Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 4171–4175, May 1993
Genetics

Rates of spontaneous mutation among RNA viruses

JOHN W. DRAKE

Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, P.O. Box 12233, Research Triangle Park, NC 27709

Communicated by Evelyn M. Witkin, January 14, 1993

ABSTRACT **Simple methods are presented to estimate rates of spontaneous mutation from mutant frequencies and population parameters in RNA viruses. Published mutant frequencies yield a wide range of mutation rates per genome per replication, mainly because mutational targets have usually been small and, thus, poor samples of the mutability of the average base. Nevertheless, there is a clear central tendency for lytic RNA viruses (bacteriophage Qβ, poliomyelitis, vesicular stomatitis, and influenza A) to display rates of spontaneous mutation of ≈1 per genome per replication. This rate is some 300-fold higher than previously reported for DNA-based microbes. Lytic RNA viruses thus mutate at a rate close to the maximum value compatible with viability. Retroviruses (spleen necrosis, murine leukemia, Rous sarcoma), however, mutate at an average rate about an order of magnitude lower than lytic RNA viruses.**

Mutation rates are expected to evolve toward a balance among forces such as the deleterious consequences of most new mutations, the adaptive consequences of a few, and the costs of reducing mutagenesis (1). Microbes encoding their genomes in DNA appear to have evolved a common spontaneous mutation rate of ≈0.003 per genome per replication (2).

Spontaneous mutation rates are reputed to be much higher among RNA viruses (e.g., ref. 3). However, this notion derives almost exclusively from measurements of mutant frequencies and rates of evolution rather than of mutation rates themselves. Mutant frequencies and rates of mutation accumulation in natural and experimental populations are related to mutation rates by several variables, the most important being selection, history of the population, mechanism of replication, and mutability at each replication step. As a result, mutant frequencies can differ from mutation rates by large factors.

The maximum tolerable deleterious mutation rate cannot be substantially greater than 1 per genome per replication (4) with or without mitigating factors such as high fecundity, large populations, and recombination. Here I estimate that median genomic mutation rates of several lytic RNA viruses do, indeed, occupy the neighborhood of 1. Retroviral mutation rates, however, appear substantially lower.

RATIONALE

The objective of the calculations is to express mutational targets in base pairs; to convert mutant frequencies into mutation rates, taking into account mechanisms of viral reproduction and population parameters; to convert those rates to probabilities of any change in RNA sequence per base per replication; and, finally, to scale up to rates per genome. Four lytic viruses (Qβ, poliomyelitis, vesicular stomatitis, and influenza A) will be considered first, and then three retroviruses (spleen necrosis, murine leukemia, and Rous sarcoma) will be considered.

Mutational Targets. The only useful measures of mutation rates are those that are well defined at the molecular level and thus can be extrapolated to the entire genome. In each experimental situation, the mutational target consists of the number of bases at which changes can be detected phenotypically. Frequently, only one or a few specific base substitutions can be scored. The rate must then be adjusted first to substitutions per base and then to all mutations per base. The first correction is made by assuming that all three possible base substitutions at a site occur equally frequently; although not necessarily true at a particular site, this assumption is likely to be more accurate, on average, than any other simplification. Thus, when only one kind of substitution is scored at a single base, multiplying by 3 converts the rate to total substitutions per base.

Adjustment for failure to detect the many other kinds of mutations can be made when a mutational spectrum is available. Suitable spectra are not yet available in RNA viral systems. For a set of DNA-based microbes, the ratio of total mutations to base substitutions among protein-encoding targets was 1.462 (2). I will therefore use this value when a correction is required. Because this correction is applied at the last stage of the calculations, the tabulated mutation rates can be adjusted easily if required by new results.

The mutational target must also be free of, or correctable for, selection (different rates of growth of mutant and parent) or phenotypic lag (phenotypic masking in RNA viral systems, in which the newly mutant genotype is not immediately expressed, for instance, because of association with parental somatic protein).

Mutation Frequencies and Rates. The measured mutant frequency will be called f regardless of the nature of the target. When several such frequencies are to be pooled, the mean will be called f_{mn}, and the median will be called f_{md}. μ_b is the average rate per bp per replication and is an average over different modes of replication—at least two modes for the lytic viruses and at least three modes for the retroviruses. For a genome of G bases, the genomic mutation rate $\mu_g = G\mu_b$.

Mutation Rates for Lytic Viruses. For purely lytic, single-stranded RNA viruses, infecting strands of sense **a** are copied to produce several strands of opposite sense **b**, which in turn are copied to make progeny strands of sense **a**; in addition, some progeny strands of sense **b** may be copied to produce additional template strands of sense **a**. Thus, there are both linear and geometrical components to virus replication and hence to mutant accumulation.

When a template strand is copied repeatedly, the linear mutation rate due to copying errors will be

$$\mu_{lin} = f_{mn}. \quad [1]$$

(Mutations may also arise randomly in time in template strands, for instance, from metabolically conditioned RNA

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: SNV, spleen necrosis virus; MuLV, Moloney murine leukemia virus; VSV, vesicular stomatitis virus.

damage by methylation, deamination, and oxidation. This random-in-time rate is $\mu_t = 2f_{mn}/c_{mn}$, where c_{mn} is the mean number of copies per lineage. Because this expression cannot yet be evaluated, it will be ignored in what follows.) To μ_{lin} must be added three geometrical components. Two of these components arise from the two steps in which strands of sense **a** are copied into strands of sense **b**, and one arises from multiple rounds of infection. For the typical case of binary geometrical replication, the mutation rate will be

$$\mu_{bin} = -f_{md}/\ln(N\mu_{bin}), \quad [2]$$

where N is the final population size (2); this equation is solved for μ_{bin} by successive approximations.

Selection against a mutant slows its accumulation. When such selection cannot be quantified, calculated mutation rates are underestimates and are so indicated by >. When mutations are lethal between but are not lethal within cycles of infection,

$$\mu_{bin} \approx -f_{md}/\ln B, \quad [3]$$

where B is the estimated burst size.

Because the reproduction of single-stranded RNA viruses is likely to consist of two linear and three geometrical modes and because a small geometrical component can produce a large mutant yield (5, 6), a realistic estimate of mutation rate is probably the mean of these two values, $\mu_m = (\mu_{lin} + \mu_{bin})/2$. (As shown below, this mean is not greatly different from μ_{lin} itself.) μ_m is then converted into the average total mutation rate per bp (μ_b) by multiplying by the correction factors discussed above.

An often superior way to determine microbial mutation rates is the null-class degeneracy of the Luria–Delbrück fluctuation test (5). This method is insensitive to the confounding effects of selection except for total lethality; however, this method is sensitive to phenotypic lag and, thus, in the case of RNA viruses, to phenotypic masking (the association of RNA of one genotype with proteins encoded by another genotype). The method involves initiating parallel cultures with inocula small enough to be mutant-free, allowing growth to proceed until roughly half of the cultures have experienced at least one mutation (as predicted by prior ranging experiments), measuring total population sizes in a representative set of the cultures, and assaying for the presence of any mutants in all cultures. (The last step is easy when the mutants can be selectively plated but otherwise can be tedious.) Assuming that mutational events are randomly distributed among cultures, the fraction of mutant-free cultures is e^{-M}, where M is the average number of events per culture. It is a remarkable and useful property of this approach that, regardless of the molecular steps in replication, very nearly N copying events are required to generate N individuals. Because the mutation rate is the average number of mutations per replication,

$$\mu = M/N. \quad [4]$$

This powerful method could be useful to students of variation in RNA viruses.

Mutation Rates for Retroviruses. Retroviral mutant frequencies have been measured over a single cycle of infection from provirus to provirus. This measurement involves three highly mutable linear steps—namely, one transcription and two reverse-transcriptions; replication of the provirus as a part of the cellular chromosome is far more accurate and is ignored here. Thus,

$$\mu = f/3, \quad [5]$$

from which μ_b is derived by correcting for mutation detection.

CALCULATIONS

Data were selected from systems that appeared free of confounding experimental difficulties (such as phenotypic masking) or circumstances where both "greater than" and "less than" constraints would apply simultaneously without a useful pinch.

Bacteriophage Qβ. G = 4220 (ref. 7 and the references therein). The A → G mutation rate at an extracistronic site was 3.5×10^{-4}, taking selection into account (8, 9). Thus, $\mu_b = 3 \times 1.462 \times 3.5 \times 10^{-4} = 1.54 \times 10^{-3}$, and $\mu_g = 6.48$.

Poliovirus. For the Mahoney type 1 strain used in these experiments, G = 7433 (10). (*i*) A → C was measured at one site in the replicase gene (11). Selection against the mutant was strong, but phenotypic masking was probably absent. Some of many stocks had no mutants at all, and a $P(0)$ calculation is thus possible. The average size of 27 populations was $N = 5.33 \times 10^4$, and 23 were mutant-free, yielding $\mu = 3.01 \times 10^{-6}$ by the null-class method. Thus $\mu_b = 3 \times 1.462 \times 3.01 \times 10^{-6} = 1.32 \times 10^{-5}$, and $\mu_g = 0.0980$. (*ii*) U → C transitions were measured at base 5310 with little or no selection (12). In two experiments with stock volumes of 5 ml, $N = 5.67 \times 10^9$ and 4.2×10^{10}, respectively, $f_{mn} = \mu_{lin} = 3.05 \times 10^{-5}$ and 2.28×10^{-5}, $f_{md} = 3.13 \times 10^{-5}$ and 2.18×10^{-5}, $\mu_{bin} = 3.19 \times 10^{-6}$ and 1.93×10^{-6}, and $\mu_m = 1.68 \times 10^{-5}$ and 1.23×10^{-5}. Thus $\mu_b = 3 \times 1.462 \times (1.68 + 1.23) \times 10^{-5} \div 2 = 6.40 \times 10^{-5}$, and $\mu_g = 0.475$. (*iii*) Mutation was measured from guanidine dependence to guanidine resistance (13). The system was free of selection and phenotypic masking. The mutations consisted of A → G at the first position of a codon and G → N at the third position; thus, they scored one and one-third base substitutions. (*a*) Among six stocks obtained by resuspending plaques, three with the lowest titers were mutant-free, whereas three others had mutants in numbers that increased with stock titer. The transition from no-mutants to mutants-present was most likely to have occurred between stocks for which the average virus contents equaled the reciprocal of the mutation rate. Thus, for transition stocks containing 0.85×10^4 and 1.3×10^4 particles, $\mu_b = (1.462 \times 2)/[(0.85 + 1.3) \times 10^4 \times 1.333] = 1.02 \times 10^{-4}$. (*b*) Six independent 4-ml stocks contained a mean $N = 4.92 \times 10^9$ plaque-forming units, $f_{mn} = \mu_{lin} = 1.11 \times 10^{-4}$, $f_{md} = 1.11 \times 10^{-4}$, $\mu_{bin} = 1.07 \times 10^{-5}$, $\mu_m = 6.11 \times 10^{-5}$, and $\mu_b = 1.462 \times 6.11 \times 10^{-5} \div 1.333 = 6.70 \times 10^{-5}$. (*c*) Eighteen independent 4-ml stocks contained a mean $N = 1.48 \times 10^{10}$ plaque-forming units, $f_{mn} = \mu_{lin} = 5.37 \times 10^{-4}$, $f_{md} = 5.57 \times 10^{-4}$, $\mu_{bin} = 4.17 \times 10^{-5}$, $\mu_m = 2.89 \times 10^{-4}$, and $\mu_b = 1.462 \times 2.89 \times 10^{-4} \div 1.333 = 3.17 \times 10^{-4}$. (*d*) From these three measurements, the mean $\mu_b = 1.62 \times 10^{-4}$, and $\mu_g = 1.21$. (*iv*) Nearly all mutations at eight guanine residues were measured using RNase T1 digestion (14), detecting $f_{mn} = \mu_{lin} = 3.63 \times 10^{-3}$ and $f_{md} = 3.65 \times 10^{-3}$ mutations per site. Because frequencies varied by <2-fold at all sites and mutations at some sites were expected to be lethal, lethality can be inferred at all eight sites. Therefore, only the last round of infection could have contributed detectable mutants. Assuming an increment of $B = 50$ in this round, $\mu_{bin} = 0.4343 \times 3.65 \times 10^{-3} \div \log(50) = 9.33 \times 10^{-4}$, $\mu_m = \mu_b = 2.28 \times 10^{-3}$, and $\mu_g = 17.0$. (*v*) The means of all four poliovirus values are $\mu_b = 6.30 \times 10^{-4}$ and $\mu_g = 4.68$. The medians are $\mu_b = 1.13 \times 10^{-4}$ and $\mu_g = 0.843$.

Vesicular Stomatitis Virus (VSV). G = 11,162 (15). (*i*) G → A frequencies were measured at two sites under conditions where phenotypic mixing and selection were negligible (16). (*a*) For eight high-titer stocks, the mean $N = 4.15 \times 10^{11}$ (J. J. Holland, personal communication), $f_{mn} = \mu_{lin} = 1.75 \times 10^{-4}$, and $f_{md} = 1.6 \times 10^{-4}$. Thus, $\mu_{bin} = 1.05 \times 10^{-5}$, and $\mu_m = 9.25$

$\times 10^{-5}$. (*b*) For six low-titer stocks, the mean $N = 3.70 \times 10^7$ (J. J. Holland, personal communication), $f_{mn} = \mu_{lin} = 2.35 \times 10^{-4}$, $f_{md} = 2.1 \times 10^{-4}$, $\mu_{bin} = 3.00 \times 10^{-5}$, and $\mu_m = 1.32 \times 10^{-4}$. (*c*) Using the mean value of μ_m, $\mu_b = 1.462 \times (9.25 + 13.2) \times 10^{-5}/(2 \times 2/3) = 2.47 \times 10^{-4}$, and $\mu_g = 2.75$. (*ii*) Nearly all kinds of mutations arising at a guanine residue were measured by using RNase T1 cleavage (17); $f_{mn} = \mu_{lin} = 6.14 \times 10^{-4}$, and $f_{md} = 6.0 \times 10^{-4}$. Because the mutations are expected to be lethal, only the last round of infection could have contributed detectable mutants. Taking $B = 50$, $\mu_{bin} = 0.4343 \times 6 \times 10^{-4} \div \log(50) = 1.53 \times 10^{-4}$, $\mu_m = \mu_b = 3.83 \times 10^{-4}$, and $\mu_g = 4.28$. (*iii*) The means of the two VSV values are $\mu_b = 3.15 \times 10^{-4}$ and $\mu_g = 3.52$.

Influenza A Virus. $G = 13{,}588$ (18). An average of 849 bases was sequenced in each of 108 clonal copies of a gene encoding a nonstructural protein (19). The population size was 1.2×10^6 plaque-forming units, and 7 base substitutions were seen among 91,708 bases, so that $f = \mu_{lin} = 7.63 \times 10^{-5}$, $\mu_{bin} = 2.30 \times 10^{-5}$, and $\mu_m = 4.97 \times 10^{-5}$. Strong selection against many mutations and complete selection against mutations other than base substitutions were likely. Thus $\mu_b > 1.462\mu_m = 7.26 \times 10^{-5}$, and $\mu_g > 0.987$.

Spleen Necrosis Virus (SNV). $G = 7800$ (H. M. Temin, personal communication). (*i*) A 288-base *lacZα* target inserted into SNV was used to sequence selectively neutral mutations arising during the three mutable replications of one provirus-to-provirus cycle with negligible selection (20). Although a mutational spectrum resulted, it cannot be compared directly with other spectra based on the *lacZα* target to determine an efficiency of mutant detection because of uncertainty about the fraction of leaky mutants detected in the SNV system. Treating two complex "hypermutations" as single mutants, $f = 2.19 \times 10^{-3}$. Of the 37 mutants sequenced, 11 were base substitutions, and two of these were nonsense mutations; scaling up from the nonsense mutations (2), the conversion factor for total mutations is $[26 + (2 \times 64)/3]/37 = 1.856$. Thus $\mu_b = (1.856 \times 2.19 \times 10^{-3})/(3 \times 288) = 4.71 \times 10^{-6}$, and $\mu_g = 0.0368$. (*ii*) The reversion of an amber mutation was monitored in a selectively neutral target during one provirus-to-provirus cycle (21). Of 17 revertants sequenced, 15 were UAG → UGG and 2 were extracodonic, so that the mutation target was $1/3 \times 15/17 = 0.294$ base. The average of three revertant frequencies was $f = 2.18 \times 10^{-5}$. Thus $\mu_b = (1.462 \times 2.18 \times 10^{-5})/(3 \times 0.294) = 3.61 \times 10^{-5}$, and $\mu_g = 0.281$. (*iii*) The means of the two SNV values are $\mu_b = 2.04 \times 10^{-5}$ and $\mu_g = 0.159$.

Moloney Murine Leukemia Virus (MuLV). (*i*) Using a MuLV vector, the reversion of an amber mutation was measured in a selectively neutral *neo* insert in one provirus-to-provirus cycle (22). All mutations were UAG to the wild-type UGG, so that the target was one-third base. The observed frequency of $f = 4.37 \times 10^{-6}$ must be divided by a relative plating efficiency of $E = 3.2$. This frequency must be further divided by the average number n of provirus targets; of 45 revertant clones, 14 had one, and 31 had multiple proviruses; assuming random infection, the Poisson distribution predicts that the average number of proviruses per infected cell was $n = 2.65$. Of 14 clones, 7 had no mutation, yielding a divisor of 2. Thus $\mu_b = 3 \times 1.462f/(2 \times 3)En = 3.77 \times 10^{-7}$. $G = 8332$ (23), so that $\mu_g = 0.00314$. (*ii*) Both RNase T1 fingerprinting and direct sequencing were used to score mutations arising in a target of 1380 bases in a single provirus-to-provirus cycle of the AKR 2A strain, detecting three mutations among 151,000 bases (24). Because of both imperfect detection and some selection, $\mu_b > f/3 = 6.62 \times 10^{-6}$. $G \approx 8372.5$ (25, 26), so that $\mu_g > 0.0554$. (*iii*) The means of the two MuLV values are $\mu_b > 3.50 \times 10^{-6}$ and $\mu_g > 0.0293$.

Rous Sarcoma Virus. $G = 9312$ (27). Mutations were measured in a 1125-base target in a single provirus-to-provirus cycle with denaturing gradient gels (28). Among 58 samples, nine mutants were found among the 65,250 nt screened, and selection was probably weak or absent. Thus $f = 1.38 \times 10^{-4}$, $\mu_b = f/3 = 4.60 \times 10^{-5}$, and $\mu_g = 0.428$.

TABULATION AND DISCUSSION

The results of the calculations appear in Table 1. Different μ_g values for a given virus vary widely, the extreme being from 0.1 to 17 for poliovirus. This result is expected from the frequent use of very small mutation targets, often as small as a single transition at a single base; in DNA-based microbes, the mutability of individual sites can vary by several orders of magnitude (29, 30). For poliovirus, the smallest value is based upon a single transversion pathway, and transversions may occur less frequently than transitions (31); on the other hand, the largest value is clearly incompatible with viability.

Both median and mean values for these rates are listed in Table 2, severely rounded to avoid overstating the accuracy of the underlying values. Those mutation rates in Table 1, which were obtained by physical methods such as RNase T1 fingerprinting or gradient gel denaturation (the final poliovirus, VSV, and MuLV values and the single Rous sarcoma virus value), were consistently higher than rates obtained by selective plating or by sequencing, suggesting a source of systematic overestimation, but their removal has little effect upon the values in Table 2.

Further discussion is conditioned on several assumptions: that the values are accurately measured and free of unrecognized artifact and that the values are representative for the virus and not overly influenced by choice of mutation target, choice of virus strain, or unrecognized modification of the mutation rate by genomic engineering or by growth in a laboratory setting. It should also be noted that all values are averages over different error sources: mutations arising in templates, **a** → **b** versus **b** → **a** copying errors, and the three quite different mutable steps of retrovirus replication.

Two facets of the compilation are immediately obvious. (*i*) The mutation rates of the lytic viruses are close to the maximum tolerable rate. For haploid organisms with small, conservative genomes, probably only a minority (perhaps only a small minority) of mutations are neutral. Thus, in addition to the low intrinsic infectivities characteristic of most RNA viruses, values of μ_g from 2 to 4 imply a further e^{-2} (7-fold) to e^{-4} (50-fold) reduction in the fraction of mutationally undamaged particles. (*ii*) The retroviruses ap-

Table 1. Spontaneous mutation rates among RNA viruses

Virus	Genome size, kb	Mutational target, base	Mutation rate per replication	
			μ_b	μ_g
Lytic virus				
Qβ	4.2	0.3	1.5×10^{-3}	6.5
Polio	7.4	0.3	1.3×10^{-5}	0.098
		0.3	6.4×10^{-5}	0.48
		1.3	1.6×10^{-4}	1.2
		8	2.3×10^{-3}	17
VSV	11.2	0.7	2.5×10^{-4}	2.8
		0.3	3.8×10^{-4}	4.3
Flu A	13.6	849	$>7.3 \times 10^{-5}$	>0.99
Retrovirus				
SNV	7.8	288	4.7×10^{-6}	0.037
		0.3	3.6×10^{-5}	0.28
MuLV	8.3	0.3	3.8×10^{-7}	0.0031
	8.4	1380	$>6.6 \times 10^{-6}$	>0.055
RSV	9.3	1125	4.6×10^{-5}	0.43

Abbreviations: Flu A, influenza A virus; RSV, Rous sarcoma virus; μ_b, average mutation rate per bp; μ_g, mutation rate per genome.

Table 2. Median and mean genomic mutation rates

Category	Set of rates	μ_g	
		Median	Mean
All RNA viruses	All 13 rates	0.5	3
	Pregrouped (7)	0.8	2
Lytic viruses	All 8 rates	2	4
	Pregrouped (4)	2	4
Retroviruses	All 5 rates	0.06	0.2
	Pregrouped (3)	0.2	0.2

Values are derived from Table 1, but > designations are omitted, and all values have been rounded to a single digit. Pregrouped means that medians or means were first determined for the several rates for a given virus and were then used for a grand median or mean.

pear no more mutable but, instead, appear roughly an order of magnitude less mutable, than the lytic viruses. Although no mutation rate is available for human immunodeficiency virus 1, available evidence suggests that even this retrovirus is not exceptionally mutable compared with the lytic RNA viruses. Assuming that the ratio of accuracies of different reverse transcriptases *in vitro* resembles the ratio of the corresponding viral mutation rates, and noting that the human immunodeficiency virus 1 enzyme is ≈14-fold less accurate than the MuLV enzyme (32), the MuLV mutation rate (Table 1) would increase to $\approx\mu_g = 0.044$ for human immunodeficiency virus 1, a value still typical of the other retroviruses.

Differences between μ_{lin} and μ_{bin} for lytic viruses become large when population sizes become much larger than $1/\mu$. The ratio μ_{lin}/μ_{bin} was 10.6, 11.6, and 3.9 for the last three poliovirus values, was 12.3 and 4.0 for the two VSV values, was 3.3 for influenza A virus, and was nonexistent elsewhere. If there were no geometrical component to mutant accumulation in lytic RNA viruses, values of μ_g would be a little less than 2-fold larger than estimated here. The result would be a median value of $\mu_g = 3.7$ for the lytic viruses, a value poorly consistent with survival ($e^{-3.7} = 0.025$).

Three of the μ_g values in Table 1 (17, 6.5, and 4.3) already appear inconsistent with survival, but there is no reason to reject the hypotheses that the corresponding three lowest μ_g values are equally misleading and that the median best represents these RNA viruses.

There is no evident association between higher μ_g values and assortment of fragmented genomes or recombination potential, the highly fragmented influenza A virus having a near-median μ_g and the similarly recombinogenic retroviruses having relatively low μ_g values. Thus, genetic exchanges do not seem to be used to mitigate the effects of high mutation rates (see ref. 33).

Among DNA-based microbes, μ_g remains constant over a 6500-fold range of genome sizes, and there is a strong inverse relation between μ_b and genome size (2). Among these RNA viruses (and treating > values as =), $\log \mu_g = -0.036 - 0.019G$, $R^2 = 0.005$, and there is no significant departure from a slope of 0 ($P = 0.88$), reinforcing the conclusion that genomic mutation rates cluster around unity. If μ_g were a constant such as 1, then the slope of $\log \mu_b$ versus $\log G$ would be −1. The observed relation is $\log \mu_b = 3.9 - 2.0 \log G$, $R^2 = 0.151$, and there is no significant departure from a slope of −1 ($P = 0.65$). Clearly, the scatter of these data is too great to provide a meaningful slope.

The contrast between the worlds of DNA and RNA becomes particularly striking when the bacteriophages M13 and Qβ are considered. Both genomes are single-stranded and similar in size (6.4 and 4.2 kb), and they savor the same host bacterium, *Escherichia coli*, but RNA-based Qβ has a spontaneous genomic mutation rate roughly 10^3-fold higher than that of DNA-based M13 (and λ, T2, and T4). Kimura (1) postulated that mutation rates would evolve to a balance mainly determined by the cost of deleterious mutations versus the cost of further reducing mutation rates. How do RNA and DNA genomes differ in this respect? The DNA bacteriophages benefit from two antimutagenic mechanisms apparently denied these RNA viruses, proofreading and mismatch repair, and either might suffice to bridge the mutation-rate gap. However, while M13 derives proofreading and perhaps some mismatch repair from host-encoded enzymes, RNA-targeted host counterparts do not appear to exist. Thus, RNA viruses would have to acquire several host genes and adapt them to RNA substrates to achieve a major reduction in spontaneous mutation rates. The result would be a substantial increase in genome size. Because the RNA phosphodiester bond is labile to hydrolysis promoted by the ribose 2′-OH group and because anecdotal evidence suggests that it is difficult to recover unbroken genomes from RNA viruses, the intrinsic lability of RNA appears to put a high cost on genome enlargement. Genomic fragmentation (as in influenza viruses) might overcome this constraint but only at the cost of ensuring accurate assortment of fragments among virus particles.

A kinetic analysis of rates of dNTP misinsertion and of extension from dNMP·rNMP mispairs by *E. coli* polymerase I using an RNA template demonstrated extraordinary accuracy even without proofreading (34). This result suggests that an RNA template need not be intrinsically error-prone and that an accuracy much higher than RNA-virus mutation rates can be achieved without proofreading. Thus, high RNA-virus mutation rates may reflect an evolutionary strategy rather than an insoluble mechanical problem.

A consequence of being a mutational limit organism is intolerance to even small increases in mutation rate. Holland *et al.* (35) observed that none of several mutagens could increase VSV or poliovirus mutation rates more than ≈2.5-fold (whereupon viability declined markedly) and sensibly suggested that reports of other, larger values were instances where multiple mutations were required to achieve the selected phenotype. A 13-fold-induced mutagenicity in the SNV retroviral system decreased the virus titer ≈20-fold (36); the induced per-genome mutation rate would have been ≈0.5 mutation per replication or ≈1.4 mutations per growth cycle. These results and Tables 1 and 2 suggest the interesting pharmacological possibility that RNA-specific mutagens might be substantially more harmful to RNA viruses than to host cells (e.g., ref. 37).

In DNA-based microbes, where strong mutator mutations are easily obtained, the situation is similar. In *E. coli*, simultaneously inactivating both proofreading and mismatch repair raises genomic mutation rates ≈3800-fold in the *lacI* gene and substantially reduces the fraction of live cells in cultures (R. M. Schaaper, personal communication). Multiplying by the genomic spontaneous mutation rate of ≈0.003 (2) suggests a rate of ≈11 per replication in the mutator strain. That the cultures are even modestly viable reflects the use of a rich medium that would support the growth of most auxotrophic mutants, the fact that most of the mutations are transitions which have the least average effect among point mutations, and probable strong selection during the growth of the culture. In *Saccharomyces cerevisiae*, simultaneously inactivating proofreading by polymerase δ and mismatch repair increases the diploid mutation rate $\approx 2 \times 10^4$-fold to ≈60 per generation and produces inviable haploids (38). Simultaneously inactivating proofreading by polymerase ε and mismatch repair increases the haploid mutation rate ≈500-fold to ≈2 per replication and produces slowly growing haploids (39).

The measurement of RNA viral mutation rates could be considerably facilitated by adopting the null-class method (described above) for lytic viruses and by using reporter

genes, as in the example of SNV (20), instead of tiny and thus potentially nonrepresentative mutational targets. It will be interesting to obtain mutation rates for the double-stranded RNA viruses, as well as general error rates of prokaryotic and eukaryotic transcription. It will also be interesting to determine whether the analysis of mutant clone size distributions can provide insights into the underlying mechanism of RNA viral replication, as it did for DNA replication (6), and can resolve differential mutability at the various replication steps.

I am grateful to Howard Temin for introducing me to the literature of RNA-virus variation, to Esteban Domingo for assuring me that someone else might be interested in this story, and to Walter Piegorsch for assistance with statistical calculations. Jim Crow, Esteban Domingo, Roger Milkman, Roel Schaaper, and Howard Temin provided helpful criticisms of earlier versions of this article.

1. Kimura, M. (1967) *Genet. Res.* **9**, 23–34.
2. Drake, J. W. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 7160–7164.
3. Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S. & VandePol, S. (1982) *Science* **215**, 1577–1585.
4. Eigen, M. & Schuster, P. (1977) *Naturwissenschaften* **64**, 541–565.
5. Luria, S. E. & Delbrück, M. (1943) *Genetics* **28**, 491–511.
6. Luria, S. E. (1951) *Cold Spring Harbor Symp. Quant. Biol.* **16**, 463–470.
7. Mills, D. R., Priano, C., Merz, P. A. & Binderow, B. D. (1990) *J. Virol.* **64**, 3872–3881.
8. Domingo, E., Flavell, R. A. & Weissmann, C. (1976) *Gene* **1**, 3–25.
9. Batschelet, E., Domingo, E. & Weissmann, C. (1976) *Gene* **1**, 27–32.
10. Kitamura, N., Semler, B., Rothberg, P. G., Larsen, G. R., Adler, C. J., Dorner, A. J., Emini, E. A., Hanecak, R., Lee, J. J., van der Werf, S., Anderson, C. W. & Wimmer, E. (1981) *Nature (London)* **291**, 547–553.
11. Sedivy, J. M., Capone, J. P., RajBhandary, U. L. & Sharp, P. A. (1987) *Cell* **50**, 379–389.
12. de la Torre, J. C., Giachetti, C., Semler, B. L. & Holland, J. J. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 2531–2535.
13. de la Torre, J. C., Wimmer, E. & Holland, J. J. (1990) *J. Virol.* **64**, 664–671.
14. Ward, C. D. & Flanegan, J. B. (1992) *J. Virol.* **66**, 3784–3793.
15. Schubert, M., Harmison, G. & Meier, E. (1984) *J. Virol.* **51**, 505–514.
16. Holland, J. J., de la Torre, J. C., Steinhauer, D. A., Clarke, D., Duarte, E. & Domingo, E. (1989) *J. Virol.* **63**, 5030–5036.
17. Steinhauer, D. A., de la Torre, J. C. & Holland, J. J. (1989) *J. Virol.* **63**, 2063–2071.
18. Murphy, B. R. & Webster, R. G. (1990) in *Virology*, eds. Fields, B. N. & Knipe, D. M. (Raven, New York), pp. 1091–1152.
19. Parvin, J. D., Moscona, A., Pan, W. T., Leider, J. M. & Palese, P. (1986) *J. Virol.* **59**, 377–383.
20. Pathak, V. K. & Temin, H. M. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6019–6023.
21. Dougherty, J. P. & Temin, H. M. (1988) *J. Virol.* **62**, 2817–2822.
22. Varela-Echavarría, A., Garvey, N., Preston, B. D. & Dougherty, J. P. (1992) *J. Biol. Chem.* **267**, 24681–24688.
23. Shinnick, T. M., Lerner, R. A. & Sutcliffe, J. G. (1981) *Nature (London)* **293**, 543–548.
24. Monk, R. J., Malik, F. G., Stokesberry, D. & Evans, L. H. (1992) *J. Virol.* **66**, 3683–3689.
25. Etzerodt, M., Mikkelsen, T., Pedersen, F. S., Kjeldgaard, N. O. & Jørgensen, P. (1984) *Virology* **134**, 196–207.
26. Herr, W. (1984) *J. Virol.* **49**, 471–478.
27. Schwartz, D. E., Tizard, R. & Gilbert, W. (1983) *Cell* **32**, 853–869.
28. Leider, J. M., Palese, P. & Smith, F. I. (1988) *J. Virol.* **62**, 3084–3091.
29. Benzer, S. (1955) *Proc. Natl. Acad. Sci. USA* **41**, 344–354.
30. Ronen, A. & Rahat, A. (1976) *Mutat. Res.* **34**, 21–34.
31. Kuge, S., Kawamura, N. & Nomoto, A. (1989) *J. Mol. Biol.* **207**, 175–182.
32. Roberts, J. D., Bebenek, K. & Kunkel, T. A. (1988) *Science* **242**, 1171–1173.
33. Chao, L. (1990) *Nature (London)* **348**, 454–455.
34. Ricchetti, M. & Buc, H. (1993) *EMBO J.* **12**, 387–396.
35. Holland, J. J., Domingo, E., de la Torre, J. C. & Steinhauer, D. A. (1990) *J. Virol.* **64**, 3960–3962.
36. Pathak, V. K. & Temin, H. M. (1992) *J. Virol.* **66**, 3093–3100.
37. Preston, B. D., Poiesz, B. J. & Loeb, L. A. (1988) *Science* **242**, 1168–1171.
38. Morrison, A., Johnson, A. L., Johnston, L. H. & Sugino, A. (1993) *EMBO J.* **12**, 1467–1473.
39. Morrison, A. & Sugino, A. (1992) *Chromosoma* **102**, S146–S148.