Neutral Mutations and Neutral Substitutions in Bacterial Genomes

Howard Ochman

Department of Biochemistry and Molecular Biophysics, University of Arizona

Molecular evolutionary biologists usually assess the underlying spectrum of mutations within a bacterial genome by examining substitutions that occur at sites believed to be under no selective constraints. Alternatively, bacterial mutation rates can also be estimated in a variety of experimental systems. The two classes of changes occurring in DNA sequences—i.e., mutations and neutral substitutions—are, in theory, identical; however, the rates and patterns of mutations in bacteria, as inferred from sequence comparisons, often differ significantly from those derived experimentally. These differences have resulted in conflicting interpretations of the nonselective forces that affect mutation rates.

Introduction

Analysis of sequences that are under no selective constraints should expose the actual frequency with which mutations occur. The influx of new mutations is the product of the mutation rate and population size, and the fixation of neutral mutations is determined by the inverse of population size; hence, substitution rates are tantamount to mutation rates at nucleotide sites that are under no selective constraints.

Among the most common applications of this comparative approach for estimating mutation rates has been the analysis of homologous noncoding regions, in particular, pseudogenes, which, unlike many nonfunctional regions, are relatively easy to align and should evolve according to the underlying rate and pattern of mutations (Gojobori, Ishii, and Nei 1982). This approach has been particularly informative in higher eukaryotes, which contain large tracts of nonfunctional DNA and are not amenable to an experimental assessment of their mutation rates.

Microbial evolutionists are not so lucky: the high gene density and paucity of pseudogenes in most bacterial genomes have imposed a reliance on degenerate codon positions in the hope that rates and patterns of evolution at synonymous sites will expose the actual mutational process. In contrast, among the advantages of working with bacteria is that their small genomes are experimentally tractable; in addition, they have always outpaced higher eukaryotes in terms of the amount of comparative sequence information available. By the mid-1980s, the sequences for more than 20 homologous genes were already available for two related enteric species, *Escherichia coli* and *Salmonella enterica* (then called *Salmonella typhimurium*), and this allowed inspection of the substitutional patterns in a diverse set of genes from around the genome.

The Comparative Approach to Studying Mutation Rates

Examining Substitutions at Synonymous Sites

Unlike the situation observed in comparisons of human and mouse homologs, the synonymous substitution rates (Ks) of *E. coli*–*S. enterica* homologs spanned nearly two orders of magnitude, which would not have been anticipated if these sites were subject to a more-or-less random process of mutation. One reason for this deviation from neutral expectations was elucidated in a now-classic paper by Sharp and Li (1987). They demonstrated that the differences in synonymous substitution rates among *E. coli*–Salmonella homologs were due to adaptive codon bias, such that highly transcribed genes utilize a very limited pool of codons, which promotes translational efficiency. Genes that are more highly expressed have stronger codon biases, resulting in lower divergence at synonymous sites due to selection acting at the level of the codon (fig. 1*a*).

Therefore, to estimate the underlying rates and patterns of mutations by examining variation at synonymous sites, the effects of adaptive codon bias need to be removed. This has been done in two ways: one can consider all available genes and numerically factor out the effects of codon bias or, alternatively, one can eliminate all high biased genes from the analysis. As discussed below, it has been possible to draw numerous conclusions about the mutational process in enteric bacteria by examining or comparing nucleotide sequences.

Chromosome Location and Substitution Rates

Aside from exposing the affects of codon usage bias, comparisons of homologous genes from *E. coli* and *S. enterica* also revealed that substitution rates increase with distance from the replication origin (fig. 1*b*). After differences in codon bias were accounted for, genes nearer to the replication origin in both species were estimated to have a substitution rate about half that of genes closer to the terminus (Sharp et al. 1989). The original recognition of this ''distance effect'' was based on relatively few ($n = 67$) pairs of homologous genes, and the significant association was due largely to the limited set of low-divergence genes near the replication origin (open circles, figure 1). However, reexamination of the distance effect using the entire complement of homologous genes whose chromosome positions are conserved in *E. coli* and *S. enterica* corroborated the initial findings, and it also revealed a similar relationship between gene location and sequence divergence in many, but not all, sequenced bacterial genomes (Mira and Ochman 2002).

The distance effect was originally thought to be the outcome of more frequent recombinational repair or biased gene conversion arising from the higher gene

Key words: mutation rates, nucleotide substitutions, bacterial evolution, *E. coli*, *S. enterica*.

E-mail: hochman@email.arizona.edu.

Mol. Biol. Evol. 20(12):2091–2096. 2003
DOI: 10.1093/molbev/msg229
Molecular Biology and Evolution, Vol. 20, No. 12,

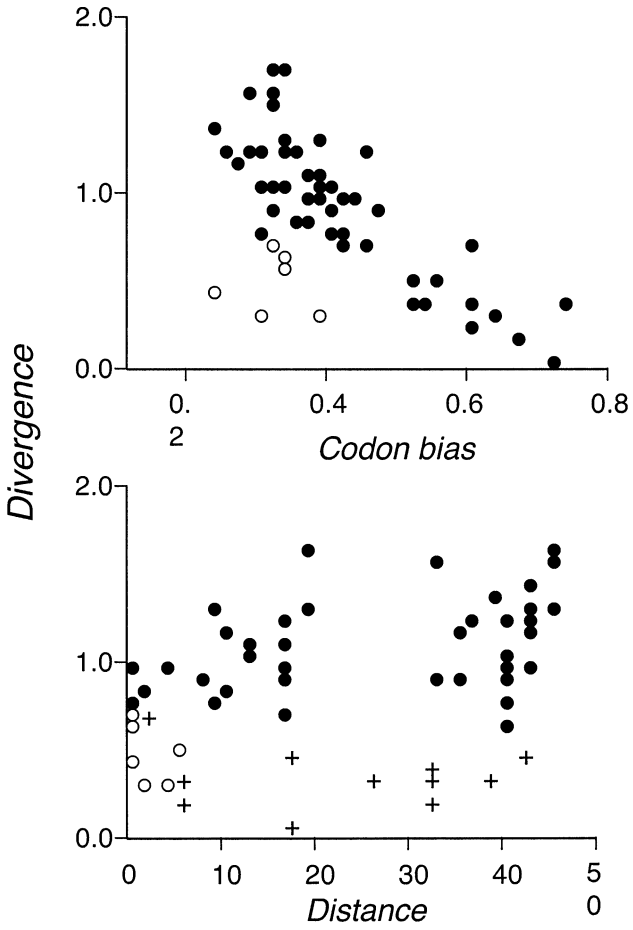

FIG. 1.—*A*. Relationship between divergence at synonymous sites and synonymous codon usage bias in comparisons of *E. coli–S. enterica* homologs. *B*. Relationship between divergence at synonymous sites and chromsome position, given as distance in minutes, from the origin of replication. Open circles represent low divergence genes with low CAI values. The affect of chromosome location on substitution rates is not apparent in highly biased genes (crosses) [from Sharp et al. 1989].

dosage near the origin, as achieved by the presence of multiple replication forks (Sharp et al. 1989; Sharp 1991; Birky and Walsh 1992). Because new rounds of replication can begin before the prior round has been completed, genes nearer the origin will be in multiple copies, allowing correction with the nearby homolog. Although this hypothesis is appealing, examination of substitutional patterns does not support such a mechanism. In the comparison of *E. coli–S. enterica* homologs, the distance effect is caused primarily by an increased rate of certain transversions near the replication terminus (Mira and Ochman 2002; Daubin and Perriere 2003), and it is difficult to see how this could result from a nondiscriminating repair process, such as gene conversion or homologous exchange. Hence, the distance effect is more likely due to the increased occurrence (or lack of repair) of particular substitutions as replication forks approach the terminus.

Transition versus Transversions

If all mutations were equally likely, transversions would occur at twice the frequency of transitions; however, the original comparisons of mitochondrial DNA sequences in primate lineages revealed that transitions greatly outnumber transversions (Brown et al. 1982). A substitutional bias toward transitions has been observed in many organisms; and in reconstructing the substitutions in several housekeeping genes sequenced in multiple strains of *E. coli* and *S. enterica*, there is an approximately 2.5:1 ratio of transitions to transversions at fourfold degenerate sites (Francino et al. 1996). Such mutational biases should be evaluated at fourfold degenerate sites because synonymous changes at twofold degenerate sites almost exclusively involve transitions. As noted above, some fourfold degenerate sites are subject to selection for translational efficiency, but they can still serve to indicate biases in the neutral pattern of substitutions—and in enteric bacteria, that bias is toward transitions.

Base Composition

Because the third positions of codons are largely degenerate—70% of changes at third codon positions are synonymous—the base composition of these sites is thought to reveal the underlying mutational bias of the genome. G + C contents at the third positions of codons are extraordinarily variable among bacterial species, ranging from 10% to 90% (Muto and Osawa 1987), and observation of the base compositional differences among bacteria served as the basis for the original formulation of the neutral theory (Sueoka 1962, 1988). In *E. coli* and *S. enterica*, the overall base composition at third codon positions ranges from 56% to 59% G + C, suggesting a tendency toward converting G·C→ A·T base pairs over the reverse mutations.

Effects of Gene Expression on Substitution Rates

Both Berg and Martelius (1995) and Eyre-Walker and Bulmer (1995) took a theoretical approach to examining the relationship between substitution rates and levels of gene expression. After accounting for the effects of selection for translational efficiency on codon usage in *E. coli–S. enterica* homologs, both studies argued that mutation rates decrease, perhaps as much as threefold, with levels of gene expression, implying that transcription lowers the frequency of certain classes of mutations.

Absolute Rate of Evolution

The lack of a fossil record for bacteria has thwarted most attempts to calculate their actual rates of substitution over an evolutionary time scale. However, for a few bacterial species, divergence times have been dated somewhat reliably from ecological or geological evidence, thus offering a means for estimating absolute rates of sequence evolution. Based on calibrated rates of ribosomal RNA divergence, the split between *E. coli* and *S. enterica* was estimated to have occurred some 100 MYA, a date that roughly coincides with the appearance of the principal niche of *E. coli*, the mammalian intestine (Ochman and Wilson 1988). A similar date of divergence for *E. coli* and *S. enterica* was obtained when assuming that universally distributed proteins evolve at the same rate in enteric

bacteria as in mammals (Doolittle et al. 1996). Given an average divergence at synonymous sites (Ks) of 0.9 for *E. coli–S. enterica* homologs, application of this date yields a per-site substitution rate of 0.45% per Myr. Although the date assigned to the split between *E. coli* and *S. enterica* can be questioned, a similar synonymous substitution rate has been computed for other bacteria, particularly endosymbionts, whose divergence times have been obtained from the fossil-based dates of divergence of their insect hosts (Moran et al. 1993).

The Experimental Approach to Studying Mutation Rates

Direct Assessment of Mutation Rates

Because *E. coli* and *S. enterica* are experimentally tractable, there is no need to rely solely on nucleotide sequence comparisons for information about the rates and patterns of mutations. Among the most commonly exploited experimental system for obtaining the frequencies of specific mutations in *E. coli* is the set of mutant *lacZ* alleles developed by Cupples and Miller (1989). Each of these mutations interrupts the *lac* gene at the identical codon position, and because the reversions needed to restore enzyme activity can occur on either strand, it is possible to assay all six possible mutation types with this reagent. The *lac* reversion rates have been tested under a variety of experimental conditions (e.g., Hall 1991; Schaaper and Dunn 1991; MacKay, Han, and Samson 1994; Fijalkowska et al. 1998), and several factors, including genetic background, environmental agents, and growth conditions, can alter the frequencies of mutations. For example, rates of point mutation estimated for the *lacZ* reversions in the *S. enterica* chromosome are more than 10 times lower than those reported for the same alleles tested on an *E. coli* episome, and mutation rates are found to be higher in nutrient-rich media. Therefore, in the following discussion, we consider only those studies in which the mutant *lacZ* alleles are encoded chromosomally and assayed in nonmutator strains grown under standard laboratory conditions (aerobically in minimal media at 37°C).

An Experimental Test of Transitions and Transversion Frequencies

Based on mutation frequencies obtained for the six *lac* revertants, which together assay all possible point mutations, the mutational spectrum in *Salmonella enterica* is biased toward transitions (black bars, fig. 2). Cumulatively, the ratio of transitions to transversions is nearly 2:1, similar to that obtained from sequence comparisons. Note, however, that several individual transversions were more frequent than the least common transition, and that C·G→ G·C mutations, which are very rare in experimental assays, occur at almost the same frequency as other transversions in sequence comparisons.

An Experimental Test of the Distance Effect

The effect of chromosome location on rates of point mutations has been assayed for two *lacZ* alleles inserted at four positions in the *Salmonella enterica* chromosome

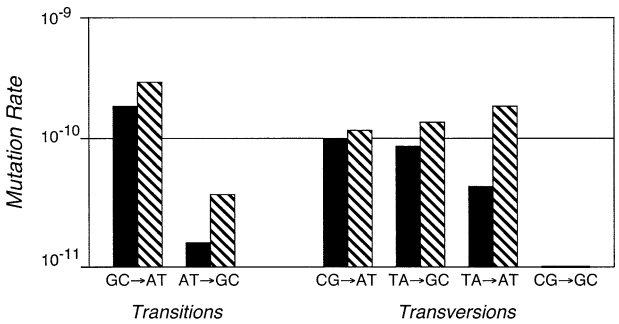

FIG. 2.—Mutation rates of *lacZ* alleles in the absence (black bars) and presence (hatched bars) of inducer (IPTG) [from Hudson, Bergthorsson, and Ochman 2003].

(Hudson et al. 2002). Despite differences in mutation rates at certain chromosomal locations, there is no systematic increase in mutation rates with distance from the replication origin, as observed in synonymous substitution frequencies. Mutation rates at an intermediate location were significantly higher than those at loci closest to and farthest from the replication origin, and this difference was observed for both A·T→ G·C transitions and T·A→ G·C transversions (fig. 3). Because these reversion rates are being assayed for mutations residing at identical positions within identical genes, but at different genomic locations, any confounding effects of either neighboring bases or local base composition on mutation rates are removed (Blake, Hess, and Nicholson-Tuell 1992).

An Experimental Test of the Transcription Effect

The inducible promoter of *lac* operon also makes it possible to directly compare the mutation rates in transcribed and nontranscribed genes. It has long been known that the spontaneous rate of G·C→ A·T mutations increase with the frequency of transcription in *E. coli*, an effect ascribed to three damage and repair processes (deamination, dimerization, and transcription coupled repair) that are specific to these nucleotides (Selby and Sancar 1993; Beletskii and Bhagwat 1996, 1998; Wright, Longacre, and Reimers 1999). When induced by the addition of the inducer IPTG, there is an increase in the rates of several of the other point mutations (fig. 2). In addition to G·C→ A·T mutations, induction significantly increases the individual rates of T·A→ A·T transversions, A·T→ G·C transitions, and the pooled rates of the other point mutations (Hudson, Bergthorsson, and Ochman 2003).

Mutational Biases Toward A + T

Based on the nucleotide contents of third codon positions, the *E. coli* and *S. enterica* genomes are thought to have a mutational bias toward G·C base pairs (Muto and Osawa 1987; Lawrence and Ochman 1998). However, in reversion assays, there is a cumulative bias toward the formation of A·T pairs, thus reducing the G + C content of the genome (fig. 2). The bias is attributable to transitions creating A·T pairs, which occur at approximately 10 times the frequency of the reverse mutation. There is no compositional bias among transversions because the rate

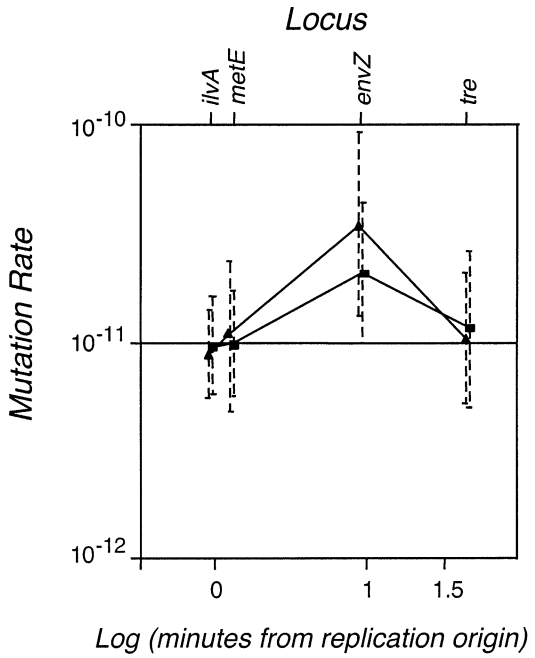

FIG. 3.—Mutation rates of *lacZ* alleles at different locations of the *Salmonella* chromosome, calculated for the T·A→ G·C transversion (▲) and the A·T→ G·C transition (■). Distance is given as the log of number of minutes from the origin of replication, and mutation rates are shown with 95% confidence intervals [from Hudson, Bergthorsson, and Ochman 2002].

of C·G→ A·T is approximately equal to that of the reverse mutation, and the other transversions do not alter the G + C contents. This trend among transitions toward A·T base pairs shown for *S. enterica* has also been observed in reversion assays of the same *lac* alleles in *E. coli* (Fijalkowska et al. 1998).

Overall Rates of Mutation

The rate of point mutations in *E. coli*–based and *S. enterica*–based *lacZ* reversions is $\sim 5 \times 10^{-10}$ per base pair per generation. Comparable studies on other loci have yielded rates in the same order of magnitude (reviewed in Drake et al. 1998). To reconcile the experimentally determined mutation rate with the neutral substitution rate derived from sequence comparisons ($\sim 5 \times 10^{-3}$ per site per year), *E. coli* and *S. enterica* would each need to average only 10 generations per year. It has been estimated that natural populations of *E. coli* have on the order of 100 to 1,000 generations per year, which implies a mutation rate over evolutionary time scales that is at least 10-fold lower than that observed in the lab (Ochman, Elwyn, and Moran 1999).

Why the Disparity?

Neutral substitutions should reveal the underlying mutational spectrum; however, neither the rate nor the pattern of neutral changes occurring in *E. coli* and *S. enterica* homologs matches the rate and pattern derived from experimental studies of mutation rates in these microorganisms. Of the five attributes analyzed by both the comparative and experimental approaches—(1) the distance effect, (2) the transcription effect, (3) the transition bias, (4) the overall mutational bias, (5) the absolute rate of mutation—only one (the preponderance of transitions over transversions) is consistent across the two types of studies. Either the experimental approach or the comparative approach, or very possibly both, is leading to incorrect perceptions about the mutational process.

There is no single factor common to either the experimental or the comparative approach that, when corrected, will reconcile the apparent differences. Certainly, the use of synonymous substitutions as markers of neutral evolution in the comparative approach is not optimal, because selection might act on these sites in ways that have not been considered. Alternatively, extrapolation from the experimental approach assumes that *E. coli* and *S. enterica* have, since they diverged, replicated unfalteringly in a warm, aerobic, nutrient-controlled environment, free of any (of the many known) factors that might alter the mutational spectrum. Fortunately, for each of the mutational features considered, there are reasonable ways to judge whether the experimental approach or the comparative approach more accurately represents the mutational process in bacteria.

Distance Effect

It this case, the evidence seems to weigh in favor of the findings from comparative studies—that mutation rates increase with distance from the replication origin. This result is based on thousands of *E. coli–S. enterica* homologs, and the same effect is observed in comparisons of other completely sequenced bacterial genomes. Despite the overall effect of chromosome location of substitution rates detected by comparative analysis, there is still quite a bit of variation in substitution rates on a more local scale. Even after correcting for codon usage biases, synonymous substitution rates can vary considerably at nearby loci, presumably because of the immediate base composition and dinucleotide frequencies. Although experimental studies can control for such local effects by examining reversion rates in identical constructs at numerous locations, it is possible that some regional factors affect mutation rates assayed by these methods. Because the experimental test of a distance effect examined reversion rates of loci at only four locations, and did not assay the specific transitions now known to contribute most to this effect in *E. coli–S. enterica* homologs, it remains to be seen if a distance effect will be evident once additional mutations and chromosome locations have been tested.

Transcription Effect

Comparative analyses of *E. coli–S. enterica* homologs suggest that transcription reduces mutation rates, whereas the experimental evidence points to an increase in mutations with transcription. Spontaneous mutations are thought to become more frequent with increased transcriptionm because DNA strands in highly-expressed genes spend more time in a single-stranded state and thus are prone to higher rates of damage caused, in part, by deamination and dimerization (Beletski and Bhagwat

1996, 1998; Francino and Ochman 2001). A likely explanation for an increase in mutation rates with transcription is that the susceptibility to such damage overrides any reductions furnished by transcription-coupled repair mechanisms (Selby and Sancar 1993; Wright, Longacre, and Reimers 1999). Aside from the described studies on *lac* reversions, similar transcription effects have recently reported for a variety of mutations in other experimental systems (Klapacz and Bhagwat 2002; Yoshiyama and Maki 2003). Because all experimental studies detect increases in mutation rates with transcription, and because many of the processes involved in the generation and repair of this damage have been identified, it may be that theoretical studies have overestimated the amount of selection acting on synonymous sites when accounting for the effects of adaptive codon bias in highly transcribed genes, As a consequence, mutation rates in highly expressed genes would be underestimated.

Transition Bias

As noted, both the comparative and experimental analyses indicate an approximately twofold bias toward transitions. Although the ratios are rather similar in the two types of studies, there are some differences in the frequencies of specific transitions and transversions. Such variation might stem from the fact that the substitution matrices used in comparative analyses are computed from heterogeneous sets of genes that differ in location, transcriptional status and base compositional features, all of which can lead to variation in the rates and patterns of certain mutations. Mutational patterns can also be derived in studies investigating the spectrum of point mutations that inactivate a particular gene. Although it is difficult to obtain actual mutation rates in such studies (because the numbers of mutational targets that inactivate a particular gene are not known), there is variation in the relative frequencies of transitions and transversions across loci, with some studies showing a preponderance of transversions (mutational spectra data complied by Hudson, Bergthorsson, and Ochman 2003).

Mutational Bias

Despite the cumulative bias toward A·T base pairs observed in experimental studies, there is thought to be an overall bias *against* A·T base pairs in *E. coli* and *S. enterica* because third codon positions are G + C rich. Although it is possible that the A·T mutational bias observed in rapidly growing, aerobic, experimental populations is not representative of the mutational process in nature, it is likely that these enteric species are undergoing a shift in base composition. In cases in which a gene has been sequenced in numerous strains or species of known phylogenetic relationships, it is possible to reconstruct the ancestral state by applying a parsimony approach, and to determine the polarity of individual substitutions (Francino et al. 1996; Duret et al. 2002; Green et al. 2003). In *E. coli* and *S. enterica*, the number of changes that convert G·C→ A·T base pairs is approximately twofold higher than the reverse substitutions (data from Francino et al. 1996). Although the bias toward higher A + T contents detected from sequence comparisons is not as pronounced as that observed experimentally, it appears that the base composition of these enteric species is not in equilibrium and that the present compositional features of a genome may not be an infallible guide to the mutational bias of a genome.

Mutation Rate

Experimentally determined mutation rates are calculated *per generation*, and those derived from comparative studies are expressed *per year*; for *E. coli* and *S. enterica*, an improbably low number of generations per year is required to make the published values agree. The disparity is likely due to differences in mutational processes under certain growth conditions, such as prolonged starvation or anaerobiosis, which are probably relevant for bacterial species that reside in ecological niches within and without animal hosts (Bjedov et al. 2003). In this case, it appears that the mutation rates obtained by both the comparative and experimental approaches could be correct, but are not compatible because they reflect mutagenesis occurring under different circumstances.

Along with the extrinsic factors that alter mutation rates, several intrinsic factors, such as neighboring nucleotides, level of transcription, and chromosome location, also modify the mutational spectrum in bacteria. Although numerous theoretical, comparative, and experimental studies have long treated the mutation rate as a static quantity for a given organism or species, the overwhelming evidence, as presented in this article, is that mutation rates cannot be considered constant, even within a genome.

Literature Cited

Beletskii, A., and A. S. Bhagwat. 1996. Transcription-induced mutations: increase in C to T mutations in the nontranscribed strand during transcription in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **93**:13919–13924.

———. 1998. Correlation between transcription and C to T mutations in the non-transcribed DNA strand. J. Biol. Chem. **379**:549–551.

Berg, O. G., and M. Martelius. 1995. Synonymous substitution-rate constants in *Escherichia coli* and *Salmonella typhimurium* and their relationship to gene expression and selection pressure. J. Mol. Evol. **41**:449–456.

Birky, C. W., Jr., and J. B. Walsh. 1992. Biased gene conversion, copy number, and apparent mutation rate differences within chloroplast and bacterial genomes. Genetics **130**:677–683.

Bjedov, I., O. Tenaillon, B. Gerard, V. Souza, E. Denamur, M. Radman, F. Taddei, and I. Matic. 2003. Stress-induced mutagenesis in bacteria. Science **300**:1404–1409.

Blake, R. D., S. T. Hess, and J. Nicholson-Tuell. 1992. The influence of nearest neighbors on the rate and pattern of spontaneous point mutations. J. Mol. Evol. **34**:189–200.

Brown, W. M., E. M. Prager, A. Wang, and A. C. Wilson. 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. **18**:225–239.

Cupples, C. G., and J. H. Miller, 1989. A set of *lacZ* mutations in *Escherichia coli* that allow rapid detection of each of the six base substitutions. Proc. Natl. Acad. Sci. USA **86**:5345–5349.

Daubin, V., and G. Perriere. 2003. G + C structuring along the genome: a common feature of prokaryotes. Mol. Biol. Evol. **20**:471–483.

Doolittle, R. F., D. F. Feng, S. Tsang, G. Cho, and E. Little. 1996. Determining divergence times of the major kingdoms of living organisms with a protein clock. Science **271**:470–477.

Duret, L., M. Semon, G. Piganeau, D. Mouchiroud, and N. Galtier. 2002. Vanishing GC-rich isochores in mammalian genomes. Genetics **162**:1837–1847.

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

Eyre-Walker, A., and M. Bulmer. 1995. Synonymous substitution rates in enterobacteria. Genetics **140**:1407–1412.

Fijalkowska, I. J., P. Jonczyk, M. M. Tkaczyk, M. Bialoskorska, and R. M. Schaaper. 1998. Unequal fidelity of leading strand and lagging strand DNA replication on the *Escherichia coli* chromosome. Proc. Natl. Acad. Sci. USA **95**:10020–10025.

Francino, M. P., L. Chao, M. A. Riley, and H. Ochman. 1996. Patterns of substitutions in enterobacterial genes are generated by transcription-coupled repair. Science **272**:107–109.

Francino, M. P., and H. Ochman. 2001. Deamination as the basis of strand-asymmetric evolution in transcribed *Escherichia coli* sequences. Mol. Biol. Evol. **18**:1147–1150.

Gojobori, T., K. Ishii, and M. Nei. 1982. Patterns of nucleotide substitution in pseudogenes and functional genes. J. Mol. Evol. **18**:360–369.

Green, P., B. Ewing, W. Miller, P. J. Thomas, and E. D. Green. 2003. Transcription-associated mutational asymmetry in mammalian evolution. Nat. Genet. **33**:514–517.

Hall, B. G. 1991. Spectrum of mutations that occur under selective and non-selective conditions in *E. coli*. Genetica **84**:73–76.

Hudson, R. E., U. Bergthorsson, J. R. Roth, and H. Ochman. 2002. Effect of chromosome location on bacterial mutation rates. Mol. Biol. Evol. **19**:85–92.

Hudson, R. E., U. Bergthorsson, and H. Ochman. 2003. Transcription increases a broad spectrum of spontaneous mutations in bacteria. Nucleic Acids Res. **31**:4517–4522.

Klapacz, J., and A. S. Bhagwat. 2002. Transcription-dependent increase in multiple classes of base substitution mutations in *Escherichia coli*. J. Bacteriol. **184**:6866–6872.

Lawrence, J. G., and H. Ochman. 1998. Molecular archaeology of the *Escherichia coli* genome. Proc. Natl. Acad. Sci. USA **95**:9413–9417.

MacKay, W. J., S. Han, and L. D. Samson. 1994. DNA alkylation repair limits spontaneous base substitution mutations in *Escherichia coli*. J. Bacteriol. **176**:3224–3230.

Mira, A., and H. Ochman. 2002. Gene location and bacterial sequence divergence. Mol. Biol. Evol. **19**:1350–1358.

Moran, N. A., M. A. Munson, P. Baumann, and H. Ishikawa. 1993. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. Proc. R. Soc. Lond. Ser. B **253**:167–171.

Muto, A., and S. Osawa. 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. Proc. Natl. Acad. Sci. USA **84**:166–169.

Ochman, H., S. Elwyn, and N. A. Moran. 1999. Calibrating bacterial evolution. Proc. Natl. Acad. Sci. USA **96**:12638–12643.

Ochman, H., and A. C. Wilson. 1988. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J. Mol. Evol. **26**:74–86.

Schaaper, R. M., and R. L. Dunn. 1991. Spontaneous mutation in the *Escherichia coli lacI* gene. Genetics **129**:317–326.

Selby, C.P., and A. Sancar. 1993. Transcription-repair coupling and mutation frequency decline. J. Bacteriol. **175**:7509–7514.

Sharp, P. M. 1991. Determinants of DNA sequence divergence between *Escherichia coli* and *Salmonella typhimurium*: codon usage, map position, and concerted evolution. J. Mol. Evol. **33**:23–33.

Sharp, P. M., and W. H. Li. 1987. The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Mol. Biol. Evol. **4**:222–230.

Sharp, P. M., D. C. Shields, K. H. Wolfe, and W. H. Li. 1989. Chromosomal location and evolutionary rate variation in enterobacterial genes. Science **246**:808–810.

Sueoka, N. 1962. On the genetic basis of variation and heterogeneity in base composition. Proc. Natl. Acad. Sci. USA **48**:582–592.

———. 1988. Directional mutation pressure and neutral molecular evolution. Proc. Natl. Acad. Sci. USA **85**:2653–2657.

Wright, B. E., A. Longacre, and J. M. Reimers. 1999. Hypermutation in derepressed operons of *Escherichia coli* K12. Proc. Natl. Acad. Sci. USA **96**:5089–5094.

Yoshiyama, K., and H. Maki. 2003. Spontaneous hotspot mutations resistant to mismatch correction in *Escherichia coli*: transcription-dependent mutagenesis involving template-switching mechanisms. J. Mol. Biol. **327**:7–18.

Simon Easteal, Associate Editor

Accepted July 16, 2003