APPLICATIONS OF NEXT-GENERATION SEQUENCING

OPINION

Genetic drift, selection and the evolution of the mutation rate

Michael Lynch, Matthew S. Ackerman, Jean-Francois Gout, Hongan Long, Way Sung, W. Kelley Thomas and Patricia L. Foster

Abstract | As one of the few cellular traits that can be quantified across the tree of life, DNA-replication fidelity provides an excellent platform for understanding fundamental evolutionary processes. Furthermore, because mutation is the ultimate source of all genetic variation, clarifying why mutation rates vary is crucial for understanding all areas of biology. A potentially revealing hypothesis for mutation-rate evolution is that natural selection primarily operates to improve replication fidelity, with the ultimate limits to what can be achieved set by the power of random genetic drift. This drift-barrier hypothesis is consistent with comparative measures of mutation rates, provides a simple explanation for the existence of error-prone polymerases and yields a formal counter-argument to the view that selection fine-tunes gene-specific mutation rates.

As mutation affects essentially every aspect of biology, the development of a unifying theory for mutation-rate evolution is highly desirable. There is much to be explained. For although the base-substitution mutation rate (u) in all organisms is low ($<10^{-7}$ mutations per nucleotide site per generation), the rates in some species are more than 1,000-fold below this level. This large range of variation implies that the accuracy of DNA replication and repair in most, if not all, species is less than what is possible at the biochemical level.

It has been argued that mutation rates, even at the single-gene level, have been fine-tuned by natural selection to maximize long-term survival and evolvability[1–4], yet there is no direct empirical or theoretical evidence that this is generally the case. If such adaptive mutation-rate arguments are valid, they will need to explain why the evolved mutation rate in microorganisms is 100- to 1,000-fold lower than that in vertebrates. Moreover, although stress-induced mutagenesis in microorganisms can sometimes provide a transient mechanism for generating an adaptive genotype in an extreme environment[5], this need not imply that the error-prone nature of the polymerases invoked during such times has been promoted by natural selection[6–8].

As whole-genome sequencing (WGS) has led to a rapid proliferation of data on mutation rates in a wide array of phylogenetic lineages, there is a need to evaluate how this information can be integrated into a general evolutionary framework. We start with a brief overview of the theory of mutation-rate evolution, followed by a comparison of the existing data with theoretical expectations. Confronted with difficulties in reconciling observations with adaptive mutation-rate hypotheses, at both the whole-genome and single-gene levels, we argue that phylogenetic variation in mutation rates reflects underlying differences in the efficiency of selection to improve replication fidelity, which in turn results from variation in the power of random genetic drift. There is still considerable room for work on the cellular determinants of replication fidelity and how these vary across phylogenetic lineages, but achieving evolutionary understanding in this area is unlikely to be served by uncritical adherence to the idea that every aspect of genome stability is refined by adaptive processes.

Selection, drift and mutation rate

A formal theoretical framework for understanding mutation-rate evolution was first presented by Kimura[9]. Noting that the vast majority of mutations are deleterious, he proposed that mutator alleles are indirectly selected against through associations with the detrimental alleles that they generate elsewhere in the genome. Under this view, a newly arisen mutator allele progressively acquires an excess equilibrium mutation load that is defined by the opposing forces of input (mutation pressure) and removal (selection and recombination). Here, we broadly define a mutator (and an antimutator) as any genomically encoded modifier of the mutation rate; for example, a variant of a DNA polymerase, a DNA repair protein or even a gene product that alters the mutagenicity of the intracellular environment. As discussed below, some genomic features can also have very localized effects on the mutation rate.

The reduction in fitness associated with a mutator allele is equal to the product of three terms: the excess genome-wide rate of production of deleterious mutations relative to the population mean (ΔU_D); the reduction in fitness per mutation (s); and the average number of generations that a mutation remains associated with the mutator ($\bar{t}$). The persistence time $\bar{t}$ is estimated by noting that an association between a mutator and an induced mutation is eliminated by selection at rate s per generation and by recombination at rate r (REF. 9).

Under asexual reproduction, the persistence time is the reciprocal of the mutational effect ($\bar{t} = 1/s$), and the selective disadvantage of a mutator allele (s_m) is simply equal to the increased rate of production of deleterious mutations ΔU_D. However, recombination weakens the selective disadvantage of a mutator allele by exporting mutations to other members of the population[9–11]. With free recombination ($r = 0.5$), mutant alleles are statistically uncoupled from their source in an average $\bar{t} = 2$ generations, so the mutator-allele disadvantage is $s_m \approx 2s\Delta U_D$, and no more than twice this with stronger linkage[6].

Substantial work on the fitness effects of mutations[12–15] suggests average values of $s < 0.1$, implying that recombination reduces the strength of selection operating on the mutation rate by at least 80%.

To understand the degree to which selection can depress the mutation rate, it must be appreciated that whenever a selection coefficient (in this case, s_m) drops below the power of random genetic drift, which is inversely proportional to the effective population size (N_e), evolution enters the realm of effective neutrality. Thus, letting λ be the fraction by which the mutation rate is reduced by an antimutator, so that $\Delta U_D = \lambda U_D$, once the genome-wide deleterious mutation rate U_D is reduced to the point that $\lambda U_D < 1/N_e$, selection for further reduction in the mutation rate in a haploid asexual population will be overwhelmed by stochastic noise (this point will be reached at an earlier point in a sexual population, owing to the multiplier 2s). A potential complication arises if λ changes with U_D, although this does not change the general principle being espoused. Simply stated, the drift-barrier hypothesis postulates that owing to the stochastic nature of evolutionary processes, and the diminishing advantages of increased perfection of molecular adaptations, there is an upper bound to the level of refinement that can be achieved by selection. This crucial point arises when the selective advantage of further improvement is overwhelmed by the power of random genetic drift and/or biased mutation pressure towards mutators[6,12,16] (FIG. 1). If this hypothesis is correct, the very process that is necessary for the production of adaptive mutations is selected against, with the fuel for evolution simply being an inadvertent by-product of an imperfect process.

The preceding analyses ignore the possibility of mutator and/or antimutator alleles having direct effects on fitness. One common argument for direct effects on fitness assumes that excessive investment in replication fidelity reduces the rate or increases the energetic cost of genome replication[9–11,17–23]; for example, as a consequence of increased investment in proofreading. However, several indirect lines of evidence suggest otherwise: the total cost of genome replication generally constitutes less than 1% of the total cellular energy budget, of which only a small fraction is associated with polymerization and proofreading[24]; the rates of DNA replication are more than tenfold faster than those for mRNA elongation[25,26], suggesting that the time to replication does not limit cell-division rates; and in bacteria, cell-division times are either unchanged or decline with increasing genome size[24,27,28]. In addition, eukaryotes minimize replication time by populating chromosomes with multiple origins of replication. Conversely, in multicellular species the direct consequences of somatic mutations can reinforce the effects of a mutator by reducing survivorship[6,29].

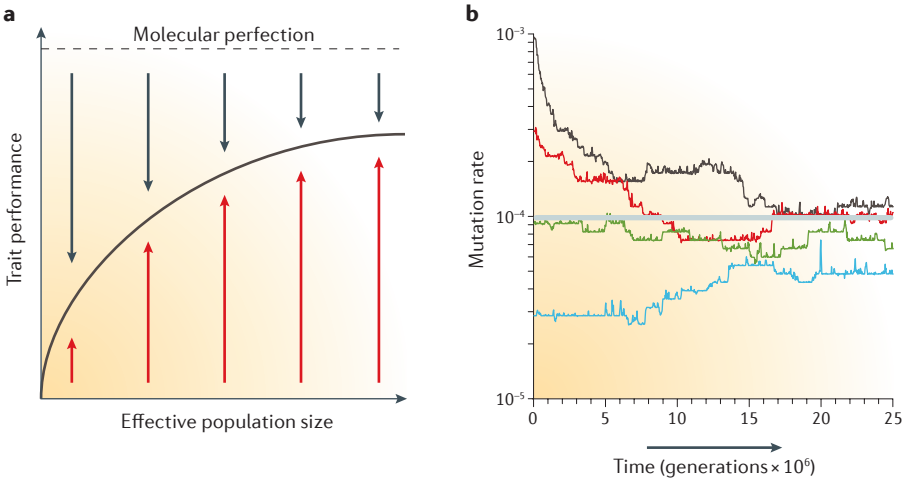

Figure 1 | **The drift-barrier hypothesis for mutation-rate evolution. a** | The level of perfection of any molecular trait is expected to become more refined in larger populations, as the efficiency of selection (red arrows) increases and the stochastic effects of genetic drift (black arrows) decrease. The achievable level of molecular perfection is also a function of the degree to which a mutation alters the trait in the upward versus downward direction; in this example, there is a downward mutation bias, so that perfection is unachievable at even the largest population sizes (asymptote of the solid black line). **b** | An example of how the evolutionary dynamics of the mutation rate are predicted to unfold under the drift-barrier hypothesis[7]. With high initial mutation rates (black, red), the advantage of antimutators is sufficiently large that the mutation rate will evolve downwardly until the strength of selection is matched by the opposing pressure from random genetic drift and mutation bias. In contrast, with a sufficiently low initial mutation rate (blue), antimutators are insufficiently advantageous to be promoted, mild mutators are insufficiently deleterious to be eliminated and the population evolves a higher mutation rate until the drift barrier (grey line) is reached. A population starting at the drift barrier (green) is expected to maintain a steady-state distribution of the mutation rate with some temporal fluctuations arising owing to the stochasticity of drift. Part **b** is adapted from REF. 7; Lynch, M. The lower bound to the evolution of mutation rates. *Genome Biol. Evol.* (2011) **3**, 1107–1118, by permission of Oxford University Press.

Estimation of mutation rates

Given its very low level, the base-substitution mutation rate (u), has historically been one of the most difficult genetic parameters to measure accurately. Until recently, most estimates were obtained either with reporter constructs that yield an obvious phenotype upon mutation or with measures of interspecific divergence of putatively neutral markers. Although readily applied to microorganisms[30,31], the reporter-construct method has two major disadvantages: uncertainty in the fraction of mutations yielding detectable phenotypes and uncertainty of whether the location, orientation and/or level of transcriptional activity influence the rate of mutation. The phylogenetic approach[32] suffers from uncertainties associated with divergence times and the degree to which the fates of mutations are modified by selection.

With the advent of high-throughput genomic sequencing, both approaches can now be avoided entirely for most species. The method of choice is a conceptually simple mutation-accumulation (MA) experiment in which a series of initially identical lines is subject to a consistent regime of extreme population bottlenecks (single individuals for clonal reproducers and self-fertilizing hermaphrodites, and single full-sib pairs for species with separate sexes) (FIG. 2). With such a protocol, the power of genetic drift is so overwhelming that nearly all mutations — except those causing complete sterility or lethality — accumulate in an effectively neutral fashion, even in the case of bacterial colonies that briefly expand before single-cell bottlenecks[33]. After an adequate number of generations of propagation, WGS reveals the unique set of mutations sequestered within the individual lines. The resultant

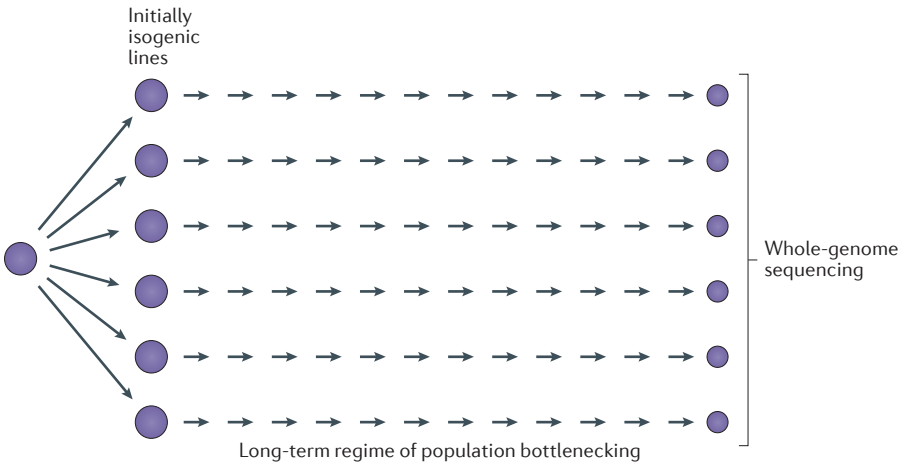

Figure 2 | **The design of a mutation-accumulation experiment.** Initially isogenic lines of organisms are subjected to a large number of sequential episodes of population bottlenecking, in which a single individual or small numbers of individuals are isolated in each round and used to seed the subsequent growth of each replicate line (shown as horizontal arrows). Under this treatment, the fitness of each line generally declines owing to the accumulation of deleterious mutations. Following the mutation accumulation (MA) procedure, the complete genomes of replicate lines are sequenced. Under the assumption of effectively neutral MA, the average rate of increase in numbers of mutations per line is equivalent to the mutation rate.

molecular spectrum of mutations covers essentially every type of alteration relevant to evolutionary change; that is, all possible base-substitutional changes, small insertions and deletions (indels), large segmental duplications and deletions, and other large-scale chromosomal rearrangements.

The MA-WGS approach is time consuming, but is readily applied to most species. However, as sequencing costs continue to drop, variants of this method are now being applied to animal species (for example, genomic sequencing of parent–offspring trios and/or members of extended pedigrees)[34–38]. Owing to the small number of generations involved, this approach can require the sequencing of additional individuals to achieve an adequate pool of mutations, but the use of multiple families also provides opportunities for studying population-level variation in mutational features.

Phylogenetic variation in mutation rate

In a first attempt to understand mutation-rate variation among species, Drake[30] argued that the base-substitution mutation rate, u, scales inversely with the number of bases in the genome, such that there is an approximately constant 0.003 mutations per genome per generation. Drake argued, "Because this rate is uniform in such diverse organisms, it is likely to be determined by deep general forces" (REF. 30). We argue below that these deep forces are to be found in the population-genetic factors that govern all evolutionary processes. Drake's conjecture was based on a very small number of taxa — three bacteriophages, the bacterium *Escherichia coli*, the budding yeast *Saccharomyces cerevisiae* and the filamentous fungus *Neurospora crassa* — only three of which are cellular in nature. Given this limited taxonomic coverage and the reliance on reporter constructs, a re-evaluation of the proposed pattern with MA-WGS results is warranted.

General scaling relationships for mutation rates. A pooled analysis of results for both bacterial and eukaryotic microorganisms yields a linear regression of u on genome size with a slope of –0.61 (standard error of the mean (s.e.m.) = 0.23) on a logarithmic scale ($r^2 = 0.32$; $p = 0.017$), roughly compatible with Drake's expectation of –1.0 (FIG. 3a). However, the negative scaling for bacteria is entirely dependent on the single value for the highly mutable *Mesoplasma florum*[39]. The slope for microbial eukaryotes alone is not statistically significant. Thus, Drake's 0.003 rule is at best a rough approximation for microorganisms. Moreover, the expected pattern is strongly violated when multicellular eukaryotes are included, as u scales positively with genome size within this group (FIG. 3a).

As there is no known molecular mechanism directly linking u with total genome size, the patterns outlined in FIG. 3a probably result indirectly from both genome sizes and mutation rates covarying with another underlying feature. One such factor is the effective population size N_e, which is positively associated with genome size in eukaryotes[40]. The N_e of a species can be estimated from the levels of standing variation (population-level heterozygosity) at silent sites in protein-coding genes, which under the assumptions of effective neutrality and drift-mutation equilibrium have expected values of xN_eu, where x is equal to two for haploid and four for diploid species[41]. Using species-wide estimates of silent-site variation, u can be factored out to yield an estimate of N_e.

For each of three major biological groupings of species, there is a significant negative relationship between u and N_e on a logarithmic scale: in bacteria, $r^2 = 0.86$, slope = –0.76 (s.e.m. = 0.10); in unicellular eukaryotes, $r^2 = 0.69$, slope = –0.70 (s.e.m. = 0.23); and in multicellular eukaryotes, $r^2 = 0.44$, slope = –0.33 (s.e.m. = 0.12) (FIG. 3b). However, the normalization constants of these three regressions are different, with bacteria having a 5- to 20-fold elevation in the mutation rate relative to unicellular eukaryotes with the same N_e.

To place these results in the context of the drift-barrier hypothesis, it is necessary to go beyond the per-site base-substitution mutation rate, u, as selection operates on the genome-wide deleterious mutation rate, U_D. Precise estimates of U_D cannot be achieved with molecular data alone, but an approximation (U_p), obtained from the product of u and the number of nucleotide sites in the exons of protein-coding genes (P), should be proportional to U_D provided that the mass of selected sites outside coding regions scales with U_D. Such a transformation leads to a reasonably unified behaviour of the mutation rate ($r^2 = 0.86$; slope = –0.74 (s.e.m. = 0.06)) (FIG. 3c), with all three phylogenetic groups falling along the same regression line. The overall data are therefore qualitatively compatible with the hypothesis that U_D decreases with increasing N_e. Contrary to a previous claim[42], P alone does not dictate mutation-rate evolution.

Can a more quantitative expectation be attached to the scaling of U_D and N_e under the drift-barrier hypothesis? In BOX 1, we present theory that specifically predicts an inverse relationship between these two parameters; that is, a slope of approximately –1.0 on a log–log plot, provided that the drift barrier itself and/or biased mutation pressure towards mutator alleles keeps the mutation rates of most species far from

the biophysical limits. The following reasoning suggests that the data are compatible with such scaling.

The analysis shown in FIG. 3c relies on protein-coding DNA as a surrogate measure of the number of nucleotides under selection (the effective genome size, G_e), but not all amino acid replacement mutations or silent-site changes are likely to be deleterious, and some non-coding sites are involved in transcription, splicing and so on. Greater than 90% of most bacterial genomes is protein coding, and there can be considerable selection on silent sites in such species with a large N_e, so P probably provides a close approximation of G_e in these species. The approximation is also relatively good in *S. cerevisiae*, as $G_e \approx 0.8P$ (REF. 43). The average estimates of G_e for nematodes, *Drosophila melanogaster* and humans are $1.1P$, $3.3P$ and $5.8P$, respectively[43–47]. Thus, there is a tendency for P to increasingly underestimate the amount of the genome under selection with increasing organismal complexity. However, accounting for these levels of bias only moves the negative scaling of the genome-wide deleterious mutation rate with N_e closer to a slope of −1, improving the overall concordance of the data with the drift-barrier hypothesis. Additional data will be required to determine whether there is a diminishing response of U_D in taxa with very high N_e, as might be expected if such species approach the biophysical limit of replication fidelity.

The theory outlined in BOX 1 also predicts detectable levels (certainly two- to threefold) of variation in the mutation rate among lineages experiencing identical levels of selection and random genetic drift, owing to the stochastic production and fixation of mutator and antimutator alleles around the expected mean. Thus, given that N_e also probably changes through time and is estimated with inaccuracies, the range of variation around the regression in FIG. 3c should not be taken as a shortcoming of the theory.

Finally, we note that although the theory suggests a specific scaling between U_D and N_e, it does not provide an explicit statement on the elevation of the curve; that is, for the absolute values of the expected mutation rates. This is because mutation-rate evolution is not simply a function of the efficiency of selection, but also of the relative rates at which mutator and antimutator alleles arise (BOX 1). It is likely that antimutator alleles arise substantially less frequently than mutator alleles, although quantitative information on this matter is lacking. Mutation bias towards mutators will oppose the evolutionary reduction in the mutation rate by selection, pushing the mutation rate to a higher level than otherwise expected (FIG. 4). Thus, species-specific variation in the relative rates of production of mutator and/or antimutator alleles may be an additional contributor to the vertical range of variation in the plots in FIG. 3.

Error-prone polymerases. The genomes of nearly all organisms encode for error-prone polymerases. As many of these are used only during times of stress, this has encouraged the assertion that high error rates associated with such enzymes have been promoted by selection as a means for generating adaptive mutations in changing environments[1,2,48–51]. As noted above, however, although increasing the mutation rate may enhance the absolute probability of a beneficial mutation, it also magnifies the background production of deleterious mutations, and thus can ultimately lead to long-term genomic deterioration[52,53].

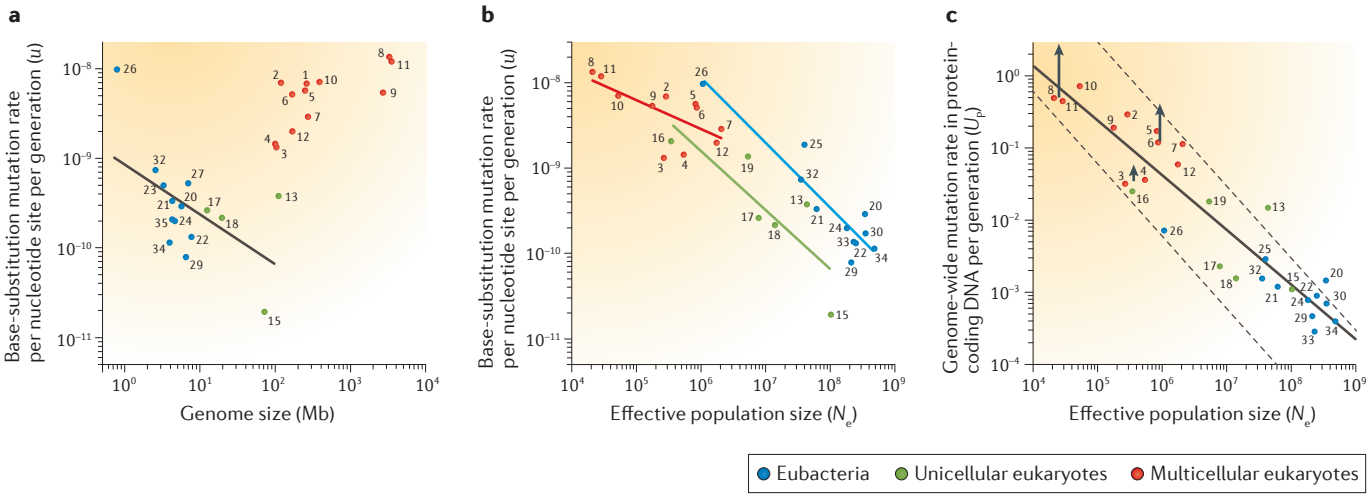

Figure 3 | **Scaling relationships involving the base-substitution mutation rate. a** | The relationship of the base-substitution mutation rate per nucleotide site per generation (u) with total haploid genome size is given for the full set of species for which data are available from mutation-accumulation whole-genome sequencing (MA-WGS) or pedigree analyses. The regression line only incorporates the data for unicellular species. **b** | The regression of u on the estimated effective population size (N_e). To increase the sample size here, the mutation rates of three bacteria (data points 25, 30 and 33) and two unicellular eukaryotes (data points 16 and 19) are based on reporter-construct estimates. **c** | The regression of the total (genome-wide) mutation rate in protein-coding DNA per generation (U_P) on N_e. The solid line is the regression fitted to the full data set, whereas the dashed lines are reference lines with slopes equal to −1.0. The arrows are the approximate degree to which the multicellular eukaryote measures are likely to move upwardly if all sites under selection are accounted for (as described in the text). All plotted data are in Supplementary information S1 (table). Numbered data points correspond to the following species: 1, *Apis mellifera*; 2, *Arabidopsis thaliana*; 3, *Caenorhabditis briggsae*; 4, *Caenorhabditis elegans*; 5, *Daphnia pulex*; 6, *Drosophila melanogaster*; 7, *Heliconius melpomene*; 8, *Homo sapiens*; 9, *Mus musculus*; 10, *Oryza sativa*; 11, *Pan troglodytes*; 12, *Pristionchus pacificus*; 13, *Chlamydomonas reinhardtii*; 14, *Neurospora crassa*; 15, *Paramecium tetraurelia*; 16, *Plasmodium falciparum*; 17, *Saccharomyces cerevisiae*; 18, *Schizosaccharomyces pombe*; 19, *Trypanosoma brucei;* 20, *Agrobacterium tumefaciens*; 21, *Bacillus subtilis*; 22, *Burkholderia cenocepacia*; 23, *Deinococcus radiodurans*; 24, *Escherichia coli*; 25, *Helicobacter pylori*; 26, *Mesoplasma florum*; 27, *Mycobacterium smegmatis*; 28, *Mycobacterium tuberculosis*; 29, *Pseudomonas aeruginosa*; 30, *Salmonella enterica*; 31, *Salmonella typhimurium*; 32, *Staphylococcus epidermidis*; 33, *Thermus thermophilus*; 34, *Vibrio cholera*; 35, *Vibrio fischeri*.

Box 1 | **Scaling features of evolved mutation rates**

Here, we consider the likely genome-wide deleterious mutation rate, $U_D = uG_e$ (where u is the base-substitution mutation rate per nucleotide site per generation, and G_e is the effective genome size) that will evolve under the drift-barrier hypothesis and how this scales with various factors. Let U_0 be the minimum possible genome-wide deleterious mutation rate (the biophysical barrier), with a series of alleles imposing mutation-rate increases by factors of $(1+\lambda)$; where λ is the fractional difference of mutation rates between adjacent classes), such that the ith class has a genome-wide deleterious mutation rate of $U_i = U_0(1+\lambda)^i$. Recall that for a sexual population, the equilibrium loss of fitness for class i owing to mutation load is $2sU_i$, where s is the deleterious effect of an individual mutation.

We want to determine the long-term evolutionary distribution of the mutation rate: as the mutation-rate modifier mutates to lower and higher positions in the allelic series, selection favours classes with lower mutation rates and drift influences the efficiency of selection. Here, we suppose that fidelity factors mutate to higher accuracy at rate μ, and to a lower accuracy at rate ν, with allelic exchange occurring only between adjacent states. Using existing methods[16], such a linear system of alleles will eventually reach an equilibrium probability distribution of occupancy of the different classes:

$$\tilde{P}(i) \approx C \cdot (\nu/\mu)^i \cdot e^{-8N_e sU_i}$$

where C is a normalization constant that ensures that the state probabilities sum to 1.0, and N_e is the effective population size. This expression assumes a sexual, diploid population; for a haploid population, a 4 should be substituted for the 8; and for an asexual population, U_0 should be substituted for $2sU_0$.

The probability distribution for alternative mutation-rate states is determined by two terms, the first involving the relative rates of production of mutator versus antimutator alleles, ν/μ. Numerical values are not available for this ratio, but it is almost certainly greater than 1.0, and perhaps substantially so. The second term defines the effectiveness of selection in promoting antimutator alleles, with the strength of selection relative to that of genetic drift being a function of $N_e sU_i$. Once a population has reached sufficiently low mutation rates, the power of selection will be overcome by drift and/or mutation bias, preventing further downward evolution of the mutation rate. By contrast, at sufficiently high mutation rates, selection will be effective enough to prevent the evolution of still higher rates.

The solution of this equation illustrates three basic features of the drift-barrier hypothesis (FIG. 4). First, unless the mutation rate (u) is driven very close to the biophysical limit (U_0), the genome-wide deleterious mutation rate (U_D) is expected to scale inversely with the effective population size (N_e). Second, an increase in the fractional difference of mutation rates between adjacent classes, λ leads to a proportional reduction in the evolved mutation rate for all N_e, but leaves the expected inverse scaling between U_D and N_e intact, provided that λ is independent of U_i. This effect arises because λ defines the selective differential between adjacent mutation-rate classes and hence the efficiency of selection. Third, the mutational bias (ν/μ) determines the absolute values of the expected mutation rates, but again does not alter the inverse scaling of U_D with N_e — an increased bias towards production of mutator alleles simply increases the evolved mutation rate by the same factor at all N_e.

An alternative explanation for the error-prone nature of some polymerases, which is consistent with the drift-barrier hypothesis, is that the net selection pressure to improve accuracy is 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 infrequently, the opportunities for mutations will necessarily be fewer than with normal replicative polymerases, and selection on accuracy will be correspondingly reduced[6–8]. This 'use it or lose it' hypothesis is also consistent with the observation that polymerases deployed in the replacement of the small RNA primers with DNA during normal replication cycles have higher error rates than the major replicative polymerases[7]. It is also notable that secondary and tertiary fidelity mechanisms (such as proofreading and mismatch repair), which necessarily involve fewer nucleotide transactions than the earlier polymerization step, have greatly elevated error rates[6].

Collectively, this combination of theory and observation supports the idea that less frequently used polymerases naturally evolve higher error rates. This view does not deny the crucial importance of error-prone polymerases as mechanisms for dealing with bulky lesions or other forms of DNA damage, and it does not deny that induced mutagenesis can play a part in generating an appropriate adaptation in extreme times, sometimes being the only means for survival.

Caveats and alternative interpretations of the data. As the relationship in FIG. 3c bears on a trait of central relevance to all organisms, potential artefacts need to be considered. There are, for example, numerous difficulties in precisely defining the effective population size, N_e, by using silent-site diversity, including the definition of the appropriate geographic range for species surveys and the assumption of effective neutrality. Such measures of N_e also reflect conditions over only the past $2N_e$ or so generations (owing to the transient nature of polymorphisms), but given the large number of loci influencing replication fidelity and the rapidity with which the equilibrium load associated with mutator and/or antimutator alleles develops, measurements reflecting recent population features are arguably preferable. It could perhaps be argued that the estimates of N_e for highly self-fertilizing species (for example, *Caenorhabditis elegans* and *Arabidopsis thaliana*) are uncharacteristically low for their phylogenetic lineages, but increasing these to levels observed in outcrossing species only increases the overall correlations (FIG. 3c).

We have raised the possibility of a spurious correlation being induced between estimates of u and N_e as a result of dividing measures of standing variation by u to estimate N_e, as sampling error can result in some negative covariance between these estimates[39]. However, the levels of sampling variance in the measures used in the preceding analyses do not appear to be adequate to cause the large-scale patterns in FIG. 3c. For example, the seven estimates of u used to obtain an average estimate for the human mutation rate range from 1.1×10^{-8} to 1.9×10^{-8}, similar to prior indirect estimates obtained through the analysis of human genetic disorders[54,55], and all estimates of human silent-site diversities are close to 0.001 (REFS 40,56). Moreover, the pattern shown in FIG. 3c remains if the dependent variable is the mutation rate for indels, which is not used in the estimation of N_e (REF. 57).

One surprising aspect of FIG. 3c is that data from organisms with widely divergent population-genetic features are largely consistent. Because the theory implies that recombination reduces the efficiency of selection for improved replication fidelity, one might expect bacteria (which lack meiotic recombination) to have unusually low mutation rates, but this is clearly not the case. It may be relevant that despite their prominent clonal nature (which is also true for a portion of the life cycles of

most unicellular eukaryotes), bacteria and eukaryotes experience roughly the same amount of recombination per nucleotide site at the population level as eukaryotes[40]. Bacterial recombination generally involves the exchange of small blocks of DNA without crossing over, although most recombination in eukaryotes also involves simple gene conversion[58].

If alternative hypotheses for mutation-rate evolution invoking selection to maximize long-term evolvability are to be entertained, they will need to explain the inverse relationship between u and N_e, and also why optimal mutation rates should be nearly 1,000 times higher in large multicellular sexual species than in most (but not all) microorganisms. Attempts to estimate theoretically optimal mutation rates for maximizing long-term rates of adaptive evolution in asexual populations do not appear to predict a negative association between u and N_e. For example, under one model, the long-term rate of adaptation is maximized when the genome-wide mutation rate equals the rate of population fixation of beneficial mutations[9,59]. Because the rate of population fixation is in part a function of the rate of mutational input, the precise predictions of this argument are not entirely clear; but, mutations arise at a higher rate in large populations and, if beneficial, fix with higher probabilities, so this hypothesis seems to imply a positive association between u and N_e. A rather different model argues that populations should evolve genome-wide mutation rates equal to the average effect of a deleterious mutation[60,61] and appears to predict an optimal mutation rate independent of population size.

It is unclear how recombination might render an adaptive mutation-rate hypothesis consistent with the data. In sexual populations, selection for a mutator by hitch-hiking with a beneficial mutation is likely to be weaker than selection for an antimutator reducing background deleterious mutations. This is because selection against deleterious alleles is recurrent, whereas hitch-hiking between a mutator and an induced beneficial allele will typically be a rare and incomplete event, owing to recombination between the mutator and the induced mutation. Moreover, as noted above, elevated mutation rates in multicellular species (those with small N_e) impose an additional, direct negative effect through somatic mutations that decrease survival[6]. Hypotheses based on the number of cell divisions per generation will not explain the data in FIG. 3, as nearly the full range of variation in mutation rates per generation is encompassed by unicellular species alone, and within multicellular species there is no simple relationship between the number of germline cell divisions and the per-generation mutation rate[6,7,29]. Generation length also does not explain the patterns, as unicellular eukaryotes have longer cell-division times but lower mutation rates than prokaryotes (bacteria and archaea)[24]. Thus, overall, both the data and the theory are inconsistent with the idea that mutation-rate evolution is guided by a population-level goal of maximizing the rate of incorporation of beneficial mutations. Rather, all lines of evidence appear to point to the conclusion that mutation rates are simply driven downward by individual-level selection operating on excess deleterious mutation loads[11,18,19,62,63].

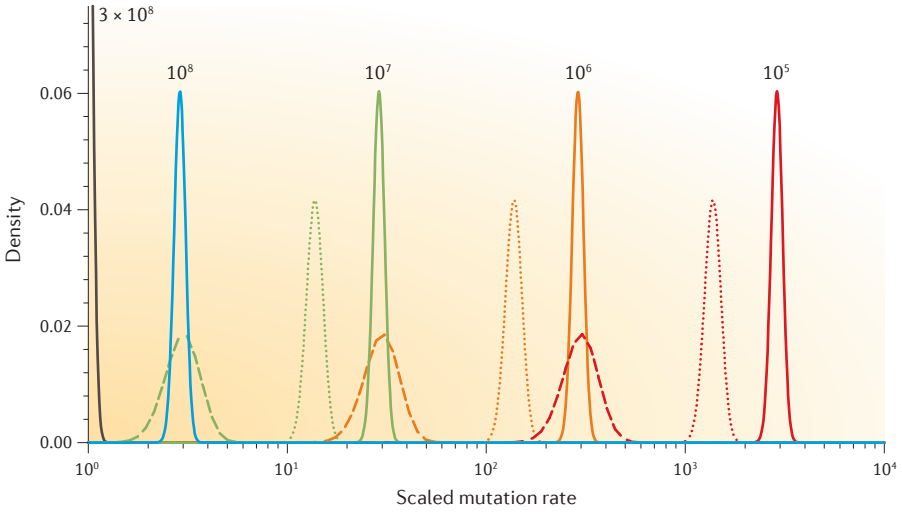

Figure 4 | **Expected evolutionary distributions of the genome-wide deleterious mutation rate.** Results, derived from the theory in BOX 1, are given for different effective population sizes (N_e), as indicated by different coloured lines (N_e values labelled at the top of the peaks). In all cases, the minimum genome-wide rate of loss of fitness per generation by mutation is $sU_0 = 10^{-7}$. Solid lines denote the situation in which the mutation bias towards mutator (ν; that is, low-accuracy mutation rate) versus antimutator (μ; that is, high-accuracy mutation rate) alleles is ν/μ = 10.0 and the increment of change in the mutation rate between adjacent classes (λ) is equal to 0.01. Dashed lines refer to the situation in which ν/μ = 10.0 and λ = 0.1, and dotted lines indicate the situation in which ν/μ = 3.0 and λ = 0.01; these special cases are shown only for the three smallest population sizes. The mutation-rate classes are simply scaled to an arbitrary baseline value of $U_0 = 1.0$. Note that ν/μ and λ determine the positions of the mutation-rate distributions, but that the scaling with N_e is retained.

Gene-level features

Influence of genomic context. The results discussed above highlight striking interspecific patterns for average genome-wide mutation rates that probably reflect *trans*-acting factors that influence mutagenic processes across the entire genome; that is, DNA polymerases and repair genes. However, *cis*-acting mechanisms can sometimes lead to gene-specific variation in the vulnerability to mutations. For example, owing to the nature of the genetic code and selection for particular amino-acid compositions, some variation in mutation rates will inevitably exist at the gene level. On average, C and G nucleotides are usually more mutable than A and T nucleotides (up to six times more)[55], and the context associated with flanking bases can cause up to 75-fold variation in site-specific mutation rates[64–66]. The folding of nascent mRNAs can also affect mutation rates in a gene-specific manner by influencing mRNA–DNA heteroduplex formation during transcription, which can leave the non-template DNA strand exposed to mutagenic effects[67]. Finally, simple-sequence repeats (SSRs), which are subject to stochastic expansion or contraction of repeat numbers at rates far above that seen for single-base substitutions, can lead to elevated rates of alterations in amino-acid runs when repeat units are three nucleotides long, or otherwise to downstream frameshifts. Although there are credible examples of adaptive variants being generated by such processes[68–70], as with all mutations, the vast majority of SSR mutations are likely to be either neutral or deleterious outcomes of stochastic mutational processes[71–72].

Context-dependent effects can become particularly important if the transcribed strand of a gene transposes between the leading strand and lagging strand of replication, as frequently happens over evolutionary time[73]. During long-term residence in a particular location, a gene will evolve towards an equilibrium usage of nucleotides based on the strand-specific mutation spectrum[40]. At equilibrium, the most mutable nucleotides will be reduced in frequency, so strand-switching will cause the nucleotide composition of a gene to suddenly deviate from the local condition, thereby elevating the genic mutation rate until sufficient time has elapsed to attain the equilibrium nucleotide composition for the strand of occupancy.

Mutation-rate variation also occurs on larger topological scales. In *E. coli*, for example, there is a symmetrical wave-like pattern of the mutation rate around the circular chromosome, with the extreme highs and lows differing by approximately 2.5-fold in average mutation rates[74]. Detection of such topographical patterns requires the cataloguing of large numbers of mutations, so few such studies have been performed. However, a collection of more than 31,000 point mutations in a mismatch-repair deficient strain of the bacterium *Pseudomonas fluorescens* revealed a chromosomal pattern of mutation-rate variation differing from that in *E. coli*: greater uniformity, except for an approximate 1.5-fold decline in the earliest 25% of the replicon[75].

Large-scale topological effects are also found in eukaryotes. For example, mutation rates in yeast are correlated with the timing and positions of the origins of replication, with a two- to fourfold range of variation on spatial scales up to approximately 100 kb (REFS 65,76). Similar observations have been made with human chromosomes[77], although here the range of variation is less than 10%. On a smaller scale, nucleosome wrapping (over lengths of ~150 bp) can protect DNA from mutagenic effects, reducing the local mutation rate by up to twofold[65,78].

The molecular mechanisms driving these large-scale patterns remain unclear, but could be associated with variation in the composition of the nucleotide pool during the cell cycle, regional variation in transcription rates and their influence on replication, and/or alterations in the rates of processivity of the DNA polymerase across different chromosomal regions. Regardless of the cause, the implications are clear: studies that use single-gene constructs to understand the mechanisms of mutation need to be cognizant of the fact that several-fold changes in mutation rates can arise as a simple consequence of the chromosomal location and/or orientation of the construct, independent of other features.

Gene expression and mutational vulnerability. Transcriptional activity can influence genic mutation rates in two opposing ways[79,80]. On the one hand, because transcription requires separation of DNA into single strands, the non-transcribed strand (which is unprotected by the growing transcript and the proteins associated with transcription) is expected to be more vulnerable to DNA damage, leading to transcription-associated mutagenesis (TAM). On the other hand, transcription-coupled repair (TCR) detects and removes damaged bases on transcribed strands.

If the influence of TAM exceeds that of TCR, genic mutation rates will increase with transcriptional activity. Some evidence suggests that this is the case, although weakly so. For example, in a mutator strain of *S. cerevisiae,* MA lines reveal an approximate twofold increase in the mutation rate in the most highly expressed genes versus the most lowly expressed genes (after accounting for nucleotide composition differences)[81]. In wild-type yeast MA lines, spontaneously mutating sites may have up to a fourfold elevation in expression level[65,82]. Comparative analyses among mammalian species provide indirect evidence for a small mutagenic effect of transcription, with a doubling in gene expression inducing an approximate 15% increase in the mutation rate[81,83,84]. Part of the reason for the small effect of transcription on the mutation rate may be that the damage-detection process associated with TCR extends beyond transcribed regions[85].

In contrast to the preceding analyses and many others[80,86], Martincorena *et al.*[3] argued that natural selection specifically promotes reduced mutation rates in highly transcribed genes in *E. coli*. However, locus-specific mutation rates in this study were not estimated directly, but were inferred from within-species nucleotide variation at silent (synonymous coding) sites. This measure is expected to scale with the base-substitution mutation rate under the assumption of neutrality, although things are not so simple in prokaryotes in which selection operates on silent sites, especially in highly expressed genes[40]. A more direct analysis using MA data indicates that the mutation rate in *E. coli* increases with the rate of transcription, as in other species, although

Glossary

Deleterious
A mutation having detrimental effects on the fitness of an organism.

Drift-barrier hypothesis
The idea that the ability of natural selection to refine a phenotype is ultimately limited by the noise created by random genetic drift, which itself is a consequence of finite population size and the stochastic effects of linked mutations.

Effective population size
(N_e). A measure of the size of a population from the standpoint of the reliability of allele-frequency transmission across generations; generally, one to several orders of magnitude below the actual population size, owing to variation in family size, a wide range of other demographic features and the hitch-hiking effects of linked mutations.

Fixation
The process by which a genetic variant at an initially polymorphic site increases in frequency until it attains a frequency of 1.0 in the population.

Full-sib pairs
Brothers and sisters sharing the same mother and father.

Gene conversion
An alteration of the nucleotide sequence at one chromosomal location resulting from the acquisition of information from a homologous sequence elsewhere in the genome during genetic recombination; such events are not always accompanied by chromosomal crossing over.

Lagging strand
A strand of nascent DNA that is synthesized in the opposite direction of the progressive opening of the DNA on a parental chromosome, resulting in discontinuous replication fragments that must be stitched together.

Leading strand
A strand of nascent DNA that is synthesized in one continuous flow in the same direction as the progression of the opening of the DNA on a parental chromosome.

Mutation–selection balance
An equilibrium allele frequency that results from the opposing pressures of natural selection and mutation, one tending to remove variation and the other creating it.

Silent sites
Genomic sites within protein-coding regions at which nucleotide substitutions have no effect on the encoded amino acid, owing to the redundancy of the genetic code.

Somatic mutations
DNA-level changes arising within the somatic cells of multicellular organisms, and therefore not transmissible across generations but having direct effects on fitness.

Standing variation
Genetic variation among individuals within a population.

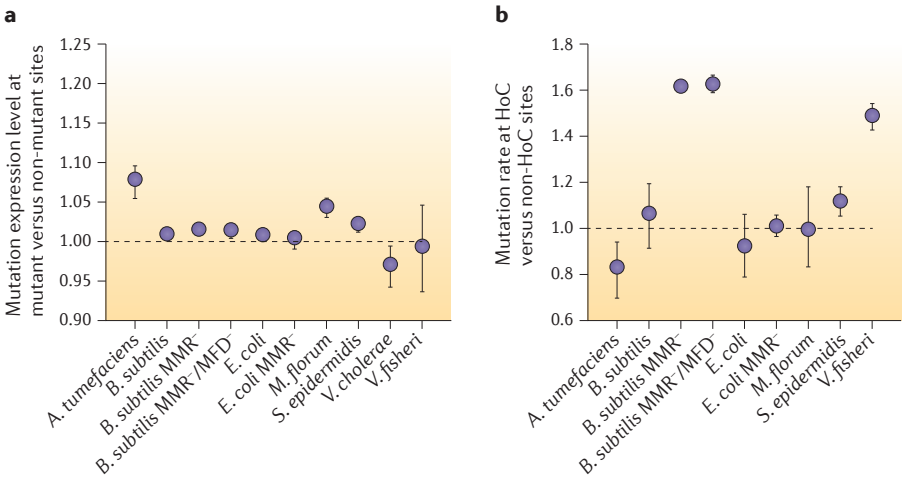

Figure 5 | **The relationships between site-specific mutagenicity, gene expression and strand occupancy. a** | The average depth of coverage of mRNA abundance for mutant sites is scaled relative to the average for non-mutant sites using data derived from mutation-accumulation (MA) studies by the authors[39,57,64] and available mRNA sequencing data. For each species, approximately 50 MA lines were sequenced, typically revealing several hundred *de novo* mutations. Vertical lines denote standard errors. MMR⁻ denotes a mismatch-repair deficient strain; and MFD⁻ denotes a strain deficient for mutation frequency decline (*mfd*), which is linked to the transcription-coupled repair machinery. **b** | Using the same MA data in part **a**, mutation rates are instead subdivided according to the relative orientations of the transcribed strand of the gene and the DNA replicon. HoC denotes a gene orientation that leads to head-on collisions between DNA and RNA polymerase. Source data for these graphs are described in Supplementary information S1 (table). *A. tumefaciens*, *Agrobacterium tumefaciens*; *B. subtilis*, *Bacillus subtilis*; *E. coli*, *Escherichia coli*; *M. florum*, *Mesoplasma florum*; *S. epidermidis*, *Staphylococcus epidermidis*; *V. cholerae*, *Vibrio cholerae*; *V. fischeri*, *Vibrio fischeri*.

the effect is again small[64,87]. Similarly, MA data from six additional bacterial species yield ratios of transcription levels for mutant and non-mutant sites that are typically very close to 1.0, with the most extreme cases showing 7–20% elevation in expression level for mutant sites (FIG. 5). Thus, virtually all direct observations support the view that transcription has mild mutagenic effects.

To understand the challenges for the arguments that invoke selection on gene-specific mutation rates, it is useful to consider the likely strength of selection at such a level. For *E. coli*, the base-substitution mutation rate is approximately 2.0×10^{-10} per nucleotide site per cell division[64], so the upper bound to the deleterious mutation rate for the average gene with length 951 bp is approximately 2×10^{-7} per gene per cell division if one assumes that all mutations are deleterious. As noted above, in the absence of recombination, this quantity would represent the maximum selective advantage of a mechanism that prevented all mutations. However, because population-level recombination rates in bacteria are not greatly different to those in eukaryotes[40], and recombination greatly reduces the magnitude of selection on mutation-rate modifiers (in the extreme, to a fraction of $s = 0.01$ to 0.1 of the preceding upper bound), the selective advantage of a newly arisen gene-specific mutation inhibitor in *E. coli* is likely to be $<<10^{-8}f$, where f is the fractional change in the mutation rate induced by the modifier; and even this full benefit will not be felt until a new mutation–selection balance has been achieved for the mutation-rate modifier.

Thus, because the observations noted above imply that transcription-associated effects are much smaller than the maximum of $f = 0.2$ advocated by Martincorena *et al.*[3], the upper bound on the selective advantage of a gene-specific antimutator mechanism associated with gene expression appears to be $<<10^{-9}$. Moreover, because most *cis*-acting modifiers of the mutation rate probably influence only a fraction of the span of a gene, even this is likely to be a conservative (high) estimate. Such a low magnitude of selection would be overwhelmed by the power of random genetic drift in a species such as *E. coli* with $N_e \approx 10^8$ (REF. 39). Genome-wide processes such as TCR can be maintained by selection, because most species have $N_e < 10^9$ (FIG. 3), although it is highly doubtful that gene-specific mutation rates are ever moulded by natural selection except in a few specialized cases.

Replication–transcription conflicts. The orientation of a gene is of particular interest in bacteria in which transcription and replication often overlap temporally[88,89]. Genes for which the transcribed strand is used as a template for the leading strand of replication have co-oriented transcription and replication, whereas genes in the opposite orientation can experience head-on collisions (HoCs) between the transcription and replication machinery. Because bacterial replisomes move at a substantially more rapid pace than RNA polymerases, clashes are nearly equally likely in either orientation, but HoCs are thought to have more substantial consequences[88,90–92]. This has led to the idea that biased chromosomal orientations of bacterial genes reflect selection for enhanced genome stability[93].

There is considerable variability among bacterial species in the fraction of genes inhabiting putatively less dangerous leading strands; approximately 75% for *Bacillus subtilis* and most other Firmicutes, but typically 55–60% for most other species[94]. Why do elevated fractions of genes have orientations leading to transcription–replication conflicts? One suggestion is that bacteria, *B. subtilis* in particular, position some of their genes to specifically promote mutagenic encounters between RNA and DNA polymerases to generate beneficial mutations[4]. Given the limitations noted above for gene-specific mutation-rate modulation, the idea that gene-strand localization can be exploited for adaptive reasons merits closer examination.

In our opinion, the evidence for elevated rates of adaptive mutation in lagging-strand genes is less than compelling. First, virtually identical average levels of silent-site variation for leading- and lagging-strand *B. subtilis* genes[4] presumably reflect similar long-term mutagenic conditions on both strands; this is consistent with MA experiment results suggesting that mutation rates for genes on leading and lagging strands are similar in this species ($3.0 \pm 0.7 \times 10^{-10}$ and $1.3 \pm 1.0 \times 10^{-10}$ base-substitution mutations per site per cell division, respectively; values given ± s.e.m.)[66]. Second, MA analyses from additional species lead to a similar conclusion: for wild-type (non-mutator) organisms, in all but one case, mutation rates on leading- and lagging-strand genes are very similar (FIG. 5). Third, elevated variation at amino-acid replacement sites in lagging-strand genes[4] need not reflect selection for protein-sequence change, a simpler explanation being that lagging-strand genes in *B. subtilis* have relatively low expression levels, which is the

strongest known correlate of elevated rates of deleterious amino-acid substitution[95–97]. Finally, the evidence invoked to support the idea of adaptive localization is based on a definition of convergent mutational change that includes the production of different amino acids[4], which is not a reflection of convergent positive selection, but an indicator of relaxed purifying selection[98].

Any attempt to understand the biased distribution of genes on leading versus lagging strands must take into consideration the physical rates of movement of genes in both directions, along with the selective advantages of alternative locations. Suppose that genes move from the lagging strand to the leading strand, and vice versa, at translocation rates v and u, respectively. If s is the selective advantage of being on the leading strand, the long-term equilibrium probability of gene residence on the leading strand is:

$$P_L = \frac{(v/u)e^{2N_e s}}{1+(v/u)e^{2N_e s}}$$

for a haploid species[40], where e is the base of natural logarithms. The composite term in the numerator is simply the ratio of the rates of fixation to leading versus lagging strand, owing to the joint forces of mutation, drift and selection.

Application of this formula shows that even with universal selection for occupancy on the leading strand (that is, a positive s), a moderate fraction of genes is still expected to reside in non-optimal locations; that is, the residence of genes on alternative strands cannot simply be assumed to reflect a strategy for optimizing mutation rates. With unbiased physical movement ($v/u = 1$), only weak selection relative to the power of random genetic drift ($2N_e s = 0.3$) is required to generate a typical microbial strand occupancy of $P_L \approx 0.57$. If the power of selection for leading-strand occupancy is equal to the power of drift for all genes ($2N_e s = 1$), approximately 73% of genes would be expected on the leading strand at equilibrium, which is the level observed in *B. subtilis*.

These results suggest that the mutational effects of altered strand orientation are unlikely to be sufficient for selectively driven gene relocations. Consistent with this view, although mutation rates may be slightly elevated when the orientations of genes change, strand relocation generates no obvious change in the strength or the pattern of selection in proteobacterial genes[73]. Moreover, most genes that switch strands actually transpose to the opposite side of the bacterial replicon, retaining the original strand orientation, and these experience the same evolutionary patterns as those switching leading- or lagging-strand status[73]. Thus, it appears that alterations in mutation rates following gene relocations do not arise so much from HoCs as from the simple context alterations noted above[99].

Closing comments

The results discussed above suggest that selection typically operates to minimize the mutation rate, with the efficiency of such downward movement eventually being overcome by the power of random genetic drift. The precise point at which this drift barrier is reached is influenced by the relative rates of production of mutator and antimutator alleles, a key issue that remains unresolved. However, the idea that the drift barrier is typically encountered before any insurmountable biophysical or biochemical limits to replication fidelity is supported by two additional sets of observations. First, bacterial populations that are founded with a mutator genotype frequently evolve lower mutation rates on relatively short timescales through compensatory molecular changes at genomic sites that are not involved in the initial mutator construct[100–103]. The rapid appearance of such modifiers implies the presence of unexploited potential for improvement in replication fidelity. Second, although vertebrates have the highest known per-generation mutation rates, the rate per germline cell division can rival the per-generation rate for unicellular species[29], implying that even species with a small N_e can achieve exceptionally high levels of replication fidelity. For example, the human germline mutation rate per nucleotide site is approximately 6×10^{-11} per cell division, approaching the lowest rates observed in unicellular species, and 10–100 times lower than rates in various human somatic tissues[29,55,104,105]. Thus, although selection operates on the per-generation mutation rate, this is accommodated by changes in replication fidelity at the cell-division level.

We have focused on the evolution of the long-term average mutation rate for species, giving little attention to the large-effect mutator alleles, which despite intrinsic disadvantages, are expected to be maintained in all populations at low frequencies through recurrent mutation pressure and to occasionally be pulled to high frequencies in extreme situations involving very strong selection and weak recombination[6,106–111]. Although rarely fixed (that is, by occasionally generating an adaptive mutation that then becomes liberated by recombination), maladapted mutator genotypes may nonetheless be prominent platforms for the origin of beneficial alleles at the population level.

Finally, it is worth noting that a potential limitation of nearly all current estimates of mutation rates and spectra is the performance of such studies in optimal laboratory conditions. We can anticipate the determination of the environmental dependence of mutation rates in the near future, as sequencing costs drop to the point that mutation analyses can be focused on large numbers of parent–offspring pairs in their natural habitats. However, evaluating the degree to which natural variation in the mutation rate is due to genes versus environment will require the application of quantitative-genetic designs with the power to reveal resemblance between relatives. For example, it has recently been argued that the mutation rate became elevated in the European human population after migration out of Africa[112], but whether this is a consequence of an altered mutational environment or altered replication machinery remains to be determined.

Michael Lynch, Matthew S. Ackerman, Jean-Francois Gout, Hongan Long, Way Sung and Patricia L. Foster are at the Department of Biology, Indiana University, Bloomington, Indiana 47401, USA.

W. Kelley Thomas is as the Department of Molecular, Cellular, and Biomedical Sciences, University of New Hampshire, Durham, New Hampshire 03824, USA.

Correspondence to M.L.
milynch@indiana.edu

doi:10.1038/nrg.2016.104
Published online 14 Oct 2016

1. Rosenberg, S. M. Evolving responsively: adaptive mutation. *Nat. Rev. Genet.* **2**, 504–515 (2001).
2. Galhardo, R. S., Hastings, P. J. & Rosenberg, S. M. Mutation as a stress response and the regulation of evolvability. *Crit. Rev. Biochem. Mol. Biol.* **42**, 399–435 (2007).
3. Martincorena, I., Seshasayee, A. S. & Luscombe, N. M. Evidence of non-random mutation rates suggests an evolutionary risk management strategy. *Nature* **485**, 95–98 (2012).
4. Paul, S., Million-Weaver, S., Chattopadhyay, S., Sokurenko, E. & Merrikh, H. Accelerated gene evolution through replication-transcription conflicts. *Nature* **495**, 512–515 (2013).
5. Ram, Y. & Hadany, L. Stress-induced mutagenesis and complex adaptation. *Proc. Biol. Sci.* **281**, 20141025 (2014).
6. Lynch, M. The cellular, developmental, and population-genetic determinants of mutation-rate evolution. *Genetics* **180**, 933–943 (2008).
This paper provides an overview of population-genetic theory for the selective disadvantage of a mutator allele associated with the indirect effects of linked deleterious mutations under arbitrary degrees of recombination, and also for the direct effects of somatic mutation. This paper also considers the expected frequencies of mutator alleles under mutation–selection balance.
7. Lynch, M. The lower bound to the evolution of mutation rates. *Genome Biol. Evol.* **3**, 1107–1118 (2011).
This article develops the theory associated with the drift-barrier hypothesis for the lower bound to mutation-rate evolution, as well as an overview of empirical observations on the error rates associated with various DNA polymerases.

8. MacLean, R. C., Torres-Barceló, C. & Moxon, R. Evaluating evolutionary models of stress-induced mutagenesis in bacteria. *Nat. Rev. Genet.* **14**, 221–227 (2013).
This study provides an overview of evolutionary theory in the context of stress-induced mutagenesis and presents supportive data for the idea that the error-prone nature of polymerases associated with such activities have arrived at such a condition by genetic drift.
9. Kimura, M. On the evolutionary adjustment of spontaneous mutation rates. *Genet. Res.* **9**, 23–34 (1967).
This is a classical paper in which the selective disadvantage of mutator alleles associated with linked mutation load was first considered.
10. Kondrashov, A. S. Modifiers of mutation-selection balance: general approach and the evolution of mutation rates. *Genet. Res.* **66**, 53–70 (1995).
11. Dawson, K. J. The dynamics of infinitesimally rare alleles, applied to the evolution of mutation rates and the expression of deleterious mutations. *Theor. Pop. Biol.* **55**, 1–22 (1999).
12. Lynch, M. *et al.* Spontaneous deleterious mutation. *Evolution* **53**, 645–663 (1999).
13. Baer, C. F., Miyamoto, M. M. & Denver, D. R. Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nat. Rev. Genet.* **8**, 619–631 (2007).
14. Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* **8**, 610–618 (2007).
This article provides a broad overview of methods for the estimation of the distribution of fitness effects of *de novo* mutations, and the implications derived from population-genetic data.
15. Hall, D. W., Fox, S., Kuzdzal-Fick, J. J., Strassmann, J. E. & Queller, D. C. The rate and effects of spontaneous mutation on fitness traits in the social amoeba, *Dictyostelium discoideum*. *G3 (Bethesda)* **8**, 1115–1127 (2013).
16. Lynch, M. Evolutionary layering and the limits to cellular perfection. *Proc. Natl Acad. Sci. USA* **109**, 18851–18856 (2012).
This paper demonstrates that when selection operates on the overall perfection of a process involving multiple levels, the alternative components are free to drift so long as the level of refinement of the entire system remains at the drift barrier; this degree of interaction can lead to an evolutionary situation in which complex systems are ultimately no more efficient than simpler systems, but maintain an illusion of adaptive robustness.
17. Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. Rates of spontaneous mutation. *Genetics* **148**, 1667–1686 (1998).
This article provides an early comprehensive overview of the rate and fitness effects of mutations in diverse organisms.
18. Sniegowski, P. D., Gerrish, P. J., Johnson, T. & Shaver, A. The evolution of mutation rates: separating causes from consequences. *Bioessays* **22**, 1057–1066 (2000).
19. Andrē, J. B. & Godelle, B. The evolution of mutation rate in finite asexual populations. *Genetics* **172**, 611–626 (2006).
20. Bessman, M. J., Muzyczka, N., Goodman, M. F. & Schnaar, R. L. Studies on the biochemical basis of spontaneous mutation. II. The incorporation of base and its analogue into DNA by wild type, mutator and antimutator DNA polymerases. *J. Mol. Biol.* **88**, 409–421 (1974).
21. Loh, E., J. Choe, J. & Loeb, L. A. Highly tolerated amino acid substitutions increase the fidelity of *Escherichia coli* DNA polymerase I. *J. Biol. Chem.* **282**, 12201–12209 (2007).
This paper is gives an empirical demonstration of the relative ease of obtaining antimutator alleles by mutations in a DNA polymerase.
22. Tian, W., Hwang, Y. T. & Hwang, C. B. The enhanced DNA replication fidelity of a mutant herpes simplex virus type 1 DNA polymerase is mediated by an improved nucleotide selectivity and reduced mismatch extension ability. *J. Virol.* **82**, 8937–8941 (2008).
23. Loh, E., Salk, J. J. & Loeb, L. A. Optimization of DNA polymerase mutation rates during bacterial evolution. *Proc. Natl Acad. Sci. USA* **107**, 1154–1159 (2010).
24. Lynch, M. & Marinov, G. F. The bioenergetic cost of a gene. *Proc. Natl Acad. Sci. USA* **112**, 15690–15695 (2015).
25. Casjens, S. The diverse and dynamic structure of bacterial genomes. *Annu. Rev. Genet.* **32**, 339–377 (1998).
26. Cox, R. A. Quantitative relationships for specific growth rates and macromolecular compositions of *Mycobacterium tuberculosis, Streptomyces coelicolor* A3(2) and *Escherichia coli* B/r: an integrative theoretical approach. *Microbiology* **150**, 1413–1426 (2004).
27. Mira, A., Ochman, H. & Moran, N. A. Deletional bias and the evolution of bacterial genomes. *Trends Genet.* **17**, 589–596 (2001).
28. Vieira-Silva, S., Touchon, M. & Rocha, E. P. No evidence for elemental-based streamlining of prokaryotic genomes. *Trends Ecol. Evol.* **25**, 319–320 (2010).
29. Lynch, M. Evolution of the mutation rate. *Trends Genet.* **26**, 345–352 (2010).
30. Drake, J. W. A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl Acad. Sci. USA* **88**, 7160–7164 (1991).
A classical paper that first suggested that there is an inverse relationship between the mutation rate per site, u, and the number of nucleotides per genome in microbes, leading to a constant expected total number of mutations per genome.
31. Rosche, W. A. & Foster, P. L. Determining mutation rates in bacterial populations. *Methods* **20**, 4–17 (2000).
This article provides a broad overview of methods for estimating microbial mutation rates using reporter constructs.
32. Nachman, M. W. & Crowell, S. L. Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**, 297–304 (2000).
This study presents a first attempt to estimate the human mutation rate from the level of molecular divergence between orthologous human and chimpanzee sequences.
33. Kibota, T. & Lynch, M. Estimate of the genomic mutation rate deleterious to overall fitness in *Escherichia coli*. *Nature* **381**, 694–696 (1996).
34. Campbell, C. D. *et al.* Estimating the human mutation rate using autozygosity in a founder population. *Nat. Genet.* **44**, 1277–1281 (2012).
35. Kong, A. *et al.* Rate of *de novo* mutations and the importance of father's age to disease risk. *Nature* **488**, 471–475 (2012).
This paper describes one of the first attempts to estimate the human mutation rate by comparing the genomic sequences of parents and offspring.
36. Venn, O. *et al.* Strong male bias drives germline mutation in chimpanzees. *Science* **344**, 1272–1275 (2014).
37. Keightley, P. D., Ness, R. W., Halligan, D. L. & Haddrill, P. R. Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. *Genetics* **196**, 313–320 (2014).
38. Keightley, P. D. *et al.* Estimation of the spontaneous mutation rate in Heliconius melpomene. *Mol. Biol. Evol.* **32**, 239–243 (2015).
This study uses population-genetic data to arrive at the conclusion that the average newborn human acquires about two new mutations.
39. Sung, W., Ackerman, M. S., Miller, S. F., Doak, T. G. & Lynch, M. The drift-barrier hypothesis and mutation-rate evolution. *Proc. Natl Acad. Sci. USA* **109**, 18488–18492 (2012).
40. Lynch, M. *The Origins of Genome Architecture* (Sinauer Assoc., 2007).
41. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, 1983).
42. Massey, S. E. The proteomic constraint and its role in molecular evolution. *Mol. Biol. Evol.* **25**, 2557–2565 (2008).
43. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
44. Halligan, D. L. & Keightley, P. D. Ubiquitous selective constraints in the *Drosophila* genome revealed by a genome-wide interspecies comparison. *Genome Res.* **16**, 875–884 (2006).
45. Keightley, P. D. Rates and fitness consequences of new mutations in humans. *Genetics* **190**, 295–304 (2012).
46. Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).
47. Rands, C. M., Meader, S., Ponting, C. P. & Lunter, G. 8.2% of the human genome is constrained: variation in rates of turnover across functional element classes in the human lineage. *PLoS Genet.* **10**, e1004525 (2014).
48. Radman, M., Taddei, F. & Matic, I. Evolution-driving genes. *Res. Microbiol.* **151**, 91–95 (2000).
49. Tenaillon, O., Taddei, F., Radman, M. & Matic, I. Second-order selection in bacterial evolution: selection acting on mutation and recombination rates in the course of adaptation. *Res. Microbiol.* **152**, 11–16 (2001).
50. Earl, D. J. & Deem, M. W. Evolvability is a selectable trait. *Proc. Natl Acad. Sci. USA* **101**, 11531–11536 (2004).
51. Foster, P. L. Stress-induced mutagenesis in bacteria. *Crit. Rev. Biochem. Mol. Biol.* **42**, 373–397 (2007).
52. Gerrish, P. J., Colato, A., Perelson, A. S. & Sniegowski, P. D. Complete genetic linkage can subvert natural selection. *Proc. Natl Acad. Sci. USA* **104**, 6266–6271 (2007).
53. Gerrish, P. J., Colato, A. & Sniegowski, P. D. Genomic mutation rates that neutralize adaptive evolution and natural selection. *J. R. Soc. Interface* **10**, 20130329 (2013).
54. Kondrashov, A. S. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum. Mutat.* **21**, 12–27 (2003).
55. Lynch, M. Rate, molecular spectrum, and consequences of spontaneous mutations in man. *Proc. Natl Acad. Sci. USA* **107**, 961–968 (2009).
56. Lynch, M. The origins of eukaryotic gene structure. *Mol. Biol. Evol.* **23**, 450–468 (2006).
57. Sung, W. *et al.* Evolution of the insertion-deletion mutation rate across the tree of life. *G3 (Bethesda)* **6**, 2583–2591 (2016).
58. Lynch, M. *et al.* Genome-wide linkage-disequilibrium profiles from single individuals. *Genetics* **198**, 269–281 (2014).
59. Leigh, E. G. Jr Natural selection and mutability. *Amer. Nat.* **104**, 301–305 (1970).
60. Orr, H. A. The rate of adaptation in asexuals. *Genetics* **155**, 961–968 (2000).
61. Johnson, T. & Barton, N. H. The effect of deleterious alleles on adaptation in asexual populations. *Genetics* **162**, 395–411 (2002).
62. Sturtevant, A. H. Essays on evolution. I. On the effects of selection on mutation rate. *Quart. Rev. Biol.* **12**, 464–476 (1937).
63. Johnson, T. Beneficial mutations, hitchhiking and the evolution of mutation rates in sexual populations. *Genetics* **151**, 1621–1631 (1999).
64. Lee, H., Popodi, E., Tang, H. & Foster, P. L. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc. Natl Acad. Sci. USA* **109**, E2774–E2783 (2012).
65. Lujan, S. A. *et al.* Heterogeneous polymerase fidelity and mismatch repair bias genome variation and composition. *Genome Res.* **24**, 1751–1764 (2014).
66. Sung, W. *et al.* Asymmetric context-dependent mutation patterns revealed through mutation accumulation experiments. *Mol. Biol. Evol.* **32**, 1672–1683 (2015).
This study uses MA-WGS data from several bacterial species to demonstrate the strong dependency of site-specific mutation rates on the identity of neighbouring nucleotides.
67. Chen, X., Yang, J. R. & Zhang, J. Nascent RNA folding mitigates transcription-associated mutagenesis. *Genome Res.* **26**, 50–59 (2016).
68. Kashi, Y. & King, D. G. Simple sequence repeats as advantageous mutators in evolution. *Trends Genet.* **22**, 253–259 (2006).
69. Moxon, R., Bayliss, C. & Hood, D. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu. Rev. Genet.* **40**, 307–333 (2006).
This paper presents an overview of the evidence suggesting that some loci may have special sequence features, potentially maintained by selection, that enhance mutagenicity.
70. Zhou, K., Aertsen, A. & Michiels, C. W. The role of variable DNA tandem repeats in bacterial adaptation. *FEMS Microbiol. Rev.* **38**, 119–141 (2014).
71. Haerty, W. & Golding, G. B. Genome-wide evidence for selection acting on single amino acid repeats. *Genome Res.* **20**, 755–760 (2010).
72. Scala, C. *et al.* Amino acid repeats cause extraordinary coding sequence variation in the social amoeba *Dictyostelium discoideum*. *PLoS ONE* **7**, e46150 (2012).

73. Lin, C. H., Lian, C. Y., Hsiung, C. A. & Chen, F. C. Changes in transcriptional orientation are associated with increases in evolutionary rates of enterobacterial genes. *BMC Bioinformatics* **12** (Suppl. 9), 19 (2011).
74. Foster, P. L. *et al.* On the mutational topology of the bacterial genome. *G3 (Bethesda)* **3**, 399–407 (2013).
This article demonstrates large-scale spatial variation in the mutation rate over the *E. coli* genome.
75. Long, H. *et al.* Mutation rate, spectrum, topology, and context-dependency in the DNA mismatch repair deficient *Pseudomonas fluorescens* ATCC948. *Genome Biol. Evol.* **7**, 262–271 (2015).
76. Lang, G. I. & Murray, A. W. Mutation rates across budding yeast chromosome VI are correlated with replication timing. *Genome Biol. Evol.* **3**, 799–811 (2011).
77. Stamatoyannopoulos, J. A. *et al.* Human mutation rate associated with DNA replication timing. *Nat. Genet.* **41**, 393–395 (2009).
78. Chen, X. *et al.* Nucleosomes suppress spontaneous mutations base-specifically in eukaryotes. *Science* **335**, 1235–1238 (2012).
79. Ganesan, A., Spivak, G. & Hanawalt, P. C. Transcription-coupled DNA repair in prokaryotes. *Prog. Mol. Biol. Transl. Sci.* **110**, 25–40 (2012).
This paper presents a broad overview of the mechanism of TCR in bacteria.
80. Jinks-Robertson, S. & Bhagwat, A. S. Transcription-associated mutagenesis. *Annu. Rev. Genet.* **48**, 341–359 (2014).
This article discusses an overview of observations on the association between transcription and mutagenesis.
81. Park, C., Qian, W. & Zhang, J. Genomic evidence for elevated mutation rates in highly expressed genes. *EMBO Rep.* **13**, 1123–1129 (2012).
This study provides evidence that highly expressed genes have elevated mutation rates.
82. Chen, X. & Zhang, J. Yeast mutation accumulation experiment supports elevated mutation rates at highly transcribed sites. *Proc. Natl Acad. Sci. USA* **111**, E4062 (2014).
83. Green, P. *et al.* Transcription-associated mutational asymmetry in mammalian evolution. *Nat. Genet.* **33**, 514–517 (2003).
84. Polak, P. & Arndt, P. F. Transcription induces strand-specific mutations at the 5′ end of human genes. *Genome Res.* **18**, 1216–1223 (2008).
85. Haines, N. M., Kim, Y. I., Smith, A. J. & Savery, N. J. Stalled transcription complexes promote DNA repair at a distance. *Proc. Natl Acad. Sci. USA* **111**, 4037–4042 (2014).
86. Eyre-Walker, A. & Bulmer, M. Synonymous substitution rates in enterobacteria. *Genetics* **140**, 1407–1412 (1995).
87. Chen, X. & Zhang, J. No gene-specific optimization of mutation rate in *Escherichia coli*. *Mol. Biol. Evol.* **30**, 1559–1562 (2013).
88. Merrikh, H., Zhang, Y., Grossman, A. D. & Wang, J. D. Replication-transcription conflicts in bacteria. *Nat. Rev. Microbiol.* **10**, 449–458 (2012).
89. Helmrich, A., Ballarino, M., Nudler, E. & Tora, L. Transcription-replication encounters, consequences and genomic instability. *Nat. Struct. Mol. Biol.* **20**, 412–418 (2013).
90. Fijalkowska, I. J., Jonczyk, P., Tkaczyk, M. M., Bialoskorska, M. & Schaaper, R. M. Unequal fidelity of leading strand and lagging strand DNA replication on the *Escherichia coli* chromosome. *Proc. Natl Acad. Sci. USA* **95**, 10020–10025 (1998).
91. Wang, J. D., Berkmen, M. B. & Grossman, A. D. Genome-wide coorientation of replication and transcription reduces adverse effects on replication in *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA* **104**, 5608–5613 (2007).
92. Srivatsan, A., Tehranchi, A., MacAlpine, D. M. & Wang, J. D. Co-orientation of replication and transcription preserves genome integrity. *PLoS Genet.* **6**, e1000810 (2010).
93. Rocha, E. P. The replication-related organization of bacterial genomes. *Microbiology* **150**, 1609–1627 (2004).
94. Rocha, E. P. Is there a role for replication fork asymmetry in the distribution of genes in bacterial genomes? *Trends Microbiol.* **10**, 393–395 (2002).
95. Drummond, D. A., Bloom, J. D., Adami, C., Wilke, C. O. & Arnold, F. H. Why highly expressed proteins evolve slowly. *Proc. Natl Acad. Sci. USA* **102**, 14338–14343 (2005).
This paper provides evidence that highly expressed genes experience a higher level of purifying selection against mutations that induce problems in translation and folding.
96. Gout, J. F., Kahn, D., Duret, L. & *Paramecium* Post-Genomics Consortium. The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. *PLoS Genet.* **6**, e1000944 (2010).
97. Park, C., Chen, X., Yang, J. R. & Zhang, J. Differential requirements for mRNA folding partially explain why highly expressed proteins evolve slowly. *Proc. Natl Acad. Sci. USA* **110**, E678–E686 (2013).
98. Chen, X. & Zhang, J. Why are genes encoded on the lagging strand of the bacterial genome? *Genome Biol. Evol.* **5**, 2436–2439 (2013).
99. Szczepanik, D. *et al.* Evolution rates of genes on leading and lagging DNA strands. *J. Mol. Evol.* **52**, 426–433 (2001).
100. McDonald, M. J., Hsieh, Y. Y., Yu, Y. H., Chang, S. L. & Leu, J. Y. The evolution of low mutation rates in experimental mutator populations of *Saccharomyces cerevisiae*. *Curr. Biol.* **22**, 1235–1240 (2012).
101. Turrientes, M. C. *et al.* Normal mutation rate variants arise in a mutator (mut S) *Escherichia coli* population. *PLoS ONE* **8**, e72963 (2013).
102. Wielgoss, S. *et al.* Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc. Natl Acad. Sci. USA* **110**, 222–227 (2013).
103. Williams, L. N., Herr, A. J. & Preston, B. D. Emergence of DNA polymerase antimutators that escape error-induced extinction in yeast. *Genetics* **193**, 751–770 (2013).
104. Lynch, M. Mutation and human exceptionalism: our future genetic load. *Genetics* **202**, 869–875 (2016).
105. Behjati, S. *et al.* Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* **513**, 422–425 (2014).
106. Denamur, E. & Matic, I. Evolution of mutation rates in bacteria. *Mol. Microbiol.* **60**, 820–827 (2006).
107. Desai, M. M. & Fisher, D. S. Beneficial mutation selection balance and the effect of linkage on positive selection. *Genetics* **176**, 1759–1798 (2007).
This article develops a general theory for considering the roles that beneficial mutations have in driving mutation-rate evolution.
108. Raynes, Y. & Sniegowski, P. D. Experimental evolution and the dynamics of genomic mutation rate modifiers. *Heredity* **113**, 375–380 (2014).
109. Giraud, A. *et al.* Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. *Science* **291**, 2606–2608 (2001).
This study considers the interplay between the short-term advantages and long-term disadvantages of mutator alleles.
110. Oliver, A., Baquero, F. & Blázquez, J. The mismatch repair system (*mutS, mutL* and *uvrD* genes) in *Pseudomonas aeruginosa:* molecular characterization of naturally occurring mutants. *Mol. Microbiol.* **43**, 1641–1650 (2002).
111. Pal, C., Maciá, M. D., Oliver, A., Schachar, I. & Buckling, A. Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* **450**, 1079–1081 (2007).
112. Harris, K. Evidence for recent, population-specific evolution of the human mutation rate. *Proc. Natl Acad. Sci. USA* **112**, 3439–3444 (2015).

Acknowledgments
Support was provided by the US Army Research Office Multidisciplinary University Research Initiative (MURI) awards W911NF-09-1-0444 to M.L., P.L.F., H. Tang and S. Finkel, and W911NF-14-1-0411 to M.L., P.L.F., A. Drummond, J. Lennon and J. McKinlay; and the US National Institutes of Health Research Project grant R01-GM036827 to M.L. and W.K.T. We thank R. Ness for providing information, and A. Kondrashov and two anonymous reviewers for their comments.

Competing interests statement
The authors declare no competing interests.

SUPPLEMENTARY INFORMATION
See online article: S1 (table)
ALL LINKS ARE ACTIVE IN THE ONLINE PDF