Evolution of the mutation rate

Michael Lynch

Department of Biology, Indiana University, Bloomington, IN 47405

Understanding the mechanisms of evolution requires information on the rate of appearance of new mutations and their effects at the molecular and phenotypic levels. Although procuring such data has been technically challenging, high-throughput genome sequencing is rapidly expanding knowledge in this area. With information on spontaneous mutations now available in a variety of organisms, general patterns have emerged for the scaling of mutation rate with genome size and for the likely mechanisms that drive this pattern. Support is presented for the hypothesis that natural selection pushes mutation rates down to a lower limit set by the power of random genetic drift rather than by intrinsic physiological limitations, and that this has resulted in reduced levels of replication, transcription, and translation fidelity in eukaryotes relative to prokaryotes.

The challenge of estimating mutation rates

Because mutation is the ultimate source of all variation, both adaptive and deleterious, a mechanistic understanding of the evolutionary process will be incomplete until a detailed account has been made of the rate of origin, molecular nature, and phenotypic consequences of spontaneous alterations for a diversity of organisms. Owing to the extreme rarity of mutational events and their frequent elimination by selection in natural environments, most prior insights into the molecular aspects of mutation have been derived from a few reporter constructs in a handful of model species [1]. This situation is now rapidly changing because the application of high-throughput genome sequencing to mutation accumulation experiments now allows the identification of *de novo* mutations in an essentially unbiased manner.

At least two broad generalizations now seem possible. First, there is a dramatic reversal in the directional relationship between the mutation rate and genome size from viruses to cellular microbes to multicellular species, with prokaryotes having higher levels of fidelity than eukaryotes at the levels of replication, transcription, and translation. Second, in multicellular species, somatic mutation rates are notably higher than germline rates, whereas on a cell division basis the latter are not much different than rates observed in unicellular species. With these observations in hand, we are now in a better position to understand the causes and consequences of mutation rate evolution in various phylogenetic lineages.

Phylogenetic scaling of the mutation rate

In one of the first attempts to understand the patterning of mutation rates across various organisms, Drake concluded

that the mutation rate/nucleotide site/generation (u) scales inversely with genome size (G) in DNA-based microbes [2], and this further implies that the mutation rate/genome/generation (uG) is essentially constant across all microbial life. Because this early analysis was based on just seven

---

Glossary

Antimutator allele: a mutant copy of any gene that confers a reduction (of arbitrary size) in the mutation rate.

Equilibrium level of heterozygosity (π_s): approximately $4N_e u$ in diploid species and $2N_e u$ in haploids under neutrality. This result is simply obtained for diploids by noting that the mutational rate of origin of new heterozygosity at a homozygous site is $2u$, whereas the fractional loss of existing heterozygosity by drift is $1/(2N_e)$, with the equilibrium being given by the ratio of these rates.

Fitness: the relative reproductive capacity of an individual (also accounting for viability to maturity), scaled to fall between values of 0 and 1.

G: genome size in number of nucleotides per haploid genome.

Mutator allele: denotes a mutant copy of any gene that confers an increase in the mutation rate.

N_e: genetic effective population size which measures the size of the population with respect to the stochastic behavior of allele frequencies relative to the situation for an ideal random-mating population. Technical details are reviewed by Lynch [23] and Charlesworth [56]. A central point is that N_e is generally much smaller than the actual size of a population, as a consequence of variation in family size, non-random mating, sex-ratio bias, and many other aspects of population structure. At small population sizes, N_e might increase in parallel with the actual population size, but at very large population sizes N_e becomes largely limited by the structure of chromosomes, because simultaneously segregating mutations at linked loci interfere with each other, thereby reducing the efficiency of selection relative to the situation for independent loci. The latter factor might result in N_e eventually asymptoting at an upper limit, regardless of the actual population size.

Random genetic drift: stochastic fluctuations in allele frequencies resulting from the sampling of a finite number of gametes in the establishment of each generation. For haploid and diploid species, the variance in allele frequency resulting from drift is proportional to $1/N_e$ and $1/(2N_e)$, respectively.

s_d: proportional reduction in fitness resulting from a heterozygous mutation.

s_r: proportional reduction in fitness resulting from the cost of replication associated with an antimutator allele.

Selection–mutation balance: refers to the equilibrium frequencies of deleterious alleles that are reached when the input of new variants by mutation is balanced by the removal by natural selection.

Silent site: a position in a protein-coding gene where nucleotide substitution has no influence on the protein sequence, owing to the redundancy in the genetic code.

Somatic mutation: a mutation arising in a non-germline cell in a multicellular species, which can influence individual fitness, but is not inherited by offspring.

u: mutation rate per nucleotide site per generation.

U or uG: genome-wide mutation rate per generation.

ΔU: the positive or negative change in the genome-wide mutation rate caused, respectively, by a mutator or antimutator allele.

Corresponding author: Lynch, M. (milynch@indiana.edu)
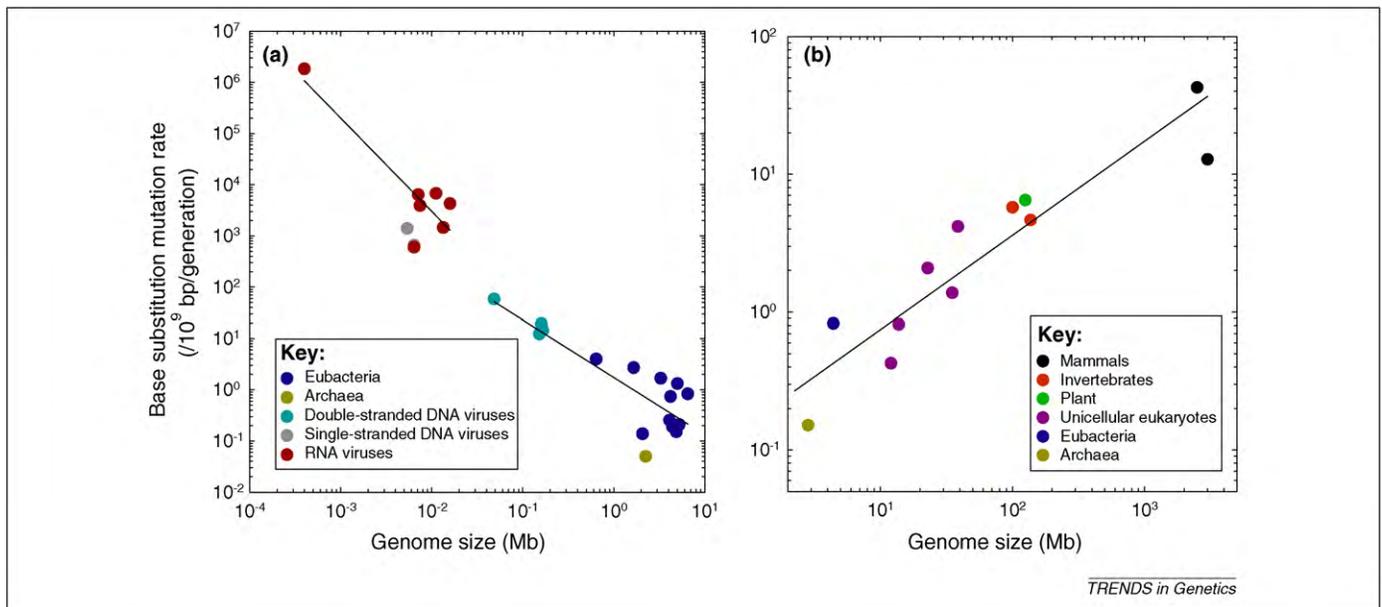
Figure 1. Scaling of base substitution rate per nucleotide site per generation with genome size. Each data point represents the average estimate for a separate taxon, although the results for most microbes are based on just one or a few reporter constructs (and hence have high sampling error), whereas those for most multicellular taxa are based on very large data sets (in several cases, whole genome sequences). **(a)** For non-eukaryotes, two separate regressions are provided, one for RNA viruses alone, and one for the pooled data from dsDNA viruses, eubacteria, and archaea. The respective regressions of the $\log_{10}$ plotted mutation rates are $-0.17 - 1.83\log_{10}(G)$ and $0.24 - 1.12\log_{10}(G)$, with G denoting the genome size in megabases, and $r^2 = 0.78$ and 0.72, respectively. **(b)** The regression for cellular organisms is $-0.81 + 0.68\log_{10}(G)$, with $r^2 = 0.80$. Here the results for various eubacteria (excluding *Buchnera*, which has an unusually small genome) are averaged into a single point. The pattern is quite similar if prokaryotes are excluded (slope $= 0.59$ and $r^2 = 0.83$).

taxa, four of which were bacteriophages, there was room for skepticism over the initial findings, but additional mutation rate assays performed in recent years have allowed for a substantial extension of this previous analysis. Although most microbial mutation rate estimates still rely on single reporter constructs, the approaches advocated by Drake [2] can be used to translate per locus rates to a per nucleotide site scale (see the supplementary material online). The focus here will be on base substitution mutations alone because considerably less work has been done on insertions and deletions.

For double-stranded DNA (dsDNA) viruses and prokaryotes, strong support for Drake's conjecture remains (Figure 1a), with the mutation rate/site/generation scaling with the -1.1 power of total genome size, although an obvious remaining concern is that the pattern is largely dependent on the inclusion of bacteriophage genomes. Mutation rates for RNA viruses are greater than those for dsDNA genomes of comparable size, but the negative scaling is retained (although this is entirely dependent on a single data point). For the prokaryotic genomes for which data are available, the sampling error of the mutation rate is so large and the range in variation for genome size is so small that only a weak scaling of u with G can be discerned. To more clearly resolve this matter, whole genome mutation accumulation assays, such as those now available for several eukaryotes [3–6], are desirable for a range of prokaryotes, particularly those with extreme genome sizes.

In striking contrast to the preceding pattern, when attention is confined to cellular species, mutation rates scale positively with genome size, with vertebrates having nearly $100 \times$ higher per generation rates than prokaryotes, and with the rates for unicellular eukaryotes,

invertebrates, and land plants being intermediate (Figure 1b). Most of the eukaryotic estimates are based on surveys of substantial genomic regions (including complete genome sequences in four cases). This greatly reduces the sampling variance associated with locus-specific peculiarities, and a statistical scaling of u with the $\sim2/3$rds power of genome size has very strong statistical support (Figure 1b). Note that on a per generation basis, the average mammalian mutation rate is nearly equal to that per replication in dsDNA viruses.

Random genetic drift as the lower limit to DNA repair fidelity

Drake suggested that the constant rate of total genomic mutation in microbes 'is likely to be determined by deep general forces, perhaps by a balance between the usually deleterious effects of mutation and the physiological costs of further reducing mutation rates' [2]. The implicit assumption here is that organisms under strong selection for rapid replication cannot maximize the fidelity of DNA replacement without limiting the rate of DNA synthesis necessary for daughter cell production. This general idea has been promoted broadly [7–14] and, although it has not been the subject of empirical investigation, it is known that microbial systems can be improved [15,16].

If the cost-of-repair hypothesis is correct, then we would infer a higher cost of replication in multicellular species (where mutation rates are high) than in prokaryotes. However, the time necessary for the replication of large eukaryotic genomes is compensated by the population of chromosomes with multiple origins of replication (in contrast to the single origin in most bacterial chromosomes). Moreover, as will be discussed below, the burden of somatic mutations imposes a downward selective pressure on

mutation rates in multicellular species which is not shared by unicellular species. Thus, an alternative explanation must be sought for the elevated rates of mutation in eukaryotes.

One possibility is that the lower bound on the mutation rate is not set by physiological or biochemical limitations, but by the intrinsic inability of selection to push the rate any lower. The power of random genetic drift ($1/2N_e$ for diploid organisms, where N_e is the genetic effective population size) ultimately constrains what natural selection can accomplish with any trait, and once the mutation rate is pushed to such a low level that any further incremental improvement conveys a fitness advantage smaller than the power of drift, selection will be incapable of reducing the rate any further. Therefore, a key to understanding mutation rate evolution is determining the degree to which evolved mutation rates approach the barriers imposed by drift.

By producing a correlated genetic load through the recurrent influx of deleterious mutations at linked and unlinked sites, even the weakest mutator allele suffers an indirect selective disadvantage associated with the excess mutational burden contained within the genomes of carrier individuals [7,8,10,17]. This disadvantage can be quite small, however, having a maximum value equal to twice the product of the average deleterious effect of a heterozygous mutation (s_d) and the diploid genome-wide increase in the deleterious mutation rate (ΔU, where U is in the range of 0.01 to 1.0 per generation for multicellular eukaryotes [14,18] and perhaps an order of magnitude lower in yeast [19,20]). The factor of two arises because most induced mutations arise on chromosomes unlinked to the mutator, retaining an association with the latter for an average of just two generations.

Two factors will reduce the selective advantage of an antimutator allele below $2 s_d \Delta U$, where ΔU is now the genome-wide reduction in the deleterious mutation rate.

First, the full long-term advantage of an antimutator is not realized until it has reached selection–mutation balance with respect to its reduced mutation load [21], making it more difficult for selection initially to promote such an allele towards fixation. Second, if there is any 'cost of replication' associated with the antimutator (s_r), the maximum selective advantage becomes $2 s_d \Delta U - s_r$. Thus, because allelic variants with selection coefficients much smaller than the power of random genetic drift evolve in an effectively neutral manner [22], an antimutator allele will be insensitive to selection unless the change in the genome-wide deleterious mutation rate is considerably greater than $[1/(2N_e) + s_r]/(2 s_d)$. Assuming that s_d and s_r are independent of N_e, this suggests that the mutation rate should scale negatively with N_e up to the point where U is so low that further incremental reductions cannot overcome the drift barrier.

Are eukaryotic mutation rates driven to such low levels? Although a definitive answer cannot yet be given, it is known that N_e is typically in the range of 10^5 to 10^6 for the nuclear genomes of multicellular species [23], and that the average value of s_d generally ranges from 10^{-3} to 10^{-2} [18]. This implies that a selectable antimutator must reduce the deleterious genome-wide mutation rate in a multicellular lineage by an amount much greater than 10^{-4} to 10^{-2}. Because these values are ~1% of the genome-wide deleterious mutation rates known for multicellular species, it follows that an antimutator allele would have to reduce U by much more than 1%, perhaps an order of magnitude more, to be promoted by selection. Although not impossible, given that DNA replication and repair are functions of dozens of loci, mutations at such loci that alter single amino acids might have such large effects only rarely. Thus, the drift hypothesis appears to be quantitatively plausible.

The drift hypothesis derives further support from the distribution of average N_e among phylogenetic lineages
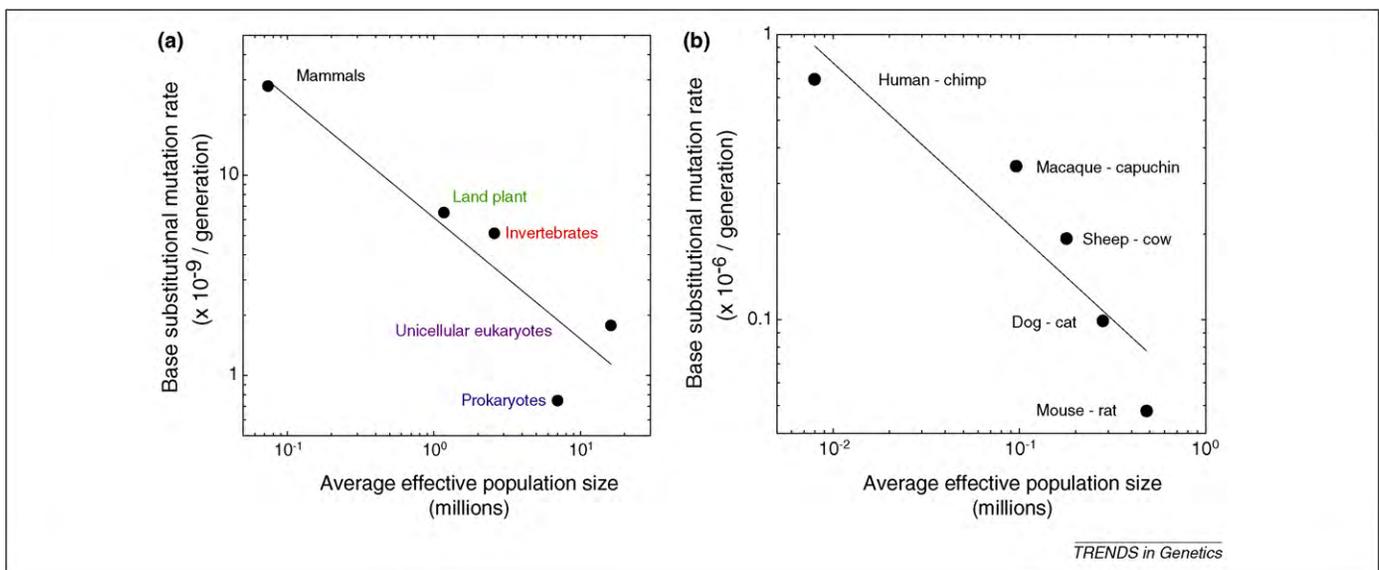
Figure 2. Scaling of the base substitution mutation rate per generation (u) and the effective number of genes per locus ($2N_e$ for diploids, and N_e for haploids). **(a)** The slope of the log–log regression for the nuclear genome of major phylogenetic groupings is −0.60 (0.16), where the number in parentheses denotes the standard error, with $r^2 = 0.84$, although if the estimated N_e for prokaryotes is assumed to be 10 times too low [23], the slope is modified to −0.52 (0.02) with $r^2 = 0.99$. **(b)** The slope of the log–log regression for the mitochondrial genome of mammalian lineages is −0.60 (0.15), with $r^2 = 0.84$. Data are average estimates from analyses assuming fixed and variable substitution rates in Piganeau and Eyre-Walker [25].

[23,24]. Under the assumption that nucleotide diversity at silent sites in natural populations is effectively neutral (due to the lack of impact at the amino acid level), the equilibrium level of heterozygosity (π_s) at such sites is $\sim 4N_e u$ in diploid species (and $2N_e u$ in haploids), where u is the mutation rate per site. Using previously summarized data on π_s from major phylogenetic groupings [24], and factoring out the average mutation rates provided in Figure 1, the average N_e in these groups can be approximated. One then finds a significant negative correlation between u and N_e in accordance with the drift hypothesis (Figure 2a).

A similar pattern is found for mammalian mitochondrial genomes using data from Piganeau and Eyre-Walker [25]. Here, N_e is the effective number of females because the mammalian mitochondrion is maternally inherited. The mutation rate was inferred indirectly from phylogenetic estimates of divergence at silent sites (assumed to be neutral), estimated times of divergence from the fossil record, and estimated mean generation times. Despite the greater degree of uncertainty in these data, the log–log regression of lineage-specific estimates of u on N_e has a slope identical to that for the nuclear data described above (Figure 2b).

Because the indirect estimates of N_e in both of these analyses are associated with a considerable (but unknown) degree of sampling error, the true scaling of u and N_e might be more extreme than the observed -0.6 power. Nevertheless, that two analyses based on different phylogenetic groups, types of data analysis, and genomic compartments yield essentially the same result provides strong support for the hypothesis that declines in N_e compromise the ability of selection to maintain high-fidelity replication and/or repair mechanisms. Still further support derives from a body of studies suggesting that several aspects of replication fidelity in eukaryotes are compromised relative to the situation in prokaryotes [17], although some aspects of DNA repair seem to be enhanced in mammals relative to microbes [26].

These observations help to explain a long-standing conundrum in evolutionary genetics – the near independence of nuclear molecular heterozygosity levels across phylogenetic groups with presumably large disparities in N_e. Lewontin [27] dubbed this pattern 'the paradox of variation,' although Nei [28] later pointed out a weak positive correlation between levels of variation and N_e. We now see that the relative phylogenetic stability of π_s across broad domains of life is not a reflection of relatively constant N_e, but of an inverse relationship between u and N_e. This inverse relationship appears also to be responsible for the relative invariance of π_s in the mitochondrial genomes of diverse animals [29–31]

The preceding arguments also provide a plausible explanation for the opposite scaling pattern of the mutation rate

Table 1. Mutation rates per nucleotide site ($\times 10^{-9}$) in different tissues[a]

Species	Tissue	Cell divisions per generation[a]	Mutation rates[b]	
			Per generation	Per cell division
Homo sapiens	Germline	216	12.85	0.06
	Retina	55	54.45	0.99
	Intestinal epithelium	600	162.00	0.27
	Fibroblast (culture)			1.34
	Lymphocytes (culture)			1.47
Mus musculus	Male germline	39	38.00	0.97
	Brain		76.94	
	Colon		83.35	
	Epidermis		90.38	
	Intestine		117.69	
	Liver		237.88	
	Lung		166.83	
	Spleen		130.00	
Rattus norvegicus	Colon		178.38	
	Kidney		167.45	
	Liver		179.92	
	Lung		223.22	
	Mammary gland		57.70	
	Prostate		448.90	
	Spleen		101.62	
Drosophila melanogaster	Germline	36	4.65	0.13
	Whole body		380.92	
Caenorhabditis elegans	Germline	9	5.60	0.62
Arabidopsis thaliana	Germline	40	6.50	0.16
Saccharomyces cerevisiae		1	0.33	0.33
Escherichia coli		1	0.26	0.26

[a]References to data on numbers of germline cell divisions: human [Crow 2000]; *D. melanogaster* and mouse [57]; *C. elegans* [58]; and *A. thaliana* [59]. Numbers of cell divisions are unknown for the mouse and rat rates.

[b]Mammalian tissue-specific rates are given only for tissues in which at least two independent estimates have been acquired. All data on human mutation rates are taken from Lynch [36]. Data for somatic mutation rates in mouse and rat are derived from references contained within the supplementary material online. References to data on germline mutation rates are: *D. melanogaster* [5], *C. elegans* [4], *A. thaliana* [Ossowski *et al.,* 2009], *S. cerevisiae* [3], and *E. coli* [24].

with genome size in viruses and prokaryotes. The case has been made that an upper-bound to N_e, in the neighborhood of 10^9 to 10^{11}, might exist in cellular species, dictated by the physical (linked) nature of the genome [23]. Assuming this upper bound is approximated in non-eukaryotic microbes, and the genome-wide deleterious mutation rate is driven to the lower limit compatible with the associated magnitude of drift, then because selection operates on the genome-wide deleterious mutation rate, any reduction in genome size would increase the lower limit of the achievable per-site mutation rate by reducing the number of mutational targets, yielding the inverse scaling suggested by Drake. Such a response is quite notable in the endosymbiotic bacterium *Buchnera aphidicola* [32], which has a highly reduced genome size and the highest known mutation rate for a prokaryote (left-most eubacterial data point in Figure 1).

It also follows that if the average effect of a deleterious mutation (s_d) were to increase, the lower limit to the achievable mutation rate would decrease. Drawing from observations that mutations that are benign at low temperatures often have elevated deleterious effects at high temperatures, Drake has argued that an elevation in s_d has promoted the evolution of reduced base-substitution mutation rates in thermophilic bacteria [33].

Finally, it should be noted that, despite the similar scaling of the per-site mutation rate with N_e in both nuclear and mitochondrial genomes, the absolute values of u are much greater for mitochondria (Figure 2). Such a pattern is also in agreement with the expectations of the drift hypothesis, because the number of mutational targets in the animal mitochondrion (e.g. just 13 protein-coding genes) is far below the number in nuclear genomes. Thus, although it is often argued that elevated mitochondrial mutation rates in metazoans are an inevitable consequence of a highly oxidative mitochondrial environment, the drift hypothesis provides an explanation based purely on the efficiency of selection. Nonetheless, a remaining puzzle with respect to organelle mutation rates concerns the apparent ~ten-fold reduction in land plant organelles relative to nuclear rates [23]. Plant organelle genomes can be up to ten-fold larger than animal mitochondrial genomes, but they are still vastly smaller than nuclear genomes, and the effective population sizes of such organelles do not appear to be unusually large [23]. Thus, to be consistent with the drift hypothesis, the average deleterious effects of organelle mutations in land plants must be unusually large, some aspects of the repair machinery must be driven by nuclear functions, and/or there must be mechanisms for reducing plant organelle mutation rates in much smaller increments than in nuclear environments.

Somatic mutation

Along with the burden of deleterious germline mutations, multicellular species experience transient somatic mutations, and these can influence the reproductive output of parental genomes via the development of cancer, senescence, and a large number of other disorders. Although almost no theory exists on the consequences of somatic mutations for the evolution of the mutation rate, because the same basic repair pathway machinery appears to be deployed in all cells, there must be a direct connection between selection to reduce the somatic mutation rate and the evolution of the germline mutation rate, and vice versa [17].

To evaluate the evolutionary consequences of somatic mutations, it is first instructive to put things on an equal footing by standardizing the germline mutation rates of multicellular lineages to a per cell division basis. Such a comparison shows that although selection has been incapable of maintaining per generation germline mutation rates for base substitutions at the levels observed in microbes, the per cell division rates have been kept low, and have perhaps even been suppressed in human (Table 1). However, this degree of conservation seems not to apply to all forms of mutation because germline mutations at microsatellite loci arise five times more frequently per cell division in *Caenorhabditis elegans* than in yeast or slime mold, with mammalian and land plant rates being ~14× those in *C. elegans* [34,35].

Although the maintenance of somatic integrity is crucial to germline transmission, metazoan somatic mutation rates are consistently greater than germline rates. In humans the average mutation rate for four somatic cell types, 1.02×10^{-9}/site/cell division (SE = 0.27×10^{-9}), is $17 \times$ higher than the germline rate and $3.5 \times$ higher than the average for yeast and *Escherichia coli* [36]. Assays of a wide range of tissue types in mouse and rat lines engineered to carry reporter constructs show that somatic cells accumulate two- to six-fold more mutations than do cells in the testes at the age of maturity, and considerably more later in life (Table 1). On an absolute time scale, somatic mutation rates are also higher than germline rates in the medaka fish [37], and in *Drosophila melanogaster* per generation somatic rates average ~80 × those in the germline [38,39]. Thus, without the advantages of germline protection, the precise nature of which remains to be determined, the heritable per generation mutation rates for animal species would be several-fold higher.

The enormity of the somatic mutation problem can be roughly estimated in humans, where the per generation rate of mutation for intestinal epithelium is ~13 × that in the germline, and by extrapolation, that in fibroblasts and lymphocytes is likely to be ~5 × higher again (Table 1). Thus, with a human germline mutation rate of ~10^{-8} base substitutions/site/generation, a site in a somatic nucleus will be mutated with a probability of 10^{-7} to 10^{-6} by the average age of reproduction, with the burden being higher in older individuals. With a diploid genome size of 6×10^9 sites and ~10^{13} cells per soma, the body of a middle-aged human might then contain >10^{16} mutations (not including insertions, deletions, or other larger scale mutations). Only about 1% of the human genome consists of coding DNA, so a substantial fraction of somatic mutations will be inconsequential, but even if just 1% of coding mutations had significant fitness effects, the total body burden of mutations would be in the order of 10^{12}. Diploidy might mask the effects of many deleterious mutations, but most mutations with small effects act in a nearly additive fashion [18] and, although processes such as apoptosis might remove some cells with major mutational defects, it is unlikely that cells with incremental levels of incapa-
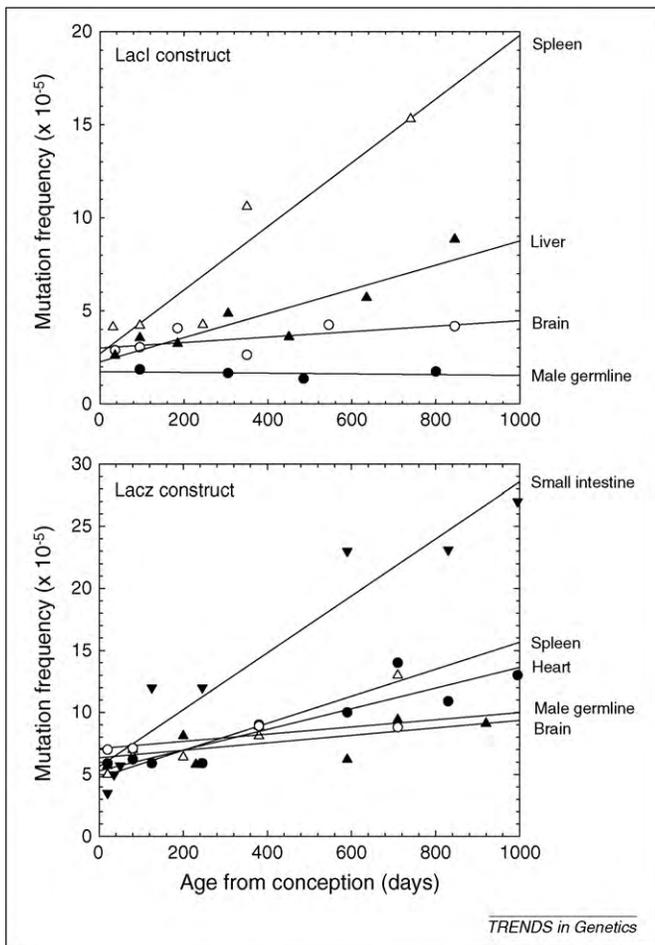
Figure 3. Tissue-specific frequencies of mutations as a function of age in mouse lines carrying Lac reporter constructs. Results are averaged over multiple studies (see the supplementary material online).

citation could be selectively eliminated. The net result is a progressive lifetime accumulation of somatic mutations, as clearly revealed in the mouse where the germline DNA remains relatively stable within an environment of degrading somatic cell genomes (Figure 3).

Without details on the absolute fitness effects of somatic mutations, only qualitative statements can be made on their consequences for the evolution of the germline mutation rate [17]. One central question is the degree to which the efficiency of selection operating on the mutation rate via the consequences of somatic mutations changes with the level of multicellularity. At low levels of organismal complexity, the reduction in individual fitness associated with somatic mutations can be described roughly as the product of four factors, $2u_sTns_s$, where $2u_s$ is the diploid somatic mutation rate per nucleotide site per generation, T is the number of sites influencing fitness, n is the number of cells influencing fitness, and s_s is the reduction in fitness per somatic mutation (a more explicit form of this expression would not treat all cells equally, but sum over independent tissues [17]).

Although T and n must increase with increasing levels of multicellularity, s_s might decrease, depending on aspects of cellular surveillance and the buffering effects of multicellularity on individual mutant cells. By contrast, as noted in Figure 2, increased multicellularity is generally

associated with a reduction in N_e, and this in turn reduces the efficiency of selection. Thus, a key to understanding the degree to which the burden of somatic mutations impacts selection on the mutation rate itself is analogous to the situation noted above for germline mutations. If the increase in $1/(2N_e)$ with increasing multicellularity exceeds the increase in $2u_sTns_s$, the ability of selection to reduce the somatic mutation rate (and probably the correlated effect on the germline rate) will become progressively compromised.

Moreover, a scenario can be envisaged whereby a critical level of multicellularity is eventually reached, beyond which the ability of selection to reduce the somatic mutation rate begins to decline [17]. Such behavior is expected because the strength of selection depends on relative rather than absolute fitness effects. Although the absolute negative consequences of somatic mutations might continue to increase indefinitely with increasing multicellularity, once a level has been reached at which the fraction of affected individuals approaches saturation, the relative selective disadvantage of a further increase in the mutation rate must begin to decline. It is unclear where organisms with various levels of multicellularity reside on this continuum. However, it is clear that if the somatic mutation load plays a role in the evolution of the germline mutation rate, it has generally been incapable of keeping the somatic rate at levels observed in unicellular species.

Transcription and translation fidelity

Although somatic nuclear mutations permanently influence a host cell and all of its descendants, two more transient forms of mutations are also of relevance – errors in transcription and translation. The average error rate estimate for the RNA polymerase of *E. coli* is 1×10^{-5} per base incorporated [40–42], whereas that for *Saccharomyces cerevisiae* RNA polymerase II (Pol II; involved in transcription of coding mRNAs) is in the range of 2×10^{-6} to 3×10^{-4} [43,44], and the single estimate for Pol II of a multicellular species is 1×10^{-3} in wheat [45]. These rough estimates are based on a variety of methodologies and have a restricted phylogenetic range. Nevertheless, they indicate that transcription error rates per nucleotide transaction are orders of magnitude higher than the replication error rates noted above. They also suggest that transcriptional fidelity is reduced in eukaryotes, perhaps substantially so in multicellular lineages.

A similar pattern is observed for translation, with the overall level of fidelity appearing to be even lower than for transcription. Although there can be considerable variation among codon types, the average translation error rate per codon in *E. coli* is 6×10^{-4} [46–48], whereas the average rates for yeast, rabbit reticulocytes, and mouse liver cells are, respectively, 2×10^{-3} [49,50], 3×10^{-4} [51], and 1×10^{-3} [52]. Taken together, with an average protein length of ~300 amino acid residues [23], these observations suggest that, without removal by post-translational surveillance, >20% of individual proteins will contain at least one inappropriate amino acid [53].

Like DNA polymerases, RNA polymerases have a proofreading capacity [54], and there is no obvious reason why they (or the translational machinery) should be intrinsically

constrained from operating at the level of efficiency of DNA polymerases. However, because individual loci generally produce multiple transcripts, and mRNAs and individual proteins have transient residence times within cells, transcriptional and translational errors are expected to have less severe effects on cell integrity than genome-level errors. Thus, the strength of selection operating on the transcriptional and translational machinery is likely to be less stringent and, consistent with the drift hypothesis, this might explain the greatly elevated error rates for these processes.

Concluding remarks

Germline mutation rate data provide a crucial basis for interpreting patterns of molecular diversity within species and divergence among species. Indeed, inferences regarding selection have been historically derived by assuming certain classes of sites (e.g. synonymous coding-region positions) to be effectively neutral and hence to evolve at the mutation rate, thereby providing clear predictions of evolutionary patterns expected in the absence of selection [22]. However, the direct estimates of mutation rates and molecular spectra obtained in the studies reviewed herein are often substantially different from those derived by indirect inference from natural populations, at least in part because selection is more pervasive than formerly believed (e.g. Ref. [55]). Thus, to be fully reliable, future molecular investigations with a goal of interpreting evolutionary mechanisms should take advantage of direct estimates of mutation rates. One might argue that laboratory estimates are subject to their own peculiar biases, but the consistent patterns noted above suggest that we are close to developing a general understanding of the rates at which base substitutions arise in various phylogenetic lineages.

Although there are strong scaling patterns between the mutation rate per generation and genome size, this pattern is not a function of direct causality – in other words large eukaryotic genomes do not intrinsically engender low fidelity of DNA replication. Instead, because there is a general insertional bias in most eukaryotic genomes, as effective population sizes decline and the efficiency of selection against excess DNA is relaxed, genome size increases in a passive fashion [23], along with the mutation rate. By contrast, effective population sizes in viruses and prokaryotes might often be so close to their maxima that the lower limit to the evolvable genome-wide mutation rate has been reached. Once this point has been reached, any events that lead to a further reduction in genome size (e.g. loss of non-essential genes in endosymbionts and parasites) will generally increase the minimum evolvable per site mutation rate, because the product of the latter and the genome-wide number of selected sites is equal to the genome-wide deleterious rate. However, why the total genomic mutation rate for microbes converges on ~0.003 per cell division [2] is a mystery that remains to be solved.

An additional unsolved problem concerns the long-term stability of the mutation rate in low-N_e lineages. Just as random genetic drift can inhibit the fixation of an antimutator with an insufficiently large effect on the genomic mutation rate, it can encourage the fixation of sufficiently mild mutator alleles. What then prevents the gradual accumulation of very mildly deleterious mutations at DNA repair loci and a slow but progressive increase in the mutation rate in multicellular lineages? Given the preceding arguments, it might be premature to assume that the mutation rate in such lineages has actually attained equilibrium.

References

1 Drake, J.W. (2006) Chaos and order in spontaneous mutation. *Genetics* 173, 1–8

2 Drake, J.W. (1991) A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl. Acad. Sci. U. S. A.* 88, 7160–7164

3 Lynch, M. *et al.* (2008) A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 105, 9272–9277

4 Denver, D.R. *et al.* (2009) A genome-wide view of *Caenorhabditis elegans* base substitution mutation processes. *Proc. Natl. Acad. Sci. U. S. A.* 106, 16310–16314

5 Keightley, P.D. *et al.* (2009) Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. *Genome Res.* 19, 1195–1201

6 Ossowski, S. *et al.* (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* 327, 92–94

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

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

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

10 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

11 Drake, J.W. *et al.* (1998) Rates of spontaneous mutation. *Genetics* 148, 1667–1686

12 Sniegowski, P.D. *et al.* (2000) The evolution of mutation rates: separating causes from consequences. *Bioessays* 22, 1057–1066

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

14 Baer, C.F. *et al.* (2007) Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nat. Rev. Genet.* 8, 619–631

15 Quiñones, A. and Piechocki, R. (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

16 Loh, E. *et al.* (2010) Optimization of DNA polymerase mutation rates during bacterial evolution. *Proc. Natl. Acad. Sci. U. S. A.* 107, 1154–1159

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

18 Lynch, M. and Walsh, J.B. (1998) *Genetics and Analysis of Quantitative Traits*, Sinauer Associates Inc.

19 Wloch, D.M. *et al.* (2001) Direct estimate of the mutation rate and the distribution of fitness effects in the yeast *Saccharomyces cerevisiae*. *Genetics* 159, 441–452

20 Joseph, S.B. and Hall, D.W. (2004) Spontaneous mutations in diploid *Saccharomyces cerevisiae*: more beneficial than expected. *Genetics* 168, 1817–1825

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

22 Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*, Cambridge University Press

23 Lynch, M. (2007) *The Origins of Genome Architecture*, Sinauer Associates Inc.

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

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

26 Saparbaev, M. *et al.* (2000) Interactions of the human, rat, *Saccharomyces cerevisiae* and *Escherichia coli* 3-methyladenine-DNA glycosylases with DNA containing dIMP residues. *Nucleic Acids Res.* 28, 1332–1339

27 Lewontin, R.C. (1974) *The Genetic Basis of Evolutionary Change*, Columbia University Press

28 Nei, M. (1983) Genetic polymorphism and the role of mutation in evolution. In *Evolution of Genes and Proteins* (Nei, M. and Koehn, R.K., eds), pp. 165–190, Sinauer Associates Inc

29 Bazin, E. *et al.* (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science* 312, 570–572

30 Nabholz, B. *et al.* (2008) Determination of mitochondrial genetic diversity in mammals. *Genetics* 178, 351–361

31 Nabholz, B. *et al.* (2009) The erratic mitochondrial clock, variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. *BMC Evol. Biol.* 10, 54

32 Moran, N.A. *et al.* (2009) The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science* 323, 379–382

33 Drake, J.W. (2009) Avoiding dangerous missense: thermophiles display especially low mutation rates. *PLoS Genet.* 5, e1000520

34 Seyfert, A.L. *et al.* (2008) The rate and spectrum of microsatellite mutation in *Caenorhabditis elegans* and *Daphnia pulex*. *Genetics* 178, 2113–2121

35 Marriage, T.N. *et al.* (2009) Direct estimation of the mutation rate at dinucleotide microsatellite loci in *Arabidopsis thaliana* (Brassicaceae). *Heredity* 103, 310–317

36 Lynch, M. (2010) The rate, molecular spectrum, and consequences of human mutation. *Proc. Natl. Acad. Sci. U. S. A.* 107, 961–968

37 Winn, R.N. *et al.* (2000) Detection of mutations in transgenic fish carrying a bacteriophage lambda cII transgene target. *Proc. Natl. Acad. Sci. U. S. A.* 97, 12655–12660

38 Garcia, A.M. *et al.* (2007) A model system for analyzing somatic mutations in *Drosophila melanogaster*. *Nat. Methods* 4, 401–403

39 Edman, U. *et al.* (2009) Lifespan extension by dietary restriction is not linked to protection against somatic DNA damage in *Drosophila melanogaster*. *Aging Cell* 8, 331–338

40 Blank, A. *et al.* (1986) An RNA polymerase mutant with reduced accuracy of chain elongation. *Biochemistry* 25, 5920–5928

41 Ninio, J. (1991) Transient mutators: a semiquantitative analysis of the influence of translation and transcription errors on mutation rates. *Genetics* 129, 957–962

42 Goldsmith, M. and Tawfik, D.S. (2009) Potential role of phenotypic mutations in the evolution of protein expression and stability. *Proc. Natl. Acad. Sci. U. S. A.* 106, 6197–6202

43 Shaw, R.J. *et al.* (2002) Use of an *in vivo* reporter assay to test for transcriptional and translational fidelity in yeast. *J. Biol. Chem.* 277, 24420–24426

44 Kireeva, M.L. *et al.* (2008) Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation. *Mol. Cell* 30, 557–566

45 de Mercoyrol, L. *et al.* (1992) Accuracy of wheat-germ RNA polymerase II. General enzymatic properties and effect of template conformational transition from right-handed B-DNA to left-handed Z-DNA. *Eur. J. Biochem.* 206, 49–58

46 Ortego, B.C. *et al.* (2007) *In vivo* translational inaccuracy in *Escherichia coli*: missense reporting using extremely low activity mutants of *Vibrio harveyi* luciferase. *Biochemistry* 46, 13864–13873

47 Kramer, E.B. and Farabaugh, P.J. (2007) The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. *RNA* 13, 87–96

48 Willensdorfer, M. *et al.* (2007) Phenotypic mutation rates and the abundance of abnormal proteins in yeast. *PLoS Comput. Biol.* 3, e203

49 Stansfield, I. *et al.* (1998) Missense translation errors in *Saccharomyces cerevisiae*. *J. Mol. Biol.* 282, 13–24

50 Salas-Marco, J. and Bedwell, D.M. (2005) Discrimination between defects in elongation fidelity and termination efficiency provides mechanistic insights into translational readthrough. *J. Mol. Biol.* 348, 801–815

51 Loftfield, R.B. and Vanderjagt, D. (1972) The frequency of errors in protein biosynthesis. *Biochem. J.* 128, 1353–1356

52 Mori, N.Y. *et al.* (1985) Analysis of translational fidelity of ribosomes with protamine messenger RNA as a template. *Biochemistry* 24, 1231–1239

53 Drummond, D.A. and Wilke, C.O. (2009) The evolutionary consequences of erroneous protein synthesis. *Nat. Rev. Genet.* 10, 715–724

54 Sydow, J.F. and Cramer, P. (2009) RNA polymerase fidelity and transcriptional proofreading. *Curr. Opin. Struct. Biol.* 19, 732–739

55 Eöry, L. *et al.* (2009) Distributions of selectively constrained sites and deleterious mutation rates in the hominid and murid genomes. *Mol. Biol. Evol.* 27, 177–192

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

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

58 Wilkins, A.S. (1992) *Genetic Analysis of Animal Development* (2nd edn), Wiley–Liss

59 Hoffman, P.D. *et al.* (2004) Rapid accumulation of mutations during seed-to-seed propagation of mismatch-repair-defective *Arabidopsis*. *Genes Dev.* 18, 2676–2685