Mutations in clusters and showers

John W. Drake*

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

Several years ago, mutants arising spontaneously in a mutation-reporter sequence in a mouse were noted to contain more mutants bearing two or more mutations than was predicted by the mutation frequency and the assumption of a random distribution of mutations among mutants (1). Many of these overloaded mutants contained two mutations that were moderately close to each other. In this issue of PNAS, Wang *et al.* (2) describe the results of a robust sequencing effort to examine the same mutants over many kilobases outside of the original mutation reporter, and they find that some of the mutants contain even more mutations farther out in genetic space.

Those who write about mutation usually either explicitly state, or else imply, that it is a random process. However, although mutations are poorly predictable, they have already been observed to be nonrandom in one important way: from the earliest decades of mutation research (the 1920s through the 1940s), mutation rates were observed to vary greatly across different eukaryotic genes. In the middle of this period, when physicists were beginning to speculate about the fundamental properties of hereditary material, target theory was adapted from atomic physics to estimate gene sizes by using the assumed randomness of mutagenic radiation "hits" (3), and the results were modeled with the Poisson distribution. Use of the Poisson to model mutagenesis persists to the present and underlies such key methods as the Luria–Delbrück fluctuation test, one of the gold standards for estimating mutation rates (4).

In the 1950s, a prepared mind and an obscure comment in a thesis led Seymour Benzer to develop the powerful T4 *rII* system to estimate the fundamental parameters of the gene on a scale approaching that of DNA base pairs (5). One of his most enduring findings was the discovery of hugely variable site-specific mutability, a well-known example being mutational hot spots (6), which are the most nonrandom of all aspects of mutability and are now ubiquitously demonstrated in scores of mutational spectra. In 1962 using yeast (7), and in 1999 summarizing decades of mouse genetics (8), meiosis or a chronologically close period was reported to sustain elevated mutation rates, so that mutations also arise nonrandomly across developmental periods. Millionfold elevated mutation rates occur within a couple of kilobases in some cells in the immune system (9), and 8-fold lower mutation frequencies have been documented in mouse male germ cells, compared with somatic cells (10). There is also considerable literature documenting transiently increased mutation rates in bacteria, as occurs after induction of the SOS response to DNA damage and in starving and stationary-phase cellular microbes.

In a seminal paper in 1991, Jacques Ninio (11) predicted that individual bacteria would vary transiently in mutability because of mistakes such as the production of faulty proteins through errors of transcription and/or translation. He estimated that the impact of such "transient mutators" would be small for mutations arising singly but large for mutants acquiring multiple mutations in a single generation. This insight would be expected to extend to the replication of genomes from riboviruses to higher eukaryotes and to encompass any possible perturbation of genome fidelity. For instance, protein folding frequently goes awry, sometimes resulting in metastable but functional states, and proteins produced in only a few copies per cell may suffer maldistribution either within cell compartments or at cell division. In addition, most cellular organisms contain specialized error-prone DNA polymerases that may contribute to localized mutation-laced tracts (12).

Although mutations are poorly predictable, they have been observed to be nonrandom.

With a mind prepared by the Ninio conjecture, I was struck a decade later by the results of studies in my own group in which a short mutation-reporter DNA was synthesized by a DNA polymerase copying a template *in vitro*: approximately an order of magnitude more mutants with two or more mutations appeared than were predicted by the assumption of a random distribution. This extension of the Ninio conjecture to even the most minimal components of DNA replication prompted the question of whether excesses of "multiples" (mutants with two or more mutations) were frequent. They were. Examples surfaced from a ribovirus, a retrovirus, a DNA bacteriophage, a herpesvirus, the bacterium *Escherichia coli*, the yeast *Saccharomyces cerevisiae*, diverse rodent and human cell lines and tissues, and a number of DNA polymerases studied *in vitro* (13).

It can be argued semiquantitatively that little of the implied hypermutagenesis occurred in mutator mutants (mutants with heritably elevated mutation rates) (2, 13). Some of the reported examples may be artifactual, but in many studies that relied on the error-prone PCR to amplify mutant DNAs directly without cloning, enough starting DNA molecules were amplified to exclude this problem. In other studies, mutation frequencies and/or numbers of multiples, together with the factors by which they exceeded the expectations of randomness, were large enough to soothe concerns about artifactual origins of excess multiples.

In a handful of studies, enough multiples accumulated to justify the analysis of their separations in units of DNA base pairs. In the case of the DNA polymerase of bacteriophage RB69 copying a reporter sequence ≈200 bases long, the mutations in doubles were a random sample of all the mutations, even when the differential mutabilities of the numerous DNA sites were taken into account (13). In the case of the human *HPRT* gene in human kidney cells, the intermutation distances were somewhat clustered (numbers of intervening bases = 1, 6, 13, 13, 214, 4,886, 5,012, 7,023, and 25,024) (14). In the present system, an *E. coli* transgene ≈1,400 bases long embedded in the mouse genome, the intermutation distances in multiples displayed an exponential distribution, with half of the pairs separated by ≤120 bases (2). The latter two observations are particularly important because they imply that the compo-

Author contributions: J.W.D. wrote the paper.

The author declares no conflict of interest.

See companion article on page 8403.

*E-mail: drake@niehs.nih.gov.

nents of those multiples tended to arise simultaneously rather than accumulating one at a time during unlocalized hypermutation persisting over multiple cell replications.

Some Clusters Are Showers

In the case of the Sommer group's (2) primary mutation-responder target of ≈1.4 kb, ≈1% of a large collection of mutants were multiples, usually doubles. A collection of 65 multiples and 130 singles harvested from diverse mouse tissues was then resequenced in two regions: 3.6 kb to the immediate left and 7.1 kb to the immediate right of the primary target (2). Because the integrity of these flanking regions is not required in the system, most mutations arising in these regions are expected to be neutral and thus detected with high efficiency by DNA sequencing. The group's central result is that 10 of the 65 multiples (9 doubles and a triple) contained one or more additional mutations in the flanking 10.7 kb, whereas only 1 of the 130 singles contained a flanking mutation. Finally, additional sequencing of the multiples in regions totaling 8.5 kb and starting ≈17 kb from the primary target revealed two more mutations in these "remote" regions. All together, the flanking mutations added from one to four new mutations to the original multiple, and the total numbers of mutations, including those in the primary target, ranged from three to six. Evidence to date suggests that these multiples extend over ≤30 kb.

Wang *et al.* (2) call these widely dispersed mutations "showers," and the term seems apt. They point out that some showers may have landed not only on their five neutral targets, but also in part on ≈13 intervening kilobases in which many mutations would prevent the recovery of the carrier genomes, which consist of bacteriophage λ sequences. However, the distribution of mutations suggests that few such undetected showers would have occurred. The authors further estimate that such showers may constitute approximately 1% of all human mutants, although many such mutants might be difficult to detect because they would contain lethal mutations. (They also point out that many of the components of showers might merely soak into introns, without significant impact.) On the other hand, if a class of mutants was produced by heavy storms rather than showers, whether localized to the 45 or so kilobases studied here or scattered throughout the transgenic λ genome, many, perhaps almost all, would escape detection because of lethal mutations.

Impact of Showers

Multiple mutations seem to be required in order to achieve carcinogenesis, and many cancers display a mutator phenotype (15). However, surveys have often noted cancers that do not display a mutator mutation. It is possible that mutational showers sometimes provide the needed impetus (2, 13), but if showers are generally limited to ≈30 kb, they would be unable to hit the widely scattered genes that must be mutated in carcinogenesis.

It has been argued since the 1930s that transitions from one adaptive peak to a higher one must pass through a valley of reduced fitness if traversed by single mutations (16). Indeed, the intermediate single mutations might even be deleterious, further slowing adaptation. Because mutator mutations produce many offspring bearing new deleterious mutations, mutators themselves have reduced fitness. Thus, transient phenotypic hypermutation is likely to be a better source of infrequent adaptive multiples than are mutator mutations. Pairs of compensating deleterious mutations that are neutral or advantageous in combination appear in evolutionary lineages, and Kondrashov *et al.* (17) recently made a case that such pairs of mutations are not only readily found but also tend to reside within the same gene, making them more readily generated by mutational showers. Finally, microbial pathogens recently isolated from individual human hosts contain mutator mutations far more often than is observed in laboratory-grown populations, although the microbes must eventually purge their mutator mutations for their lineages to survive. Thus, mutational showers may contribute to the serial adaptations of pathogens as they move between hosts of the same, or even different, species.

How many more nonrandom aspects of mutagenesis remain to be revealed? In 1923, before he discovered how to induce mutations with ionizing radiation, Hermann Muller (18) wrote, "Beneath the imposing building called 'Heredity' there has been a dingy basement called 'Mutation.' Parts of the basement are still dingy but, on the other hand, Lewis Thomas (19) has written more famously, "The capacity to blunder slightly is the real marvel of DNA. Without this special attribute, we would still be anaerobic bacteria and there would be no music." But there will be fun, because students of the mutation process have much to do to discover how multiples arise.

1. Buettner VL, Hill KA, Scaringe WA, Sommer SS (2000) *Mutat Res* 452:219–229.
2. Wang J, Gonzalez KD, Scaringe WA, Tsai K, Liu N, Gu D, Li W, Hill KA, Sommer SS (2007) *Proc Natl Acad Sci USA* 104:8403–8408.
3. Timofeeff-Ressovsky NW, Zimmer KG, Delbrück M (1935) *Nachr Ges Wiss Gott Biol* 1:189–245.
4. Luria SE, Delbrück M (1943) *Genetics* 28:491–511.
5. Holmes FL (2006) *Reconceiving the Gene: Seymour Benzer's Adventures in Phage Genetics* (Yale Univ Press, New Haven, CT).
6. Benzer S (1961) *Proc Natl Acad Sci USA* 47:403–415.
7. Magni GE, von Borstel RC (1962) *Genetics* 47:1097–1108.
8. Russell LB (1999) *Proc Natl Acad Sci USA* 34: 16–23.
9. Weigert MG, Cesari IM, Yonkovich SJ, Cohn M (1970) *Nature* 228:1045–1047.
10. Walter CA, Intano GW, McCarrey JR, McMahan CA, Walter RB (1998) *Proc Natl Acad Sci USA* 95:10015–10019.
11. Ninio J (1991) *Genetics* 129:957–962.
12. Kunkel TA (2004) *J Biol Chem* 279:16895–16898.
13. Drake JW, Bebenek A, Kissling GE, Peddada S (2005) *Proc Natl Acad Sci USA* 102:12849–12854.
14. Colgin LM, Hackmann AFM, Emond MJ, Monnat RJ (2002) *Proc Natl Acad Sci USA* 100:1437–1442.
15. Loeb LA, Loeb KR, Anderson JP (2003) *Proc Natl Acad Sci USA* 100:776–781.
16. Wright S (1932) in *Proceedings of the Sixth International Congress of Genetics*, ed Jones DF (Brooklyn Botanic Garden, Brooklyn, NY), Vol 1, pp 356–366.
17. Kondrashov AS, Sunyaev S, Kondrashov FA (2002) *Proc Natl Acad Sci USA* 99:14878–14883.
18. Muller HJ (1923) *Eugenics, Genetics, and the Family* (Williams & Wilkins, Baltimore), Vol 1, pp 106–112.
19. Thomas L (1979) *The Medusa and the Snail: More Notes of a Biology Watcher* (Bantam, New York).