Mutation Pressure and the Evolution of Organelle Genomic Architecture

Michael Lynch, *et al.*
Science **311**, 1727 (2006);
DOI: 10.1126/science.1118884

Mutation Pressure and the Evolution of Organelle Genomic Architecture

Michael Lynch,† Britt Koskella,* Sarah Schaack*

The nuclear genomes of multicellular animals and plants contain large amounts of noncoding DNA, the disadvantages of which can be too weak to be effectively countered by selection in lineages with reduced effective population sizes. In contrast, the organelle genomes of these two lineages evolved to opposite ends of the spectrum of genomic complexity, despite similar effective population sizes. This pattern and other puzzling aspects of organelle evolution appear to be consequences of differences in organelle mutation rates. These observations provide support for the hypothesis that the fundamental features of genome evolution are largely defined by the relative power of two nonadaptive forces: random genetic drift and mutation pressure.

The evolution of eukaryotes, and subsequently of multicellularity, was accompanied by dramatic changes in the nuclear genome, including expansions in sizes and numbers of introns, proliferation of mobile elements, and increases in lengths of intergenic regions. The continuity in scaling of these architectural features with genome size across major phylogenetic groups suggests that cellular and developmental features are not the primary driving forces in genome evolution, and the hypothesis has been raised that expansions in genome complexity are largely driven by two nonadaptive processes, random genetic drift and mutation (1, 2). If this hypothesis is correct, it ought to apply to all genomic regions.

However, in contrast to the shared patterns of evolution in the nuclear genomes of animals and plants, the organelle genomes of these lineages have evolved in radically different directions. Animal mitochondrial genomes are highly streamlined, whereas plant mitochondrial genomes contain large amounts of noncoding DNA. Is the theory less general than supposed, or do unique features of various organelle lineages encourage different evolutionary trajectories? Here we argue that when differences in mutation rates are accounted for, patterns of variation in organelle genome architecture support the theory that multiple aspects of genomic complexity owe their origins to nonadaptive processes.

Scaling of Mitochondrial Genome Content

Over the range of eukaryotic diversity, the scaling of mitochondrial genome content with genome size is quite similar to that in nuclear genomes (1, 2). The largest genome-size expansions are only weakly associated with gene number and primarily reflect increases in intronic and intergenic DNA [Fig. 1 and (3)]. However, in contrast to the situation with nuclear genomes, animals and plants occupy positions at the opposite ends of this gradient. The diminutive mitochondrial genomes of animals generally fall in the range of 14 to 20 kb, whereas plant mitochondrial genome sizes range from ~180 to 600 kb. Most unicellular species have intermediate aspects of mitochondrial genomic architecture and contain many genes absent from animal mitochondria (4). Thus, mitochondrial genomic architecture does not show overlap between animals and plants; this incongruity appears to be a consequence of contrasting evolutionary pressures unique to each lineage, with a strong ancestral component.

To put these results in a broader perspective, the average fractions of intergenic DNA in the nuclear genomes of vertebrates [0.65 (SEM = 0.05)], invertebrates [0.64 (0.03)], and plants [0.68 (0.10)] (1) are comparable to that for plant mitochondria, 0.72 (0.07). In contrast, the fraction of noncoding DNA in most animal mitochondria is just 0.05 to 0.10, less than that in any eukaryotic nuclear genome, and even below the average for prokaryotes, 0.12 (0.01) (1).

Mutation Rate

The two primary nonadaptive forces influencing genomic evolution are mutation, which defines the excess vulnerability of genes with complex structural features, and random genetic drift, which defines the magnitude of stochasticity in the evolutionary process (1). Any attempt to explain organelle genome diversity must address these issues. Comparative analysis of mitochondrial protein-coding genes implies substantial mutation-rate differences among major phylogenetic groups (Table 1). Rates of silent-site divergence range from 15 to 34 substitutions per site per billion years for all bilaterian-animal groups, whereas the average for plants is just 1/100th as much. In contrast, mutation rates

are fairly similar in animal and plant nuclear genomes (1, 5). Mitochondrial mutation-rate estimates in bilaterians are ~9 to 25 times those for the nuclear genomes in the same lineages, whereas the rates for most plants are ~0.05 times the nuclear rate (Table 1). This estimated disparity in mitochondrial mutation rates may be downwardly biased, as the only direct measures of the mitochondrial rate in animals are ~10 times the phylogenetically derived estimates, possibly because silent sites are not entirely neutral (6, 7).

The fact that unicellular eukaryotes have similar mitochondrial and nuclear mutation rates (Table 1) suggests that animal and plant mitochondria, respectively, acquired higher and lower mutation rates, rather than one of these lineages retaining the ancestral condition. At least three factors may promote elevated mutation rates in animal mitochondria. First, mitochondria generate free oxygen radicals, producing an internal environment with an exceptionally high mutagenic potential (8). Second, in contrast to nuclear DNA, mitochondrial DNA (mtDNA) is continuously replicated within nondividing cells, and the base-misincorporation rate (before proofreading) is ~10^3 to 10^4 times that in the nuclear genome (9). Third, few mitochondrial genomes encode DNA repair proteins, although some mitochondrial repair genes were apparently transferred to the nucleus during the establishment of the primordial mitochondrion. Mitochondrial nucleotide-excision repair may have been entirely lost, and mismatch repair is greatly curtailed in mammalian cells relative to yeast (10). Less clear are the reasons for the dramatic reduction in plant mitochondrial mutation rates, although this feature is not entirely invariant (11).

Genetic Effective Population Size

The genetic effective size of a population (N_e), which defines the power of random genetic drift, is a function of the absolute number of individuals in the population, the mating system, the degree of genetic linkage, and the background mutation rate (1, 12). Although there is substantial variation within lineages, the average N_e for nuclear genomes is substantially reduced in multicellular species; it is ~10^7 for unicellular eukaryotes, ~10^6 for invertebrates and annual plants, and ~10^4 for vertebrates and trees (1). Thus, from the standpoint of drift, the population-genetic environments of animal and plant nuclear genomes are quite similar. Does this conclusion extend to organelles?

It is commonly argued that haploidy and uniparental inheritance reduce the effective number of organelle genes per locus (N_g) in a diploid population to about one-quarter that for nuclear genes (13). (N_g equals the effective number of segregating units at the population level $2N_e$ for a nuclear locus and is approximately the effective number of females for a maternally inherited organelle.) This argument overlooks

Department of Biology, Indiana University, Bloomington, IN 47405, USA.

*These authors contributed equally to the gathering of data.

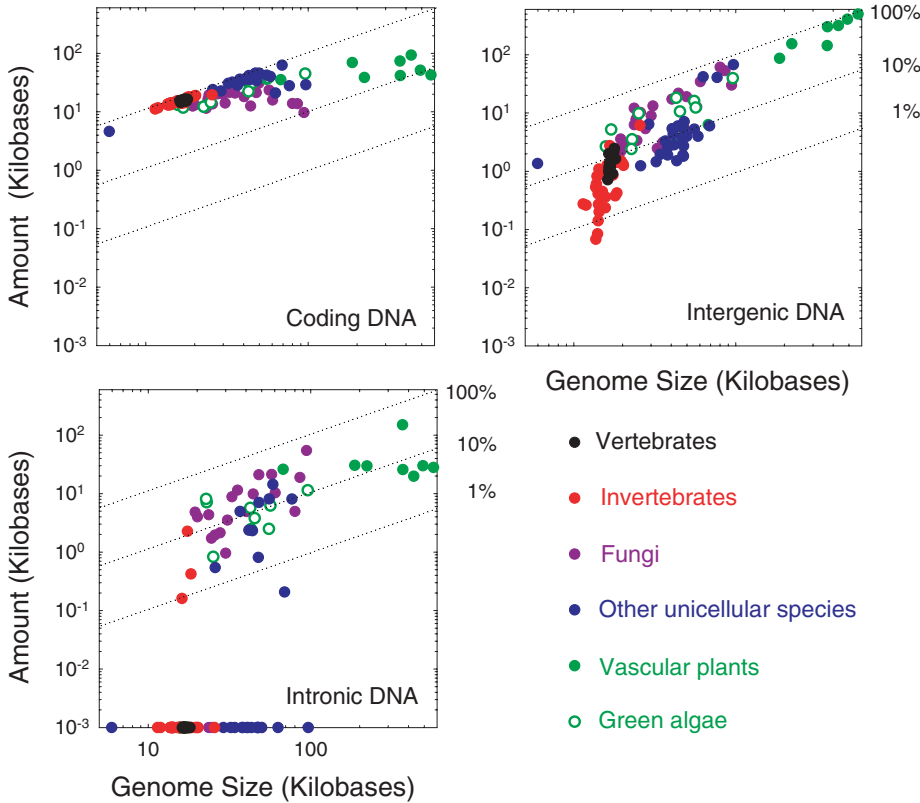†To whom correspondence should be addressed. E-mail: milynch@indiana.edu

Fig. 1. Scaling of genome content with mitochondrial genome size, color coded according to major organismal groups (*3*). Diagonal lines denote points of constant fractional genomic contributions. Data points at the base of the graph (i.e., 10^{-3}) denote zero content.

Table 1. Rates of mutation per nucleotide site in mitochondria estimated from phylogenetic comparisons under the assumption of neutral silent sites (u_m, in units of 10^{-9} base substitutions per site per year), and ratios of mutation rates (μ_m/μ_n) and effective number of genes (N_{gm}/N_{gn}) in mitochondrial (*m*) versus nuclear (*n*) genomes (*3*). Plants are defined to be multicellular members of the chlorophyte lineage. SEM in parentheses, a convention used throughout the paper.

Phylogenetic group	u_m	μ_m/μ_n	N_{gm}/N_{gn}
Mammals	33.88 (6.11)	24.60 (5.80)	1.27 (0.43)
Birds	17.34 (4.88)	13.72 (2.86)	
Reptiles/amphibians	15.43 (3.70)	24.68 (8.12)	2.18 (0.37)
Fish		23.11 (12.70)	
Bilaterian invertebrates	16.86 (8.70)	8.84 (3.17)	0.31 (0.14)
Plants	0.34 (0.07)	0.05 (0.01)	
Uni/oligocellular species		1.58 (0.48)	0.49 (0.16)

two key issues. First, the "one-quarter rule" assumes an identical level of selective interference in nuclear- and organelle-housed genes. The absence of recombination in animal mitochondria is one reason why this might not be true [e.g., (*14*)], but there are other complicating factors: Nuclear chromosomes contain many more potential targets for selective sweeps; the distributions of mutational effects driving selective interference may differ between the two types of genomes; and the organelle genomes of some unicellular species and plants do recombine (*15*). Second, the one-quarter rule assumes that males and females are equivalent with respect to progeny production. In principle,

a low ratio of male to female participants in mating (common in animals) can reduce the effective number of nuclear genes below that of maternally inherited organelle genes (*16*). Given these complexities, the degree to which the population-genetic environment differs among organelle and nuclear genes can only be resolved by empirical study.

From observations on mutation rates (Table 1) and within-population silent-site variation for mitochondrial versus nuclear genes (π_{sm} versus π_{sn}), the ratio of N_g for mitochondrial versus nuclear loci can be estimated (*3*). The average ratios for invertebrates and unicellular species are not significantly different from 0.25,

consistent with the "one-quarter" rule, whereas N_{gm} in vertebrates is generally one to two times N_{gn}. For plant mitochondrial genes, within-population polymorphisms are usually almost entirely absent, so few attempts have been made to estimate π_{sm}. However, if we take 0.001 to be a conservative upper bound (*3*), and note that the range of π_{sn} for plant nuclear genes is ~0.003 to 0.04 (*1*), then N_{gm}/N_{gn} is <0.5 and <6.6, respectively. All of these results imply that N_{gm} and N_{gn} within species are generally within a factor of 4 or so from each other. Thus, given the similarity of N_{gn} in animals and plants, the altered patterns of mitochondrial genome evolution in these lineages do not appear to be a consequence of a radical change in the power of random genetic drift. This leaves mutation as the likely determinant.

The Mutational Barrier to Organelle Genome Evolution

A key determinant of many aspects of genomic evolution is the ratio of the per-generation rate of mutation per nucleotide site (μ) to the power of random genetic drift ($1/N_g$), i.e., $N_g\mu$ (*1*), and it is useful that the within-population sequence divergence at silent sites (π_s) has an expected value equal to twice this quantity under drift-mutation equilibrium. In the nuclear genome, π_s is generally <0.01 in animals and plants and severalfold higher in unicellular species with elevated N_g (*1*). In contrast, π_s for animal mitochondrial genomes is generally higher than that in unicellular lineages and >100 times that in plants (Table 2), in accordance with the reduction in μ in the latter. These mutation-rate driven differences in $N_g\mu$ provide a potentially unifying explanation for several previously unexplained and disconnected observations on organelle genomes.

Noncoding DNA is a genomic liability from the standpoint of mutational vulnerability. For example, introns increase the mutational target size of their host genes, which must maintain specific nucleotide sequences for splice-site recognition during mRNA processing (*17*). Likewise, intergenic DNA is a mutational substrate for the appearance of inappropriate transcription factor–binding sites, core promoters, premature initiation codons, etc. (*18*, *19*). Theory suggests that significant intron proliferation requires $2N_g\mu n < 1$, where n is the number of nucleotides reserved for splice-site recognition (*17*), or equivalently $\pi_s < 1/n$. For nuclear spliceosomal introns, $n > 20$ implies a threshold π_s for intron proliferation of ~0.05, which is consistent with the disparities in intron abundance between multicellular and unicellular species (*1*, *2*). Because organelle introns are self-splicing (i.e., do not rely on an external spliceosome), they must retain a larger number of nucleotides critical to proper splicing, implying a threshold π_s for organelle intron proliferation lower than 0.05 by a factor of perhaps 3 to 5. Consistent with the theory, this condition is generally violated in

Table 2. Average silent-site nucleotide diversity (π_s), in units of numbers of substitutions per site between random pairs of sequences. The sample size (n) denotes the number of pooled genera from which the averages were computed (3). Nuclear data are from (1).

Phylogenetic group	Mitochondrion	n	Nucleus	n
Mammals	0.0406 (0.0087)	12	0.0036 (0.0010)	10
Birds	0.0169 (0.0053)	4	0.0060 (0.0012)	4
Reptiles/amphibians	0.0516 (0.0128)	5	0.0013 (0.0008)	2
Fish	0.0362 (0.0150)	6	0.0046 (0.0012)	5
Arthropods	0.0276 (0.0056)	17	0.0292 (0.0060)	8
Molluscs	0.0135 (0.0068)	6	0.0229 (0.0132)	2
Nematodes	0.0677 (0.0084)	8	0.0272 (0.0168)	2
Fungi	0.0120 (0.0046)	3	0.0507 (0.0202)	12
Plants	<0.0004 (0.0004)	4	0.0152 (0.0027)	24

intron-free animal mitochondria but easily met in intron-rich plant mitochondria (Table 2).

Two additional observations support the hypothesis that high $N_g\mu$ imposes a barrier to organelle intron colonization. First, the only animal mitochondria known to harbor introns are those of cnidarians (20, 21), which, like plant mitochondria, have such low mutation rates that within-species nucleotide polymorphisms are essentially unobservable, i.e., $\pi_s < 0.001$ (22). Second, in contrast to land-plant mitochondria, which generally contain 20 to 30 group II introns, all observed green-algal mitochondria have 0 to 8 mitochondrial introns (Fig. 1) (3). It is not known whether the per-generation rate of mutation per nucleotide site for green algae is similar to that for vascular plants, but an elevated N_g in the former is expected to promote a less permissive environment for intron proliferation.

Another unexplained aspect of mitochondrial genome evolution concerns the genetic code. Whereas the mitochondria of most unicellular lineages have experienced no more than two mitochondrial code changes, those of all bilaterians have between 3 and 5, with at least 12 unique changes occurring throughout the bilaterian phylogeny (23). In contrast, no reassignments have been found in plant mitochondria, and just one has been found in a cnidarian. Thus, there is an apparent association between the incidence of genetic-code alterations and the mutation rate.

The key first step in genetic-code evolution is a transient period during which a codon is entirely unused (24). The likelihood of such an event is miniscule in nuclear genomes with thousands of genes, but nontrivial in organelle genomes, ~70% of which completely lack one or more codons (25). There are still substantial impediments to genetic-code alterations in diminutive organelle genomes, but a central point is that codon reassignments must involve a series of fortuitous mutational events in the same linked genome, including modifications of transiently unassigned transfer RNAs (tRNAs) and reappearance of lost codons. Thus, the inverse scaling between the mutation rate and the waiting time for multiple mutations provides a reasonable explanation for the uneven inci-

dence of mitochondrial genetic-code changes in animals, unicellular species, and plants.

A third peculiar feature of organelle genome evolution is the phylogenetic distribution of mRNA editing (26). Although a few animals use editing to restore mismatches in mitochondrial tRNA stems [e.g., (27)], mRNA editing appears to be absent from animal mitochondria. In contrast, plant mitochondria use mRNA editing extensively. For example, 441 editing sites are present in *Arabidopsis* mitochondria (28), and similar levels of mitochondrial editing are found in other plants (29). The absence of mRNA editing from the organelles of green algae suggests a dramatic expansion of editing after the origin of multicellular plants (29).

The vast majority of mRNA editing in plant organelles ensures the maintenance of amino acids at sites that are conserved across distantly related species (30). Although this observation motivates the idea that editing provides a genomic buffer against the accumulation of deleterious mutations (26), three observations raise doubts about this interpretation. First, there appear to be no phylogenetic barriers to editing (26), and yet under the buffering hypothesis, editing is expected to be most common in genomes with high mutation rates, contrary to the pattern seen with animals and plants. Second, the buffering hypothesis ignores the complex requirements of the editing process itself. Plant mitochondrial mRNA editing relies on cis-binding sites for trans-acting editing-site–specific proteins encoded in the nucleus (31, 32). It is difficult to imagine a net advantage to editing if the processing of each site depends on numerous cis and trans sequences. Third, editing in plant organelles produces a heterogeneous pool of transcripts, some incompletely edited and others containing erroneous changes (33). Finally, mutations that restore the proper nucleotide at a previously edited site should accumulate at the neutral rate under the buffering hypothesis, but actually occur at four times the rate of silent-site substitution, which suggests a selective disadvantage to editing (34).

The mutation-pressure hypothesis helps explain these paradoxical observations by postulating that the maintenance of proper edito-

some recognition sites imposes a mutational burden on an allele. The minimal mutational disadvantage of an editing site is approximated by the total mutation rate over the nucleotide sites reserved for editing-site recognition, >23 for plants (31, 32), which implies a threshold $\pi_s < 0.04$ for the maintenance of editing sites. Thus, the absence of mRNA editing in animal mitochondria is in accordance with the hypothesis that the mutation-associated disadvantages are simply too great to allow its establishment, whereas π_s for plant mitochondria is well below the barrier to the accumulation of editing sites.

These observations on the attributes of mitochondrial genomes, combined with prior analyses of nuclear genomes (1, 2), lend generality to the conclusion that the primary factors driving genome architectural evolution are nonadaptive in nature. Although analyses spanning all of eukaryotes leave little room for independent hypothesis testing, a third opportunity is provided by the more phylogenetically limited chloroplast lineage. Studies of silent-site divergence in plants suggest that chloroplast mutation rates are about two to four times those in mitochondria and about 1/10th those in nuclei (35, 36), and the limited data for species with silent-site diversities measured jointly in chloroplast and nuclear genomes (3) suggest a ratio of N_g of 1.03 (0.45). In addition, the average π_s for plant chloroplasts, 0.0037 (0.0011) (3), is >10 times that for plant mitochondria but about 1/10th of that for animal mitochondria (Table 1). These observations suggest that, although the power of random genetic drift is roughly comparable in all three compartments of the plant genome, the efficiency of selection in the chloroplast is intermediate to that for animal and plant mitochondria, although much closer to the latter.

In accordance with the mutation-pressure hypothesis, intron densities per protein-coding gene and fractional contributions of intergenic DNA in plant chloroplasts are about one-third those in plant mitochondria (3). In addition, plant chloroplasts have experienced no genetic-code changes, and although editing is much less extensive than in plant mitochondria, there are still ~25 to 30 editing sites per genome (30). With the exception of euglenoids, which may be obligately asexual and highly vulnerable to selective interference, the chloroplast genomes of the main algal groups (with presumably larger N_g than plants) are completely lacking in introns or nearly so and also tend to have much lower levels of intergenic DNA (green algae being exceptions) (3).

Concluding Comments

Because mutation and random drift are universal genetic forces, before invoking natural selection as the underlying determinant of an observed pattern of biodiversity, an evaluation of the expectations under a purely nonadaptive scenario is desirable. Natural selection is clearly

a significant force on organelle gene-sequence evolution (*37, 38*), but selective arguments for the architectural features of organelle genomes have remained elusive. Although it has been suggested that an intracellular ''race to replication'' is responsible for the streamlining of animal mitochondrial genomes (*38*), it is unclear whether broader phylogenetic patterns in organelle evolution can be explained by variation in intracellular competition. Perhaps differential metabolic demands and/or organelle turnover rates are involved, but this remains to be demonstrated. The arguments presented above help explain not just the phylogenetic variation in noncoding organelle DNA, but also the peculiar distribution of genetic code changes and mRNA editing. Thus, while serving as a useful null model, the hypothesis that genome evolution is strongly influenced by nonadaptive forces appears to have broad explanatory power, with variation in nuclear-genome architecture being primarily driven by variation in N_e (*1, 2*), and differences in μ making a major contribution to organelle evolution.

References and Notes
1. M. Lynch, *Mol. Biol. Evol.* **23**, 450 (2006).
2. M. Lynch, J. S. Conery, *Science* **302**, 1401 (2003).
3. Materials and methods are available as supporting material on *Science* Online.
4. M. W. Gray *et al.*, *Nucleic Acids Res.* **26**, 865 (1998).
5. W.-H. Li, *Molecular Evolution* (Sinauer Associates, Sunderland, MA, 1997).
6. D. R. Denver, K. Morris, M. Lynch, L. L. Vassilieva, W. K. Thomas, *Science* **289**, 2342 (2000).
7. N. Howell *et al.*, *Am. J. Hum. Genet.* **72**, 659 (2003).
8. R. S. Balaban, S. Nemoto, T. Finkel, *Cell* **120**, 483 (2005).
9. A. A. Johnson, K. A. Johnson, *J. Biol. Chem.* **276**, 38097 (2001).
10. P. A. Mason, R. N. Lightowlers, *FEBS Lett.* **554**, 6 (2003).
11. Y. Cho, J. P. Mower, Y. L. Qiu, J. D. Palmer, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 17741 (2004).
12. J. H. Gillespie, *Genetics* **155**, 909 (2000).
13. S. R. Palumbi, F. Cipriano, M. P. Hare, *Evol. Int. J. Org. Evol.* **55**, 859 (2001).
14. S. Berlin, H. Ellegren, *Nature* **413**, 37 (2001).
15. N. W. Gillham, *Organelle Genes and Genomes* (Oxford Univ. Press, Oxford, UK, 1994).
16. C. W. Birky Jr., T. Maruyama, P. M. Fuerst, *Genetics* **103**, 513 (1983).
17. M. Lynch, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6118 (2002).
18. M. W. Hahn, J. E. Stajich, G. A. Wray, *Mol. Biol. Evol.* **20**, 901 (2003).
19. M. Lynch, D. G. Scofield, X. Hong, *Mol. Biol. Evol.* **22**, 1137 (2005).
20. M. J. H. van Oppen *et al.*, *J. Mol. Evol.* **55**, 1 (2002).
21. G. Pont-Kingdon *et al.*, *J. Mol. Evol.* **46**, 419 (1998).
22. T. L. Shearer, M. J. H. van Oppen, S. L. Romano, G. Worheide, *Mol. Ecol.* **11**, 2475 (2002).
23. R. D. Knight, S. J. Freeland, L. F. Landweber, *Nat. Rev. Genet.* **2**, 49 (2001).
24. T. H. Jukes, S. Osawa, *Comp. Biochem. Physiol. B* **106**, 489 (1993).
25. J. Swire, O. P. Judson, A. Burt, *J. Mol. Evol.* **60**, 128 (2005).
26. T. L. Horton, L. F. Landweber, *Curr. Opin. Microbiol.* **5**, 620 (2002).
27. D. V. Lavrov, W. M. Brown, J. L. Boore, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13738 (2000).
28. P. Giegé, A. Brennicke, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 15324 (1999).
29. R. Hiesel, B. Combettes, A. Brennicke, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 629 (1994).
30. T. Tsudzuki, T. Wakasugi, M. Sugiura, *J. Mol. Evol.* **53**, 327 (2001).
31. T. Miyamoto, J. Obokata, M. Sugiura, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 48 (2004).
32. E. Kotera, M. Tasaka, T. Shikanai, *Nature* **433**, 326 (2005).
33. C. G. Phreaner, M. A. Williams, R. M. Mulligan, *Plant Cell* **8**, 107 (1996).
34. D. C. Shields, K. H. Wolfe, *Mol. Biol. Evol.* **14**, 344 (1997).
35. K. H. Wolfe, W. H. Li, P. M. Sharp, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 9054 (1987).
36. B. S. Gaut, B. R. Morton, B. C. McCaig, M. T. Clegg, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10274 (1996).
37. M. Lynch, *Mol. Biol. Evol.* **14**, 914 (1997).
38. D. M. Rand, *Annu. Rev. Ecol. Syst.* **32**, 415 (2001).
39. Supported by grants from the NIH and NSF to M.L., an NSF predoctoral fellowship to B.K., and an NSF Integrative Graduate Education and Research Traineeship Program (IGERT) fellowship to S.S. Some key comments from J. Palmer led us to pursue this work. We are grateful to M. Neiman, J. Palmer, A. Richardson, and the reviewers for helpful comments.

The Nature and Dynamics of Bacterial Genomes

Howard Ochman* and Liliana M. Davalos

Though generally small and gene rich, bacterial genomes are constantly subjected to both mutational and population-level processes that operate to increase amounts of functionless DNA. As a result, the coding potential of bacterial genomes can be substantially lower than originally predicted. Whereas only a single pseudogene was included in the original annotation of the bacterium *Escherichia coli*, we estimate that this genome harbors hundreds of inactivated and otherwise functionless genes. Such regions will never yield a detectable phenotype, but their identification is vital to efforts to elucidate the biological role of all the proteins within the cell.

The organization of bacterial genomes is simple and elegant. These genomes are small, ranging from 500 to 10,000 kb, and are tightly packed with genes and other functional elements. The coding regions themselves are intronless and short, averaging a scant 1 kb, and are aligned almost contiguously along the chromosome. The common view is that the streamlining of bacterial genomes is the result of selection acting on replication efficiency and growth rates. Although this idea is warranted by the relatively low ceiling on bacterial genome size, there is no clear association between chromosome length and cell division rates either within or across bacterial species, implying that factors other than selection on overall chromosome size contribute to the compactness of bacterial genomes (*1*).

The elucidation of complete sequences has helped define the forces that shape bacterial genomes. Early research showed bacterial genomes to be tightly packed with functional elements, but unprecedented discoveries from genome analyses have shown that the genetic information encoded within bacterial genomes decays over evolutionary time scales (*2–4*). At first, this feature seems at odds with the high gene density observed in most bacterial genomes, but it is actually one of the primary determinants of their streamlined organization. All organisms accumulate mutations that can disrupt and degrade functional regions, but in bacteria (as well as in several eukaryotes) there is a mutational bias toward deletions over insertions (*1, 5–7*). When disruptions occur in genes that are no longer required, the nonfunctional regions can be maintained in the genome for some time, but they gradually erode and are eventually eliminated, as is evident from comparisons of bacterial pseudogenes with their functional counterparts. In this manner, bacterial genomes maintain high densities of functional genes.

The primary force countering the erosion of genomes is natural selection, which serves to maintain the functional regions. The degree of selection varies along a continuum depending on the role of a gene in cell survival and replication: Genes with little contribution to fitness are more susceptible to inactivating and deletional mutations, whereas those that are critical will resist such mutation. Moreover, the degree of selection acting on any particular gene can change over time and according to a specific ecological context. For example, the inactivation or loss of one or more genes can increase the value of others, and changes in bacterial ecology or lifestyle might render some genes redundant (*8*).

As important as the intensity of selection is the effectiveness of selection, which depends on population size and structure. In very large populations, deleterious mutations in beneficial genes are not likely to become fixed by chance; but in small populations, even useful genes can

Department of Biochemistry and Molecular Biophysics, University of Arizona, Tucson, AZ 85721, USA.

*To whom correspondence should be addressed. E-mail: hochman@email.arizona.edu