COMMENTARY

Transcriptional errors and the drift barrier

David M. McCandlish[a] and Joshua B. Plotkin[a,1]

Population genetics predicts that the balance between natural selection and genetic drift is determined by the population size. Species with large population sizes are predicted to have properties governed mainly by selective forces; whereas species with small population sizes should exhibit features governed by mutational processes alone. This "drift-barrier hypothesis" has been successful in explaining extensive variation in genome size, mutation rate, transposable element abundance, and other molecular features across diverse taxa (1–3). However, in PNAS Traverse and Ochman (4) report a striking exception to this theory by showing that transcriptional error rates are nearly equal across several bacterial species with very different population sizes.

Although the term "drift barrier" was coined in the context of mutation rates (5), the drift-barrier hypothesis applies to any heritable trait (e.g., ref. 6), and so it can provide a simple explanation for patterns of many molecular traits across the diversity of life. The explanation relies on the fact that mutations whose effects on fitness are smaller than the inverse of the population size behave like neutral mutations. As a result, natural selection will tend to optimize a trait until the selective benefits of further optimization are smaller than the inverse of the population size (7): that is, until the population hits the drift barrier. Because the location of the drift barrier depends on the population size, so too should the values of these phenotypic traits. Thus, differences in population size among species can explain a concerted syndrome of traits at the cellular and molecular levels (1–3, 5, 6).

In the context of broad empirical support for the drift-barrier hypothesis, the results of Traverse and Ochman (4) are particularly surprising. Using a CircSeq (8) strategy (based on the sequencing of short, circularized fragments of mRNA that are copied multiple times by rolling-circle amplification before sequencing) to reduce errors in high-throughput sequencing, Traverse and Ochman measured transcriptional error rates in *Escherichia coli* along with two endosymbiotic prokaryotes, *Buchnera aphidicola* and *Carsonella ruddii*. These endosymbionts feature dramatically reduced genome sizes, increased mutation rates, and other features typical of their small population sizes, including the loss of several transcriptional fidelity factors (4). Nevertheless, Traverse and Ochman report less than twofold variation in the rates of transcriptional errors among these three species. This result is in sharp contrast to mutation rates, which differ by orders of magnitude between species (5).

Possible Explanations for Conserved Error Rates

What could explain the apparent constancy of transcriptional error rates across species with such different population sizes? One hypothesis we must consider is errors in the sequencing procedure itself. If the rate of sequencing error exceeds the true transcriptional error rates, then it would give the impression of constant error rates. However, this does not appear to be the explanation. The CircSeq protocol (8) sequences the same nucleotide fragments several times in tandem, allowing errors in sequencing to be identified by differences among these tandem copies, rather than between the genome and the tandem copies. This procedure produces error rates of less than 10^{-6} per base (8), much lower than the transcriptional error rates of 3.4×10^{-5} to 5.1×10^{-5} reported by Traverse and Ochman (4). Furthermore, as the quality score cutoff used in their analysis increases, the error rates inferred by Traverse and Ochman decrease and then asymptote. This observation suggests that the sequencing error rate for low-quality scores exceeds the transcription error rate, but that at high-quality score sequencing errors are reduced to a level that permits stable, accurate measurement of transcriptional errors.

If we rule out experimental noise, then we must face the difficult task of providing a genuine biological explanation for Traverse and Ochman's (4) perplexing results. One possible explanation for constant error rates is that all three species studied have achieved the minimum transcription error rate

[a]Department of Biology, University of Pennsylvania, Philadelphia, PA 19104

Author contributions: D.M.M. and J.B.P. wrote the paper.

The authors declare no conflict of interest.

See companion article on page 3311.

[1]To whom correspondence should be addressed. Email: jplotkin@sas.upenn.edu.

www.pnas.org/cgi/doi/10.1073/pnas.1601785113

possible, given the bio-physics of RNA polymerases. This hypothesis seems unlikely, however, in light of the known effects of mutants and elongation factors that improve transcriptional fidelity (9, 10). Furthermore, a recent study in *Caenorhabditis elegans* has reported even smaller transcriptional error rates (11) than those reported by Traverse and Ochman (4) in these bacterial species.

A second possibility is that all three species studied by Traverse and Ochman (4) exhibit the maximum transcriptional error rate compatible with cellular life, and that any further increase would be so deleterious as to be selected against, even in *Buchnera* and *Carsonella*. However, this alternative explanation also seems unlikely because translational error rates differ by orders of magnitude between conditions and between species (5, 12). The total rate of error in protein production is simply the sum of transcriptional and translational error rates, and so this hypothesis would make sense only if the catastrophic effects of increased transcriptional errors were mediated by nonprotein-based features, such as the production of aberrant noncoding RNAs.

More complicated hypotheses to explain the results of Traverse and Ochman (4) are also possible. For example, Rajon and Masel (13) have distinguished between global and local solutions to the problem of molecular errors. Examples of global solutions include accurate polymerases and proof-reading factors, whereas local solutions include the evolution of individual coding sequences to mitigate the phenotypic effects of errors. Rajon and Masel predicted that local solutions are more likely to evolve in large populations. Traverse and Ochman (4) provide some evidence that *E. coli* has indeed evolved such local solutions. The authors report that although *E. coli* and *Buchnera* have similar transcriptional error rates, the fraction of these errors that would result in amino acid changes is substantially lower in *E. coli*, and indeed, lower than expected if errors occurred at random across the transcriptome. Thus, *E. coli* coding sequences may have evolved to ameliorate the effects of transcription errors on proteins, especially in highly expressed genes, an effect that would be analogous, in the context of transcription, to codon bias minimizing translational errors (14).

A fourth hypothesis to reconcile Traverse and Ochman (4) with the drift-barrier hypothesis is that the effects of transcriptional errors are more deleterious in *Buchnera* and *Carsonella*, per nucleotide, than they are in *E. coli*. This possibility is not without merit. *Buchnera* and *Carsonella* have so severely compressed genomes that they have likely lost the molecular complexity, including chaperones, that confers robustness in *E. coli*, and they may be unable to recover robustness from their host. Loss of robustness combined with a constant transcriptional error rate could arise in a model of gene loss following a rapid decrease in population size, because the mutational target size for gene loss is large compared with the number of mildly deleterious point mutations in RNA polymerases. However, this scenario still fails to explain why the rate of transcriptional errors is unchanged despite the loss of transcriptional fidelity factors (4).

How Do Transcriptional Errors Influence Fitness?

All of the hypotheses above rest on the assumption that natural selection acts to minimize the per base transcriptional error rate, but there may be other selective factors at play, such as a speed-accuracy trade-off. More generally, the target of selection depends on how exactly transcriptional errors influence fitness. If cellular function depends primarily on the proportion of codons in the transcriptome that are free of errors, then selection would indeed directly target the per base transcriptional error rate. On the other hand, if fitness defects are primarily a result of the energetic cost of producing nonfunctional proteins, then the strength of selection would scale with the total amount of nonfunctional protein produced per cell division, which depends on both the per base transcriptional error rate and also total protein production per cell division. By way of comparison, patterns in mutation rates predicted by the drift-barrier hypothesis best fit empirical data when selection is assumed to act on the number of errors per exome, rather than the per base error rate (3).

Another possibility is that the per base transcriptional error rate is indeed the direct target of natural selection, but selection pushes this rate toward an intermediate value, rather than minimizing it. By producing a diverse set of proteins from a single genomic sequence, transcriptional errors may sometimes produce "error" products that contribute positively to fitness (15). Similarly, environmental fluctuations may select for bet-hedging behavior (16), with epigenetic switches triggered by errors in transcription (17). Either mechanism could result in selection for some degree of error in transcription.

If we rule out experimental noise, then we must face the difficult task of providing a genuine biological explanation for Traverse and Ochman's perplexing results.

Any explanation for conserved rates of transcriptional errors must be consistent with the known variability in rates of translational errors. Here it is worth noting that the effects of transcriptional errors differ in subtle and important ways from the effects of translational errors. Although it is true, as Traverse and Ochman (4) point out, that a given mRNA transcript may be translated several times so that each transcriptional error may produce more altered proteins than each translational error, both sources of error have the same effect on the probability of error per amino acid in the proteome. The difference is not in the total probability of error but rather in the spectrum of errors: translation errors produce a diverse ensemble of proteins, whereas transcription errors produce many copies of the same, incorrect protein. Thus, the fitness effects of the two types of error may differ greatly, depending on the dose-dependence of the burden (or benefit) caused by noisy protein production. For example, translational errors will tend to be more deleterious if a single incorrect protein is capable of inducing aggregation or toxicity, whereas transcriptional errors will be more deleterious if multiple identical incorrect proteins are required to nucleate aggregation. On top of this, the specific types of errors in protein synthesis differ as well, because transcriptional errors are predominantly a result of deamination of cytosine (4), whereas the spectrum of mistranslation errors depends on the abundances of tRNAs (18). Thus, simply estimating the overall rates of transcriptional and translational errors may be insufficient to understand the effects of such errors.

These complexities in how errors in transcription are transmuted into cellular fitness make it difficult to formulate a strong, a priori hypothesis about how selection should shape these error rates across species. This uncertainty, in turn, makes the striking consistency of transcriptional error rates observed across bacterial taxa (4) all the more surprising.

1 Lynch M (2007) *The Origins of Genome Architecture* (Sinauer Associates, Sunderland, MA).
2 Kuo CH, Moran NA, Ochman H (2009) The consequences of genetic drift for bacterial genome complexity. *Genome Res* 19(8):1450–1454.
3 Sung W, Ackerman MS, Miller SF, Doak TG, Lynch M (2012) Drift-barrier hypothesis and mutation-rate evolution. *Proc Natl Acad Sci USA* 109(45):18488–18492.
4 Traverse CC, Ochman H (2016) Conserved rates and patterns of transcription errors across bacterial growth states and lifestyles. *Proc Natl Acad Sci USA* 113:3311–3316.
5 Lynch M (2010) Evolution of the mutation rate. *Trends Genet* 26(8):345–352.
6 Lynch M (2012) Evolutionary layering and the limits to cellular perfection. *Proc Natl Acad Sci USA* 109(46):18851–18856.
7 Hartl DL, Dykhuizen DE, Dean AM (1985) Limits of adaptation: The evolution of selective neutrality. *Genetics* 111(3):655–674.
8 Acevedo A, Brodsky L, Andino R (2014) Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* 505(7485):686–690.
9 Strathern JN, Jin DJ, Court DL, Kashlev M (2012) Isolation and characterization of transcription fidelity mutants. *Biochem Biophys Acta* 1819(7):694–699.
10 Zhou YN, et al. (2013) Isolation and characterization of RNA polymerase rpoB mutations that alter transcription slippage during elongation in *Escherichia coli*. *J Biol Chem* 288(4):2700–2710.
11 Gout JF, Thomas WK, Smith Z, Okamoto K, Lynch M (2013) Large-scale detection of in vivo transcription errors. *Proc Natl Acad Sci USA* 110(46):18584–18589.
12 Barak Z, Gallant J, Lindsley D, Kwieciszewki B, Heidel D (1996) Enhanced ribosome frameshifting in stationary phase cells. *J Mol Biol* 263(2):140–148.
13 Rajon E, Masel J (2011) Evolution of molecular error rates and the consequences for evolvability. *Proc Natl Acad Sci USA* 108(3):1082–1087.
14 Akashi H (1994) Synonymous codon usage in *Drosophila melanogaster*: Natural selection and translational accuracy. *Genetics* 136(3):927–935.
15 Whitehead DJ, Wilke CO, Vernazobres D, Bornberg-Bauer E (2008) The look-ahead effect of phenotypic mutations. *Biol Direct* 3:18.
16 King OD, Masel J (2007) The evolution of bet-hedging adaptations to rare scenarios. *Theor Popul Biol* 72(4):560–575.
17 Gordon AJ, Satory D, Halliday JA, Herman C (2015) Lost in transcription: Transient errors in information transfer. *Curr Opin Microbiol* 24:80–87.
18 Kramer EB, Farabaugh PJ (2007) The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. *RNA* 13(1):87–96.