

Mutation rates among RNA viruses

John W. Drake*[†] and John J. Holland*

*Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709-2233; and [†]Institute for Molecular Genetics and Department of Biology, University of California at San Diego, La Jolla, CA 92093

Edited by Walter M. Fitch, University of California, Irvine, CA, and approved October 5, 1999 (received for review August 18, 1999)

The rate of spontaneous mutation is a key parameter in modeling the genetic structure and evolution of populations. The impact of the accumulated load of mutations and the consequences of increasing the mutation rate are important in assessing the genetic health of populations. Mutation frequencies are among the more directly measurable population parameters, although the information needed to convert them into mutation rates is often lacking. A previous analysis of mutation rates in RNA viruses (specifically in riboviruses rather than retroviruses) was constrained by the quality and quantity of available measurements and by the lack of a specific theoretical framework for converting mutation frequencies into mutation rates in this group of organisms. Here, we describe a simple relation between ribovirus mutation frequencies and mutation rates, apply it to the best (albeit far from satisfactory) available data, and observe a central value for the mutation rate per genome per replication of $\mu_g \approx 0.76$. (The rate per round of cell infection is twice this value or about 1.5.) This value is so large, and ribovirus genomes are so informationally dense, that even a modest increase extinguishes the population.

Rates of spontaneous mutation are critical to understanding the genetic structure of populations over time and thus to understanding the course of evolution. Mutation provides the prime variation on which selection, recombination, and genetic drift operate. Recombination and its enabling partner, sex, probably persist primarily because of the deleterious consequences of mutation (1). In a species capable of contemplating its own genetic health, the products of mutation are viewed as major threats to public health, and the consequences of increasing the mutation rate are viewed with alarm (2).

When rates of spontaneous mutation are expressed per genome per genome replication, different broad groups of organisms display characteristic values (3): roughly 0.2 for retroelements; close to 0.0034 for DNA-based microbes (including both viral and cellular organisms); and roughly 0.01 for higher eukaryotes. Riboviruses (RNA viruses exclusive of retroviruses) tend to display very high mutation rates (4); however, quantifying those rates has proved difficult, because the mutational targets have been too small to sample the genome reliably and because it was uncertain how to combine mutation frequencies and population history to calculate mutation rates (5).

Here, we describe a robust relationship among the mutation frequency f , the easily determined growth parameter c , and the mutation rate μ . Applying this formulation to the available data provides the best estimate to date of the rate of spontaneous mutation in riboviruses. In addition, the formulation also suggests ways to measure hitherto opaque parameters of viral replication.

Theory

The replication of riboviruses whose chromosomes are composed of single-stranded RNA follows a simple scheme (Fig. 1). A cell is infected with one (or more) virus particles. Each infecting genome is copied iteratively such that complementary strands accumulate. Subsequently, the complementary strands are themselves copied iteratively, producing final strands of the same polarity as the infecting strand, and these final strands are packaged and released. We make the reasonable assumption that "final" strands rarely or never reenter the beginning of the

cycle within a single infection. The key attribute of iterative replication is that the mutation frequency f equals the mutation rate μ per copying event: if n complementary strands are copied from a template and if μ is the mutation rate per copying event, then the number of mutations will be $n\mu$, and $f = n\mu/n = \mu$.

When a ribovirus infects a cell, the population of complementary strands accumulating from the first round of repeated copying has $f_1 = \mu_1$. Ignoring rare precise back mutations and in the absence of selection, this mutant subpopulation will persist at the frequency f_1 . In the second round of repeated copying, a new fraction of mutations will be contributed, amounting to $f_2 = \mu_2$. Thus, in one cycle of cell infection, $f = f_1 + f_2 = \mu_1 + \mu_2$. It is presently unknown whether μ_1 and μ_2 differ, and we must combine them into $f = 2\mu$. This process of linear accumulation will continue as a virus stock grows through c cycles of cell infection; c will generally be a small (but not necessarily an integer) number, from one for a single round of infection (high multiplicity of infection), to perhaps four or five for a large population of host cells and a small inoculum. For stocks grown to a size sufficient to accumulate numerous mutations, the general formulation is $f = 2c\mu$. Note that this expression holds only in the absence of selection, which is likely to be weak or absent during the first cycle of infection but may become strong in subsequent cycles. The mutational targets used in the studies we cite are believed to be largely free of selection. Given specific information about selection coefficients, it is possible to describe their effects on mutation accumulation over multiple cycles.

An alternative approach, used frequently with DNA-based microbes and occasionally with riboviruses, is the null-class method (6). In this method, numerous parallel cultures are each seeded with a small amount of virus and grown to a population size such that only roughly half of the cultures have accumulated any mutants. Each culture is then screened simply for the presence or absence of mutants, and the average number N of virus particles per culture is also determined. On the assumption that mutational events are randomly distributed among replication events and because the number of replication events per culture is close to N , the proportion of cultures with no mutants $P(0) = e^{-N\mu}$.

Calculations

General Method. By using either the accumulation method or the null-class method described above, a published mutation frequency f for a particular trait can be converted into a mutation rate μ . A mutation rate μ_b per base can be obtained by dividing μ by the mutational target size T (the number of bases at which the event can occur) and multiplying by a correction factor for mutations other than base substitutions. For mutations studied in riboviruses, T is usually very small; for instance, if the mutations consist exclusively of a single base substitution (such as G \rightarrow A) at a single site, then $T = 1/3$. (Strictly speaking, T is

This paper was submitted directly (Track II) to the PNAS office.

Abbreviation: VSV, vesicular stomatitis virus.

[†]To whom reprint requests should be addressed. E-mail: drake@niehs.nih.gov.

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.

templates that is characteristic of these organisms. The mutational consequences of iterative replication have not been explored much in the several decades since the primordial analysis of a “stamping machine” model (22). An exception is the single-stranded DNA bacteriophage ϕ X174. This phage replicates in a complicated but fundamentally linear manner (23, 24), such that the mutation equations (25, 26) are similar to the equation used here for riboviruses, although the observed rates are far lower than those of riboviruses and are instead characteristic of DNA-based microbes (3). However, uncertainty persists concerning the extent to which second-generation strands reenter the beginning of the cycle within a single ϕ X174 infection (24).

Applying mainly mutation-accumulation or null-class analyses to published data yields a set of nine rates of mutation per genome per replication that vary fairly smoothly over about an order of magnitude (Table 1). Just such variation is expected from the tiny mutational target sizes characteristic of the cited measurements: 1/3 base for seven entries, 2/3 base for two entries, and roughly 7.5 bases for the measles-virus entry. Most mutational spectra display a wide range of site-specific mutabilities, and even a specific substitution such as G:C \rightarrow A:T in DNA can vary by over 2,000-fold depending on the neighboring sequences (27). The smooth spread of values in Table 1 and the similarity of mean and median lead us to propose the median (0.76) as the best current value for the rate of spontaneous mutation in riboviruses generally. A similar value, 1.2, was recently obtained by applying the fundamentally different Bateman–Mukai analysis to data obtained with VSV (28). These values are very high in comparison to values in other organisms but were anticipated in classical studies of mutation in an RNA phage (15, 16) and have long been known to be generally high (29).

Bacteriophage ϕ 6 has a segmented, double-stranded RNA genome and thus differs profoundly from the viruses in Table 1. Nevertheless, the ϕ 6 mutation rate must be much higher than that of DNA-based microbes, because stocks accumulate about 0.5% temperature-sensitive mutations (30). Nonsense mutations in ϕ 6 display revertant frequencies $\geq 10^{-4}$ in stocks where $c \approx 5$ (L. Chao, personal communication). For ϕ 6, $G = 13,379$ (31). Assuming negligible selection and $T = 1.5$, $\mu_g \geq 0.13$, a value compatible with the values shown in Table 1. Thus, the high ribovirus mutation rate seems to encompass both animal viruses and phages.

The viral particles emerging from individual infected cells contain occasional mutant clones descended from new mutations. Mutation in riboviruses should produce mutant clone size distributions very different from those characteristic of exponentially replicating chromosomes. In riboviruses, fewer mutations will occur in the first round of genome copying than in the second round, because fewer copying events occur in the first round. Therefore, a few clones will contain several mutants, whereas most clones will contain only one mutant. This predic-

tion has already been verified for the single-stranded DNA phage ϕ X174, where roughly 5 large clones and about 309 single mutants (clones of size 1) were observed; the uncertainties reflect probable coincidences of clones of size 1 and a few preexisting mutations producing large clones (23). In riboviruses, the mutant clone size distribution can be described in terms of the mutation rates characteristic of the first and second rounds of copying (μ_1 and μ_2 , respectively) and the mean numbers of genomes produced per template in the first and second rounds (n_1 and n_2 , respectively). Specifically, the first round contributes $\mu_1 n_1$ mutant clones of mean size n_2 ; the second round contributes $\mu_2 n_2$ mutant clones of size 1; and the total virus progeny per cell is $n_1 n_2$. These are experimentally resolvable parameters.

Because a single cycle of infection produces a mutant frequency twice that of the rate per genome replication, few progeny viruses escape mutation. In general, for 1.5c mutations distributed randomly among genomes, the mutation-free fraction of a population would be $e^{-1.5c}$ (0.22 for $c = 1$; 0.05 for $c = 2$; etc.). However, many or most mutations in riboviruses are deleterious, and selection against deleterious mutations will reduce both mutant frequency and virus yield even within a single cycle of infection. Their high mutation rate renders the riboviruses particularly vulnerable to the consequences of rate increases, and both poliovirus and VSV populations are extinguished by chemical mutagenesis sufficient to increase the rate by a mere 2.5-fold (32). Merely tripling the viral mutation rate, if this increase could be achieved without harming the host, could cure ribovirus infections. Conversely, the normal mutation rate does not seem to be limiting for adaptation (29, 33).

Another consequence of this high rate of mutation is that all but the most mild mutator mutations will be lethal. Mutation pressure alone is likely to generate many mutator mutations, and an analysis of the VSV polymerase gene, encompassing almost 0.6 of the genome, suggested considerable polymerase compositional variation within a population (34). Influenza virus populations seem to accumulate a high frequency ($\approx 12\%$) of mutators, but these increase mutation rates only by 2- to 4-fold (35); stronger mutators probably die out rapidly. The strength of a VSV mutator (36) cannot be quantitated.

Because recombination can regenerate unmutated genomes, it can reduce the danger of a very high mutation rate. Crossing over is frequent among many riboviruses, and independent assortment of segmented genomes also occurs (30). Perhaps most importantly, however, riboviruses have high fecundity, with yields of 100–10,000 per infected cell, such that the product of yield times the proportion of mutation-free progeny tends to exceed unity.

We thank Lin Chao, Jim Crow, Dave Denhardt, Esteban Domingo, Dmitry Gordenin, Mike Resnick, Roel Schaaper, and Charlie Steinberg for their critical comments on evolving stages of this article.

- Barton, N. H. & Charlesworth, B. (1998) *Science* **281**, 1986–1990.
- Muller, H. J. (1950) *Am. J. Hum. Genet.* **2**, 111–176.
- Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. (1998) *Genetics* **148**, 1667–1686.
- Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S. & VandePol, S. (1982) *Science* **215**, 1577–1585.
- Drake, J. W. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 4171–4175.
- Luria, S. E. & Delbrück, M. (1943) *Genetics* **28**, 491–511.
- Drake, J. W. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 7160–7164.
- de la Torre, J. C., Giachetti, C., Semler, B. L. & Holland, J. J. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 2531–2535.
- de la Torre, J. C., Wimmer, E. & Holland, J. J. (1990) *J. Virol.* **64**, 664–671.
- Holland, J. J., de la Torre, J. C., Steinhauer, D. A., Clarke, D., Duarte, E. & Domingo, E. (1989) *J. Virol.* **63**, 5030–5036.
- Wang, W., Lee, W.-M., Mosser, A. G. & Rueckert, R. R. (1998) *J. Virol.* **72**, 1210–1218.
- Lee, W.-M., Wang, W. & Rueckert, R. R. (1995) *Virus Genes* **9**, 177–181.
- Schrag, S. J., Rota, P. A. & Bellini, W. J. (1999) *J. Virol.* **73**, 51–54.
- Griffin, D. E. & Bellini, W. J. (1996) in *Field's Virology*, eds. Field, B. N., Knipe, D. M. & Howley, P. M. (Lippincott–Raven, Philadelphia), 3rd Ed., pp. 1267–1312.
- Domingo, E., Flavell, R. A. & Weissmann, C. (1976) *Gene* **1**, 3–25.
- Batschelet, E., Domingo, E. & Weissmann, C. (1976) *Gene* **1**, 27–32.
- Steinhauer, D. A., de la Torre, J. C. & Holland, J. J. (1989) *J. Virol.* **63**, 2063–2071.
- Ward, C. D. & Flanagan, J. B. (1992) *J. Virol.* **66**, 3784–3793.
- Parvin, J. D., Moscona, A., Pan, W. T., Leider, J. M. & Palese, P. (1986) *J. Virol.* **59**, 377–383.
- Sedivy, J. M., Capone, J. P., RajBhandary, U. L. & Sharp, P. A. (1987) *Cell* **50**, 379–389.
- Dixon, W. J. & Massey, F. J., Jr. (1969) *Introduction to Statistical Analysis* (McGraw–Hill, New York).

22. Luria, S. E. (1951) *Cold Spring Harbor Symp. Quant. Biol.* **16**, 463–470.
23. Denhardt, D. T. & Silver, R. B. (1966) *Virology* **30**, 10–19.
24. Denhardt, D. T. (1999) in *Encyclopedia of Virology*, eds. Granoff, A. & Webster, R. G. (Academic, London), 2nd Ed., pp. 274–281.
25. Loeb, L. A., Kunkel, T. A. & Schaaper, R. M. (1980) *ICN-UCLA Symp. Mol. Cell. Biol.* **19**, 735–751.
26. Fersht, A. R. & Knill-Jones, J. W. (1983) *J. Mol. Biol.* **165**, 633–654.
27. Ronen, A. & Rahat, A. (1976) *Mutat. Res.* **34**, 21–34.
28. Elena, S. F. & Moya, A. (1999) *J. Evol. Biol.* **12**, 1078–1088.
29. Domingo, E. & Holland, J. J. (1997) *Annu. Rev. Microbiol.* **51**, 151–178.
30. Chao, L., Tran, T. T. & Tran, T. T. (1997) *Genetics* **147**, 953–959.
31. Gottlieb, P., Metzger, S., Romantschuk, M., Carton, J., Strassman, J., Bamford, D. H., Kalkkinen, N. & Mindich, L. (1988) *Virology* **163**, 183–190.
32. Holland, J. J., Domingo, E., de la Torre, J. C. & Steinhauer, D. A. (1990) *J. Virol.* **64**, 3960–3962.
33. Lee, C. H., Gilbertson, D. L., Novella, I. S., Huerta, R., Domingo, E. & Holland, J. J. (1997) *J. Virol.* **71**, 3636–3640.
34. Schubert, M., Harmison, G. & Meier, E. (1984) *J. Virol.* **51**, 505–514.
35. Suárez, P., Valcárcel, J. & Ortín, J. (1992) *J. Virol.* **66**, 2491–2499.
36. Pringle, C. R., Devine, V., Wilkie, M., Preston, C. M., Dolan, A. & McGeoch, D. J. (1981) *J. Virol.* **39**, 377–389.