Another step toward quantifying spontaneous mutation

Alexey S. Kondrashov*

Life Sciences Institute and Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109

Spontaneous mutation, the occurrence of random changes in the nucleotide sequences, is an unavoidable result of the basic laws of nature and the key property of DNA. Together with natural selection, mutation is a *sine qua non* of evolution, and the genome of a species can be thought of as a record of mutations, beneficial, neutral, or very slightly deleterious, that became fixed, over a very long time, in an evolving lineage. Moreover, at the level of complete genomes, spontaneous mutations are quite common. A eukaryotic genome may acquire, on average, one or more new mutations each generation, and the genotype of an individual carries a number mutations, some of them substantially deleterious, that occurred in its not-too-distant ancestors. Still, at the level of individual nucleotides, and even of individual loci, mutations are very rare. Thus, quantitative studies of mutation should naturally be performed at the genome level, but up to now, this had been technically impossible. Fortunately, this is no longer the case, and, in this issue of PNAS, Lynch *et al.* (1) report the first successful application of whole-genome sequencing to detection of *de novo* mutations.

Let us put this achievement into historical context. Mutations were discovered in pre-DNA times through sudden changes of phenotypes, and only mutations that have drastic impacts on morphological traits and/or fitness were recognized and studied in the first 20 years. The next important step was the discovery that drastic mutations represent only the tip of the iceberg, and that mild mutations that cannot be detected individually at the level of phenotypes are more common than drastic mutations (2).

However, it took >30 years to perform the first quantitative analysis of mild mutations (3, 4). At that time, although the role of DNA as the molecule of inheritance had already been recognized, no effective methods of sequencing yet existed. Thus, the analysis relied on measuring the decline of fitness caused by accumulation of mutations in the course of experiments that lasted for 10–100 generations. For >20 years, the conclusions of these studies, which detected high mutation rates, were widely accepted, but later they became controversial (5). Indeed, precise measurements of fitness are notoriously difficult, and ascertaining new mutations through their impact on fitness involves many complications. Also, many mutations affect fitness only very slightly, if at all. There is a general agreement now that any quantitative study of mutation must involve detection of the exact sequence change for each event.

So far, three approaches that satisfy this requirement have been widely used. First, new mutations at one locus can be detected in an old-fashioned way, through drastic spontaneous phenotypic changes, after which each mutation is characterized at the sequence level. It is essential that a mutation must be detected immediately, or, at most, after just a few generations after it occurs, because otherwise cryptic negative selection can cause underestimation of the mutation rates. Thus, only dominant or X-linked (in humans) recessive mutations can be used. Knowledge of the molecular organization of the locus makes it possible to determine the mutational target for the phenotypic changes, that is, the number of nucleotides whose mutations are capable of causing such changes. Then, simple calculations can produce estimates of the rates of mutations of various kinds. However, screening for new, mutant phenotypes is very difficult, because of per locus rarity of mutations. So far, this approach has been successfully applied only to humans, when such screenings for loci that cause Mendelian diseases are performed in countries with comprehensive health care, by countless physicians (6).

Only a small fraction of new mutations lead to fitness effects.

Second, mutations can be detected by comparing homologous genome segments of the two related species. In this case, every mutation that occurred within this segment and became fixed in an evolving lineage after the two species diverged from their most recent common ancestor will be seen. Obviously, observing all of the mutations that accumulated over an evolutionary time makes it possible to estimate parameters of the mutation process from even a relatively short segment of the genome. However, two factors severely restrict the utility of this approach. First, the number of generations that separate two modern species is usually known only very approximately. Second, the rate of evolution coincides with the mutation rate only for those segments of the genome that are strictly selectively neutral. Otherwise, even a very weak selection against new mutations can drastically reduce the rate of evolution, and thus lead to a drastic underestimation of the mutation rate (7). So far, measuring mutation rates through interspecies divergence has led to credible results only for the human-chimpanzee species pair, because mammalian genomes contain a lot of genetic junk, and the timing of divergence and generation lengths are known with some certainty for these two species (8).

Obviously, these two approaches represent the two extremes, short-term and long-term, at the scale of time. The third approach is middle-term: mutations are allowed to accumulate under laboratory conditions in a number of lines that originated from the common ancestor [mutation-accumulation (MA) lines], for some number of generations (of the order of 100), after which genotypes of these MA lines are determined. Obviously, in this case, the number of generations is known exactly. Selection can also be effectively eliminated under laboratory conditions, at least for less crucial genome segments. Finally, a substantial number of generations during which mutations are allowed to accumulate makes it possible to detect them directly by DNA sequencing, without looking at phenotypes first. However, because this number is not huge, a large amount of sequencing is necessary to detect enough new mutations in MA lines. So far, this has been done by conventional methods, and the application of Lynch *et al.* (1) is the first of the new-generation whole-genome sequencing to this purpose. This has led to several results.

Most importantly, Lynch *et al.* (1) establish that whole-genome sequencing is already mature enough to be used for detecting mutations in MA lines. Pyro-

Author contributions: A.S.K. wrote the paper.

The author declares no conflict of interest.

See companion article on page 9272.

*E-mail: kondrash@umich.edu.

sequencing the complete genomes of only four MA lines of yeast *Saccharomyces cerevisiae* in an experiment that lasted 200 single-cell bottlenecks and ≈4,800 cell division generations yielded 33 nucleotide substitutions and a number of other mutations. Although pyrosequencing is relatively error-prone, resequencing all of the suspected mutations can definitely eliminate false positives. In fact, all of the 33 putative nucleotide substitutions detected by pyrosequencing were confirmed to be real by PCR. There seems to be no possibility for systematic false-negative results. Second, Lynch *et al*. report a relatively high mutation rate in *S. cerevisiae*. In particular, only a small fraction of new mutations lead to fitness effects that could possibly be detected in the laboratory. Thus, any attempt to measure the mutation rate through the decline of fitness of MA lines will lead to its very substantial underestimation. Third, Lynch *et al*. produce interesting insights into peculiarities of the mutation process in yeast, as far as relative rates of mutations of different kinds are concerned. Some key properties of spontaneous mutations appear to be surprisingly uniform across a variety of species. In particular, the per nucleotide rate of all kinds of mutations is close to 10^{-8} in humans (6, 8), *Caenorhabditis elegans* (9), and *Drosophila melanogaster* (10). However, many other parameters exhibit substantial variability, and data on more species are needed.

What next? I believe that Lynch *et al*. (1) is the last stepping stone on our way to the brave new world of direct, single-generation measurements of mutation rates by (improved) whole-genome sequencing. Indeed, MA lines are not without problems. First, in obligately outbreeding organisms, such as *Drosophila*, such lines must be maintained by close inbreeding, which may affect mutation rate. Second, MA lines are not an option in organisms with generation times longer than days or weeks.

Thus, the ultimate method of studying mutation at the molecular level is to sequence the genomes of a mother, a father, and their child. It seems very likely that the required refinements in whole-genome sequencing, in terms of its cost and fidelity, will be in place in a few years. The advent of whole-genome sequencing that will make it practical to detect events that occur as rarely as once per 100 million nucleotides will lead to the end of a long journey toward quantifying spontaneous mutation at the molecular level. As it is often the case, a tough long-standing problem will probably be solved by the brute-force application of a new technology.

1. Lynch M, *et al*. (2008) A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci USA* 105:9272–9277.
2. Timofeeff-Ressovsky NW (1935) Auslosung von Vitlitatsmutationen durch Rontgenbestrahlung bei *Drosophila melanogaster. Nachr Gess Wiss Gottingen Biol N F* 1:163–180.
3. Mukai T (1964) The genetic structure of natural populations of *Drosophila melanogaster.* I. Spontaneous mutation rate of polygenes controlling viability. *Genetics* 50:1–19.
4. Mukai T, Chigusa SI, Mettler LE, Crow JF (1972) Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster. Genetics* 72:335–355.
5. Keightley PD, Eyre-Walker A (1999) Terumi Mukai and the riddle of deleterious mutation rates. *Genetics* 153:515–523.
6. Kondrashov AS (2003) Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum Mutat* 21:12–27.
7. Baer CF, Miyamoto MM, Denver DR (2007) Mutation rate variation in multicellular eukaryotes: Causes and consequences. *Nat Rev Genet* 8:619–631.
8. Nachman MW Crowell SL (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics* 156:297–304.
9. Denver DR, *et al.* (2004) High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. *Nature* 430:679–682.
10. Haag-Liautard C, *et al.* (2007) Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila. Nature* 445:82–85.