Conserved rates and patterns of transcription errors across bacterial growth states and lifestyles

Charles C. Traverse and Howard Ochman

Department of Integrative Biology, University of Texas, Austin, TX 78712

PNAS | March 22, 2016 | vol. 113 | no. 12 | 3311–3316

Errors that occur during transcription have received much less attention than the mutations that occur in DNA because transcription errors are not heritable and usually result in a very limited number of altered proteins. However, transcription error rates are typically several orders of magnitude higher than the mutation rate. Also, individual transcripts can be translated multiple times, so a single error can have substantial effects on the pool of proteins. Transcription errors can also contribute to cellular noise, thereby influencing cell survival under stressful conditions, such as starvation or antibiotic stress. Implementing a method that captures transcription errors genome-wide, we measured the rates and spectra of transcription errors in Escherichia coli and in endosymbionts for which mutation and/or substitution rates are greatly elevated over those of E. coli. Under all tested conditions, across all species, and even for different categories of RNA sequences (mRNA and rRNAs), there were no significant differences in rates of transcription errors, which ranged from 2.3 × 10⁻³ to 5.2 × 10⁻⁵ per nucleotide in mRNA of the endosymbiont Buchnera aphidicola to 5.2 × 10⁻⁵ per nucleotide in RNA of the endosymbiont Carsonella ruddii. The similarity of transcription error rates in these bacterial endosymbionts to that in E. coli (4.63 × 10⁻³ per nucleotide) is all the more surprising given that genomic erosion has resulted in the loss of transcription fidelity factors in both Buchnera and Carsonella.

Among the multiple types of information processing errors, the majority of research has focused on mutations that occur during DNA replication because such errors are heritable and form the basis of evolutionary change. However, errors that occur during transcription and translation can also have substantial effects on gene function by producing misfolded and malfunctioning proteins. The rate of translation errors is typically an order of magnitude higher than the rate of transcription errors (1–6). However, errors occurring during transcription often elicit more dire consequences than those occurring during translation because individual mRNAs can be translated up to 40 times (7, 8), resulting in a burst of flawed proteins. Therefore, a single transcription error can result in many flawed proteins, whereas a translation error will disrupt only a single protein.

Because deleterious transcription errors are not transmitted to subsequent generations, they can occur more frequently than mutations to DNA but still infrequently enough to ensure the cell is not overburdened with faulty proteins. Estimates of the rate of transcription errors in Escherichia coli have been determined in vitro by measuring the misincorporation of radiolabeled nucleotides into repeating dinucleotide tracts (1, 9) and in vivo by quantifying the reversion frequencies of nonsense mutations in lacZ (2, 3). These assays yielded variable estimates of transcription error rates of 10⁻⁷–10⁻⁵ per nucleotide, several orders of magnitude higher than the mutation rate (10–12). Studies that assay individual loci are often not representative of the genome as a whole because sequence- or genome-specific features, such as base composition (12, 13) or sequence motifs (14), affect the incidence of information processing errors. Moreover, transcription error rate reversion assays based on the recovery of functional proteins might also include translation errors, if these occur at a sufficiently high rate.

RNAseq offers an approach to both disentangle transcription errors from translation errors and provide an error rate for every transcribed gene in a genome. Unfortunately, the high error rates both of cDNA synthesis (3–6 × 10⁻³ per nucleotide) (15–17) and of high-throughput sequencing technologies (possibly as high as 10⁻²–10⁻³ per nucleotide) (18, 19) renders the transcription errors obtained by conventional RNAseq indistinguishable from sequencing artifacts. Two recently developed methods offer ways to circumvent these problems by allowing transcription errors to be distinguished from sequencing and cDNA synthesis errors. Through the use of altered library preparation protocols, these methods reduce the overall error rate of RNAseq to less than 10⁻⁶ (20) and 10⁻⁶ (21, 22) per nucleotide, making it possible to measure error rates across the entire transcriptomes of viruses and other organisms.

In this study, we implement both of these RNAseq-based methods in E. coli to examine whether transcription error rates vary according to growth state and physiological condition, as has been reported for translation error rates (23–26) and for the combined transcription and translation error rate (27). Moreover, we asked whether transcription error rates are increased in the endosymbiotic bacteria Buchnera aphidicola and Carsonella ruddii—species that have lost known transcription fidelity factors and whose mutation rates, substitution rates, and rates of protein sequence evolution are all amplified as a result of genetic drift and the loss of repair enzymes (Fig. S1). We show that transcription error rates are remarkably similar across organisms, even for broad categories of RNA on which the cell is known to selectively degrade malfunctioning rRNA (28).

Results

Resource Limitation and Growth Phase Do Not Alter Transcription Error Rates. We tested the effects of different growth conditions—all of which have been associated with altered mutation rates and/or

Significance

Organisms rely on accurate transcription for proper cellular function. Whereas errors incurred during replication are transmitted to subsequent generations, those that occur during transcription are transient and affect only a subset of the encoded proteins. Although transcription errors may increase survival in stressful conditions, the majority of these errors are harmful, and their rates must be minimized. By assessing the transcription errors genome-wide in Escherichia coli and in two bacterial endosymbionts, we discovered that all species had remarkably similar transcription error rates. This conservation is unexpected given that both endosymbiotic species lack orthologs of several E. coli RNA fidelity factors and that lifestyle differences among these species have led to vast differences in their mutation and substitution rates.

Author contributions: C.C.T. and H.O. designed research; C.C.T. performed research; C.C.T. and H.O. analyzed data; and C.C.T. and H.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. See Commentary on page 3136.

1To whom correspondence should be addressed. Email: howard.ochman@austin.utexas.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1525329113/-/DCSupplemental.
translation error rates—on rates of transcription errors. Using a deep-sequencing approach to identify errors, we measured the transcription error rate in *E. coli* when grown under four growth conditions (tryptic soy broth (TSB) complex media or M9 glucose minimal media, each sampled at midlog and at stationary phase). Note that these errors include both base substitutions during the process of transcription and any damage to the mRNA after transcription. Each of the four conditions were assayed in duplicate, and in total, we detected 2,621 transcription errors, with the number of errors per sample ranging from 156 to 681. In neither of the nutrient sources was there significant differences in transcription error rates for cells harvested at midlog phase or at 8 h after entering stationary phase (Fig. 1; paired Wilcoxon test, *P* > 0.30). Similar to what we observed for *E. coli* assayed at different growth phases, transcription error rates do not differ significantly in nutrient-rich (TSB) and nutrient-poor (M9) growth media (Fig. 1; paired Wilcoxon test, *P* = 0.3429). Furthermore, there are no significant differences in overall transcription error rates between any pair of individual conditions tested [Fig. 1; two-tailed *t* tests, *t* (2) < 2.3, *P* > 0.14], and the average transcription error rate over all conditions is 4.63 ± 0.34 (SEM) × 10⁻⁵ for *E. coli* mRNA.

**Distribution of Transcription Errors.** The use of a high-throughput sequencing method to detect transcription errors (as opposed to a reporter-gene method) enables analysis of transcription errors genome-wide as well as the localization of errors to individual sites in each transcript. Starting at the scale of whole genomes, we analyzed the fluctuation in transcription error rates and found that the 95% of measurements made for 50-kb nonoverlapping windows across the entire *E. coli* genome varies threefold among genomic regions, ranging from 2.3 to 7.2 × 10⁻⁵ (Fig. S2). Regions containing highly expressed genes had an increased number of transcription errors (Fig. S3), resulting from increased coverage enabling the discovery of more errors relative to areas in the genome with low coverage.

Transcription proceeds in the direction of DNA replication on the leading strand and in the opposite direction on the lagging strand, in which case there can be collisions between the replication and transcription machineries. Despite an increased likelihood of collision-induced errors on the lagging strand, there is no significant difference in the transcription error rates between genes encoded on the two strands (Wilcoxon test, *P* > 0.90; Fig. S4). Next, we tested whether adjacent nucleotides affected the occurrence of transcription errors and found that neither a particular preceding nor succeeding nucleotide induced transcription errors. Only when both the preceding and succeeding nucleotides are guanine residues do we observe a significant increase in transcription error frequency (Fisher’s exact test, *P* < 0.02). Taken together, transcription errors occur without regard for genome location, direction of transcription, or for the vast majority of neighboring nucleotides.

**Biases in *E. coli* Transcription Errors.** Measuring transcription errors using a sequencing-based approach provides information about the absolute frequencies of each of the possible base substitutions. C→U errors were most common, occurring at a significantly higher frequency than all other transcription errors (Fig. 2A), presumably attributable to high rates of cytosine deamination after the RNA is transcribed. It has previously been reported that transcription errors incur a higher rate of transitions than transversions (20, 29), the same overall pattern that we observe in *E. coli* (Wilcoxon test, *P* < 0.05). This trend, however, is driven solely by high incidence of C→U changes and no longer reaches significance after removing these transitions from the analysis (Wilcoxon test, *P* > 0.50). Next, we tested the effect of individual nucleotides on the frequency of transcription errors in *E. coli* and found that G/C→N errors occur at higher frequencies than do A/U→N errors (Wilcoxon test, *P* < 0.02; Fig. 2B). Additionally, N→A/U errors occurred at a significantly higher rate than do N→G/C errors (Fig. 2C; Wilcoxon test, *P* < 0.02). These effects are not due solely to the high frequency of C→U errors: even after the removal of C→U errors (Methods), G/C→N errors remain significantly more frequent than A/U→N errors (Fig. 2B), and N→A/U errors remain significantly more frequent than N→G/C errors (Fig. 2C).

**Transcription Error Rates in Host-Restricted Bacteria with Reduced Genomes.** The bacterial endosymbionts, *B. aphidicola* and *C. ruddii*, harbor small genomes (450 and 190 kb, respectively) and have very high substitution rates, as a consequence of both their lack of several repair mechanisms (Fig. S1) and the reduced efficacy of selection due to their small effective population sizes. These features are also expected to augment rates of transcription errors, so we assayed the transcription error rates in these endosymbionts using methods identical to those used for *E. coli*. For the replicate samples of *B. aphidicola*, we detected a total of 169 transcription errors in total mRNA, yielding a transcription error rate of 2.69 ± 0.73 (SEM) × 10⁻⁵, which is not significantly different from the rate that we obtained for *E. coli* mRNA [two-tailed *t* test, *t* (8) = 2.527, *P* > 0.05; Fig. 3A].

Transcription errors in *C. ruddii* mRNA could not be assigned unequivocally because the *C. ruddii* RNA was extracted from a natural population of individuals, rendering it difficult to distinguish between transcription errors and the polymorphisms that might be present in the population. Instead, we quantified transcription error rates for 16S and 23S ribosomal RNA in both *C. ruddii* and *B. aphidicola* because these operons are present in single copy, have high read-coverage (despite the rRNA removal step), and are not polymorphic within a species. Unlike *C. ruddii* and *B. aphidicola*, the *E. coli* genome possesses multiple polymorphic tRNA operons, making it unfeasible to estimate tRNA transcription error rates in *E. coli*. We detected a total of 1,014 errors in *C. ruddii* rRNAs and 4,377 errors in *B. aphidicola* tRNAs, yielding tRNA transcription error rates of 5.13 × 10⁻⁵ for *C. ruddii* and 3.37 × 10⁻⁵ for *B. aphidicola* (Fig. 3A). Our estimates of bacterial transcription error rates are, in descending order, 5.13 × 10⁻⁵ for *C. ruddii* tRNA, 4.63 × 10⁻⁵ for *E. coli* mRNA, 3.37 × 10⁻⁵ for Buchnera tRNA, and 2.69 × 10⁻⁵ for Buchnera mRNA. The transcription error rates for *B. aphidicola* mRNA and tRNA do not differ significantly from one another.

**Biases in Endosymbiont Transcription Error Rates.** Assessing the transcription errors occurring in both Buchnera mRNA and tRNA allowed us to determine whether there are any observable differences in the error rates for two RNA substrates, as might be caused by base compositional biases or selection. All possible nucleotide substitutions, as attributable to transcription errors, were detected in both the mRNA and tRNA samples (although one of the *B. aphidicola* mRNA replicates lacked any A→C changes). There were no significant differences for any of the
individual substitution classes between mRNA and rRNA or among any of individual substitution classes (Fig. 3B).

Effects of Transcription Errors on Protein Sequences. Given that each transcript can be translated—perhaps multiple times—into protein, we determined which transcription errors result in an amino acid substitution. On average, $68 \pm 1.46\%$ (SEM) of transcription errors cause an amino acid substitution in *E. coli*, whereas 80% of the transcription errors in *Buchnera* result in amino acid substitutions. If errors were to occur at random over the *E. coli* transcriptome, the probability of changing an amino acid is significantly higher than that actually incurred by transcription errors (76% vs. 68%; pairwise Wilcoxon test, $P < 0.008$).

**Discussion**

Considering the range of variation in replication error rates and in translation error rates both within and among bacterial species, our finding that transcription error rates are similar for different species and for different classes of RNA sequences and under different physiological conditions within a species is bewildering. The mutation (i.e., DNA replication error) rates for bacteria span by several orders of magnitude (10); for the specific organisms that we
consider, spontaneous mutation rates vary nearly 50-fold, from $8.9 \times 10^{-11}$ per site per generation in *E. coli* (10) to $4.0 \times 10^{-10}$ for *Buchnera aphidicola* (29). In contrast, based on our genome-wide deep-sequencing approach, the transcription error rates of these two species differ by less than twofold ($2.7 \times 10^{-10}$ vs. $4.3 \times 10^{-10}$), with *E. coli* having the slightly higher rate. Our initial prediction was that endosymbionts would have higher transcription error rates because they are subject to high levels of genetic drift and would therefore sustain more deleterious mutations; however, neither of the studied endosymbionts had elevated transcription error rates.

We reasoned that differential regulation of transcription fidelity factors, such as *greA* (30, 31), *greB* (31), or *dksA* (31, 32), operating during transcription, translation, or protein degradation, could provide a mechanism for *E. coli* to modulate its transcription error rate across bacteria, selections may have increased. The intrinsic error rate of transcription error rates among species is all of the more surprising given that these endosymbionts lack homologs for several of these transcription fidelity factors (Fig. S1). Endosymbionts possess the most highly reduced bacterial genomes (33), and the genome sizes of *Buchnera* and *Carsonella* are only 641 and 160 kb, respectively (34, 35), in contrast to the 4,640-kb genome of the *E. coli* MG1655. Genome reduction in endosymbionts results from the loss of genes that are no longer necessary in the host environment but also involves the loss of apparently beneficial genes, such as DNA repair systems. The efficiency of translation is also a deciding factor such as DNA repair, transcription, and translation (Fig. S1). The lack of certain DNA repair enzymes in endosymbionts have been implicated in their extreme base compositions and increased mutation rates (36–38); however, loss of multiple RNA fidelity factors, such as *greB* in *Buchnera* (Fig. S1) and *greA*, *greB* (Fig. S1) and *dksA* in *Carsonella*, seems not to affect transcription error rates.

These endosymbionts are missing transcription fidelity factors, but their transcription error rates are unchanged, implying that there are mutations within RNAP that can increase the fidelity of transcription. If there is indeed an optimal transcription error rate across bacteria, selections may have increased. The intrinsic error rate in the endosymbionts RNA聚合酶s after they lost the transcription fidelity factors. However, neither of the RNAPs of the endosymbionts possess a mechanism known to increase transcription fidelity in *E. coli* (39). It is possible that endosymbionts do not require rapid transcription and can tolerate slow but accurate transcription (39). The presence of these fidelity factors in *E. coli* could allow its RNAP to make fewer errors (which are then corrected), as a result of selection for increased transcription speeds and increased growth rates.

Not only were transcription error rates similar in proteobacterial taxa of vastly different lifestyles, population structures, genomes sizes, and mutation rates, but the error rates were comparable across organisms for different broad categories of RNAs. Because structural RNAs (16S and 23S rRNAs) persist longer than mRNAs, they can incur more damage (due to oxidative stress or deamination), thereby leading to an increase in our estimates of the error rate for ribosomal RNAs. On the other hand, one might anticipate rRNAs to have lower error rates than mRNAs, because subfunctional molecules would be preferentially targeted for degradation (28), leaving only those rRNAs that do not contain errors. It should be noted that under both scenarios, the error rate during transcription does not change, but rather the variation in the estimated error rates is caused by differences in the rate of rRNAs after transcription. We were only able to measure transcription error rates for both mRNA and rRNA in *Buchnera*. The average error rate for *Buchnera* mRNA was slightly lower than for rRNA, but this estimate was based on the detection of many fewer errors, and there is no significant difference between the two categories of RNAs (Fig. 3). It is not possible to measure transcription error rates in rRNAs of *E. coli* because selection methods are not possible. This suggests that the differences in transcription error rates in *Buchnera* may be due to the higher copy number of rRNA, although many rRNA molecules may simply generate errors and do not discriminate between those caused by base misincorporations occurring during transcription and by damage to the RNA after transcription. Nonetheless, chemical damage occurring after transcription is biologically relevant because ribosomes can still translate the damaged base.

Many of the initial measurements of transcription errors in bacteria were restricted to single reporter genes and assay the combined effects of transcription and translation errors by assessing how frequent functional proteins were produced from a mutant gene (2, 3, 14). These assays considered errors in translation to be a normal part of the system, but it is thought that only the first ribosome on a transcript would be capable of mistranslation and that most errors could be ascribed to the process of transcription (2, 3). However, transcription errors in *E. coli* can occur at rates between $10^{-4}$ and $10^{-2}$ per codon (4–6), suggesting that many of the original measurements of transcription errors are confounded by the inclusion of translation errors. Furthermore, the error rates varied by up to an order of magnitude for different stop codons (3), indicating that these fluctuations may be attributed to different transcription error rates for different codons (55); therefore, the rates derived from these studies require validation by methods that exclusively consider transcription errors.

Previous studies reported that the combined transcription/translation error rate, as inferred from the frequency of errors in protein sequences, increases both in stationary phase and under starvation conditions (4, 6, 27). Because we detected no differences in transcription error rates between these different growth conditions, we reason that this variation manifests during translation and is most likely caused by rRNA scarcity during stationary phase (5, 55). Although decreases in ribonucleotide concentration occur during stationary phase (57), this has little effect on the overall fidelity of gene expression. Decreases in ribonucleotide concentration have been shown to increase the frequency of transcriptional pausing (58), which is closely associated with base misincorporations during transcription (39, 59, 60), so it seems that either (i) ribonucleotide concentration does not decrease enough under our experimental conditions to significantly alter the transcription error rate or (ii) that ribonucleotide concentration-induced pausing does not result from transcription errors. Nonetheless, it is curious that cellular growth conditions modify both the rate of RNA mutations and the rate of protein translation; caution errors but not the transcription error rate.

Rates of transcription errors have been estimated as being at least an order of magnitude higher than rates of transcription errors, but because most transcripts are translated multiple times, the realized number of modified proteins originating from transcription errors will equal or exceed the number caused by
translation errors. This amplification of individual transcription errors into multiple proteins is likely to account for the reduction of transcription vs. translation error rates (10^{-4} vs. 10^{-3}).

It has been suggested that errors in proteins, as caused by transcription and translation errors, contribute to survivability in the face of external stresses by the production of novel proteins or metabolites (27, 61, 62) or by inducing the general stress response (63). Such effects could not be accomplished through genomic mutations because such mutations can incur permanent decrements to fitness after the stress is removed. Although transcription errors can increase cellular noise and confer a benefit under certain temporary conditions, most variation introduced by errors will not be advantageous. Thus, the predominant direction of selection is to lower error rates because too many errors will overload the proteome with deleterious proteins. Whether or not the above argument is tenable, our findings, showing a remarkable consistency of transcription and translation errors across ecologically diverse bacterial species, different RNA categories, and under a variety of stress and nonstress growth conditions indicate that transcription errors would contribute very little to such transient protein errors. Transcription is a much less accurate process than DNA replication, and because transcription errors are not heritable (and the vast majority of RNAs are transcribed faithfully under any set of conditions), there appears to be little selection to modulate the overall transcription error rate.

**Methods**

**Strains and Growth Conditions.** Transcription errors were enumerated for *E. coli* MG1655 grown at 37 °C in (i) 15 g TSb or (ii) M9 minimal media supplemented with 0.4% glucose. Bacterial cultures were preconditioned in either TSb or M9 minimal media for 24 h before inoculation for sampling. Overnight cultures were diluted to OD_{600} = 0.05 into fresh media and sampled at midlog phase (4 h for TSb; 6 h for M9) and stationary phase (18 h for TSb; 24 h for M9).

Transcription errors were enumerated for *B. aphidicola*, an insect endosymbiont recovered directly from its aphid host, *Acarus siro*, and another insect endosymbiont, *C. rudii*, an insect endosymbiont recovered directly from its aphid host, *Pachypsylla venusta* collected locally from galls present on a hackberry tree. Bacterioocytes were stored in buffer A at −20 °C before RNA extraction.

**RNA Extractions.** RNA was extracted from *E. coli* following the RNASnap protocol for gram-negative bacteria (65). Roughly 10^6 bacterial cells were harvested by centrifugation at 16,000 × g for 30 s, the supernatant was removed by aspiration, and pelleted cells were immediately transferred to liquid nitrogen to halt transcription. Samples were transferred to ice, mixed with TRIzol RNA extraction solution (1/8 mL EDTA, 0.025% SDS, 1% 2-mercaptoethanol, and 95% (vol/vol) formamide), briefly vortexed, and incubated for 7 min at 4 °C, and the homogenate was centrifuged at 1,500 × g for 15 min. Pellets were resuspended in 15 mL buffer A (25 mM KCl, 35 mM Tris-HCl, 100 mM EDTA, and 250 mM sucrose, pH 8.0) at 4 °C, and the supernatant was removed by centrifugation at 10,000 × g for 15 min. Nucleic acids were washed twice with 70% (vol/vol) ethanol, resuspended in 50 μL 100% isopropanol, and RNA was eluted from crushed gel slices by overnight incubation in a solution containing 600 mM sodium acetate, 0.017% (wt/vol) SDS, and 1.67 mM EDTA at 4 °C. RNA was recovered from the eluent by ethanol precipitation, washed in 70% (vol/vol) EtOH, resuspended in 14 μL dH_2O, and analyzed for quality on an Agilent Bioanalyzer RNA chip. RNA fragments were circularized by incubating the entire sample volume with 1 μL T4 polynucleotide kinase (NEB), 1 μL T4 RNA ligase (NEB), 2 μL 10X T4 RNA ligase buffer (NEB), and 2 μL 10 mM ATP for 30 min at 37 °C. Samples were purified by PCI extraction and ethanol precipitation, and libraries were prepared for Illumina sequencing by following the protocol accompanying the NEBNext Ultra RNA Library Prep Kit through completion of the second strand synthesis step. After this step, samples were repurified by PCI extraction and ethanol precipitation and analyzed with an Agilent Bioanalyzer RNA chip to determine the degree of rolling-circle amplification, which occurred during the cDNA synthesis step of the NEB protocol. After confirming amplification status, dH_2O was added to a final volume of 200 μL, and samples were subjected to 12 min of pulsed sonication (15 s on, 15 s off, amplitude 20%) in a Qsonica sonicator to obtain fragments for sequencing. After harvesting nucleic acids by EtOH precipitation, we resumed the NEBNext Ultra RNA Library Prep Kit protocol for a target insert size of 300 bp. Samples were barcoded using NEBNext Multiplex Oligos (Index Primers Set 1), and the resulting libraries were sequenced on an Illumina MiSeq using 300 nt reads. Sequencing files were discriminated based on their identifying barcodes and analyzed using the CirSeq_v2 pipeline (21).

**Data Processing and Analysis.** After the sequences were processed by the CirSeq_v2 pipeline with an average quality score cutoff of 20 (Fig. S5 and SI Methods), we removed the duplicate and multiple reads determined by the CirSeq_v2 pipeline to determine the identity of rolC. We aligned the reads to the *E. coli* genome (e.g., structural RNA genes, ompP and ompC, and tufA and tufB) because the source of variation cannot be unequivocally assigned. Transcription error rates were adjusted for base composition of the sample using the weighted average of the occurrence of each nucleotide in the particular individual transcriptome being considered.

We developed custom Python scripts to determine the following: (i) transcription errors, calculated by tabulating the total number of errors identified by the CirSeq_v2 pipeline within the protein coding regions of the genome (SI Methods); (ii) nucleotide coverage, calculated by adding the overall coverage of each base within the protein coding regions of the genome; (iii) error rates, calculated by tabulating the total number of errors and base coverage of all coding regions within 50-kb nonoverlapping windows across an entire genome and dividing the number of errors by the coverage, yielding an error rate (SI Methods); (iv) leading/lagging strand error rates, calculated by tabulating the errors and coverage of all genes situated on either the leading or lagging strands and calculating the error rate as above; and (v) the number of errors that would result in an amino acid replacement by chance, calculated by randomly generating simulated transcription errors from each sequenced transcriptome and determining their effects on the amino acid sequence. All statistics were performed in Prism Graphpad or R.

The list of nucleic acid information processing genes and the associated functions were curated using EcoCyc (66). Orthologs of these genes in the endosymbionts were determined using BLASTP from the National Center for
1. Springgate CF, Loeb LA (1975) On the fidelity of transcription by Escherichia coli ribonucleic acid polymerase. J Mol Biol 97(4):577–591.
2. Rosenberg RF, Fossett G (1987) The estimate of the frequency of in vivo transcriptional errors at a nonsense codon in Escherichia coli. Mol Gen Genet 183(3):561–563.
3. Rosenberg RF, Hinjton J (1983) The frequency of transcriptional and translational errors at nonsense codons in the lacZ gene of Escherichia coli. Mol Gen Genet 191(2):207–212.
4. O’Farrell PH (1978) The suppression of defective translation by pPPp and its role in the stringent response. Cell 14(3):545–557.
5. Bouadloun F, Donner D, Kurland CG (1983) Codon-specific misense errors in vivo. EMBO J 2(8):1351–1356.
6. Ballesteros M, Fredriksson A, Henriksson J, Nyström T (2001) Bacterial senescence: Proteomes of declining E. coli cells is dictated by the accuracy of the ribosomes. EMBO J 20(18):5280–5289.
7. Kennel D, Riezman H (1977) Translation and transcription initiation frequencies of the Escherichia coli lac operon. J Mol Biol 114(1–2):1–21.
8. Golding I, Paulson J, Zawilski SM, Cox EC (2005) Real-time kinetics of gene activity in individual bacterial cells. Cell 123(6):1025–1036.
9. Blank A, Gallant JA, Burgess RR, Loeb LA (1986) An RNA polymerase mutant with reduced accuracy of chain elongation. Biochemistry 25(20):5920–5928.
10. Drake JW (1991) The spontaneous mutation rate of spontaneous mutation in DNA-based microbes. Proc Natl Acad Sci USA 88(16):7160–7164.
11. Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutation. Genetics 148(4):1667–1668.
12. Lee H, Popodi T, Spudich JA, Foster PL (2012) Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing. Proc Natl Acad Sci USA 109(41):E2774–E2783.
13. Wielgosz S, et al. (2011) Mutation rate inferred from synonymous substitutions in a long-term evolution experiment with Escherichia coli. G3 (Bethesda) 12(1):183–186.
14. Zhou YN, et al. (2013) Isolation and characterization of RNA polymerase rpoM mutations that alter transcription slippage during elongation in Escherichia coli. J Biol Chem 288(27):2700–2710.
15. Aresti V, Hogrefe F (2007) Escherichia coli DNA polymerase III α: subunits increase Moloney murine leukemia virus reverse transcriptase fidelity and accuracy of RT-PCR procedures. Anal Biochem 360(1):84–91.
16. Skasko M, et al. (2005) Mechanistic differences in RNA-dependent DNA polymerization and fidelity between murine leukemia virus and HIV-1 reverse transcriptases. J Biol Chem 280(13):12190–12200.
17. Banaraukas A, et al. (2012) Generation and characterization of new highly thermostable and processive M-MuLV reverse transcriptase variants. Eng Prot Enz 24(3):195–207.
18. Loman NJ, et al. (2012) Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 30(5):434–439.
19. Meacham F, et al. (2011) Identification and correction of systematic error in high-throughput sequence data. BMC Bioinformatics 12:364.
20. Gout JF, Thomas PA, Z, Okamoto K, Mesbah M (2013) Large-scale detection of in vivo transcription errors. Proc Natl Acad Sci USA 110(46):18584–18589.
21. Acevedo A, Brodsky L, Andino R (2014) Mutational and fitness landscapes of an RNA virus revealed through population sequencing. Nat Methods 11(5):397–400.
22. Acevedo A, Andino R (2014) Library preparation for highly accurate population sequencing of RNA viruses. Nat Protoc 9(7):1760–1769.
23. Fredriksson A, et al. (2007) Decline in ribosomal fidelity contributes to the accumulation and stabilization of the master stress response regulator signal on carbon starvation. Genes Dev 21(7):862–874.
24. Barak Z, Gallant J, Lindsay D, Kwicizewski B, Heidel D (1996) Enhanced ribosome frameshifting in stationary phase cells. J Mol Biol 263(2):140–148.
25. Fu C, Parker J (1994) A ribosomal frameshifting error during translation of the argT mRNA of Escherichia coli. Mol Gen Genet 243(4):434–441.
26. Wentzel AM, Stancek M, Isaksson LA (1998) Growth phase dependent stop codon readthrough and shift of translation reading frame in Escherichia coli. FEBS Lett 421(3):237–242.
27. Phylogenomics M, Mamou G, Ben-Yehuda S (2010) Visualizing high error levels during gene expression in living bacterial cells. Proc Natl Acad Sci USA 107(25):11543–11548.
28. Cheng JF, Deutscher MP (2003) Quality control of ribosomal RNA mediated by polynucleotide phosphorylase and RpsL. Proc Natl Acad Sci USA 100(11):6388–6393.
29. Iriuchishimizu M, Oshima T, Lubkovskaya L, Kashlev M (2013) Direct assessment of transcription fidelity by high-resolution RNA sequencing. Nucleic Acids Res 41(19):9090–9104.
30. Eire DA, Hajiseyedjavadi O, Young MC, von Hippel PH (1993) Multiple RNA polymerase core and G50 decisions of the fidelity of transcription. Science 262(5153):867–873.
31. Zenkin N, Yuzenkova Y (2015) New insights into the functions of transcription factors that bind the RNA polymerase secondary channel. Biomolecules 5(3):1195–1209.
32. Rognan D, Lachaine E, Yuzenkova Y (2014) Computational and experimental characterization of DksA and Tppp interaction with RNA polymerase. Nucleic Acids Res 43(15):1529–1536.

ACKNOWLEDGMENTS. We thank Dianne Lou and Sandie Shan for technical support. Ashley Anthony provided essential clarifications. Alejandro Caro-Quintero for isolation of C. ruddii, and Kim Hammond for the preparation of figures. This research was supported by National Institutes of Health Award R01GM108657 (to H.O.).