Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 7160–7164, August 1991
Genetics

A constant rate of spontaneous mutation in DNA-based microbes

JOHN W. DRAKE

Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709

Communicated by Evelyn M. Witkin, May 22, 1991

ABSTRACT **In terms of evolution and fitness, the most significant spontaneous mutation rate is likely to be that for the entire genome (or its nonfrivolous fraction). Information is now available to calculate this rate for several DNA-based haploid microbes, including bacteriophages with single- or double-stranded DNA, a bacterium, a yeast, and a filamentous fungus. Their genome sizes vary by ≈6500-fold. Their average mutation rates per base pair vary by ≈16,000-fold, whereas their mutation rates per genome vary by only ≈2.5-fold, apparently randomly, around a mean value of 0.0033 per DNA replication. The average mutation rate per base pair is inversely proportional to genome size. Therefore, a nearly invariant microbial mutation rate appears to have evolved. Because this rate is uniform in such diverse organisms, it 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.**

Rates of spontaneous mutation per base pair vary hugely within and between organisms, as if chaotically. Both the kinds of mutations and the processes that generate them are diverse and only partially discovered; the catalogue of errors is large compared with the manual of correct procedure. The rate of even a single well-defined pathway such as G·C → A·T can vary by more than 2000-fold at different sites within a single gene (1), presumably under the still largely mysterious influences of local DNA sequence. Is there any underlying order to these mutation rates? If so, it is likely to be found in the most macroscopic measure, the mutation rate per entire genome per round of DNA replication.

A dependable measure of the genomic mutation rate should satisfy several criteria. The experimental mutational target should be large enough to sufficiently sample the genome, and particularly to sample the diverse kinds of mutations that may arise. The mutants should grow at the same rate as their progenitors, or else differential growth should be measured and taken into account in the calculations. (In practice, selection is largely absent from most systems adopted for mutation studies.) The mutations should be expressed in a short fraction of a generation or, if substantial phenotypic lag occurs, it, too, should be quantified and mathematically nullified. The system should have been sufficiently explored so that its limits and artefacts are discovered. For instance, plating density artefacts are universal among microbial systems, although variable in magnitude and direction. Residual growth within screening systems is also common, so that actual population sizes tend to be larger than naively measured. Finally, the pattern of mutation—the mutational spectrum—should have been described at the molecular level. This is important in two respects. First, a spectrum will reveal whether the pattern is reasonably comparable to that in other systems: all kinds of mutations should be represented, and "hotspotting" (high mutability at particular sites within a gene), although typically present, should not overwhelm the response of the target gene. Second, a spectrum enables calculations to estimate what fraction of mutations escape detection.

I am now aware of approximations to such data sets in six organisms (taking the sibling coliphages T2 and T4 to be a single organism). All employ DNA to encode their genomes. Three are cellular organisms and three represent diverse types of bacteriophages. In most of the examples the data are imperfect and compromises had to be made while performing the calculations. Nevertheless, these genomic mutation rates are remarkably similar, while the corresponding average mutation rates per base pair (bp) vary by more than 10^4-fold. This genomic mutation rate, ≈0.0033 per DNA replication, has interesting implications for the evolution of mutation rates.

METHODS AND CALCULATIONS

Two methods are commonly used to calculate mutation rates in microbes. One is the fluctuation test of Luria and Delbrück (2). Many replicate cultures are grown from small mutant-free inocula until roughly half have experienced a mutation, and the cultures are then scored for average population size and for mutants. The mutation rate can then be calculated either by using the entire distribution (method F) or by using only the fraction of cultures without mutants (method F_0). Another method is that of mutant accumulation (3). In one protocol (method A), a population is grown until large enough to contain many mutants and the mutant frequency is then followed as a function of population size. Then $\mu = (f_2 - f_1)/\ln(N_2/N_1)$, where μ = mutation rate per replication, f = mutant frequency (at time 1 or 2) and N = population size (at time 1 or 2). In the more common experimental situation (method A_0), several cultures are initiated from small mutant-free inocula and grown extensively, and their mutant frequencies are then determined. Here N_1 is not the initial inoculum but rather the value when the population reaches the size when mutation is likely to occur, namely $1/\mu$. Thus, taking $f_1 = 0$, $\mu = f/\ln(N\mu) = 0.4343f/\log(N\mu)$, where f is the *median* frequency rather than the mean, the latter being overly sensitive to "jackpot" cultures (2). This equation must be solved by iteration; several significant figures can be quickly obtained with a hand calculator and the following algorithm: calculate $0.4343f$ and enter into memory; select a test $\mu \sim f/5$; calculate $N\mu$ and take its $\log_{10}$; execute ÷, then memory return, then =, then reciprocal; choose a new μ and recalculate.

It is also necessary to estimate the correction factor C that converts an observed mutation frequency f_o or rate μ_o into the underlying frequency or rate: $f = Cf_o$ or $\mu = C\mu_o$. (C is the reciprocal of the efficiency of mutation detection.) I usually performed these calculations approximately as follows. Because almost all of the mutational targets considered here are protein-encoding genes, mutations other than base-pair substitutions (BPSs) are well detected because they profoundly disrupt protein structure; these comprise frame-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: BPS, base-pair substitution; CT, chain-terminating.

shift mutations (almost always the addition or deletion of one or two base pairs), insertion mutations (almost always of mobile genetic elements), gene-disrupting chromosomal mutations (duplications and rearrangements, which are infrequent), and deletions. Among BPSs, the nonsense or chain-terminating (CT) mutations, which generate internal UAG, UAA, and UGA codons, are well detected because they generate prematurely terminated proteins. The remaining BPSs generate missense mutations (which cause amino acid substitutions) or synonymous mutations (which alter codons but not amino acids); in laboratory screens the former are detected inefficiently, the latter hardly at all. A mutational spectrum will thus reveal almost all of the non-BPS mutations and the CT BPSs but only a small and highly variable fraction of missense and synonymous mutations. The latter comprise 61 of the 64 codons, so that total BPSs can be estimated as 64/3 times CT mutations. (This calculation depends somewhat upon local (A+T):(G+C) ratios and codon usage patterns, but such second-order effects are small in comparison with other uncertainties in most of the calculations and I have usually ignored them.)

The typical procedure is to estimate C, then to calculate the mutant frequency f, then the mutation rate μ_t of the measured target sequence, then the mutation rate μ_{bp} of the average base pair (dividing μ_t by the size of the target sequence), then the mutation rate μ_g of the entire genome (multiplying μ_{bp} by the number of base pairs per genome). Values were rounded only after the calculations were completed. The mean value of C was 3.12 (range 1.34–8.01). The average fraction of BPSs for protein-encoding genes (thus excluding *Saccharomyces cerevisiae SUP4*) was $0.684 \approx 2/3$ (range 0.267–0.980).

Bacteriophage M13. The genome size is 6407 bp (4). The mutational target is 258 bp of an inserted *Escherichia coli lacZα* sequence (5). The spontaneous mutant frequency is 6.4×10^{-4} but only 117 of 128 sequenced mutants revealed a mutation in the target sequence, the other 11 displaying an equivocal phenotype (5); thus, $f_o = (6.4 \times 10^{-4})(117/128) = 5.85 \times 10^{-4}$. Of the 258 target bp, only about 130 bp are detectable as BPSs (ref. 6; T. A. Kunkel, personal communication), so that 0.5 of the BPSs go undetected. Of the 117 spontaneous mutants, 50 contained non-BPS mutations and 67 contained BPSs; thus, $C = (50 + 2 \times 67)/117 = 1.57$ and $f = 9.20 \times 10^{-4}$. Assuming geometrical replication and using method A_0 with $N = 10^{15}$ (T. A. Kunkel, personal communication), $\mu_t = 0.4343f/\log(N\mu_t) = 3.78 \times 10^{-5}$. However, the M13 mode of DNA replication is single strand (SS) → double strand (DS) [= replicative form (RF)] → pool of RF → rolling circle, then continuous release of phage particles from slowly replicating cells (7). For M13, RF → RF and cell → cell are geometrical modes but RF → SS is linear. Here the experimental population is obtained by sampling a plaque, growing the cells and harvesting supernatant phages; thus, there is a substantial component of linear replication. For purely linear replication with a nonmutating template, $\mu_t = f = 9.20 \times 10^{-4}$; but a mutating template would reduce this value. To estimate the mutation rate, I took the geometric mean of the two values, obtaining $\mu_t = 1.86 \times 10^{-4}$. Then $\mu_{bp} = 7.23 \times 10^{-7}$ and $\mu_g = 0.00463$.

Bacteriophage λ. The genome size is 48,502 bp (8). The mutational target is the *cI* prophage repressor gene, which includes 714 codon and CT bp + 74 regulatory bp = 788 bp (9); the target was scored after lytic growth, when mutations are reported to generate a mutant clone size distribution characteristic of exponential rather than linear DNA replication (10). The uncorrected forward mutation rate is 2×10^{-5} per DNA replication (11). Two collections of *cI* mutants have been characterized; both, however, were obtained from mutants arising during prophage replication and are likely to differ in sometimes harboring insertion-element mutations and in other ways difficult to predict. In one collection of >300 mutants (12), 5% were UAG; among 25 non-UAG mutations, 15 were IS insertions, 6 were frameshift mutations or small deletions, 3 were polar (probably non-UAG CT mutations), and 1 was missense. Thus, 100 mutations would comprise 5 UAG CT, 57 insertion, 22.8 frameshift and/or deletion, 11.4 non-UAG CT, and 3.8 missense mutations; ignoring the insertion mutations, $C = [22.8 + 16.4(64/3)]/43 = 8.67$, probably an upper limit. In another collection (ref. 13; F. Hutchinson, personal communication), the *cI* genes were sequenced from 49 mutants; 21 were non-BPS, 27 were missense, and 1 was non-UAG CT. Thus, $C = (21 + 63/2)/49 = 1.071$, certainly a lower limit because the host strain carried the *supE* amber suppressor and the *cI857* temperature-sensitive mutation (14), the latter probably increasing the detection of missense mutations (ref. 15; K. R. Tindall, personal communication). The geometric mean of these two values is $C = 3.05$. Therefore, $\mu_t = 6.1 \times 10^{-5}$, $\mu_{bp} = 7.74 \times 10^{-8}$, and $\mu_g = 0.00375$.

Bacteriophage T2. The genome size is 160 kbp (16). The mutational target is the *rII* locus, whose size is taken to be the same as in the closely related phage T4 (see below), namely 3136 bp. When method F was used (17), 87.6 newly arisen *r* clones (corrected for coincidences) were observed among 22,620 bursts that yielded 1,850,000 progeny. Therefore, the number of DNA replications was 1,850,000 − 22,620 = 1,827,380 and the uncorrected mutation rate was $87.6/1,827,380 = 4.79 \times 10^{-5}$. Turning again to phage T4 (see below), the *rII*/*r* ratio is 0.596 and $C = 2.96$; thus, $\mu_t = 4.79 \times 10^{-5} \times 0.596 \times 2.96 = 8.46 \times 10^{-5}$. Then $\mu_{bp} = 2.70 \times 10^{-8}$ and $\mu_g = 0.00432$.

Bacteriophage T4. The genome size is 166 kbp (16). The mutational target is the 3136 bp of the *rII* locus (18–21). Stocks grown from small inocula to $N \approx 3 \times 10^{11}$ particles contain a median of $\approx 6 \times 10^{-4}$ *r* mutants (unpublished results). Pooling indistinguishable collections (22–24), 203 *r* mutants comprised 55 *rI* mutants, 121 *rII* mutants, and 27 mostly rapidly reverting or leaky *rII* mutants; thus, there are 121/203 = 0.596 conventional *rII*/*r* and the median *rII* frequency is $f_o = 3.58 \times 10^{-4}$. Among 121 *rII* mutants, the above three collections give 27 revertible by base analogues (= BPSs) and 94 not (of which 37/39 were revertible by proflavin and thus contained frameshift mutations, the other two probably containing frameshift mutations not revertible by proflavin). In bp 220–654 there are 15 CT and 21 missense mutations (18, 25); with 66% A+T we expect 0.073 of random BPS to be CT; 15/0.073 = 205.6, so that only 21/205.6 = 0.102 of non-CT BPS are scored, in general agreement with a previous study showing that *rII* missense mutations are inefficiently detected (15). Thus, $C = [(27/0.102) + 94]/121 = 2.96$. Thus, $f = 3.58 \times 10^{-4} \times 2.96 = 1.06 \times 10^{-3}$. Then, by method A_0, $\mu_t = 0.4343f/\log(N\mu_t) = 6.32 \times 10^{-5}$, $\mu_{bp} = 2.02 \times 10^{-8}$, and $\mu_g = 0.00334$.

E. coli lacI. The genome size is 4704 kbp (26). The *lacI* target contains 1110 bp (27, 28) and resides either chromosomally or on an F′ element, its mutation rate being similar in the two locations (R. M. Schaaper, personal communication). The efficiency of mutant detection was determined as follows. A total of 108 BPS comprised 32 CT and 76 missense mutations (ref. 29; J. Halliday and B. W. Glickman, personal communication); thus, the BPS correction factor = $32 \times (64/3)/108 = 6.321$. Of 174 total spontaneous mutants, 154 were non-BPS and 20 were BPS (29); thus, $C = (154 + 6.321 \times 20)/174 = 1.61$. Two measurements of the mutation rate by method A_0 are available. In the first (F. Allen and B. W. Glickman, personal communication), $N = 4 \times 10^8$ and $f_o = 1.46 \times 10^{-6}$; thus, $f = 2.35 \times 10^{-6}$, $\mu_t = 0.4343f/\log(N\mu_t) = 4.53 \times 10^{-7}$, $\mu_{bp} = 4.08 \times 10^{-10}$, and $\mu_g = 0.00192$. In the second (R. M. Schaaper, personal communication), $N = 2 \times 10^9$ and $f_o = 3.5 \times 10^{-6}$. Then $f = 5.64 \times 10^{-6}$, $\mu_t =$

$0.4343f/\log(N\mu_t) = 7.69 \times 10^{-7}$, $\mu_{bp} = 6.93 \times 10^{-10}$, and $\mu_g = 0.00326$.

E. coli hisGDCBHAFE. The *his* operon contains 7389 target bp (30). The mutation rate to histidine auxotrophy by method F is 1.17×10^{-6} (31). In the closely related bacterium *Salmonella typhimurium*, 226 histidine-requiring mutants comprised 74 non-BPS and 152 BPS and, of 140 BPS tested, 28 were CT and 112 were missense (32). Thus, $C = [74 + 152(28 \times 64/3)/140)]/226 = 3.20$. Then $\mu_t = 3.74 \times 10^{-6}$, $\mu_{bp} = 5.06 \times 10^{-10}$, and $\mu_g = 0.00238$.

S. cerevisiae URA3. The genome contains 12,537 kbp of non-rRNA-encoding DNA (rDNA) (33) and 1274 kbp of rDNA (34) for a total of 13,811 kbp; there is little or no "junk" DNA (35, 36). The *URA3* gene contains 804 bp (37). The average mutation rate from three fluctuation tests (method F_0) is 2.77×10^{-8} (E. A. Savage, G. S.-F. Lee, R. G. Ritzel, and R. C. von Borstel, personal communication). Of 106 mutations, 17 were non-BPS and 89 were BPS (39 CT and 50 missense) (ref. 38; R. C. von Borstel, personal communication). Thus, $C = [17 + 39 \times (64/3)]/106) = 8.01$. Then $\mu_t = 2.22 \times 10^{-7}$, $\mu_{bp} = 2.76 \times 10^{-10}$, and $\mu_g = 0.00381$.

S. cerevisiae SUP4. The target is a plasmid-borne tRNA gene containing 75 exon bp and 14 intron bp whose mutational responses are well explored (ref. 39; B. A. Kunz, personal communication). Because few mutations are detected in the intron, only the exons will be considered. In 307 sequenced mutants bearing exon mutations, 56 non-BPS and 253 BPS mutations were detected (39); but the BPSs included two close doubles, so that 251 BPSs were scored. The screen can detect 171 of 3×75 possible BPS. Thus $C = [56 + 251(3 \times 75/171)]/307 = 1.26$. The measured mutation frequency was 2.01×10^{-6}; thus $f = 2.53 \times 10^{-6}$. The population size, including 3.5 residual divisions on the plate, was $N = 1.2 \times 10^8$. Then, by method A_0, $\mu_t = 0.4343f/\log(N\mu_t) = 5.93 \times 10^{-7}$, $\mu_{bp} = 7.91 \times 10^{-9}$, and $\mu_g = 0.109$.

S. cerevisiae CAN1. The *CAN1* canavanine-resistance gene is composed of a 1773-bp open reading frame (40, 41) and about 258 bp of regulatory sequences (42) for a total of 2031 bp. Some 20 mutation rates have been measured, usually by fluctuation tests (method F), in 14 strains in four laboratories (refs. 43–48; E. A. Savage, G. S.-F. Lee, R. G. Ritzel, and R. C. von Borstel, personal communication). When corrected for residual growth in the screening system, and then averaged first by strain and then by laboratory and then overall (R. C. von Borstel, personal communication), the mean *CAN1* mutation rate is 1.13×10^{-7}. There is no mutational spectrum to use in calculating C, so the average of the preceding six well-determined values was used: $C = 3.12$. Then $\mu_t = 3.51 \times 10^{-7}$, $\mu_{bp} = 1.73 \times 10^{-10}$, and $\mu_g = 0.00238$.

Neurospora crassa ad-3AB. The genome size of 41.9 Mbp is the average of an estimate of 40.9 based on microfluorimetry (49) and 42.9 based on pulsed-field gel electrophoresis of intact chromosomal DNA (ref. 50; M. J. Orbach, personal communication) calibrated, in both cases, against values from yeast. The target sizes of the "purple-adenine" genes *ade-3A* and *ade-3B* were estimated from the coding sequences of the homologous yeast genes: 921 bp for *ADE1* (5-aminoimidazole-4-carboxylate ribonucleotide synthetase) (51) + 1713 bp for *ADE2* (5-aminoimidazole ribonucleotide carboxylase) (K. G. Skryabin, personal communication) = 2634 bp rounded up to 2700 bp to include regulatory sequences. The mean uncorrected mutant frequency was 3.9×10^{-7} (52). Detailed spectral information is not available to calculate C, but 100 *ade-3B* mutants comprised 41 probable missense mutations and 59 probable CT, frameshift, and/or small-deletion mutations (52). Thus, non-BPS mutants probably make up a substantial fraction of the whole and missense mutations seem relatively well represented, as in the analogous *E. coli lacI* system; therefore, I used the *lacI* value of $C = 1.6$. The mutations arose in a two-component balanced heterokaryon (53). Stocks were grown through sporulation to 10^7 conidia and were then screened for mutants under conditions where sectoring could be detected to about ⅛ colony (F. J. de Serres, personal communication). Counts of nuclei (D. R. Stadler, personal communication) revealed 0 conidia with 0 nuclei, 49 with 1, 85 with 2, 72 with 3, 25 with 4, and 10 with ≥5; the average was 2.43 per conidium. Renormalizing this nonrandom distribution to conidia with ≥2 nuclei, assorting nuclei in each class using binomial coefficients, and assuming scoring after a number of synchronous divisions producing ≤8 nuclei per conidium generated the equivalent of an overall 3.55-fold residual growth in the plating system. Thus, $N = 3.55 \times 10^7$ and $f = 3.9 \times 10^{-7} \times 1.6 \div 3.55 = 1.76 \times 10^{-7}$. Then, by method A_0, $\mu_t = 0.4343f/\log(N\mu_t) = 1.21 \times 10^{-7}$, $\mu_{bp} = 4.47 \times 10^{-11}$, and $\mu_g = 0.00187$.

N. crassa mtr. The *mtr* (methyltryptophan resistance) region generates a 2.3-kb transcript that includes a small upstream open reading frame (uORF) plus the structural ORF for the neutral amino acid permease (54). Because *mtr* mutations arise within the permease ORF and up to several hundred bp upstream, including near or within the uORF (D. R. Stadler, personal communication), the mutational target appears to extend at least from the beginning of the uORF through the permease ORF polyadenylylation signal, or 1529 bp. Two differently obtained mutation rates are available (D. R. Stadler, personal communication). With method A_0, $N = 4 \times 10^8$ and $f_o = 2.7 \times 10^{-7}$ mutants per conidium. However, the mutations are recessive and only mononuclear mutant conidia are scored; these are 20% of the conidia, and the mutant frequency must therefore be increased 5-fold. As with the *S. cerevisiae CAN1* locus, let $C = 3.12$. Thus, $f = 2.7 \times 10^{-7} \times 3.12 \times 5 = 4.21 \times 10^{-6}$. Then $\mu_t = 0.4343f/\log(N\mu_t) = 7.39 \times 10^{-7}$, $\mu_{bp} = 4.64 \times 10^{-10}$, and $\mu_g = 0.0195$. With method A, the regression slope (when 0, 0 was the obligate intercept) was 9.78×10^{-9} mutants per conidium per division. As above, this value must be adjusted by 5×3.12 to obtain $\mu_t = 1.52 \times 10^{-7}$. Then $\mu_{bp} = 9.96 \times 10^{-11}$ and $\mu_g = 0.00417$.

RESULTS

Thirteen spontaneous genomic mutation rates could be estimated, scattered among three different kinds of bacteriophages, a bacterium, a yeast, and a filamentous fungus (Table 1). It is perhaps surprising that more examples are not available, several favorite experimental microbes or genes being conspicuously absent, but the requisite data are lacking.

The two value-pairs given in parentheses in Table 1 are outliers by both biological and statistical criteria and should be considered to be nonrepresentative. The *S. cerevisiae SUP4* gene encodes a suppressor tRNA and resides on a plasmid. Such suppressor loci tend to be highly mutable in both yeast and bacteria, perhaps because of their high potential for secondary structure; residence on a plasmid may further affect the mutation rate. The other two yeast values are central. The high *N. crassa mtr* value is countered by two other *N. crassa* values, one from *mtr* itself. These two value-pairs are also statistical outliers by a conservative Dixon-type discordancy test (ref. 55, pp. 99 and 174): the hypothesis of no discordancy is rejected ($P < 0.01$) with a 95% confidence interval of 0.0024–0.0041 around $\mu_g = 0.0033$ not conditional on the outcome of the discordancy test (the asymmetry being a consequence of the robustness of the procedure to identify outliers).

Table 1 reveals a remarkable clustering of genomic mutation rates: while average mutation rates per base pair varied by about 16,000-fold, 11 of the 13 genomic mutation rates (excluding the outliers) varied by only 2.5-fold (58-fold in-

Table 1. Rates of spontaneous mutation in DNA-based microbes

Organism	Genome size, bp	Target	Mutation rate: Per bp (μ_{bp})	Mutation rate: Per genome (μ_g)
Bacteriophage M13	6.41×10^3	*lacZα*	7.2×10^{-7}	0.0046
Bacteriophage λ	4.85×10^4	*cI*	7.7×10^{-8}	0.0038
Bacteriophage T2	1.60×10^5	*rII*	2.7×10^{-8}	0.0043
Bacteriophage T4	1.66×10^5	*rII*	2.0×10^{-8}	0.0033
E. coli	4.70×10^6	*lacI*	4.1×10^{-10}	0.0019
			6.9×10^{-10}	0.0033
		hisGDCBHAFE	5.1×10^{-10}	0.0024
S. cerevisiae	1.38×10^7	*URA3*	2.8×10^{-10}	0.0038
		SUP4	(7.9×10^{-9})	(0.11)
		CAN1	1.7×10^{-10}	0.0024
N. crassa	4.19×10^7	*ad-3AB*	4.5×10^{-11}	0.0019
		mtr	(4.6×10^{-10})	(0.019)
			1.0×10^{-10}	0.0042
		Median excluding outliers		0.0033
		Arithmetic mean excluding outliers		0.0033
		Geometric mean excluding outliers		0.0031

cluding the outliers). Average mutation rates per base pair are reciprocally proportional to genome size (Fig. 1) with slope ± SE = −0.96 ± 0.06, the slope being indistinguishable from −1 excluding none of the values (P = 0.55 based on a two-tailed t test). (The slope becomes −1.06 when the outliers are excluded.)

Thus, pending the appearance of new results to the contrary, I conclude that DNA-based microbes share a spontaneous mutation rate of about 0.003 per genome per replication.

DISCUSSION

In a group of DNA-based microbes whose genome sizes vary by 6500-fold and whose average mutation rates per base pair vary by 16,000-fold, mutation rates per genome vary by only 2.5-fold (excepting two well-defined outliers). It is remarkable that the range of genomic mutation rates is so small in view of the difficulties encountered in making the estimates and because of uncertainties as to whether these organisms are fair samples of populations undoubtedly polymorphic for mutation rate. Indeed, a much more crude 1968 sample (56) exhibited much more variability.

A common mutation rate for such diverse organisms, whose genomes are composed of either single-stranded (phage M13) or double-stranded DNA, and which include both lysogenic and lytic bacteriophages, bacteria, and microbial eukaryotes, strongly implies that this rate is highly evolved. Indeed, this rate must have been shaped in response to evolutionary forces of a very general nature, forces independent of kingdom and niche. Students of the evolution of mutation rates have postulated a number of driving forces (57, 58). Natural selection in favor of newly arisen mutants will select indirectly for organisms with higher-than-average mutation rates, although recombination will subsequently separate the selected locus from the up-modifier of mutation rate and will thus relax the selection for increased mutation rates. Conversely, organisms with higher-than-average mutation rates will produce more mutationally defective offspring, so that selection will favor down-modifiers of mutation rates. However, the reduction of mutation rates by down-modifiers is achieved only at some physiological cost, such as the expenditure of resources to encode and operate more systems for preventing mutation, or the slowing of replication to permit extant systems to operate more efficiently. These factors are of the desired generality, and some closely approached equilibrium among them—particularly between the deleterious impact of mutations and the cost of further reducing mutation rates—is likely to be responsible for the observed value.

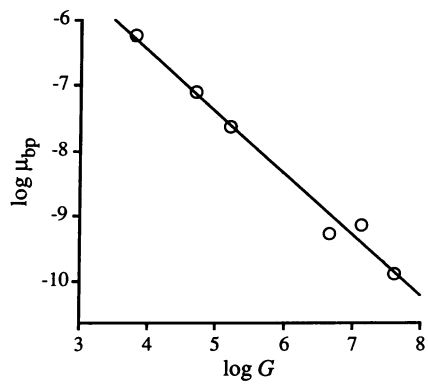

FIG. 1. Average mutation rate μ_{bp} per base pair as a function of genome size G in bp. The logs of the rates for each organism were averaged and all 13 values are included. Phages T2 and T4 were treated as a single organism.

The large (10^4-fold) range in mutation rates per base pair among these microbes seems inconsistent with the operation of any simple, universal molecular evolutionary clock (a constant rate per year of neutral evolution), as is much discussed for metazoans and sometimes suggested for bacteria (59, 60). Rates of molecular evolution are usually calibrated with BPSs, which make up variable fractions of the genomic rates in Table 1, but this variation is sporadic and much smaller than 10^4-fold. For such a clock to operate, it would have to be stabilized by an extraordinary balance among total mutation rate, fraction of neutral mutations, and average generation time, parameters not obviously related to microbial genome size.

The mutation rates in Table 1 were obtained by using cells growing in an often rich mixture of nutrients in a usually aerobic environment. Thus, the mutations will have arisen as a simple mixture of unforced errors of DNA replication plus the consequences of DNA damage directly attendant upon ordinary cellular metabolism and tightly linked temporally with DNA replication itself. The contrasting condition, more likely to obtain in nature, is that of a microbe spending most of its time malnourished and only occasionally encountering a burst of nutrients (61). Such a microbe would be subject to a barrage of mutagens of natural origin and its mutation rate

might thus contain a substantial replication-independent, time-dependent component; it might even deliberately increase its mutation rate at such times (62–64). It would therefore be interesting to learn whether time-dependent spontaneous mutation rates are subject to constraints as strong as those shaping replication-dependent rates.

Finally, I would note that the regularity of the relationships manifest in Table 1 may provide a rational basis for extrapolating spontaneous mutation rates among diverse microbial organisms.

I am grateful to the many people who provided me with unpublished data. Numerous colleagues provided vigorous, invaluable suggestions concerning the calculations and their interpretation: Carl Barrett, John Cairns, Allan Campbell, Jim Crow, Fred de Serres, Barry Hall, Burke Judd, Bernie Kunz, Chuck Langley, Mike Lynch, John Maynard Smith, Roel Schaaper, Dave Stadler, Charlie Steinberg, and Jack von Borstel. The statistical calculations were cheerfully performed by Maja Kricker and Walter W. Piegorsch.

1. Ronen, A. & Rahat, A. (1976) *Mutat. Res.* **34,** 21–34.
2. Luria, S. E. & Delbrück, M. (1943) *Genetics* **28,** 491–511.
3. Drake, J. W. (1970) *The Molecular Basis of Mutation* (Holden-Day, San Francisco).
4. van Wezenbeek, P. M. G. F., Hulsebos, T. J. M. & Schoenmakers, J. G. G. (1980) *Gene* **11,** 129–148.
5. Kunkel, T. A. (1985) *J. Biol. Chem.* **260,** 5787–5796.
6. Kunkel, T. A. & Alexander, P. S. (1986) *J. Biol. Chem.* **261,** 160–166.
7. Kornberg, A. (1980) *DNA Replication* (Freeman, San Francisco).
8. Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F. & Petersen, G. B. (1982) *J. Mol. Biol.* **162,** 729–773.
9. Sauer, R. T. (1978) *Nature (London)* **276,** 301–302.
10. Jacob, F., Fuerst, C. R. & Wollman, E. L. (1957) *Ann. Inst. Pasteur* **93,** 724–753.
11. Dove, W. F. (1968) *Annu. Rev. Genet.* **2,** 305–340.
12. Lieb, M. (1981) *Mol. Gen. Genet.* **184,** 364–371.
13. Skopek, T. R. & Hutchinson, F. (1982) *J. Mol. Biol.* **159,** 19–33.
14. Tindall, K. R., Stein, J. & Hutchinson, F. (1988) *Genetics* **118,** 551–560.
15. Koch, R. E. & Drake, J. W. (1970) *Genetics* **65,** 379–390.
16. Kim, J.-S. & Davidson, N. (1974) *Virology* **57,** 93–111.
17. Luria, S. E. (1951) *Cold Spring Harbor Symp. Quant. Biol.* **16,** 463–470.
18. Pribnow, D., Sigurdson, D. C., Gold, L., Singer, B. S., Napoli, C., Brosius, J., Dull, T. J. & Noller, H. F. (1981) *J. Mol. Biol.* **149,** 337–376.
19. Sugino, A. & Drake, J. W. (1984) *J. Mol. Biol.* **176,** 239–249.
20. Huang, W. M. (1986) *Nucleic Acids Res.* **14,** 7379–7390.
21. Daegelen, P. & Brody, E. (1990) *Genetics* **125,** 237–248.
22. Drake, J. W. (1966) *J. Bacteriol.* **91,** 1775–1780.
23. Drake, J. W. (1966) *J. Bacteriol.* **92,** 144–147.
24. Meistrich, M. L. & Drake, J. W. (1972) *J. Mol. Biol.* **66,** 107–114.
25. Shinedling, S., Singer, B. S., Gayle, M., Pribnow, D., Jarvis, E., Edgar, B. & Gold, L. (1987) *J. Mol. Biol.* **195,** 471–480.
26. Smith, C. L., Econome, J. G., Schutt, A., Klco, S. & Cantor, C. R. (1987) *Science* **236,** 1448–1453.
27. Farabaugh, P. J. (1978) *Nature (London)* **274,** 765–769.
28. Reznikoff, W. S. & Abelson, J. N. (1978) in *The Operon*, eds. Miller, J. H. & Reznikoff, W. S. (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 221–243.
29. Schaaper, R. M., Danforth, B. N. & Glickman, B. W. (1986) *J. Mol. Biol.* **189,** 273–284.
30. Carlomagno, M. S., Chiariotti, L., Alifano, P., Nappo, A. G. & Bruni, C. B. (1988) *J. Mol. Biol.* **203,** 585–606.
31. Lieb, M. (1951) *Genetics* **36,** 460–477.
32. Hartman, P. E., Hartman, Z. & Stahl, R. C. (1971) *Adv. Genet.* **16,** 1–34.
33. Mortimer, R. K., Schild, D., Contopoulou, C. R. & Kans, J. A. (1989) *Yeast* **5,** 321–403.
34. Petes, T. D. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 410–414.
35. Coleman, K. G., Steensma, H. Y., Kaback, D. B. & Pringle, J. R. (1986) *Mol. Cell. Biol.* **6,** 4516–4525.
36. Steensma, H. Y., Crowley, J. C. & Kaback, D. B. (1987) *Mol. Cell. Biol.* **7,** 410–419.
37. Rose, M., Grisafi, P. & Botstein, D. (1984) *Gene* **29,** 113–124.
38. Lee, G. S.-F., Savage, E. A., Ritzel, R. G. & von Borstel, R. C. (1988) *Mol. Gen. Genet.* **214,** 396–404.
39. Kunz, B. A., Kohalmi, L., Kang, X. & Magnusson, K. A. (1990) *J. Bacteriol.* **172,** 3009–3014.
40. Hoffman, W. (1985) *J. Biol. Chem.* **260,** 11831–11837.
41. Ahmad, M. & Bussey, H. (1986) *Curr. Genet.* **10,** 587–592.
42. Wilke, C. M., Heidler, S. H., Brown, N. & Liebman, S. W. (1989) *Genetics* **123,** 655–665.
43. Magni, G. E., von Borstel, R. C. & Sora, S. (1964) *Mutat. Res.* **1,** 227–230.
44. Magni, G. E. (1964) *J. Cell. Comp. Physiol.* **64,** Suppl. 1, 165–172.
45. Magni, G. E. & Sora, S. (1969) in *Mutation as a Cellular Process*, eds. Wolstenholme, G. E. W. & O'Conner, M. (Churchill, London), pp. 186–198.
46. Gottlieb, D. J. C. & von Borstel, R. C. (1976) *Genetics* **83,** 655–666.
47. Whelan, W. L., Gocke, E. & Manney, T. R. (1979) *Genetics* **91,** 35–51.
48. Paquin, C. & Adams, J. (1983) *Nature (London)* **302,** 495–500.
49. Durán, R. & Gray, P. M. (1989) *Mycotaxon* **36,** 205–219.
50. Orbach, M. J., Vollrath, D., Davis, R. W. & Yanofsky, C. (1988) *Mol. Cell. Biol.* **8,** 1469–1473.
51. Myasnikov, A. N., Plavnik, U. A., Sasnauskas, K. V., Gedminene, G. K., Yanulaitis, A. A. & Smirnov, M. N. (1989) *Yeast* **4,** 348 (abstr.).
52. de Serres, F. J. (1988) *Environ. Mol. Mutagen.* **11,** Suppl. 11, 28–29 (abstr.).
53. de Serres, J. J. & Malling, H. V. (1971) in *Chemical Mutagens: Principles and Methods for Their Detection*, ed. Hollaender, A. (Plenum, New York), Vol. 2, pp. 311–342.
54. Koo, K. & Stuart, W. D. (1991) *Genome* **34,** 644–651.
55. Barnett, V. & Lewis, T. (1978) *Outliers in Statistical Data* (Wiley, New York), 2nd Ed.
56. Drake, J. W. (1969) *Nature (London)* **221,** 1132.
57. Sturtevant, A. H. (1937) *Q. Rev. Biol.* **12,** 464–467.
58. Kimura, M. (1967) *Genet. Res.* **9,** 23–34.
59. Ochman, H. & Wilson, A. C. (1987) *J. Mol. Evol.* **26,** 74–86.
60. Ochman, H. & Wilson, A. C. (1987) in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, ed. Neidhardt, F. C. (Am. Soc. Microbiol., Washington), pp. 1649–1654.
61. Koch, A. L. (1971) *Microb. Physiol.* **6,** 147–217.
62. Cairns, J., Overbaugh, J. & Miller, S. (1988) *Nature (London)* **335,** 142–145.
63. Hall, B. G. (1990) *Genetics* **126,** 5–16.
64. Drake, J. W. (1991) *Annu. Rev. Genet.* **25,** in press.