertions and deletions (Dataset EV1).

Owing to its much smaller genome size, there are few reliable estimates of the mutation rate for mitochondrial genomes, although the rates are invariably inflated relative to those in the nuclear genome. For the most reliable estimates (which still have high standard errors), the degrees of inflation are 6× for the yeast *S. cerevisiae*, 53× for the nematode *C. elegans*, 101× for the microcrustacean *D. magna*, and 14× for the fruit fly *D. melanogaster* (Dataset EV1).

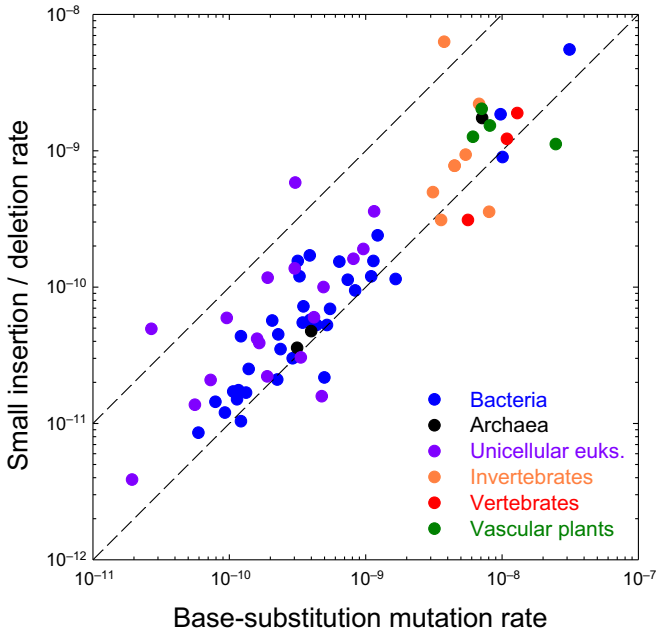

Figure 3. Scaling of insertion/deletion and base-substitution mutation rates.
Relationship between the mutation rate for insertion/deletions < 50 bp in length, u_{id}, and for base substitutions, u_{bs} across the Tree of Life, both in units of nucleotide site^{-1} · generation^{-1}. The upper and lower dashed lines denote references for 1.0 and 0.1 ratios of the two rates. The three species with ratios > 1.0 are the yeast *Zymoseptoria tritici*, the slime mold *Dictyostelium*, and the nematode *Caenorhabditis elegans*. Numerical results are tabulated in Dataset EV1.

These elevations are qualitatively consistent with the DBH in that the effective proteome size of the mitochondrial genome is ~ 1% of that of the nuclear genome in yeast, and ~ 0.05% in the invertebrates. The less than proportional decline in mitochondrial u_{bs} with proteome size may be a consequence of the unique population-genetic environment for organelle genomes. Given the central roles in bioenergetics played by the small number of mitochondrial genes, and the absence of recombination in mitochondrial genomes, it is likely that the strength of selection operating on the mutation rate in such settings is much greater than in nuclear genomes; mutations with very large deleterious effects will be effectively purged from all lineages and so will not contribute to a gradient in the drift barrier.

Evolutionary fluidity of mutational/surveillance mechanisms

Many cellular features retain relatively constant functions across the Tree of Life, despite the turnover of the underlying molecular constituents, e.g., regulatory networks, signal-transduction pathways, and the components of mitosis and meiosis (Lynch, 2023). Although the mechanisms allowing for such rewiring are not always known, the general principles appear to extend to mutation-rate evolution. Under the theory of evolutionary layering in a multi-component system for mutational surveillance, the component parts are expected to wander in terms of individual efficiency, provided that the net outcome of the total system remains the same (Lynch, 2012).

Phylogenetic variation in the molecular spectra of mutations provides one line of indirect evidence for the evolutionary divergence of DNA-repair and/or replication-fidelity mechanisms. For example, population-genomic analyses suggest that significant differences in the molecular spectra of mutations have arisen in the human population over space and time (Harris & Pritchard, 2017; Seoighe & Scally, 2017; DeWitt *et al*, 2021; Milligan *et al*, 2022; Gao *et al*, 2023). More direct evidence for such shifts derives from two summary features of mutation spectra obtained from mutation-accumulation and pedigree analyses.

First, the ratio of transition (purine ↔ purine and pyrimidine ↔ pyrimidine substitutions) to transversion (purine ↔ pyrimidine) mutations has an expected value of 0.5 for the situation in which all 12 mutation types arise at equal rates, yet the ts:tv ratio exceeds 1.0 in almost all species, in a few cases rising to values as high as 10.0 (Fig 4A). Although some of this dispersion is a simple consequence of errors associated with numbers of observed mutations, most of it is not (see legend to Fig 4A). Second, the AT mutation bias (β, the ratio of G/C → A/T to A/T → G/C mutation rates) exceeds the null expectation of 1.0 in most species, with some values ranging as high as 4.0 (Fig 4A). The average CV of these AT mutation-bias estimates is on the order of twice the CVs noted above for total mutation rates, so again a substantial fraction of the observed dispersion is due to real biological differences.

Knowledge of the AT mutation bias provides insight into the mechanisms driving genome-wide nucleotide composition in different species. In the absence of selection, the fractional AT composition of a genome is expected to equilibrate at $\beta/(1+\beta)$. However, nearly every genome-wide estimate of AT composition is near or below this expectation (Fig 4B). Some of this apparent selection for GC is simply associated with the need to utilize amino acids with GC

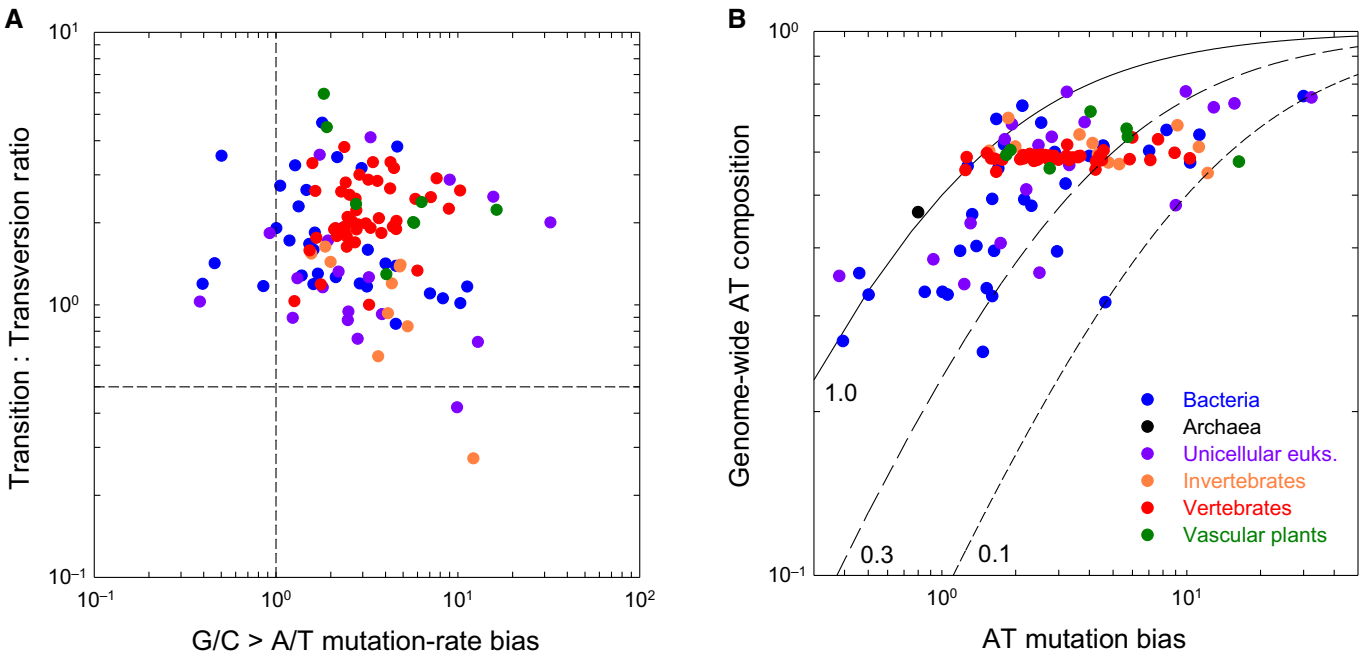

Figure 4. Aspects of the molecular spectra of base-substitution mutation rates across the Tree of Life.

(A) The phylogenetic dispersion of transition:transversion (ts:tv) ratios and AT mutation bias (β) across the Tree of Life. The dashed lines denote the respective null values of 0.5 and 1.0 under the hypothesis that all 12 possible base-substitution mutation types arise at equal rates. Data are only shown for species with $n \geq 20$ observed base substitution mutations. From equation A1.19b in Lynch & Walsh (1998), the coefficient of variation (CV) of a ts:tv estimate equal to x is $\sim (1+x)/(n\sqrt{x})$, which is < 0.10 for $x = 1$ and < 0.175 for $x = 10$. (B) Genome-wide AT compositions as a function of the AT mutation bias. The solid line denotes the null expectation under neutrality resulting from the balance between bidirectional mutational pressures, whereas the dashed lines denote the expectations for cases in which selection reduces the fixation probabilities of G/C → A/T mutations by factors of 0.3 and 0.1 relative to A/T → G/C mutations (as described in the text).

content in their codons. However, in numerous vertebrates and land plants, > 90% of the genome is intergenic and/or intronic, so selection on coding DNA alone cannot explain these prevailing patterns (Long *et al*, 2018b). When selection operates on nucleotide variation, the expected AT composition is approximately $\beta\varphi/(1+\beta\varphi)$, where φ is a measure of the ratio of fixation probabilities for G/C → A/T relative to A/T → G/C mutations, which in turn is a function of the strength of selection relative to drift (Li, 1987; Bulmer, 1991). The data suggest that average φ is in the range of 0.1 to 1.0 for nearly all observed species. Thus, although there is a weak positive association between the magnitude of AT mutation bias and genome-wide AT composition, selection pressure in favor of GC nucleotides also prevails across the Tree of Life. More fine-scale resolution than provided here is necessary to ascertain how selection varies among sites with different degrees of functional significance (Long *et al*, 2018b).

Some variation in mutation spectra might be due to environmental variation experienced by different study species (e.g., Maharjan & Ferenci, 2017; Shewaramani *et al*, 2017), and the molecular-genetic basis for such variation is generally unknown. However, a clear mechanistic example of the evolutionary wandering of mutational features involves the mismatch-repair (MMR) pathway, which detects and eliminates base-pair mismatches in DNA. The molecular machinery underlying this pathway is shared in model genetic systems such as *E. coli*, *S. cerevisiae*, *C. elegans*, and mouse, all of which exhibit a 50- to 100-fold increase in the mutation rate when MMR is deleted (Long *et al*, 2018a; Lujan & Kunkel, 2021). Nevertheless, a number of bacterial species appear not to harbor the conventional MMR machinery, and yet experience no apparent inflation in the mutation rate. For example, *Corynebacterium glutamicum* (Takemoto *et al*, 2018) and *Mycobacterium smegmatis* (Castañeda-Garc a *et al*, 2020) have completely novel mismatch-repair machineries, and yet do not have unusual mutation rates relative to other bacteria with similar genome sizes.

The preceding observations make clear that the mutation rate is evolutionarily malleable, but what is the timescale of such change? The mutational target size for the mutation rate is likely to be very large, as it includes multiple DNA polymerases, DNA-repair proteins, and essentially all genes whose products alter the mutagenicity of the intracellular environment (including those influencing the production of free oxygen radicals via metabolic activity and those modulating the relative abundances of free nucleotides). Thus, with the expectation that both mutators and anti-mutators are recurrently introduced into all populations (Denamur & Matic, 2006; Lynch, 2008, 2011; Raynes & Sniegowski, 2014), virtually all natural populations likely harbor polymorphisms for the mutation rate. This supposition is corroborated by several-fold ranges of variation for u_{bs} observed for natural isolates of species ranging from *Chlamydomonas* (López-Cortegano *et al*, 2021) to *Saccharomyces* (Gou *et al*, 2019; Jiang *et al*, 2021) to *Drosophila* (Schrider *et al*, 2013; Wang *et al*, 2023) to mice (Sasani *et al*, 2022) to humans (Rahbari *et al*, 2016; Kessler *et al*, 2020).

The idea that the mutation rate is capable of rapid change is also supported by a diversity of experimental observations. For example, Boe *et al* (2000) estimate that *E. coli* cells with mutation rates elevated by 20–80× arise at rates of 5×10^{-6} per cell division, and one can imagine even higher rates of origin of milder (and less easily detected) mutators (as well as antimutators). Indeed, in an *E. coli* mutation-accumulation experiment initiated with a mutator strain that allowed the neutral accumulation of diversity over a period of 1,250 generations, numerous lines evolved mutation rates $< 10\%$ of the baseline rate (antimutators), while a small fraction of them experienced up to 10× increases in the mutation rate (Singh *et al*, 2017). In laboratory evolution experiments in which selection is allowed, bacterial populations founded with a mutator genotype frequently evolve lower mutation rates on relatively short time scales via compensatory molecular changes at genomic sites not involved in the initial mutator construct (Turrientes *et al*, 2013; Wielgoss *et al*, 2013; Williams *et al*, 2013; McDonald *et al*, 2016; Wei *et al*, 2022). Given this potential for rapid change in the level of replication fidelity, it is not surprising that clonally reproducing microbes transiently evolve mutator genotypes when confronted with strong selective challenges (such as antibiotic treatment). For example, Swings *et al* (2017) found that in lethally stressful environments, laboratory *E. coli* populations evolve substantial elevations in the mutation rate on time scales of $\sim$ 100 generations, but then revert to background rates once adaptation has been achieved.

The common appearance of antimutators implies the presence of substantial unexploited potential for improvement in replication fidelity, as expected under the drift-barrier hypothesis. As further evidence that evolved mutation rates are not constrained by biophysical limitations, consider the fact that although mammals harbor some of the highest known eukaryotic mutation rates per generation, the germline mutation rate per cell division rivals the very low rates for unicellular species. The human germline mutation rate per nucleotide site, $\sim 6 \times 10^{-11}$ per cell division, is lower than any rate observed in bacterial species, and 10 to 100× lower than rates in various human somatic tissues (Lynch, 2010; Behjati *et al*, 2014; Milholland *et al*, 2017; Cagan *et al*, 2022). The key point here is that although selection operates on the per-generation mutation rate, in multicellular species this is accommodated by changes in replication fidelity at the cell division level—an increased number of germline cell divisions is balanced by enhanced replication fidelity per division.

Error-prone polymerases

In all known organisms, almost all DNA replication is carried out by one or two major polymerases, each of which has a high baseline level of accuracy, with a substantial fraction of the few errors arising at the polymerization step being removed secondarily via a proofreading step, and thereafter by mismatch repair. However, nearly all genomes also encode for one or more error-prone polymerases, whose usage is restricted mostly to times of stress or to dealing with bulky lesions in DNA. The phenomenon of stress-induced mutagenesis (SIM) has been found in virtually all organisms that have been examined, for example, *E. coli* and many other bacteria (Kang *et al*, 2006; Foster, 2007; Kivisaar, 2010); yeast (Heidenreich, 2007); *Chlamydomonas* (Goho & Bell, 2000); *Caenorhabditis* (Matsuba *et al*, 2013); and *Drosophila* (Sharp & Agrawal, 2012). An elevation in error rates under extreme environmental situations should not be too surprising, as physiological breakdown can be expected for virtually all traits. Nonetheless, some have argued that high mutation rates associated with error-prone polymerases have been promoted by selection as a means for generating adaptive responses to changing environments (Radman *et al*, 2000; Rosenberg, 2001; Tenaillon *et al*, 2001; Earl & Deem, 2004; Galhardo *et al*, 2007; Rosenberg *et al*, 2012). Direct empirical support for such an argument is lacking, although special scenarios have been shown in theory to encourage selection for SIM (Ram & Hadany, 2014; Lukačišinová *et al*, 2017).

A simpler and more compelling explanation for the error-prone nature of some polymerases follows directly from the drift-barrier hypothesis—the net selection pressure to improve accuracy is expected to be proportional to the average number of nucleotide transactions that a DNA polymerase engages in per generation. Because error-prone polymerases generally replicate only small patches of DNA and do so quite infrequently, the strength of selection on accuracy will be correspondingly reduced (Lynch, 2008, 2011; MacLean *et al*, 2013). This "use it or lose it" hypothesis is also consistent with the high error rates for polymerases deployed in the replacement of small RNA primers used in replication initiation (Lynch, 2011). In addition, the secondary and tertiary fidelity mechanisms associated with replication (proof-reading and mismatch repair), which necessarily involve far fewer nucleotide transactions than the earlier polymerization step, have greatly elevated error rates (Lynch, 2008; Lujan & Kunkel, 2021). These latter observations cannot be explained as specific adaptations to stress, as all of the factors are fundamental to normal replication cycles.

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

The nonrandom nature of mutations

The postulate that mutations arise randomly with respect to the forces imposed by natural selection has survived decades of scrutiny. However, it is sometimes argued that selection refines mutation rates on a gene-by-gene basis to enhance evolvability and/or to protect essential or highly expressed genes from damage (Galhardo *et al*, 2007; Martincorena *et al*, 2012; Paul *et al*, 2013). Such arguments ignore the strength of selection necessary to advance a molecular adaptation, and relatively simple calculations suggest that the strength of selection operating on a gene-specific mutation-rate modifier is unlikely to ever be high enough to promote its establishment (Chen & Zhang, 2013; Lynch *et al*, 2016; Liu & Zhang, 2021). The formal arguments involved here do not deny the possibility of differences in average mutabilities among genes but indicate instead that where they exist, such properties are inadvertent by-products of other gene-specific features such as nucleotide composition and transcription rate.

Although mutations are random with respect to the essentiality of gene targets, they are nonrandom in numerous other ways. For example, in some bacteria, there is a symmetrical wave-like pattern of the mutation rate around the circular chromosome (Foster *et al*, 2013; Long *et al*, 2015), although the amplitude of differences does not exceed 2.5× and the pattern differs among species. Up to twofold mutation-rate differences have also been found among locations on eukaryotic chromosomes on spatial scales ranging from 200 to 100 kb (Stamatoyannopoulos *et al*, 2009; Lang & Murray, 2011; Chen *et al*, 2012; Poetsch *et al*, 2018; Li & Luscombe, 2020). The molecular mechanisms driving these large-scale patterns remain unclear but may be associated with variation in the nucleotide-pool composition during the cell cycle, regional variation in transcription rates and their influence on replication, alterations in the rates of processivity of DNA polymerase across different chromosomal regions, and/or protection by nucleosomes in eukaryotes.

In some filamentous fungi and land plants, mutation rates in gene bodies are ~30% lower than in surrounding noncoding regions (Habig *et al*, 2021; preprint: Monroe *et al*, 2022a; Monroe *et al*, 2022b; see Liu & Zhang (2022) and Charlesworth & Jensen (2023) for some caveats). Notably, in both cases, as well as in mammals, nucleosomes associated with gene bodies are frequently marked by specific histone modifications that recruit mismatch-repair proteins (Li *et al*, 2013; Frigola *et al*, 2017; Supek & Lehner, 2017; Huang *et al*, 2018). The phylogenetic reach of this effect and the degree to which it is promoted by selection remains unclear, and the benefits of sequestering repair machinery to transcribed regions may be of minor significance in genomes with minimal intergenic DNA. In yeast, for example, mutation rates appear to be elevated in highly expressed genes (Park *et al*, 2012), and in a variety of lineages, certain classes of genes, such as tRNAs, have elevated rates of mutation (Thornlow *et al*, 2018). These latter counter-examples may reflect the consequences of transcription-associated mutagenesis (Cho & Jinks-Robertson, 2017).

Finally, the appearance of mutations at adjacent sites can be temporally correlated. The usual view is that if mutations arise at an average rate u per nucleotide site, the rate of simultaneous origin of mutations at two specific sites would be u^2, at three sites would be u^3, etc. Given that average u is on the order of 10^{-9}, this would imply that double mutants would rarely ever occur except in large microbial populations. However, data from mutation-accumulation experiments suggest that on spatial scales of < 100 bp, multinucleotide mutations commonly comprise 1 to 3% of mutational events in diverse lineages (Drake, 2007; Schrider *et al*, 2011; Terekhanova *et al*, 2013; Harris & Nielsen, 2014; Uphoff *et al*, 2016). Potential reasons for mutational clusters include local patches of DNA damage, the occasional use of a defective DNA polymerase molecule, accidental deployment of an innately error-prone polymerase, and mutagenic repair of double-strand breaks (Drake, 2007; Hicks *et al*, 2010; Malkova & Haber, 2012; Chan & Gordenin, 2015; Seplyarskiy *et al*, 2015). The key point is that transient, localized hypermutation is common enough that rates of occurrence of double mutations are often many orders of magnitude above the u^2 expectation under independent occurrence, and more commonly on the order of u/1,000 to u/100. Such high incidences of mutation clusters have major implications for the evolution of complex features, as modifications requiring multiple nucleotide changes on small spatial scales (e.g., within genes) need not await the sequential fixation of individual mutations, but can arise and be promoted together.

Implications for population-genomic and phylogenetic analyses

Drift of mutation rates both within and among phylogenetic lineages raises significant challenges for studies attempting to draw phylogenetic and/or demographic inferences from patterns of molecular variation. For example, point estimates of u_{bs} from single, contemporary individuals, like those noted above, are often applied to silent-site divergences among species to infer times of phylogenetic separation. The logic underlying such applications is the conclusion that for neutrally evolving sites, the rate of sequence divergence between two lineages is equal to $2u_{bs}$ per generation (Kimura, 1983). When dates inferred from molecular data are inconsistent with fossil evidence, arguments are often made that this is due to persistent historical changes in mutation rates in specific lineages (e.g., in primates; Elango *et al*, 2006; Scally & Durbin, 2012; Amster & Sella, 2016; Wu *et al*, 2020). However, given that such discordancies often involve factors of only 2 to 3, it is difficult to rule out the possibility that they are simple and essentially unavoidable results of the stochastic temporal wandering of mutation rates within lineages, rather than reflections of deterministic processes.

In addition, there has been long-standing interest in what has popularly become known as Lewontin's paradox (Lewontin, 1974), the observation that molecular genetic diversity increases only weakly with population size contrary to the expectations under neutral theory. Numerous attempts have been made to explain such weak scaling with various population-genetic arguments invoking the influence of selection on the maintenance of variation (Nei, 1983; Buffalo, 2021; Charlesworth & Jensen, 2022). However, such a pattern is entirely expected under the DBH, as there is a strong negative association between u_{bs} and N_e in all organismal groups. As justification for ignoring this explanation, examples are sometimes invoked in which two species with different N_e have mutation-rate differences that deviate from the average expectation under the DBH. However, such arguments ignore the fact that the drift barrier to mutation-rate evolution is not absolutely deterministic. Rather, as for any quantitative trait, there is drift around the drift barrier (Walsh & Lynch, 2018). One would never take a morphological measure of a single individual to provide an accurate estimate of a species-wide average, and the same caution should be applied to complex molecular traits such as the mutation rate.

Discussion

Although some skepticism has been raised over the claim that mutation rates scale negatively with N_e (Galtier & Rousselle, 2020; Liu & Zhang, 2021; Charlesworth & Jensen, 2022), the accumulation of many dozens of genome-wide mutation-rate estimates over the past few years has only reinforced the generality of this pattern and done so consistently for each organismal group with adequate data—prokaryotes, unicellular eukaryotes, vascular plants, invertebrates, and vertebrates. There is also approximate continuity in scaling across major groups, once differences in proteome sizes are taken into

consideration. The observed patterns of negative scaling are not statistical artifacts or consequences of systematic sampling errors. As inconvenient as they may be for certain kinds of population-genetic analyses, the strong negative scaling of u with N_e and the substantial level of within-population variation for u are now established facts.

This being said, although the DBH makes a fairly robust prediction that mutation rates should decline with increasing N_e, without knowledge of the distribution of selection effects for *de novo* mutations and of the relative incidence of mutator versus antimutator alleles, one cannot predict in advance the slope or the elevation of the regression (Lynch, 2011; Devi *et al*, 2023). However, the data suggest a power-law relationship between u_{bs} and N_e, with a slope near -0.75 to -1.0 on a log–log plot once differences in proteome size are taken into consideration. This kind of scaling implies a constant fractional increase in the power of downward directional selection on u_{bs} with each fractional increase in N_e. That is, averaged across the Tree of Life, there is an $\sim$ 6- to 10-fold decline in the genome-wide deleterious mutation rate with each order of magnitude increase in N_e. How and why biology is structured in such a way as to yield such a phylogenetic pattern remains to be worked out.

One lingering concern may be that, after accounting for effective proteome and population sizes, bacteria have no lower mutation rates than unicellular eukaryotes. Theory predicts that the efficiency of selection on the mutation rate increases with reduced levels of recombination (Kimura, 1967; Lynch, 2008), and bacteria are commonly viewed as being clonal in nature. However, although bacteria lack meiotic recombination, they commonly experience roughly the same average amount of recombination per nucleotide site per generation (via other mechanisms) as do eukaryotes (Lynch, 2007; Shapiro, 2016).

Is there an alternative to the DBH as an explanation for the phylogenetic reduction in u with increasing N_e? Mutational processes generate a large fraction of detrimental variants that must be removed by natural selection, and it is the linked association of such mutations with mutator alleles that persistently drives selection against the latter. However, this onslaught of deleterious mutations overlies a smaller fraction of beneficial mutations essential for adaptation in changing environments. Under the DBH, the supply of these beneficial variants is assumed to come essentially for free as a consequence of the nonzero mutation rate. Nonetheless, substantial attention has been given to the idea that natural selection might fine-tune the mutation rate so as to maximize the long-term rate of adaptive evolution in the face of a background dominated by deleterious mutations. Under this optimization perspective, selection does not constantly push the mutation rate to the lowest achievable level but instead promotes specific levels of mutation via indirect effects associated with the small pool of beneficial mutations. This is a difficult area for theory development as the relative merits of increasing versus decreasing the mutation rate depend on the distribution of mutational effects, the population size, the recombination rate, and the pattern of environmental change.

The primary problem here is that in a sexual population, hitch-hiking of a mutator allele with a linked beneficial mutation will generally be thwarted by their dissociation by recombination (on average in just two generations when the two loci are on different chromosome arms in sexual species). The continuous reinforcement necessary for the promotion of a mutator allele then requires a substantial rate of input of closely linked beneficial mutations. In contrast, because the vast majority of mutations are deleterious, there will be a steady-state background deleterious load associated with all mutator alleles, regardless of whether the mutator is involved in a transient (and most likely incomplete) beneficial sweep. Additional limitations in multicellular organisms are the direct negative effects that mutators impose via the production of somatic mutations with immediate detrimental effects on fitness (Lynch, 2008, 2010).

Several attempts have been made to estimate theoretically optimal mutation rates for maximizing long-term rates of adaptive evolution in nonrecombining populations, but the resultant models do not explain why, if optimized, mutation rates are nearly 1,000× higher in large multicellular species than in most microbes. Where they are not completely silent on the matter, most models concerned with optimal mutation rates in persistently changing environments appear to imply a positive association between N_e and u_{bs} (Kimura, 1967; Leigh, 1970; Orr, 2000; Johnson & Barton, 2002; Desai & Fisher, 2007; Good & Desai, 2016). Thus, one could argue that the utility of these models is not that they explain the data, but that they highlight the difficulties of accounting for interspecific differences in u_{bs} with a stabilizing selection model. Although they may be helpful in limited contexts, hypotheses based on number of cell divisions per generation (Thomas *et al*, 2018) do not explain the patterns outlined in Fig 1 either, as nearly the full range of variation in mutation rates per generation is encompassed by unicellular species alone. Nor does generation length explain the patterns, as unicellular eukaryotes have longer cell division times but lower mutation rates than prokaryotes.

To help bridge the gap between theory and data, empirical tests of alternative hypotheses are desirable. Drawing from observations on the yeast *S. cerevisiae*, Liu & Zhang (2021) argued that the facile production of antimutator strains and the low amount of standing genetic variation for the mutation rate relative to the neutral expectation are inconsistent with the predictions of the DBH, and instead support a stabilizing selection hypothesis. However, neither of these arguments is correct. As noted above, the DBH actually predicts the recurrent production of antimutator alleles and simply postulates that at equilibrium, selection cannot generally promote these to permanent fixation. In addition, virtually all models of directional selection predict a strong reduction in standing levels of variation for quantitative traits (Walsh & Lynch, 2018), in many cases stronger than the expectation under stabilizing selection.

One possible approach to the problem is to perform laboratory experiments with lines engineered to have varying initial mutation rates and maintained at different effective population sizes. Under the DBH, the mutation rates for populations exposed to identical N_e are expected to converge on particular average mutation rates, the levels of which should increase with decreasing N_e. In contrast, under a stabilizing selection model, the average outcomes should be relatively invariant with respect to N_e. Owing to the stochastic nature of mutations, such experiments would need to be based on large numbers of replicates and proceed for thousands of generations. Ideally, such experiments would also be carried out in stable and fluctuating environments. Some work of this sort has been done with experimental populations of *E. coli*, with the results being broadly consistent with the DBH, but also with the

In need of answers

Despite the enormous progress that has been made in our understanding of mutation rates over the past two decades, a number of difficult questions remain to be answered.

- What are the mechanistic determinants of the slopes and elevations of the scaling relationships outlined in Fig 1? Under the DBH, the strength of selection operating on modifiers of the mutation rate is a function of the distribution of fitness effects of deleterious mutations. However, the effects of most mutations are very small in magnitude and hence difficult to quantify by direct observation, thereby necessitating indirect statistical inference with substantial uncertainties (Eyre-Walker & Keightley, 2007). The observed power-law relationships imply a relatively constant proportional increase in the recurrent deleterious mutation load with decreasing population size, which may in turn be related to changes in genome architecture that accompany changes in N_e. In addition, the location of the drift barrier depends on the relative rates of appearance of mutator and antimutator alleles and their magnitudes of effects, of which we know almost nothing.
- What is the molecular source of mutations at the molecular level? In principle, most mutations may simply be sporadic accidents of fully functional DNA polymerases and repair enzymes. However, because transcription- and translation-error rates are orders of magnitude higher than those at the DNA level (Lynch, 2023), it is possible that many mutations arise when DNA engages with erroneous proteins. Among other things, this would help explain the relatively frequent occurrence of mutational events involving multiple nucleotide sites.
- How do the direct effects of mutations on somatic tissues influence the evolution of the germline mutation rate in multicellular species? Nearly all existing theory on mutation-rate evolution views the process as being an indirect outcome of inherited mutations linked to mutation-rate modifiers. However, in multicellular species, high mutation rates will have direct effects via the damage that they produce in somatic tissues, which in turn will influence germline success. As the same polymerases and repair enzymes are deployed in all cells, and somatic mutation rates are typically at least 10× higher than those in the germline, such direct effects must elevate the strength of selection against mutator alleles, yet the per-generation rates of mutation in multicellular species are in rough accordance with those in microbes once differences in N_e and proteome sizes are taken into consideration.
- What is the appropriate measure of N_e for understanding the response to directional selection on a trait like the mutation rate? The measure of N_e used herein is more precisely referred to as a coalescent N_e, in that it defines the amount of standing variation at neutral sites—under an ideal Wright–Fisher model of population genetics, the average number of generations separating two randomly sampled alleles from a diploid population is $2N_e$, and with each branch incurring mutations at rate u, the average heterozygosity $\simeq .5em4N_eu$. A potential problem here is that there are many possible measures of N_e, all of which are idealized parameters that can differ depending upon the population-genetic feature that is focused upon. For traits under persistent directional selection, the N_e that governs the probability of fixation of mutations is not necessarily the coalescence N_e (Devi *et al*, 2023). In principle, the two may be highly correlated in some settings, but further work is required to determine the degree to which the former is a good surrogate of the latter in different types of evolutionary analyses.
- Finally, if the drift-barrier hypothesis is not the correct explanation for the negative phylogenetic scaling between mutation rates and the coalescent N_e, then we require an alternative narrative for the general patterns outlined in Fig 1B and C. For example, if a stabilizing selection hypothesis is to become a viable alternative, arguments will be required to explain why the optimum mutation rate should increase in species with decreasing N_e but decrease with increasing size of the functional genome.

degree of environmental variation playing a significant role in the degree to which mutator alleles can transiently increase (Wei *et al*, 2022).

Expanded View for this article is available online.

Acknowledgements

This work was supported by National Institutes of Health grant R35-GM122566-01, US Department of Army MURI award W911NF-14-1-0411, National Science Foundation grants MCB-1518060 and DBI-2119963, grant 735927 from the Moore and Simons Foundations to ML, by National Natural Science Foundation of China grant 31961123002, and the Young Taishan Scholars Program of Shandong Province grant tsqn201812024 to HL.

Author contributions

Michael Lynch: Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration. **Farhan Ali:** Data curation; formal analysis; methodology. **Tongtong Lin:** Data curation; formal analysis. **Yaohai Wang:** Data curation; formal analysis. **Jiahao Ni:** Data curation; formal analysis. **Hongan Long:** Data curation; formal analysis; funding acquisition; investigation; methodology.

Disclosure and competing interests statement

The authors declare that they have no conflict of interest.

References

Amster G, Sella G (2016) Life history effects on the molecular clock of autosomes and sex chromosomes. *Proc Natl Acad Sci USA* 113: 1588–1593

Banerjee K, Kolomeisky AB, Igoshin OA (2017) Elucidating interplay of speed and accuracy in biological error correction. *Proc Natl Acad Sci USA* 114: 5183–5188

Behjati S, Huch M, van Boxtel R, Karthaus W, Wedge DC, Tamuri AU, Martincorena I, Petljak M, Alexandrov LB, Gundem G *et al* (2014) Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* 513: 422–425

Bergeron LA, Besenbacher S, Zheng J, Li P, Bertelsen MF, Quintard B, Hoffman JI, Li Z, St. Leger J, Shao C *et al* (2023) Evolution of the germline mutation rate across vertebrates. *Nature* 615: 285–291

Boe L, Danielsen M, Knudsen S, Petersen JB, Maymann J, Jensen PR (2000) The frequency of mutators in populations of *Escherichia coli*. *Mutat Res* 448: 47–55

Buffalo V (2021) Quantifying the relationship between genetic diversity and population size suggests natural selection cannot explain Lewontin's paradox. *Elife* 10: e67509

Bulmer M (1991) The selection-mutation-drift theory of synonymous codon usage. *Genetics* 129: 897–907

Cagan A, Baez-Ortega A, Brzozowska N, Abascal F, Coorens THH, Sanders MA, Lawson ARJ, Harvey LMR, Bhosle S, Jones D *et al* (2022) Somatic mutation rates scale with lifespan across mammals. *Nature* 604: 517–524

Castañeda-Garc a A, Mart n-Blecua I, Cebrián-Sastre E, Chiner-Oms A, Torres-Puente M, Comas I, Blázquez J (2020) Specificity and mutagenesis bias of the mycobacterial alternative mismatch repair analyzed by mutation accumulation studies. *Sci Adv* 6: eaay4453

Chan K, Gordenin DA (2015) Clusters of multiple mutations: incidence and molecular mechanisms. *Annu Rev Genet* 49: 243–267

Charlesworth B (2009) Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat Rev Genet* 10: 195–205

Charlesworth B, Jensen JD (2022) How can we resolve Lewontin's paradox? *Genome Biol Evol* 14: evac096

Charlesworth B, Jensen JD (2023) Population genetic considerations regarding evidence for biased mutation rates in *Arabidopsis thaliana*. *Mol Biol Evol* 40: msac275

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

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

Cho JE, Jinks-Robertson S (2017) Ribonucleotides and transcription-associated mutagenesis in yeast. *J Mol Biol* 429: 3156–3167

Dawson KJ (1999) The dynamics of infinitesimally rare alleles, applied to the evolution of mutation rates and the expression of deleterious mutations. *Theor Popul Biol* 55: 1–22

Denamur E, Matic I (2006) Evolution of mutation rates in bacteria. *Mol Microbiol* 60: 820–827

Desai MM, Fisher DS (2007) Beneficial mutation selection balance and the effect of linkage on positive selection. *Genetics* 176: 1759–1798

Devi A, Speyer G, Lynch M (2023) The divergence of mean phenotypes under persistent directional selection. *Genetics* 224: iyad091

DeWitt WS, Harris KD, Ragsdale AP, Harris K (2021) Nonparametric coalescent inference of mutation spectrum history and demography. *Proc Natl Acad Sci USA* 118: e2013798118

Drake JW (1991) A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci USA* 88: 7160–7164

Drake JW (2007) Too many mutants with multiple mutations. *Crit Rev Biochem Mol Biol* 42: 247–258

Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutation. *Genetics* 148: 1667–1686

Earl DJ, Deem MW (2004) Evolvability is a selectable trait. *Proc Natl Acad Sci USA* 101: 11531–11536

Elango N, Thomas JW, NISC Comparative Sequencing Program, Yi SV (2006) Variable molecular clocks in hominoids. *Proc Natl Acad Sci USA* 103: 1370–1375

Eyre-Walker A, Keightley PD (2007) The distribution of fitness effects of new mutations. *Nature Rev Genet* 8: 610–618

Foster PL (2007) Stress-induced mutagenesis in bacteria. *Crit Rev Biochem Mol Biol* 42: 373–397

Foster PL, Hanson AJ, Lee H, Popodi EM, Tang H (2013) On the mutational topology of the bacterial genome. *G3* 3: 399–407

Frigola J, Sabarinathan R, Mularoni L, Muiños F, Gonzalez-Perez A, López-Bigas N (2017) Reduced mutation rate in exons due to differential mismatch repair. *Nat Genet* 49: 1684–1692

Galhardo RS, Hastings PJ, Rosenberg SM (2007) Mutation as a stress response and the regulation of evolvability. *Crit Rev Biochem Mol Biol* 42: 399–435

Galtier N, Rousselle M (2020) How much does N_e vary among species? *Genetics* 216: 559–572

Gao Z, Zhang Y, Cramer N, Przeworski M, Moorjani P (2023) Limited role of generation time changes in driving the evolution of the mutation spectrum in humans. *Elife* 12: e81188

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

Good BH, Desai MM (2016) Evolution of mutation rates in rapidly adapting asexual populations. *Genetics* 204: 1249–1266

Gou L, Bloom JS, Kruglyak L (2019) The genetic basis of mutation rate variation in yeast. *Genetics* 211: 731–740

Habig M, Lorrain C, Feurtey A, Komluski J, Stukenbrock EH (2021) Epigenetic modifications affect the rate of spontaneous mutations in a pathogenic fungus. *Nat Commun* 12: 5869

Haldane JBS (1935) The rate of spontaneous mutation of a human gene. *J Genet* 31: 317–326

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

Harris K, Pritchard JK (2017) Rapid evolution of the human mutation spectrum. *Elife* 6: e24284

Heidenreich E (2007) Adaptive mutation in *Saccharomyces cerevisiae*. *Crit Rev Biochem Mol Biol* 42: 285–311

Hicks WM, Kim M, Haber JE (2010) Increased mutagenesis and unique mutation signature associated with mitotic gene conversion. *Science* 329: 82–85

Hopfield J (1974) Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc Natl Acad Sci USA* 71: 4135–4139

Huang Y, Gu L, Li GM (2018) H3K36me3-mediated mismatch repair preferentially protects actively transcribed genes from mutation. *J Biol Chem* 293: 7811–7823

Jiang P, Ollodart AR, Sudhesh V, Herr AJ, Dunham MJ, Harris K (2021) A modified fluctuation assay reveals a natural mutator phenotype that drives mutation spectrum variation within *Saccharomyces cerevisiae*. *Elife* 10: e68285

Johnson T, Barton NH (2002) The effect of deleterious alleles on adaptation in asexual populations. *Genetics* 162: 395–411

Kang JM, Iovine NM, Blaser MJ (2006) A paradigm for direct stress-induced mutation in prokaryotes. *FASEB J* 20: 2476–2485

Kessler MD, Loesch DP, Perry JA, Heard-Costa NL, Taliun D, Cade BE, Wang H, Daya M, Ziniti J, Datta S *et al* (2020) *De novo* mutations across 1,465 diverse genomes reveal mutational insights and reductions in the Amish founder population. *Proc Natl Acad Sci USA* 117: 2560–2569

Kimura M (1967) On the evolutionary adjustment of spontaneous mutation rates. *Genet Res* 9: 23–34

Kimura M (1983) *The neutral theory of molecular evolution*. Cambridge: Cambridge Univ Press

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

Kondrashov AS (1995) Modifiers of mutation-selection balance: general approach and the evolution of mutation rates. *Genet Res* 66: 53–70

Kondrashov AS (2003) Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum Mutat* 21: 12–27

Lang GI, Murray AW (2011) Mutation rates across budding yeast chromosome VI are correlated with replication timing. *Genome Biol Evol* 3: 799–811

Leigh EG Jr (1970) Natural selection and mutability. *Amer Natur* 104: 301–305

Lewontin RC (1974) *The genetic basis of evolutionary change*. New York, NY: Columbia University Press

Li WH (1987) Models of nearly neutral mutations with particular implications for non-random usage of synonymous codons. *J Mol Evol* 24: 337–345

Li C, Luscombe NM (2020) Nucleosome positioning stability is a modulator of germline mutation rate variation across the human genome. *Nat Commun* 11: 1363

Li F, Mao C, Tong D, Huang J, Gu L, Yang W, Li GM (2013) The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSα. *Cell* 153: 590–600

Liu H, Zhang J (2021) The rate and molecular spectrum of mutation are selectively maintained in yeast. *Nat Commun* 12: 4044

Liu H, Zhang J (2022) Is the mutation rate lower in genomic regions of stronger selective constraints? *Mol Biol Evol* 39: msac169

Long H, Sung W, Miller SF, Ackerman MS, Doak TG, Lynch M (2015) Mutation rate, spectrum, topology, and context-dependency in the DNA mismatch-repair deficient *Pseudomonas fluorescens* ATCC948. *Genome Biol Evol* 7: 262–271

Long H, Miller SF, Williams E, Lynch M (2018a) Specificity of the DNA mismatch repair system (MMR) and mutagenesis bias in bacteria. *Mol Biol Evol* 35: 2414–2421

Long H, Sung W, Kucukyildirim S, Williams E, Miller SF, Guo W, Patterson C, Gregory C, Strauss C, Stone C *et al* (2018b) Evolutionary determinants of genome-wide nucleotide composition. *Nat Ecol Evol* 2: 237–240

López-Cortegano E, Craig RJ, Chebib J, Samuels T, Morgan AD, Kraemer SA, Böndel KB, Ness RW, Colegrave N, Keightley PD (2021) *De novo* mutation rate variation and its determinants in *Chlamydomonas*. *Mol Biol Evol* 38: 3709–3723

Lujan SA, Kunkel TA (2021) Stability across the whole nuclear genome in the presence and absence of DNA mismatch repair. *Cell* 10: 1224

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

Lynch M (2007) *The origins of genome architecture*. Sunderland, MA: Sinauer Assocs, Inc

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

Lynch M (2009) Rate, molecular spectrum, and consequences of spontaneous mutations in man. *Proc Natl Acad Sci U S A* 107: 961–968

Lynch M (2010) Evolution of the mutation rate. *Trends Genet* 26: 345–352

Lynch M (2011) The lower bound to the evolution of mutation rates. *Genome Biol Evol* 3: 1107–1118

Lynch M (2012) Evolutionary layering and the limits to cellular perfection. *Proc Natl Acad Sci USA* 109: 18851–18856

Lynch M (2023) *Evolutionary cell biology: the origins of cellular features*. Oxford: Oxford University Press

Lynch M, Walsh JB (1998) *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer Assocs, Inc

Lynch M, Ackerman M, Gout JF, Long H, Sung W, Thomas WK, Foster PL (2016) Genetic drift, selection, and evolution of the mutation rate. *Nat Rev Genet* 17: 704–714

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

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

Malkova A, Haber JE (2012) Mutations arising during repair of chromosome breaks. *Annu Rev Genet* 46: 455–473

Martincorena I, Seshasayee AS, Luscombe NM (2012) Evidence of non-random mutation rates suggests an evolutionary risk management strategy. *Nature* 485: 95–98

Matsuba C, Ostrow DG, Salomon MP, Tolani A, Baer CF (2013) Temperature, stress and spontaneous mutation in *Caenorhabditis briggsae* and *Caenorhabditis elegans*. *Biol Lett* 9: 20120334

McDonald MJ, Rice DP, Desai MM (2016) Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature* 531: 233–236

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

Milligan WR, Amster G, Sella G (2022) The impact of genetic modifiers on variation in germline mutation rates within and among human populations. *Genetics* 221: iyac087

Monroe JG, Murray KD, Xian W, Carbonell-Bejerano P, Fenster CB, Weigel D (2022a) Report of mutation biases mirroring selection in *Arabidopsis thaliana* unlikely to be entirely due to variant calling errors. *bioRxiv* https://doi.org/10.1101/2022.08.21.504682 [PREPRINT]

Monroe JG, Srikant T, Carbonell-Bejerano P, Becker C, Lensink M, Exposito-Alonso M, Klein M, Hildebrandt J, Neumann M, Kliebenstein D *et al* (2022b) Mutation bias reflects natural selection in *Arabidopsis thaliana*. *Nature* 602: 101–105

Nei M (1983) Genetic polymorphism and the role of mutation in evolution. In *Evolution of genes and proteins*, Nei M, Koehn RK (eds), pp 165–190. Sunderland, MA: Sinauer Assocs, Inc.

Ninio J (1975) Kinetic amplification of enzyme discrimination. *Biochimie* 57: 587–595

Orr HA (2000) The rate of adaptation in asexuals. *Genetics* 155: 961–968

Park C, Qian W, Zhang J (2012) Genomic evidence for elevated mutation rates in highly expressed genes. *EMBO Rep* 13: 1123–1129

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

Poetsch AR, Boulton SJ, Luscombe NM (2018) Genomic landscape of oxidative DNA damage and repair reveals regioselective protection from mutagenesis. *Genome Biol* 19: 215

Qian H (2006) Reducing intrinsic biochemical noise in cells and its thermodynamic limit. *J Mol Biol* 362: 387–392

Radman M, Taddei F, Matic I (2000) Evolution-driving genes. *Res Microbiol* 151: 91–95

Rahbari R, Wuster A, Lindsay SJ, Hardwick RJ, Alexandrov LB, Turki SA, Dominiczak A, Morris A, Porteous D, Smith B *et al* (2016) Timing, rates and spectra of human germline mutation. *Nat Genet* 48: 126–133

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

Raynes Y, Sniegowski PD (2014) Experimental evolution and the dynamics of genomic mutation rate modifiers. *Heredity* 113: 375–380

Rosche WA, Foster PL (2000) Determining mutation rates in bacterial populations. *Methods* 20: 4–17

Rosenberg SM (2001) Evolving responsively: adaptive mutation. *Nat Rev Genet* 2: 504–515

Rosenberg SM, Shee C, Frisch RL, Hastings PJ (2012) Stress-induced mutation via DNA breaks in *Escherichia coli*: a molecular mechanism with implications for evolution and medicine. *Bioessays* 34: 885–892

Sasani TA, Ashbrook DG, Beichman AC, Lu L, Palmer AA, Williams RW, Pritchard JK, Harris K (2022) A natural mutator allele shapes mutation spectrum variation in mice. *Nature* 605: 497–502

Scally A, Durbin R (2012) Revising the human mutation rate: implications for understanding human evolution. *Nat Rev Genet* 13: 745–753

Schrider DR, Hourmozdi JN, Hahn MW (2011) Pervasive multinucleotide mutational events in eukaryotes. *Curr Biol* 21: 1051–1054

Schrider DR, Houle D, Lynch M, Hahn MW (2013) Rates and genomic consequences of spontaneous mutational events in *Drosophila melanogaster*. *Genetics* 194: 937–954

Seoighe C, Scally A (2017) Inference of candidate germline mutator loci in humans from genome-wide haplotype data. *PLoS Genet* 13: e1006549

Seplyarskiy VB, Bazykin GA, Soldatov RA (2015) Polymerase ζ activity is linked to replication timing in humans: evidence from mutational signatures. *Mol Biol Evol* 32: 3158–3172

Shapiro BJ (2016) How clonal are bacteria over time? *Curr Opin Microbiol* 31: 116–123

Sharp NP, Agrawal AF (2012) Evidence for elevated mutation rates in low-quality genotypes. *Proc Natl Acad Sci USA* 109: 6142–6146

Shewaramani S, Finn TJ, Leahy SC, Kassen R, Rainey PB, Moon CD (2017) Anaerobically grown *Escherichia coli* has an enhanced mutation rate and distinct mutational spectra. *PLoS Genet* 13: e1006570

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

Stamatoyannopoulos JA, Adzhubei I, Thurman RE, Kryukov GV, Mirkin SM, Sunyaev SR (2009) Human mutation rate associated with DNA replication timing. *Nat Genet* 41: 393–395

Sturtevant AH (1937) Essays on evolution. I. On the effects of selection on the mutation rate. *Q Rev Biol* 12: 464–476

Sung W, Ackerman MS, Miller SF, Doak TG, Lynch M (2012) The drift-barrier hypothesis and mutation-rate evolution. *Proc Natl Acad Sci USA* 109: 18488–18492

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

Supek F, Lehner B (2017) Clustered mutation signatures reveal that error-prone DNA repair targets mutations to active genes. *Cell* 170: 534–547.e23

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

Takemoto N, Numata I, Su'etsugu M, Miyoshi-Akiyama T (2018) Bacterial EndoMS/NucS acts as a clamp-mediated mismatch endonuclease to prevent asymmetric accumulation of replication errors. *Nucleic Acids Res* 46: 6152–6165

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

Terekhanova NV, Bazykin GA, Neverov A, Kondrashov AS, Seplyarskiy VB (2013) Prevalence of multinucleotide replacements in evolution of primates and *Drosophila*. *Mol Biol Evol* 30: 1315–1325

Thomas GWC, Wang RJ, Puri A, Harris RA, Raveendran M, Hughes DST, Murali SC, Williams LE, Doddapaneni H, Muzny DM *et al* (2018) Reproductive longevity predicts mutation rates in primates. *Curr Biol* 28: 3193–3197.e5

Thornlow BP, Hough J, Roger JM, Gong H, Lowe TM, Corbett-Detig RM (2018) Transfer RNA genes experience exceptionally elevated mutation rates. *Proc Natl Acad Sci USA* 115: 8996–9001

Turrientes MC, Baquero F, Levin BR, Mart nez JL, Ripoll A, González-Alba JM, Tobes R, Manrique M, Baquero MR, Rodr guez-Dom nguez MJ *et al* (2013) Normal mutation rate variants arise in a mutator (Mut S) *Escherichia coli* population. *PloS One* 8: e72963

Uphoff S, Lord ND, Okumus B, Potvin-Trottier L, Sherratt DJ, Paulsson J (2016) Stochastic activation of a DNA damage response causes cell-to-cell mutation rate variation. *Science* 351: 1094–1097

Walsh JB, Lynch M (2018) *Evolution and selection of quantitative traits*. Oxford: Oxford University Press

Wang Y, Obbard DJ (2023) Experimental estimates of germline mutation rate in eukaryotes: a phylogenetic meta-analysis. *Evol Lett* 7: 216–226

Wang Y, McNeil P, Abdulazeez R, Pascual M, Johnston SE, Keightley PD, Obbard D (2023) Variation in mutation, recombination, and transposition rates in *Drosophila melanogaster* and *Drosophila simulans*. *Genome Res* 33: 587–598

Wei W, Ho WC, Behringer MG, Miller SF, Bcharah G, Lynch M (2022) Rapid evolution of mutation rate and spectrum in response to environmental and population-genetic challenges. *Nat Commun* 13: 4752

Wielgoss S, Barrick JE, Tenaillon O, Wiser MJ, Dittmar WJ, Cruveiller S, Chane-Woon-Ming B, Médigue C, Lenski RE, Schneider D (2013) Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc Natl Acad Sci USA* 110: 222–227

Williams LN, Herr AJ, Preston BD (2013) Emergence of DNA polymerase ε antimutators that escape error-induced extinction in yeast. *Genetics* 193: 751–770

Wu FL, Strand AI, Cox LA, Ober C, Wall JD, Moorjani P, Przeworski M (2020) A comparison of humans and baboons suggests germline mutation rates do not track cell divisions. *PLoS Biol* 18: e3000838