DOI: 10.1534/genetics.108.090456

The Cellular, Developmental and Population-Genetic Determinants of Mutation-Rate Evolution

Michael Lynch[1]

Department of Biology, Indiana University, Bloomington, Indiana 47405

Manuscript received April 20, 2008
Accepted for publication July 18, 2008

ABSTRACT

Although the matter has been subject to considerable theoretical study, there are numerous open questions regarding the mechanisms driving the mutation rate in various phylogenetic lineages. Most notably, empirical evidence indicates that mutation rates are elevated in multicellular species relative to unicellular eukaryotes and prokaryotes, even on a per-cell division basis, despite the need for the avoidance of somatic damage and the accumulation of germline mutations. Here it is suggested that multicellularity discourages selection against weak mutator alleles for reasons associated with both the cellular and the population-genetic environments, thereby magnifying the vulnerability to somatic mutations (cancer) and increasing the risk of extinction from the accumulation of germline mutations. Moreover, contrary to common belief, a cost of fidelity need not be invoked to explain the lower bound to observed mutation rates, which instead may simply be set by the inability of selection to advance very weakly advantageous antimutator alleles in finite populations.

ALTHOUGH considerable uncertainties remain about the rate of origin and phenotypic consequences of spontaneously arising mutations, it is clear that the vast majority of mutations with effects on fitness are mildly deleterious (Lynch *et al.* 1999; Crow 2000; Baer *et al.* 2007; Eyre-Walker and Keightley 2007). Three lines of defense serve to minimize the accumulation of such mutations. First, most replication polymerases have a strong tendency to incorporate bases complementary to those on template strands, while also harboring a proofreading capacity for removing a substantial fraction of the few base misincorporations that do initially occur (Friedberg *et al.* 2005; Kornberg and Baker 2005; McCulloch and Kunkel 2008). Second, errors remaining after proofreading are scrutinized by postreplicative mismatch-repair (MMR) pathways (Harfe and Jinks-Robertson 2000; Li 2008). Third, natural selection serves as the final arbiter, operating at the population level and eliminating the subset of deleterious germline mutations with selection coefficients large enough to offset the vagaries of random genetic drift (Hartl and Clark 2007).

Although the relative roles of the various factors molding the evolution of the mutation rate in different organisms remain unresolved, some generalizations seem clear. First, because multicellular species experience numerous germline cell divisions per developmental cycle, their per-generation rate of mutation is expected to be magnified relative to that in unicellular species, unless there is a compensatory increase in the efficiency of recognition and repair of premutations at the DNA level. Second, mutationally aggressive genotypes can experience significant levels of somatic damage in multicellular species, as dramatically illustrated by the numerous hereditary forms of cancer in humans (Weinberg 2006). Third, alleles whose products increase the mutation rate are expected to develop statistical associations with detrimental mutations at linked and unlinked loci, while also inducing the origin of more mutator alleles at the same locus in heterozygous carriers. Finally, because effective population sizes are greatly reduced in multicellular relative to unicellular species, the efficiency of selection for mutation-avoidance mechanisms is expected to be reduced (Lynch 2006, 2007).

It is frequently argued that mutation rates are optimized by natural selection to enhance the long-term rate of adaptive change (*e.g.*, Wilke *et al.* 2001; André and Godelle 2006; Denamur and Matic 2006). However, the logic underlying this view applies mainly to asexual populations, where beneficial mutations remain permanently linked to the backgrounds in which they arise (Johnson 1999a; Sniegowski *et al.* 2000). For sexual populations, it has proved difficult to avoid the conclusion that mutation rates are predominantly driven downward by transient linkage of mutator alleles to their deleterious side effects (Sturtevant 1937; Leigh 1970, 1973; Johnson 1999b). Here we explore the extent to which phylogenetic variation in rates of mutation can emerge passively in response to factors defined by the cellular, developmental, and population-genetic environments, purely in the context of deleterious-mutation

[1]*Author e-mail:* milynch@indiana.edu

management. Together, these three levels of biological organization define the power of mutation, selection, and drift operating on the molecular machinery responsible for mutational screening. In the following discussion, the term "mutator" allele is used in a generic sense, in that the theory applies to variants at any locus that alter the mutagenic state of the intracellular environment to any degree.

THE EVOLUTIONARY LABILITY OF REPLICATION/REPAIR-PATHWAY GENES

Despite the vital nature of genomic replication fidelity in all organisms, substantial changes in the molecular determinants of mutation rates have occurred in various phylogenetic lineages, including the complete loss of specific genes for DNA-damage repair. For example, in baker's yeast *Saccharomyces cerevisiae* and mammals, msh2 and msh3 form heterodimers that are involved in the removal of small insertions and deletions during MMR, but msh3 is absent from the Caenorhabditis and Drosophila genomes (Eisen 1998). In addition, although msh3 functions to guard against nonhomologous recombination in the fission yeast *Schizosaccharomyces pombe*, it plays little role in damage repair in this species (Marti *et al.* 2003). Msh1, the mitochondrial MMR gene in yeast and slime mold, appears to be absent from animal and land-plant genomes (Lin *et al.* 2007). Two key base-excision repair pathways in prokaryotes are entirely absent from all well-characterized eukaryotic genomes, and nearly all other such pathways exhibit patchy distributions in both prokaryotes and eukaryotes (Aravind and Koonin 2000; Denver *et al.* 2003; Hardeland *et al.* 2003). Pyrimidine-dimer photolyases, which are involved in the elimination of UV-light-induced damage, appear to have been lost in the ancestor leading to placental mammals (Kanai *et al.* 1997). Many other examples of loss could be cited (Eisen and Hanawalt 1999; Lin *et al.* 2007). Although horizontal transfer may lead to the resurrection of some pathways in prokaryotes (Denamur *et al.* 2000), losses from eukaryotic lineages may be essentially permanent.

Even when retained, the proteins involved in replication fidelity readily undergo significant structural changes. For example, the human complexes involving msh2, msh3, and msh6 not only fail to complement loss-of-function mutations in MMR-deficient yeast, but substantially increase the mutation rate of MMR-proficient yeast, apparently because after localizing correctly to mismatches, they interact inappropriately with the downstream yeast proteins essential to repair (Clark *et al.* 1999). Likewise, the MMR proteins from Pseudomonas increase the mutation rate of MMR-proficient *Escherichia coli*, again presumably by competing with the native MMR machinery for mismatch sites (Oliver *et al.* 2002). Even within a single species, *S. cerevisiae*, *mlh1* from one strain can exhibit strong negative epistasis with the *pms1* gene (the products of which form heterodimers to carry out MMR repair) from another, apparently due to non-complementary amino acid substitutions in each (Heck *et al.* 2006). Although compensatory mutations may be involved in the retention of MMR activity in individual lineages over long spans of evolutionary time, these observations suggest the possibility of intermediate stages in which MMR efficiency is compromised.

Further insight into the magnitude of lineage-specific differences in mutational vulnerability can be gleaned from *in vitro* studies on the replication fidelity of DNA polymerases. A large number of such studies have been performed with a variety of methods, and although they are confined to a small number of model species, the average polymerase error rate (after proofreading, but prior to mismatch repair) appears to be about an order of magnitude lower in prokaryotes than in eukaryotes (Figure 1). (Although the standard errors are too large to discriminate between unicellular and multicellular eukaryotes on a per-cell-division basis, total polymerase error rates/generation are much higher for multicellular species.) The overall pattern appears to be a consequence of a lower proofreading efficiency for eukaryotic polymerases and less clearly associated with differences in base-misincorporation rates. In addition, mismatch repair is about twice as efficient in prokaryotes as in *S. cerevisiae*, with the level for animals appearing to be intermediate (Figure 2).

Direct *in vivo* estimates of the mutation rate support the idea that replication fidelity in eukaryotes is generally no greater than that in prokaryotes. For example, summarizing data from Lynch (2007), the average base-substitutional mutation rate for several prokaryotes is $0.5\ (\mathrm{SE} = 0.2) \times 10^{-9}$/site/replication, whereas that for four unicellular eukaryotes is $1.6\ (0.4) \times 10^{-9}$. After accounting for the number of germline cell divisions (Drost and Lee 1995; Kimble and Ward 1998), the rates for *Caenorhabditis elegans* and *Drosophila melanogaster* are 1.0 and 0.6×10^{-9}/site/replication, whereas those for mice and humans are 0.9 and 0.1×10^{-9}. Although the rate for humans may be exceptionally low on a per-cell division basis, with an average 200 germline cell divisions per generation, the per-generation rate for humans is ~40 times greater than that for a prokaryote. Moreover, for microsatellite loci, the mutation rates for unicellular eukaryotes (*S. cerevisiae* and the slime mold Dictyostelium), *C. elegans*, and mammals scale as 1:5:67 on a per-cell division basis (Seyfert *et al.* 2008).

The preceding observations are collectively consistent with the hypothesis that the accuracy of DNA replication and subsequent error removal is reduced in most eukaryotes relative to prokaryotes and generally more so in multicellular species. However, compromises in repair capacity are by no means restricted to eukaryotes, as mutations in the relevant wild-type proteins in *E. coli* and bacteriophage, some involving single-amino-acid changes, can increase the fidelity of replication as much

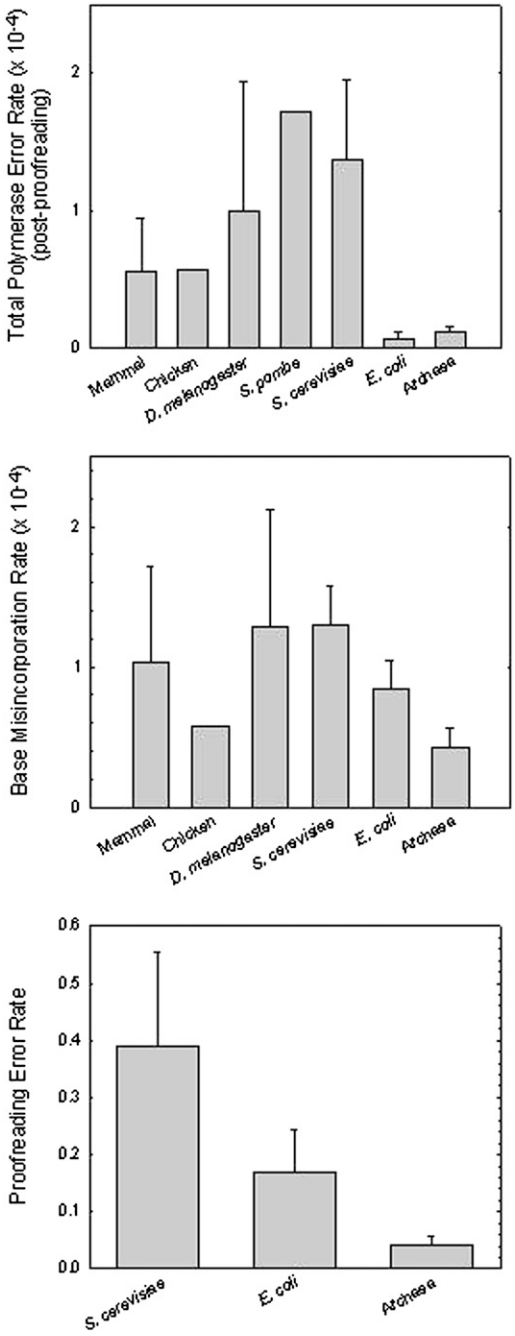

Figure 1.—Estimates of the error rates associated with the primary polymerases involved in chromosomal replication (polymerases α, δ, and ε for eukaryotes and polymerases I, II, and III for prokaryotes). Results are given for base-substitutional changes, averaged over a diversity of *in vitro* experimental studies (details in supplemental material). Top: the total error rate (not including postreplicative mismatch repair). Center: the baseline misincorporation rate, prior to proofreading and mismatch repair. Bottom: error rate associated with proofreading (the ratio of mutation rates with proofreading-proficient polymerase *vs.* that with variants lacking the proofreading domain). The latter measures are more phylogenetically restricted than the former.

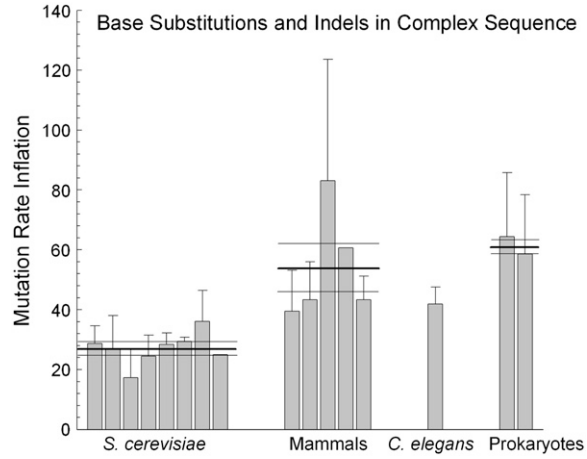

Figure 2.—Average estimates of mismatch-repair efficiency for four phylogenetic groups, reported as the inflation in the mutation rate in experimental constructs in which the MMR pathway has been knocked out (relative to control values). Individual bars for different species denote the results for complete MMR knockouts elicited by the removal of alternative essential genes (or pairs of them) in the MMR pathway. Large horizontal bars denote the mean and SEs of these independent lineage-specific estimates. The results are summarized over a variety of *in vivo* studies involving diverse reporter constructs (details in supplemental material).

as 50-fold (Quiñones and Piechocki 1985; Schaaper 1998; Loh *et al.* 2007).

These types of observations are puzzling for at least two reasons. First, on the basis of the theory cited above, from an adaptational standpoint, the mutation rate is expected to be driven to a lower level in recombining eukaryotic species than in nonrecombining prokaryotes. Second, although many have argued that the cost of high-replication fidelity imposes an ultimate constraint on mutation-avoidance mechanisms (Kimura 1967; Kondrashov 1995; Dawson 1998, 1999; Drake *et al.* 1998; Sniegowski *et al.* 2000; André and Godelle 2006), under this hypothesis the observed pattern of higher mutation rates in eukaryotes would imply a higher cost of eukaryotic replication. There is no evidence for the latter point, and many have actually argued the contrary, *i.e.*, that prokaryotes are under especially strong selection for high rates of genomic replication, although there is no direct evidence for this either (Lynch 2007). An alternative to the view that mutation rates have been optimized to intermediate levels, explored below, is that the selective disadvantages of mutations influencing replication fidelity are often small enough that mild mutator alleles have a high probability of fixation by random genetic drift and/or mutation pressure.

THE SELECTIVE DISADVANTAGE OF MUTATOR ALLELES

In sexual populations, the ability of mutator alleles to rise to appreciable frequencies and/or fixation ulti-

mately depends on the magnitude of their selective disadvantage. Aside from the possibility of pleiotropic physiological effects (which are not considered here), these fitness disadvantages derive from two sources. First, mutator alleles experience transient statistical associations with defective germline mutations that they induce at linked and unlinked loci. Second, in multicellular species, mutator alleles can create loss-of-function mutations in somatic cells, imperiling germline transmission. The consequences of these effects are explored in the following two sections, where we assume a diploid genome.

Gametic-phase disequilibrium: Consider a locus relevant to fitness recombining at rate r with respect to the repair locus (where $r = 0.5$ denotes free recombination). Suppose that in the heterozygous state, the mutator induces mutations at the fitness locus at a rate elevated by Δu per gene, with a heterozygous reduction in fitness equal to hs per mutation (s being the effect of a homozygous mutation, and h being the coefficient of dominance, with $h = 0.5$ implying additive effects). The selective disadvantage of the mutator allele (when at low frequency and only in heterozygous carriers) induced by gametic-phase disequilibrium with the fitness locus is then

$$s_{\rm d} \simeq \frac{hs \cdot \Delta u}{1 - (1 - hs)(1 - r)}, \tag{1a}$$

assuming $\Delta u \ll hs$ (Kimura 1967; Dawson 1999), which for unlinked loci reduces to

$$s_{\rm d,u} \simeq \frac{2hs \cdot \Delta u}{1 + hs}. \tag{1b}$$

A similar result was obtained by Kondrashov (1995).

For fitness loci on the same chromosome as the mutator, the average induced effect can be obtained by applying to Equation 1a Haldane's mapping function, $r = (1 - e^{-2m})/2$, where m is the distance between the mutator and the fitness locus (in morgans). Drawing from data summarized in Lynch (2007), the average chromosome lengths (SE) in fungi, invertebrates, vertebrates, and land plants are 1.87 (0.60), 1.08 (0.19), 1.24 (0.26), and 1.28 (0.10) M, respectively. None of these estimates are greatly different from 1.0, the expectation for chromosomes with single crossover events per arm. Thus, assuming a chromosome length of 1 M, letting c be the chromosomal location of the mutator locus (on the scale of 0.0–1.0 M), and integrating Equation 1a over a uniform distribution of possible fitness-locus locations (summing integrals from 0 to c and c to 1), the average induced effect of a fitness locus on the same chromosome as the mutator is

$$s_{\rm d,l} = s_{\rm d,u}\left[1 + \ln\left(\frac{(1 - ae^{-2c})^{0.5}(1 - ae^{-2(1-c)})^{0.5}}{1 - a}\right)\right], \tag{1c}$$

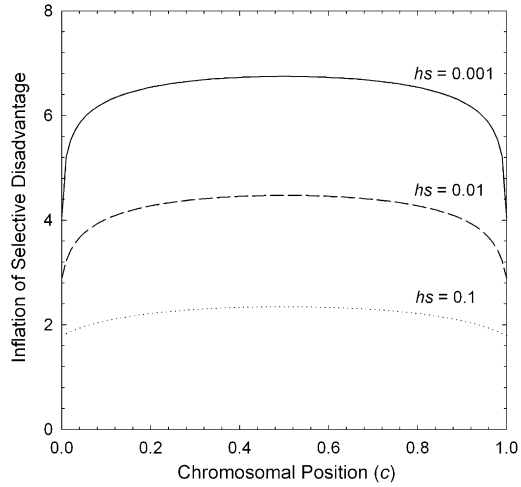

Figure 3.—Inflation of the average induced selective disadvantage of a mutator allele resulting from deleterious mutations at a linked fitness locus relative to that for unlinked loci. hs is the heterozygous effect of the deleterious mutations, and c is the position of the repair locus on a chromosome of 1.0 M. Mutator alleles at the tips of chromosomes are less harmful because they are bounded on only one side by linked mutations.

where $a = (1 - hs)/(1 + hs)$ (assuming that $\Delta u \ll s$). The term in brackets, which represents the elevation in the average fractional induced fitness effect of mutations on the mutator-bearing chromosome relative to that on all others, is close to

$$\phi = 1 + \ln\left(\frac{1 - ae^{-1}}{1 - a}\right), \tag{2}$$

provided the mutator locus is within the interior 80% of the chromosome (Figure 3). For values of hs in the range of 0.001–0.1, which fully covers the range of average effects derived from mutation-accumulation experiments (Lynch and Walsh 1998), the induced effect of individual linked deleterious mutations is approximately two to seven times that for unlinked mutations, as a consequence of their longer associations with the mutator locus.

Assuming L chromosomes of equal length (1 M each), and a haploid genomewide increase in the deleterious mutation rate of ΔU, the preceding results yield a total induced selection coefficient for the mutator allele of

$$s_{\rm d} \simeq \frac{2hs \cdot \Delta U(L - 1 + \phi)}{L(1 + hs)}. \tag{3}$$

This shows that the selective disadvantage of a mutator allele is close to twice the product of the heterozygous fitness effect of new mutations (hs) and the haploid genomewide increase in the deleterious mutation rate (ΔU), unless the chromosome number is very small, and even then not likely to be much more than a few-fold higher (Figure 4). The factor by which $s_{\rm d}$ exceeds $hs \cdot \Delta U$ is equivalent to the average number of generations that

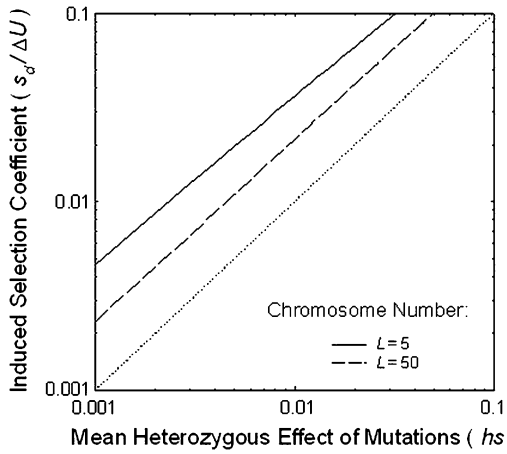

FIGURE 4.—The total selective disadvantage of a mutator allele associated with induced mutations, both linked and unlinked, as a function of the deleterious effect of heterozygous mutations (hs) and the number of chromosomes (L), scaled by the inflation in the mutation rate. The actual selective disadvantage of the mutator allele (s_m) is obtained by multiplying the plotted values by the increase in the haploid genomewide deleterious mutation rate (ΔU). The ratio of the plotted values to the dotted line is the average number of generations that a mutator allele remains associated with the mutations it creates, approximately two when the number of chromosomes is large.

an induced deleterious mutation remains associated with the mutator responsible for its origin (two for unlinked loci). A central point to appreciate here is that because hs for newly arising mutations is generally well below 0.05 (LYNCH and WALSH 1998) and single-amino-acid substitutions in DNA-processing proteins may have arbitrarily small effects on the mutation rate, s_d for many mutator alleles may be sufficiently small (relative to the power of genetic drift) to render them effectively immune to the eyes of natural selection.

Somatic mutations: Quantification of the selective consequences of somatic mutations is difficult in the absence of knowledge of the way in which numbers of somatic-cell mutations translate into the fitness of their bearers. Among other things, the latter issue must depend on the timing of somatic mutations in development; the degree to which aberrant cells can be detected, eliminated, and replaced; the rate of replenishment by proliferating stem cells; and the interactions among cells with independently arising mutations. Moreover, defective cells can lead to a reduction in fitness in numerous ways, *e.g.*, inadequate structural support for tissue function, faulty signaling to neighboring cells, and tumor progression. None of these transient effects have been previously factored into analyses of mutation-rate evolution.

One qualitative way to describe the risk to fitness arising from somatic mutations is to consider an organism with discrete generations characterized by a certain effective number of mutational risks for inviability and/or sterility, T_m. This quantity is a function of the number of cell divisions in various tissues, each weighted by the number of key genes essential for fitness, and also allows for the possibility that more than one mutation is necessary for effective lethality/infertility. The underlying principle here is that nonzero fitness requires that all mutational risks be avoided, leading to the general definition of relative fitness as $e^{-u_{MM}T_m}$ and $e^{-u_{Mm}T_m}$ for the nonmutator and heterozygous mutator genotypes. The selection coefficient opposing the mutator allele is then

$$s_s \simeq e^{-u_{MM}T_m} - e^{-u_{Mm}T_m}. \quad (4)$$

As a heuristic example of how T_m might be defined, consider a developmentally simple organism with determinate growth involving a homogeneous population of cells, with the completion of development occurring at the point at which each cell has divided t times. In this case, with a constant number (n) of mutational targets being exposed to mutation at each cell division, $T_m = nd$, where $d = \sum_{i=1}^{t} 2^i$ is the total number of cell divisions per individual at the completion of development. More complex developmental scenarios can be accommodated by nesting groups of cells with common developmental schedules and mutational vulnerabilities (*e.g.*, cells within specific tissues). For example, for an organism in which cells differentiate into two types following an initial t_0 divisions, one type having n_1 mutational targets and undergoing t_1 additional divisions and the other having features n_2 and t_2, $T_m = n_0d_0 + n_1d_1 + n_2d_2$, where $d_j = \sum_{i=1}^{t_j} 2^i$. With N total cell types in an organism (*e.g.*, three in the preceding example), this expression generalizes to $T_m = N(\bar{n}\,\bar{d} + \sigma_{n,d})$, where $\bar{n}$ is the mean number of mutational targets, $\bar{d}$ is the mean number of cell divisions, and $\sigma_{n,d}$ is the covariance between the number of cell cycles and mutational targets for individual cell types. Clearly, the mutational costs of multicellularity will be magnified if highly proliferating cell types also harbor elevated numbers of key mutational targets (positive $\sigma_{n,d}$). Although these types of constructs subsume many aspects of an organism's molecular, cellular, and developmental biology into a few summary parameters, the general approach may prove useful for understanding the vulnerability of alternative developmental architectures to somatic mutation. For example, a good deal might be learned about the role of somatic mutation in the evolution of developmental complexity by evaluating whether $\sigma_{n,d}$ is negative.

Clearly, the fitness of *Mm* individuals declines with both u_{Mm} and T_m. More relevant, however, is the fact that the selection coefficient associated with somatic mutations, s_s, is not monotonic in T_m, but reaches a maximum at an intermediate level of T_m (Figure 5). For $u_{Mm}T_m \ll 1$, the probability of a mutation-free individual is essentially linear with T_m, and $s_s \simeq (u_{Mm} - u_{MM})T_m$, but at larger T_m, individuals begin to acquire multiple muta-

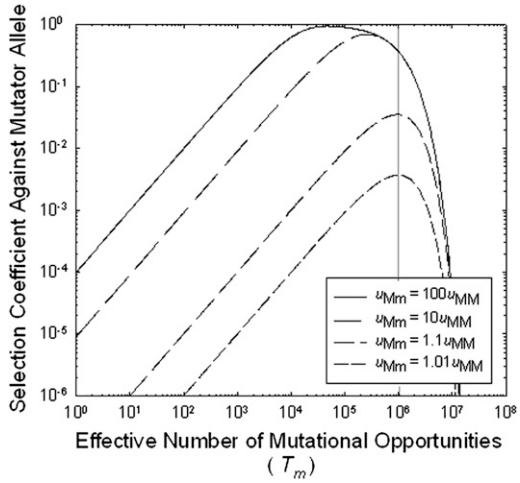

FIGURE 5.—The selection coefficient for a mutator allele induced by somatic mutations, as a function of the number of mutational risks prior to the completion of development. u_{MM} is assumed to equal 10^{-6}. The vertical dotted line denotes $1/u_{MM}$.

tional hits. Because just a single *effective* mutation entirely eliminates fitness under the proposed model, this saturating effect results in a reduction in the magnitude of the fitness difference between *MM* and *Mm* genotypes, even though the absolute fitness of each continues to decline.

Letting $\theta = u_{Mm}/u_{MM}$ be the proportional inflation of the mutation rate in mutator heterozygotes, s_s attains a maximum of $[(\theta - 1)/\theta]e^{(\ln\theta)/(1-\theta)}$ when $T_m = (\ln\theta)/[u_{MM}(\theta - 1)]$. For weak mutator alleles ($u_{Mm} < 1.1u_{MM}$), the maximum selective disadvantage associated with somatic mutations, $s_{s,max} \simeq (\theta - 1)/e$, appears at $T_m \simeq 1/u_{MM}$, *i.e.*, when the average effective number of mutational hits in the nonmutator background is equal to one, and then drops off precipitously with higher T_m. Thus, any selective forces on developmental architecture that cause T_m to surpass this point will enhance the likelihood of invasion of weak mutator alleles.

EXPECTED FREQUENCY OF DETRIMENTAL ALLELES FOR REPLICATION/REPAIR GENES

In the traditional framework of population genetics, the mutation rate is assumed to be independent of the genotypic background. However, such treatment is at best a first-order approximation, because as noted above, any mutation that negatively influences the functionality of a replication/repair locus will magnify the mutation rate of not just all other cosegregating loci but of the replication/repair locus itself. In principle, such a condition can lead to a positive feedback that elevates the mutation rate at the population level.

For diploid species, a key determinant of the evolutionary dynamics of mutant alleles at loci under selection, including those influencing the mutation rate, is the degree to which detrimental alleles are recessive with respect to fitness. For loci involved in replication fidelity, we must also consider the extent to which the mutation rate is elevated in heterozygous carriers of defective alleles. Most observations suggest that loss-of-function alleles at MMR loci are only partially recessive. For example, whereas homozygotes for knockout mutations in genes essential for MMR generally have zero fitness (CHAO and LIPKIN 2006), in mammals, heterozygous carriers exhibit Lynch syndrome, a predisposition to cancer resulting from an elevated rate of mutation in somatic cell lineages (generally in the colon) (MITCHELL *et al.* 2002; DE LA CHAPELLE 2005). Because tumor development generally occurs late in life, the effects on reproductive fitness are small to moderate. Homozygous knockouts of MMR genes can elevate mutation rates by as much as 100-fold, whereas the effects in heterozygotes are much more modest, again implying partial dominance ($\simeq$ 10% penetrance) (BORGDORFF *et al.* 2005; HEGAN *et al.* 2006; BURR *et al.* 2007). On the other hand, consistent with the observations on heterologous systems noted above, the degree of dominance for missense mutations with moderate effects on repair efficiency can be much more pronounced (PARSONS *et al.* 1995; DROTSCHMANN *et al.* 1999), with some data suggesting that heterozygous mammalian carriers of missense mutations at MMR loci have mutation rates elevated by factors of 5–10 (COOLBAUGH-MURPHY *et al.* 2004; ALAZZOUZI *et al.* 2005).

A similar pattern exists for mutations in polymerase genes and those for proteins involved in other aspects of DNA maintenance, with partially defective alleles exhibiting weak but significant mutation-rate effects in heterozygotes ($\simeq$ 10% dominance), as well as significant effects on cancer susceptibility and life span (PAVLOV *et al.* 2001; THOMPSON *et al.* 2005; CABELOF *et al.* 2006; VENKATESAN *et al.* 2007). Effectively lethal, knockout mutations for such genes are more nearly completely recessive with respect to the mutation rate (GOLDSBY *et al.* 2001; TRIFUNOVIC *et al.* 2004). All of these observations are quite consistent with a large body of data on other quantitative trait loci indicating that deleterious mutations with large effects tend to be nearly completely recessive, while those with moderate effects act in a more additive fashion (SIMMONS and CROW 1977; LYNCH and WALSH 1998). However, there is a clear need for more work on the mutation-rate/fitness aspects of heterozygous carriers of mutator alleles.

To obtain insight into the power of mutators to drive the overall mutational properties of a population, consider the situation that is least conducive to mutator-allele proliferation—the classical case of a population with a large enough effective size that the power of selection is in substantial excess of the power of random genetic drift. Under these conditions, a deleterious mutator allele will either be maintained at an intermediate frequency by mutation–selection balance or driven to

fixation by mutation pressure. Letting the fitnesses of the *MM*, *Mm*, and *mm* genotypes be 1, $1 - h_m s_m$, and $1 - s_m$, respectively, and assuming no back mutation of *m* to *M*, the cross-generational dynamics of the frequency of the mutator allele *m* are defined by

$$q' = \frac{1}{\overline{W}}[(1-q)^2 u_{MM} + (1-q)q(1-h_m s_m)(1+u_{Mm}) + q^2(1-s_m)], \quad (5)$$

where $\overline{W}$ is the mean population fitness. At equilibrium, $q' = q$, and the expected frequency of the functional allele, $p = 1 - q$, is the solution to

$$\hat{p}^2[s_m(1-2h_m)] + \hat{p}[(1-u_{MM}) - (1-u_{Mm})(1-h_m s_m) - 2s_m(1-h_m)] + [(1-u_{Mm})(1-h_m s_m) - (1-s_m)] = 0. \quad (6)$$

Provided $h_m \leq 0.5$, the likely situation for most mutations, there is one equilibrium solution for any set of mutation and selection parameters.

To connect the following results with the points made in the previous section, the heterozygous fitness effect of a mutator allele, $h_m s_m$, can be viewed as the sum of the effects from gametic-phase disequilibrium and somatic mutation, $s_d + s_s$. Except in the case of highly aggressive and highly penetrant mutators, it appears likely that $h_m s_m$ will generally be <0.1. Some special cases are instructive.

First, if the mutator allele *m* is completely recessive ($h_m = 0$; $0 < s_m \leq 1$), its equilibrium frequency is

$$\hat{q} = (\Delta u/2s_m) + \sqrt{(u_{MM}/s_m) + (\Delta u/2s_m)^2}, \quad (7a)$$

where as above, $\Delta u = u_{Mm} - u_{MM}$ is the elevation of the mutation rate in heterozygous carriers. This expression reduces to the classical result, $\hat{q} = \sqrt{u/s_m}$, when the mutation rate is assumed to be independent of the genetic background ($\Delta u = 0$). For recessive, lethal mutations ($h_m = 0$; $s_m = 1$),

$$\hat{q} = (\Delta u/2) + \sqrt{u_{MM} + (\Delta u/2)^2}. \quad (7b)$$

This reduces to the classical result for a recessive lethal, $\hat{q} = \sqrt{u}$, when the mutation rate is independent of the genetic background. However, when the mutation rate in *Mm* individuals is sufficiently high ($\Delta u \gg \sqrt{u_{MM}}$), the equilibrium frequency of a recessive lethal mutator allele approaches the mutation rate for heterozygotes, $\hat{q} \simeq u_{Mm}$. These results show that the equilibrium frequency of a defective repair allele with recessive fitness effects can be elevated up to an order of magnitude if the heterozygous carriers experience increased mutation rates relative to the "wild-type" background (Figure 6). In extreme cases (where s_m is small), the mutator allele can be driven to fixation by mutation pressure alone, leading to the elevation of the average population mutation rate from $\sim u_{MM}$ to u_{mm}.

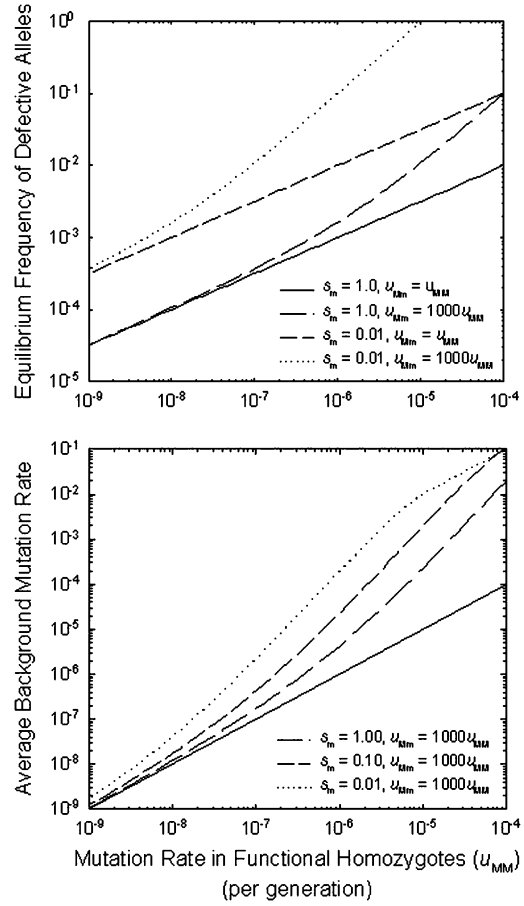

FIGURE 6.—Top: equilibrium frequencies of mutator alleles with recessive fitness effects, for populations of effectively infinite size, given for lethal fitness effects ($s_m = 1.0$) and for mildly deleterious effects ($s_m = 0.01$), for situations in which the mutation rate to defective alleles is independent of the repair-locus genotype ($u_{MM} = u_{Mm}$) and when the rate is 1000 times higher in heterozygous repair-locus carriers of single defective alleles. Bottom: average mutation rates at a reference locus for populations in mutation–selection equilibrium, obtained from the results in the top assuming random mating. Here it is assumed that the mutation rates of both the mutant heterozygotes and the homozygotes are equal to u_{Mm}. The solid line is the reference for the situation in which the mutation rate is independent of the genotypic background.

Second, if the fitness effects of the defective repair allele are additive ($h_m = 0.5$),

$$\hat{q} \simeq \frac{2u_{MM}}{s_m - 2\Delta u}, \quad (8)$$

which reduces to the classical result, $\hat{q} = 2u/s_m$, when $\Delta u = 0$ and closely approximates it when $\Delta u \ll s_m$ (Figure 7). Thus, if the mutation rate in heterozygotes is elevated to a level of $s_m/2$, the null repair allele will be driven to a frequency of 1.0 by mutation pressure alone; *e.g.*, if $s_m = 10^{-4}$, fixation will occur if $\Delta u > 0.5 \times 10^{-4}$. A particularly simple and illuminating result follows for unicellular species, where somatic mutations can be ignored. In this case, assuming a large number of chromosomes and

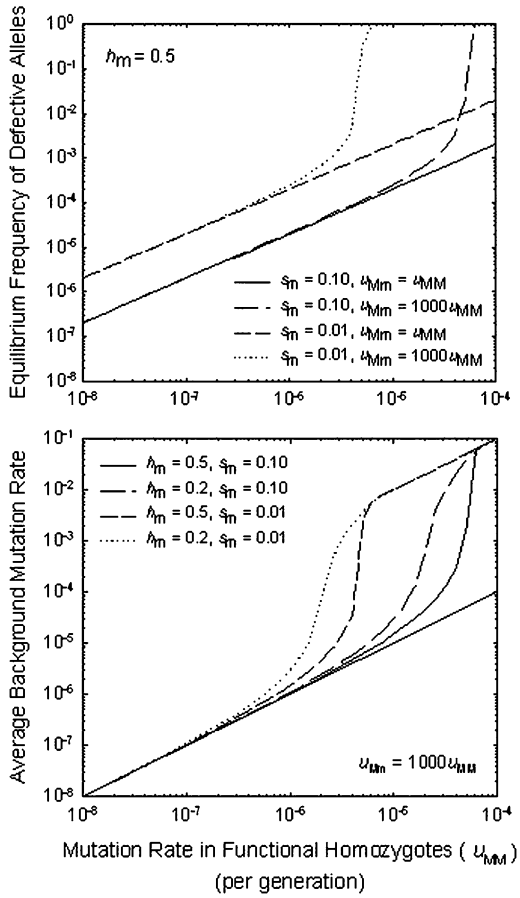

FIGURE 7.—Equilibrium frequencies of nonfunctional repair alleles with nonrecessive effects on fitness, for populations of effectively infinite size (as in Figure 6).

mutations with additive effects, $s_m/2 \simeq sn\Delta u$, where n is the total number of loci critical to fitness, which leads to $\hat{q} \simeq u_{MM}/[\Delta u(sn-1)]$. Making the reasonable assumption that $sn \gg 1$ (the simultaneous mutation of all loci critical to fitness is well beyond lethality), this implies that a mutator will drive itself to fixation if the increase in the genomic deleterious mutation rate ($\Delta U = n\Delta u$) is smaller than u_{MM}/s, where the latter quantity is approximately half the frequency of a deleterious allele at an average locus under selection–mutation equilibrium in the nonmutator background.

As genomic deleterious mutation rates (U) are typically in the range of 0.1–1.0 (LYNCH and WALSH 1998), and mutator alleles can increase the mutation rate by severalfold or more, the preceding results suggest that the conditions for population-level expansions of mild mutator alleles are by no means prohibitive, even in relatively large populations. Indeed, because all of the preceding derivations assume the selective disadvantage of a mutator allele to be its expected value when selection–mutation balance has been achieved at all associated loci, the actual conditions for spread must be less than those outlined above, and perhaps considerably so. This is because mutator alleles initially arise on relatively mutation-free genetic backgrounds and hence require many generations to realize their maximum deleterious effects (JOHNSON 1999a).

EVOLUTIONARY IMPLICATIONS

The preceding analyses raise a number of issues about mutation-rate evolution that appear to have been previously underappreciated. All other things being equal, the mutational cost of development in a multicellular species is a function of the product of the number of divisions in various cell lineages and the mutation rate per cell division, so, in principle, the cost of increasing multicellularity can be eliminated by reducing the mutation rate. However, such compensation is not observed, with eukaryotes appearing to have lower levels of replication fidelity than do prokaryotes, on both a per-cell-division and a per-generation basis. Thus, the evolution of complex multicellularity appears to have been accompanied by an increase in mutational cost, both in terms of induced somatic damage and in terms of increased deleterious-mutation accumulation in the germline. At least three factors may have contributed to such a condition.

First, because replication is mutagenic, any increase in the number of germline-cell divisions will encourage a higher per-generation mutation rate, causing the latter to gradually approach the average selective consequence of cumulative induced mutations (denominator of Equation 8). As this point is approached, the appearance of mild mutator alleles with sufficient strength to offset the selective disadvantage of associated mutations becomes increasingly likely, at which point a mutator allele would drive itself to fixation by its own mutation pressure. Because this phase transition in the behavior of partially dominant mutators typically occurs over a very narrow range of the mutation rate for the nonmutator (Figure 7), developmental factors that cause an increase in the per-generation germline mutation rate in nonmutators might sometimes suddenly pave the way for further loss of repair capacity.

Second, a threshold of developmental complexity appears to exist, beyond which the relative selective disadvantage of mutator alleles resulting from somatic mutations begins to decline, despite the increased absolute cost of multicellularity. Such nonlinear behavior results because cell lineages experiencing large numbers of divisions have increasingly high probabilities (asymptotically approaching 1.0) of accumulating degenerative mutations in their terminal members, even when they contain only "nonmutator" alleles. For weak mutator alleles, the selective disadvantage associated with somatic mutations reaches a maximum at the point where the average effective number of mutational hits in the nonmutator background is equal to one per generation per individual. Although this particular result may highlight a general principle regarding the mutational consequences

of multicellularity, where extant multicellular organisms reside with respect to the critical mutational threshold remains to be determined.

Third, an apparently unavoidable impediment to the evolution of low mutation rates in multicellular species is simply the general reduction in effective population size (N_e) in organisms with large somas (Lynch 2007), which will necessarily magnify the vulnerability of mutator alleles to fixation by random genetic drift. Any degenerative mutation in an allele at a replication/repair locus that causes a change in fitness $<(2N_e)^{-1}$ is expected to proceed to fixation at a rate in excess of the neutral expectation (because of the mutational drive of the mutator itself). Likewise, weak "antimutators" with selective advantages $<(2N_e)^{-1}$ cannot be advanced by selection. From the standpoint of the deleterious mutational load resulting from gametic-phase disequilibrium, the upper limit to the selective advantage of an antimutator allele (which accrues with a mutation rate of zero) can be found by setting ΔU equal to the prevailing genomic deleterious mutation rate (U) in Equation 3, which yields $\simeq 2hsU$ (the expected genome-wide reduction in average fitness from a single generation of mutations). Thus, even for unicellular species with large N_e, the idea that mutation rates are often driven to minimal levels defined by the cost of replication fidelity appears to be unnecessary. Once the mutation rate is reduced to a sufficiently low level by selective forces, further small reductions will have insufficiently large selective consequences to offset the vagaries of random genetic drift, leading to a sort of quasi-equilibrium minimal mutation rate defined by the population-genetic environment alone. Palmer and Lipsitch (2006) have also alluded to this possibility in asexual populations.

Paradoxically then, multicellularity not only magnifies the likelihood of fitness reduction associated with somatic damage, but also actually encourages the accumulation of mutator alleles that cause such damage. Thus, although plausible arguments have been made that the origins and maintenance of multicellularity are driven by external forces of natural selection (Buss 1987; Maynard Smith and Szathmáry 1995; Michod 1999), the conclusion that multicellularity imperils the maintenance of mechanisms for deleterious-mutation management seems inescapable. To resolve the above issues in a more quantitative sense, information will be required on the distribution of Δu for random mutational changes to genes involved in replication and repair, as well as on the numbers and distributions of mutational effects in various somatic cellular environments.

Dating back to Haldane (1937), there has been a long-standing interest in the load that recurrent deleterious mutations impose on the fitness of a population. For mutations with partial expression in heterozygotes and nonepistatic fitness effects, relative to the situation in which all loci are fixed for beneficial alleles, mean population fitness associated with loci kept in selection–mutation balance is equal to e^{-2U}, where U is the haploid genomic deleterious mutation rate. Thus, under this model, the fractional reduction in mean fitness due to segregating mutations is $\sim 2U$, provided $U < 1$, independent of the effects of individual mutations. However, the preceding results indicate that most species are likely to harbor a "hidden" mutation load associated with the fixation of mildly deleterious mutations at replication/repair loci. Each such mutation will further reduce mean population fitness by its homozygous effect, s_m.

Finally, it is notable that when mutator alleles rise to moderate frequencies, considerable heterogeneity in the mutation rate is expected among individuals, with most of the mutations at the population level arising in genetic backgrounds transiently associated with mutator heterozygotes. For example, for the case in which the mutator allele is a recessive lethal with the mutation rate to defective alleles being inflated 100-fold in mutator heterozygotes (relative to the wild-type rate of $u_{MM} = 10^{-5}$), the equilibrium frequency of the null repair allele is ~ 0.0037. In that case, just 0.74% of the population (the heterozygous mutator carriers) accounts for $\sim 42\%$ of the mutations arising in the population each generation. Such heterogeneity has considerable implications for studies that rely on small numbers of natural isolates to estimate species-specific mutation rates.

I am grateful to Matthew Hahn, Ignasi Lucas, and Paul Sniegowski for helpful comments. This work was funded by National Institutes of Health grant GM36827 to the author and W. Kelley Thomas.

LITERATURE CITED

Alazzouzi, H., E. Domingo, S. González, I. Blanco, M. Armengol *et al.*, 2005 Low levels of microsatellite instability characterize *MLH1* and *MSH2* HNPCC carriers before tumor diagnosis. Hum. Mol. Genet. **14:** 235–239.

André, J. B., and B. Godelle, 2006 The evolution of mutation rate in finite asexual populations. Genetics **172:** 611–626.

Aravind, L., and E. V. Koonin, 2000 The alpha/beta fold uracil DNA glycosylases: a common origin with diverse fates. Genome Biol. **1**(4): RESEARCH0007.

Baer, C. F., M. M. Miyamoto and D. R. Denver, 2007 Mutation rate variation in multicellular eukaryotes: causes and consequences. Nat. Rev. Genet. **8:** 619–631.

Borgdorff, V., S. van Hees-Stuivenberg, C. M. Meijers and N. de Wind, 2005 Spontaneous and mutagen-induced loss of DNA mismatch repair in *Msh2*-heterozygous mammalian cells. Mutat. Res. **574:** 50–57.

Burr, K. L., A. van Duyn-Goedhart, P. Hickenbotham, K. Monger, P. P. van Buul *et al.*, 2007 The effects of MSH2 deficiency on spontaneous and radiation-induced mutation rates in the mouse germline. Mutat. Res. **617:** 147–151.

Buss, L. W., 1987 *The Evolution of Individuality.* Princeton University Press, Princeton, NJ.

Cabelof, D. C., Y. Ikeno, A. Nyska, R. A. Busuttil, N. Anyangwe *et al.*, 2006 Haploinsufficiency in DNA polymerase beta increases cancer risk with age and alters mortality rate. Cancer Res. **66:** 7460–7465.

Chao, E. C., and S. M. Lipkin, 2006 Molecular models for the tissue specificity of DNA mismatch repair-deficient carcinogenesis. Nucleic Acids Res. **34:** 840–852.

Clark, A. B., M. E. Cook, H. T. Tran, D. A. Gordenin, M. A. Resnick *et al.*, 1999 Functional analysis of human MutSα and MutSβ complexes in yeast. Nucleic Acids Res. **27:** 736–742.

Coolbaugh-Murphy, M., A. Maleki, L. Ramagli, M. Frazier, B. Lichtiger *et al.*, 2004 Estimating mutant microsatellite allele frequencies in somatic cells by small-pool PCR. Genomics **84:** 419–430.

Crow, J. F., 2000 The origins, patterns and implications of human spontaneous mutation. Nat. Rev. Genet. **1:** 40–47.

Dawson, K. J., 1998 Evolutionarily stable mutation rates. J. Theor. Biol. **194:** 143–157.

Dawson, K. J., 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.

de la Chapelle, A., 2005 The incidence of Lynch syndrome. Fam. Cancer **4:** 233–237.

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

Denamur, E., G. Lecointre, P. Darlu, O. Tenaillon, C. Acquaviva *et al.*, 2000 Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. Cell **103:** 711–721.

Denver, D. R., S. L. Swenson and M. Lynch, 2003 An evolutionary analysis of the helix-hairpin-helix superfamily of DNA repair glycosylases. Mol. Biol. Evol. **20:** 1603–1611.

Drake, J. W., B. Charlesworth, D. Charlesworth and J. F. Crow, 1998 Rates of spontaneous mutation. Genetics **148:** 1667–1686.

Drost, J. B., and W. R. Lee, 1995 Biological basis of germline mutation: comparisons of spontaneous germline mutation rates among *Drosophila*, mouse, and human. Environ. Mol. Mutagen. **25**(Suppl. 26): 48–64.

Drotschmann, K., A. B. Clark and T. A. Kunkel, 1999 Mutator phenotypes of common polymorphisms and missense mutations in *MSH2*. Curr. Biol. **9:** 907–910.

Eisen, J. A., 1998 A phylogenomic study of the MutS family of proteins. Nucleic Acids Res. **26:** 4291–4300.

Eisen, J. A., and P. C. Hanawalt, 1999 A phylogenomic study of DNA repair genes, proteins, and processes. Mutat. Res. **435:** 171–213.

Eyre-Walker, A., and P. D. Keightley, 2007 The distribution of fitness effects of new mutations. Nat. Rev. Genet. **8:** 610–618.

Friedberg, E. C., G. C. Walker, W. Siede, R. D. Wood, R. A. Schultz *et al.*, 2005 *DNA Repair and Mutagenesis*. ASM Press, Herndon, VA.

Goldsby, R. E., N. A. Lawrence, L. E. Hays, E. A. Olmsted, X. Chen *et al.*, 2001 Defective DNA polymerase-delta proofreading causes cancer susceptibility in mice. Nat. Med. **7:** 638–639.

Haldane, J. B. S., 1937 The effect of variation on fitness. Am. Nat. **71:** 337–349.

Hardeland, U., M. Bentele, J. Jiricny and P. Schär, 2003 The versatile thymine DNA-glycosylase: a comparative characterization of the human, *Drosophila* and fission yeast orthologs. Nucleic Acids Res. **31:** 2261–2271.

Harfe, B. D., and S. Jinks-Robertson, 2000 DNA mismatch repair and genetic instability. Annu. Rev. Genet. **34:** 359–399.

Hartl, D. L., and A. G. Clark, 2007 *Principles of Population Genetics*, Ed. 4. Sinauer Associates, Sunderland, MA.

Heck, J. A., J. L. Argueso, Z. Gemici, R. G. Reeves, A. Bernard *et al.*, 2006 Negative epistasis between natural variants of the *Saccharomyces cerevisiae MLH1* and *PMS1* genes results in a defect in mismatch repair. Proc. Natl. Acad. Sci. USA **103:** 3256–3261.

Hegan, D. C., L. Narayanan, F. R. Jirik, W. Edelmann, R. M. Liskay *et al.*, 2006 Differing patterns of genetic instability in mice deficient in the mismatch repair genes *Pms2*, *Mlh1*, *Msh2*, *Msh3* and *Msh6*. Carcinogenesis **27:** 2402–2408.

Johnson, T., 1999a The approach to mutation-selection balance in an infinite asexual population, and the evolution of mutation rates. Proc. Biol. Sci. **266:** 2389–2397.

Johnson, T., 1999b Beneficial mutations, hitchhiking and the evolution of mutation rates in sexual populations. Genetics **151:** 1621–1631.

Kanai, S., R. Kikuno, H. Toh, H. Ryo and T. Todo, 1997 Molecular evolution of the photolyase-blue-light photoreceptor family. J. Mol. Evol. **45:** 535–548.

Kimble, J., and S. Ward, 1998 Germ-line development and fertilization, pp. 191–213 in *The Nematode Caenorhabditis elegans*, edited by W. B. Wood. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

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

Kondrashov, A. S., 1995 Modifiers of mutation-selection balance: general approach and the evolution of mutation rates. Genet. Res. **66:** 53–70.

Kornberg, A., and T. A. Baker, 2005 *DNA Replication*. University Science Books, Sausalito, CA.

Leigh, Jr., E. G., 1970 Natural selection and mutability. Am. Nat. **104:** 301–305.

Leigh, Jr., E. G., 1973 The evolution of mutation rates. Genetics **73**(Suppl.): 1–18.

Li, G. M., 2008 Mechanisms and functions of DNA mismatch repair. Cell Res. **18:** 85–98.

Lin, Z., M. Nei and H. Ma, 2007 The origins and early evolution of DNA mismatch repair genes—multiple horizontal gene transfers and co-evolution. Nucleic Acids Res. **35:** 7591–7603.

Loh, E., J. Choe and L. A. Loeb, 2007 Highly tolerated amino acid substitutions increase the fidelity of *Escherichia coli* DNA polymerase I. J. Biol. Chem. **282:** 12201–12209.

Lynch, M., 2006 The origins of eukaryotic gene structure. Mol. Biol. Evol. **23:** 450–468.

Lynch, M., 2007 *The Origins of Genome Architecture*. Sinauer Associates, Sunderland, MA.

Lynch, M., and J. B. Walsh, 1998 *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.

Lynch, M., J. Blanchard, D. Houle, T. Kibota, S. Schultz *et al.*, 1999 Spontaneous deleterious mutation. Evolution **53:** 645–663.

Marti, T. M., A. A. Mansour, E. Lehmann and O. Fleck, 2003 Different frameshift mutation spectra in non-repetitive DNA of MutSα- and MutLα-deficient fission yeast cells. DNA Repair **2:** 571–580.

Maynard Smith, J., and E. Szathmáry, 1995 *The Major Transitions in Evolution*. W. H. Freeman, San Francisco.

McCulloch, S. D., and T. A. Kunkel, 2008 The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. Cell Res. **18:** 148–161.

Michod, R. E., 1999 *Darwinian Dynamics: Evolutionary Transitions in Fitness and Individuality*. Princeton University Press, Princeton, NJ.

Mitchell, R. J., S. M. Farrington, M. G. Dunlop and H. Campbell, 2002 Mismatch repair genes *hMLH1* and *hMSH2* and colorectal cancer: a HuGE review. Am. J. Epidemiol. **156:** 885–902.

Oliver, A., F. Baquero and J. Blázquez, 2002 The mismatch repair system (*mutS*, *mutL* and *uvrD* genes) in *Pseudomonas aeruginosa*: molecular characterization of naturally occurring mutants. Mol. Microbiol. **43:** 1641–1650.

Palmer, M. E., and M. Lipsitch, 2006 The influence of hitchhiking and deleterious mutation upon asexual mutation rates. Genetics **173:** 461–472.

Parsons, R., G. M. Li, M. Longley, P. Modrich, B. Liu *et al.*, 1995 Mismatch repair deficiency in phenotypically normal human cells. Science **268:** 738–740.

Pavlov, Y. I., P. V. Shcherbakova and T. A. Kunkel, 2001 *In vivo* consequences of putative active site mutations in yeast DNA polymerases alpha, epsilon, delta, and zeta. Genetics **159:** 47–64.

Quiñones, A., and R. Piechocki, 1985 Isolation and characterization of *Escherichia coli* antimutators: a new strategy to study the nature and origin of spontaneous mutations. Mol. Gen. Genet. **201:** 315–322.

Schaaper, R. M., 1998 Antimutator mutants in bacteriophage T4 and *Escherichia coli*. Genetics **148:** 1579–1585.

Seyfert, A. L., M. E. A. Cristescu, L. Frisse, S. Schaack, W. K. Thomas *et al.*, 2008 The rate and spectrum of microsatellite mutation in *Caenorhabditis elegans* and *Daphnia pulex*. Genetics **178:** 2113–2121.

Simmons, M. J., and J. F. Crow, 1977 Mutations affecting fitness in *Drosophila* populations. Annu. Rev. Genet. **11:** 49–78.

Sniegowski, P. D., P. J. Gerrish, T. Johnson and A. Shaver, 2000 The evolution of mutation rates: separating causes from consequences. BioEssays **22:** 1057–1066.

Sturtevant, A. H., 1937 Essays on evolution. I. On the effects of selection on mutation rate. Q. Rev. Biol. **12:** 464–476.

Thompson, D., S. Duedal, J. Kirner, L. McGuffog, J. Last *et al.*, 2005 Cancer risks and mortality in heterozygous ATM mutation carriers. J. Natl. Cancer Inst. **97:** 813–822.

Trifunovic, A., A. Wredenberg, M. Falkenberg, J. N. Spelbrink, A. T. Rovio *et al.*, 2004 Premature ageing in mice expressing defective mitochondrial DNA polymerase. Nature **429:** 417–423.

Venkatesan, R. N., P. M. Treuting, E. D. Fuller, R. E. Goldsby, T. H. Norwood *et al.*, 2007 Mutation at the polymerase active site of mouse DNA polymerase delta increases genomic instability and accelerates tumorigenesis. Mol. Cell. Biol. **27:** 7669–7682.

Weinberg, R. A., 2006 *The Biology of Cancer*. Garland Science, London.

Wilke, C. O., J. L. Wang, C. Ofria, R. E. Lenski and C. Adami, 2001 Evolution of digital organisms at high mutation rates leads to survival of the flattest. Nature **412:** 331–333.

Communicating editor: J. A. Birchler