LETTER

Yeast mutation accumulation experiment supports elevated mutation rates at highly transcribed sites

Because mutation is the ultimate source of genetic variation, understanding mutational patterns and mechanisms is of fundamental importance. Previous studies provided genomic evidence that, in the bacteria *Escherichia coli* and *Salmonella typhimurium*, yeast *Saccharomyces cerevisiae*, and human germ line, the mutation rate of a gene generally increases with the expression level of the gene, likely because of the transcription-associated mutagenesis (TAM) that trumps transcription-coupled repair (1–3). In PNAS, Zhu et al. sequence the genomes of 145 yeast mutation accumulation (MA) lines but detect no influence of transcriptional level on mutability (4). To investigate this discrepancy, we reanalyzed their data.

Zhu et al. identify 867 single nucleotide mutations (SNMs) (4), 559 of which are located in exons. We quantified the relative mRNA expression level at an exonic position by the number of mRNA sequencing (mRNA-seq) reads covering the nucleotide, estimated from a diploid yeast strain of an appropriate genetic background grown in the same medium as in the MA experiment (NCBI accession no. SRR1002821). The 559 SNM sites have on average 5.49 mRNA-seq reads, significantly greater than the random expectation of 1.20 ($P = 0.0013$, randomization test). Here, the random expectation was estimated by picking 10,000 random sets of 559 nucleotide sites from all exons in the yeast genome. Because the MA lines are diploid, mutation at any exonic site was assumed to be selectively permitted (4). To control the different mutabilities between G:C and A:T positions, we required that the randomly picked sites contained the same numbers of G:C nucleotides and A:T nucleotides as observed in the 559 SNMs.

TAM occurs because, during transcription, the nontranscribed strand of DNA is exposed as a single-stranded DNA, which is susceptible to damage. That is, TAM at a site should increase with the amount of exposure time. To test TAM, we estimated the relative exposure time of a site by counting the number of nascent transcript sequencing (NET-seq) reads that end at the site (5). The 559 SNM sites have on average 3.37 sense reads, significantly greater than the random expectation of 1.43 ($P = 0.0007$, randomization test), which was also estimated by the aforementioned method. Thus, TAM is supported.

There are two primary differences between the analysis of Zhu et al. and our analysis. First, they used a microarray-based dataset of nascent transcripts (4), which is expected to be less reliable than the NET-seq data. Second, they used only 256 SNMs, a subset of the 867 SNMs with relatively high genomic sequencing coverage (4). We found that 181 of these 256 SNMs are located in exons. Although these 181 sites also have significantly more mRNA-seq reads ($P = 0.0008$) and NET-seq reads ($P = 0.0013$) than their respective random expectations, these results are less reliable than those from the 559 SNMs due to potential Illumina sequencing biases that are shared among genome sequencing, mRNA-seq, and NET-seq. Regardless, appropriate analysis of the MA data of Zhu et al. reconfirms that the mutation rate of a nucleotide increases with its transcriptional level and illustrates the value of this large set of spontaneous mutations from nonmutator cells for uncovering mutational patterns and mechanisms.

Xiaoshu Chen and Jianzhi Zhang[1]
Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109

1 Park C, Qian W, Zhang J (2012) Genomic evidence for elevated mutation rates in highly expressed genes. *EMBO Rep* 13(12):1123–1129.
2 Chen X, Zhang J (2013) No gene-specific optimization of mutation rate in Escherichia coli. *Mol Biol Evol* 30(7):1559–1562.
3 Lind PA, Andersson DI (2008) Whole-genome mutational biases in bacteria. *Proc Natl Acad Sci USA* 105(46):17878–17883.
4 Zhu YO, Siegal ML, Hall DW, Petrov DA (2014) Precise estimates of mutation rate and spectrum in yeast. *Proc Natl Acad Sci USA* 111(22):E2310–E2318.
5 Churchman LS, Weissman JS (2011) Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 469(7330):368–373.

Author contributions: J.Z. designed research; X.C. performed research; X.C. analyzed data; and J.Z. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. Email: jianzhi@umich.edu.

www.pnas.org/cgi/doi/10.1073/pnas.1412284111