# Unique Cost Dynamics Elucidate the Role of Frameshifting Errors in Promoting Translational Robustness

Tobias Warnecke*[,1], Yang Huang[2], Teresa M. Przytycka[2], and Laurence D. Hurst[1]

[1]Department of Biology and Biochemistry, University of Bath, Bath, United Kingdom

[2]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland

*Corresponding author: E-mail: T.Warnecke@bath.ac.uk.

## Abstract

There is now considerable evidence supporting the view that codon usage is frequently under selection for translational accuracy. There are, however, multiple forms of inaccuracy (missense, premature termination, and frameshifting errors) and pinpointing a particular error process behind apparently adaptive mRNA anatomy is rarely straightforward. Understanding differences in the fitness costs associated with different types of translational error can help us devise critical tests that can implicate one error process to the exclusion of others. To this end, we present a model that captures distinct features of frameshifting cost and apply this to 641 prokaryotic genomes. We demonstrate that, although it is commonly assumed that the ribosome encounters an off-frame stop codon soon after the frameshift and costs of mis-elongation are therefore limited, genomes with high GC content typically incur much larger per-error costs. We go on to derive the prediction, unique to frameshifting errors, that differences in translational robustness between the 5′ and 3′ ends of genes should be less pronounced in genomes with higher GC content. This prediction we show to be correct. Surprisingly, this does not mean that GC-rich organisms necessarily carry a greater fitness burden as a consequence of accidental frameshifting. Indeed, increased per-error costs are often more than counterbalanced by lower predicted error rates owing to more diverse anticodon repertoires in GC-rich genomes. We therefore propose that selection on tRNA repertoires may operate to reduce frameshifting errors.

**Key words:** translational error, GC content, translational accuracy, codon usage bias, tRNA repertoire.

## Introduction

A growing body of evidence supports the idea that codon usage patterns partially reflect selection to avoid errors during translation (reviewed in Drummond and Wilke 2009). But what types of error are being selected against and why? Misincorporation errors have arguably received a lion's share of recent attention but inserting the wrong amino acid is by no means the only and perhaps not even the most common or costly mishap that can occur during translation. For instance, the ribosome can also abandon the nascent polypeptide before completion (drop-off, premature termination error) or leave the correct reading frame and elongate the peptide chain based on nucleotide triplets never meant to serve as a template for protein synthesis (frameshifting error) (Parker 1989).

A failure to accurately decode the underlying codon lies at the heart of all of these translational errors. Consequently, detecting biased usage of more efficiently decoded ("trans-lationally optimal") synonymous codons is not in itself sufficient to implicate a particular error process. To go beyond diagnosing translational selection and attribute adaptive features of gene anatomy to specific error processes, we need to develop critical tests that can implicate one type of error to the exclusion of others. In this context, it is interesting to note that different translational errors exhibit variation with regard to the fitness costs involved (detailed below). Understanding and exploiting divergent cost dynamics might therefore hold the key to devising critical tests and assessing the relative evolutionary importance of different error processes.

Mistranslation events can be costly for a variety or reasons. Some cost models are focused on the erroneous "product" and propose that errors are deleterious because they abrogate function or because the mistranslated product elicits dominant negative effects downstream of translation (Drummond and Wilke 2009). For example, mistranslated proteins might misfold and, consequently,

disrupt a variety of cellular processes, by interacting promiscuously with other proteins and forming toxic aggregates (Drummond and Wilke 2008) or by occupying quality control capacity (chaperones, proteases, etc.), thereby interfering with normal protein homeostasis.

Other cost models are centered on the notion that the act of generating an erroneous product can be costly in itself ("process cost"). Fitness costs here may arise through nonproductive occupation of ribosomal capacity, which can be rate limiting for growth (Shachrai et al. 2010) or through sequestration of other translational resources (amino acids, tRNAs, etc.), which may prevent other proteins from being made in a timely fashion (Stoebel et al. 2008). In addition, it has been suggested that the energy wasted in futile synthesis and degradation may constitute a relevant evolutionary cost (Wagner 2005, 2007).

One key prediction of error models that focus on process costs is that such costs should strongly covary with the length of the erroneous product because residency time at the ribosome, the level of resource sequestration and the amount of energy wasted in protein synthesis and degradation should all increase with length. In line with this prediction, Stoebel et al. (2008) found, when they induced *lac* genes in a lactose-free environment (i.e., expressing a protein without any functional benefit to the cell), that longer genes were associated with greater costs.

The strong theoretical link between product length and process-related fitness cost can inform strategies to pinpoint particular error processes behind adaptive codon usage patterns because different translational errors have stereotypically different effects on the length of the erroneous product. Misincorporation errors do not alter the length of the polypeptide relative to the wild-type protein. In contrast, premature termination errors lead to truncation of variable severity depending on where along the mRNA the error occurs. This has led to the prediction that nonsense errors should become increasingly more costly toward the 3′ end of the mRNA and that, concomitantly, selection should be more powerful in promoting accurate decoding toward the 3′ end (Eyre-Walker 1996). Consistent with this prediction, optimal codon usage increases toward the 3′ end of coding sequences in *Escherichia coli* (Qin et al. 2004; Stoletzki and Eyre-Walker 2007). Importantly, this constitutes a critical test for translational selection against errors other than missense errors because—unless missense errors also promote drop-off—misincorporation errors do not predict a gradient in the leverage of selection increasing toward the 3′ end of the mRNA.

In this study, we ask whether frameshifting errors show process cost dynamics that discriminate them from other types of translational error and can thus help us gain a better understanding of the role of frameshifting avoidance in shaping gene anatomy.

Building on previous work (Huang et al. 2009), we present a simple quantitative model of frameshifting cost centered on genome-specific tRNA concentrations and relative binding affinities. Comparing process cost estimates across 641 prokaryotic genomes, we demonstrate that frameshifting errors exhibit process cost dynamics that are different from both missense and premature termination errors and can be exploited to establish support for the hypothesis that selection against frameshifting at least in part explains differential codon adaptation at the 5′ and 3′ termini of mRNAs. Furthermore, our study highlights that comparative genomic estimates of the costs of translational error can be highly misleading when mRNA sequences are considered in isolation or with disregard to species-specific biology. This is principally because there are strong interactions between process cost, GC content, tRNA repertoire, and error rates that generate considerable variability in average expected frameshifting costs across prokaryotic genomes.

## Materials and Methods

### Prokaryotic Genomes and tRNA Repertoires

We downloaded protein-coding sequences for 1,035 complete prokaryotic genomes from the *National Center for Biotechnology Information* (NCBI) (ftp://ftp.ncbi.nih.gov/genomes/Bacteria/) in February 2010. Applying custom scripts, we filtered the data to limit analysis to genes with a multiple of three nucleotides ($n = 4$ genes excluded based on this criterion), without ambiguous nucleotides or internal in-frame stop codons, and with a proper stop according to the relevant NCBI translation table, either table 11 (TGA, TAG, TAA) or table 4 (TAG, TAA) where TGA is decoded as tryptophan.

For 756 of these genomes, we could obtain information on the copy number and diversity of different tRNA isoacceptors from tRNADB-CE (Abe et al. 2009). For the purpose of this study, we make the reasonable assumption (Dong et al. 1996; Kanaya et al. 1999) that tRNA copy numbers are an adequate proxy for cellular tRNA concentrations in unicellular organisms. For reasons described in the supplementary methods (Supplementary Material online), we discarded an additional 115 genomes to yield a final set of 641 genomes (supplementary table 1, Supplementary Material online).

### A Comparative Model for the Process Cost of Accidental Frameshifting

We suggest that the genomic process cost of accidental frameshifting ($C_G$) is approximated by

$$C_G = \sum^{G} \sum_{i=1}^{L} t_i p_i (n_{\text{pre}} + n_{\text{post}}), \qquad (1)$$

where $p_i$ is the probability that a frameshift occurs at codon $i$ (detailed below); $n_{\text{pre}}$ and $n_{\text{post}}$ are the number of peptide
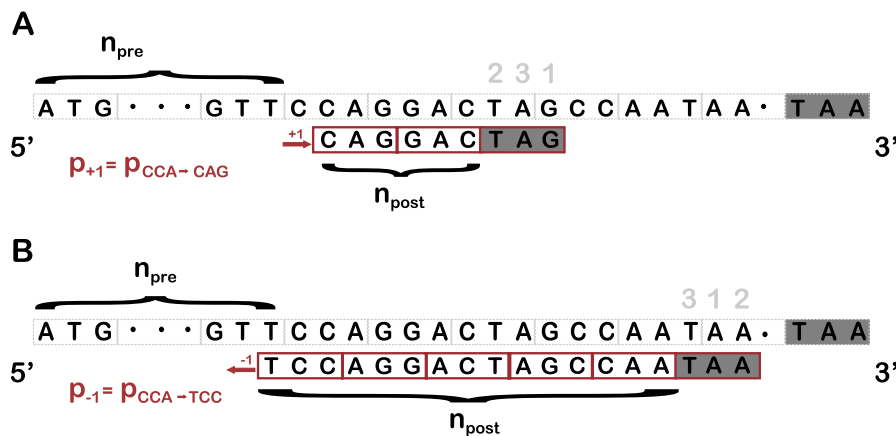
Fig. 1.—Schematic representation of frameshifting cost. $n_{pre}$ ($n_{post}$) is the number of codons translated before (after) the frameshift occurs, either in (A) the +1 or (B) the −1 direction. $n_{post}$ is determined as the number of codons translated until an off-frame stop (highlighted in gray) is encountered, which minimally requires a T in the 2nd (3rd) position of the original frame for +1 (−1) events. In this example, $p_{i+1}$ ($p_{i-1}$) is the probability of shifting from codon CCA onto CAG (TCC).

bonds made before and after an error occurs at codon $i$ (fig. 1), respectively; and $t_i$ is the number of times codon $i$ is translated.

The model is nested so that we can obtain a per-gene estimate through summing across the entirety of codons ($L$) in a given mRNA, and a per-genome estimate through summing per-gene estimates across the entirety of mRNAs ($G$). Below, we focus on average cost per site or per gene because summing across all genes to determine genomic cost is likely to be misleading in the absence of information on translation levels. Note that, for the purpose of this analysis, we assume that every frameshifting error will yield a completely nonfunctional product. Although there may be an argument that functionality is more likely to be preserved when frameshifting occurs at the 3′ end of the mRNA, it is difficult to see how to systematically discount costs in a biologically relevant manner without detailed, gene-specific information on the impact of truncation and mis-elongation on functionality. In addition, there is evidence that translational selection operates even at the very 3′ end of mRNAs (Tuller et al. 2010), strongly suggesting that these regions are typically not functionally dispensable.

We define $n_{post}$ as the number of codons translated before the ribosome encounters the first off-frame stop codon or the coding sequence ends. Note that in the latter case $n_{post}$ represents a conservative estimate because the translation unit may not end at the 3′ end of the coding sequence. This is particularly true for bacteria, where mRNAs are often polycistronic (Sorek and Cossart 2010).

## Modeling Site-Specific Frameshifting Probabilities ($p_i$)

Error propensity can differ considerably across sites and also depends on the state of the translational machinery. For example, homomeric nucleotide runs appear much more li-

able to frameshifting (Farabaugh 1996) than other sequence contexts, and there is ample evidence that the composition of the cellular tRNA pool is a critical determinant of decoding accuracy and, consequently, the propensity for frameshifting. Increasing the concentration of a particular tRNA results in reduced frameshifting frequencies at the corresponding codons (Atkins et al. 1979; Curran and Yarus 1989; Sipley and Goldman 1993). Conversely, codons matched by rare tRNAs are particularly liable to frameshifting (Sipley and Goldman 1993; Farabaugh and Björk 1999) and amino acid starvation can substantially increase the likelihood of frameshifting at codons read by the affected tRNA (Gallant and Lindsley 1992, 1993; Kolor et al. 1993).

Farabaugh and Björk (1999) have suggested that tRNA-mRNA interactions at the ribosome can, in fact, provide a unifying model to understand accidental frameshifts, where frameshifting probability is principally a function of relative tRNA concentrations and binding affinities. Briefly, the authors proposed that frameshifting can occur when a near-cognate tRNA erroneously binds to the codon in the ribosomal A site—more likely when there is a relative shortage of cognate tRNAs—and, after translocation to the P site, the weak anticodon:codon interaction permits downstream (+1) or upstream (−1) slippage by one nucleotide if a sufficiently stable interaction can be formed in the new reading frame. Huang et al. (2009) recently presented a quantitative formulation of the Farabaugh and Björk model, where the probability $p_i$ of (+1) frameshifting at any one codon $i$ is determined as

$$p_i^{+1} = \frac{b \sum_{t \in V_i^{+1}} n_t}{b \sum_{t \in V_i^{+1}} n_t + b \sum_{t \in R_i^{+1}} n_t + n_{tc_i}}, \quad (2)$$

where $V_i^{+1}$ and $R_i^{+1}$ are the sets of near-cognate tRNAs able and unable to slip one nucleotide downstream, respectively;

$n_t$ represents tRNA gene copy number; $n_{tci}$ the number of cognate tRNA genes of codon $c_i$, and $b$ a positive constant <1, denoted "weak binding coefficient" by Huang et al. (2009), which models the fact that binding of near-cognate tRNAs is less stable than binding of cognate tRNAs. For each genome, we derived $V_i$ and $R_i$ for all codon contexts based on a set of parsimonious anticodon:codon matching strategies proposed by Grosjean et al. (2010) (for details, see supplementary methods, Supplementary Material online).

The parameter $p_i$ captures an important aspect of decoding accuracy, namely that error rate is intrinsically dependent on the relative (rather than absolute) concentration of cognate, near-cognate, and noncognate codons, so that it is critical to consider the diversity and relative abundance of tRNAs to assess tRNA-dependent translation parameters (Fluitt et al. 2007).

## Results

### Individual Frameshifting Errors Are Typically More Costly in GC-Rich Genomes

Different types of translational error are associated with different stereotypical process costs. Although premature termination errors incur costs approximately proportional to the number of residues translated before the error occurred ($n_{pre}$, fig. 1), frameshifting errors incur an additional cost ($n_{post}$) because the ribosome carries on translating until it encounters an off-frame stop codon or the mRNA ends.

It is widely assumed that $n_{post}$ is typically small, courtesy of a high chance of encountering an off-frame stop codon in the immediate downstream neighborhood (Parker 1989; Farabaugh 1996; Farabaugh and Björk 1999; Itzkovitz and Alon 2007). Itzkovitz and Alon (2007) reported that, for an "average" genome (uniform codon usage and amino acid frequencies averaged over 134 genomes from all three kingdoms), the ribosome encounters a fortuitous off-frame stop on average only 15 codons downstream of the frameshifting error.

Figure 2A demonstrates that this figure can be profoundly misleading when genomic GC content is high. Standard stop codons (TGA, TAA, TAG) are AT-rich and the probability of encountering AT-rich in-frame codons, required to specify the off-frame stop, decreases with increasing GC content. This is all the more pronounced for −1 frameshifts where a T at the 3rd codon position is required to yield an off-frame stop (fig. 1). In contrast to the first two codon positions, where A/T nucleotides may be required to specify amino acid identity, GC variability is much more extreme at 3rd sites (Muto and Osawa 1987) so that encountering a 3rd site T in a high-GC genome is comparatively rare.

$n_{post}$, however, only represents part of the process cost of an individual frameshifting error because it ignores the number of amino acids translated before the error occurred ($n_{pre}$). To approach a more realistic estimate of average
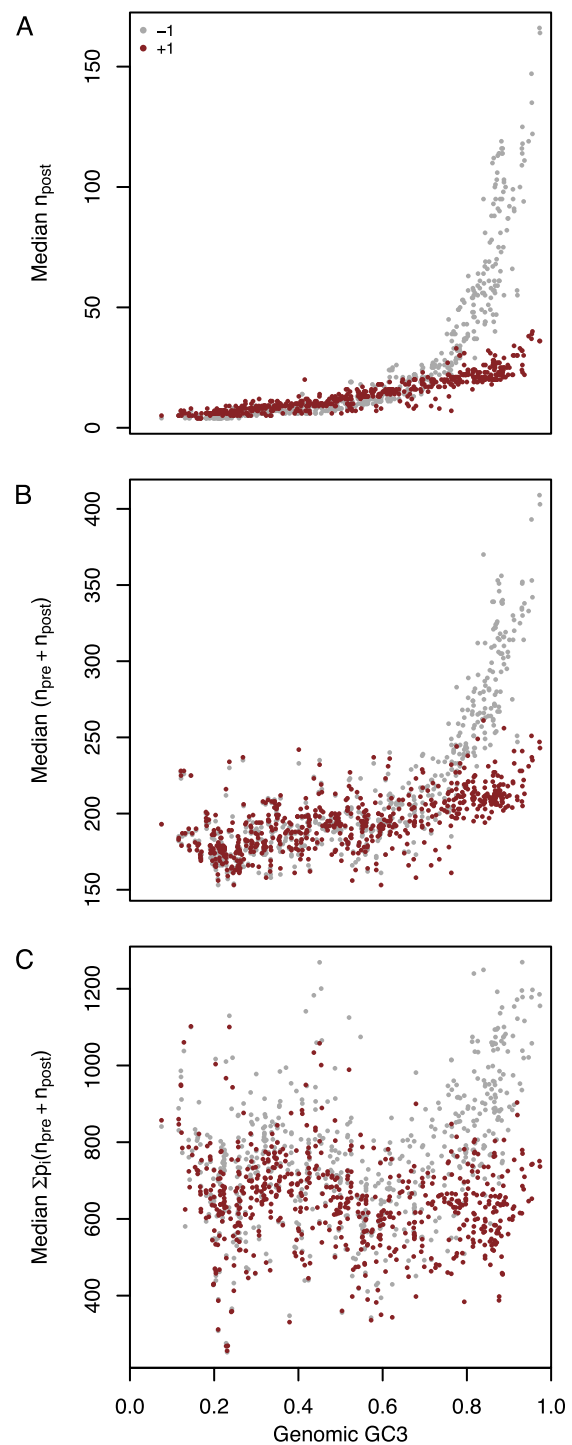


**Fig. 2.**—The relationship between GC3 content and components of frameshifting cost. (A) $n_{post}$ (median number across sites), (B) $n_{pre} + n_{post}$ (median number across sites), (C) $\sum p_i (n_{pre} + n_{post})$ (median cost across genes, i.e., for each gene, we sum cost across all codons in that gene to obtain a gene-specific cost and then plot the median by-gene cost across all genes). Genomic GC3 content is the proportion of G and C nucleotides across all 3rd sites across all coding sequences analyzed. Each data point represents one genome.

genome-specific frameshifting cost, we computed $n_{pre} + n_{post}$ for every codon in every gene. Results suggest that GC-dependent differences in average cost between genomes might not be as pronounced as suggested by $n_{post}$ considered in isolation (fig. 2B). This is principally because $n_{post}$ typically contributes less than 20% (40%) of the total process cost ($n_{pre} + n_{post}$) of +1 (−1) frameshifts even at extremely high GC (supplementary fig. 1, Supplementary Material online). At the same time, average $n_{pre}$ varies across GC content only in as far as proteins tend to be slightly longer on average in genomes with higher GC content (supplementary fig. 2, Supplementary Material online).

## GC-Rich Genomes Are Buffered Against +1 but not −1 Frameshifts

This reduction in between-genome variability notwithstanding, the average process cost still appears to be higher in genomes with high GC content. But do GC-rich genomes really shoulder a greater fitness burden in relation to frameshifting? Clearly, that depends on whether any one particular error actually occurs and, if so, how frequently. This is a function of the probability $p_i$ that the error occurs at the focal codon $i$, and the number of times that site is translated ($t_i$). Although by-gene estimates of $t_i$ are not available for the vast majority of genomes, we can derive relative frameshifting probabilities for every possible codon context with reference to genome-specific tRNA competition at the ribosome (see Materials and Methods). Incorporating genome- and context-specific frameshifting probabilities into our model of process cost, we unexpectedly find the positive correlation between GC content and average +1 frameshifting cost reversed (rho = −0.18, $P = 5.81 \times 10^{-06}$, fig. 2C). This is despite protein length increasing slightly with GC content (linear regression estimate of average protein length in genomes with 20% GC3: 256 amino acids [90% GC3: 278], supplementary fig. 2, Supplementary Material online). The average cost of −1 frameshifts, however, remains highest for high-GC genomes (rho = 0.35, $P = 2.21 \times 10^{-20}$). Considering only one prokaryotic species per genus name to reduce phylogenetic nonindependence does not affect overall trends (data not shown).

## More Diverse tRNA Repertoires in GC-Rich Genomes Counterbalance Larger Per-Error Costs of +1 Frameshifting

Why does incorporation of genome-specific frameshifting probabilities transform the relationship between GC content and estimates of frameshifting cost?

Comparing $p_i$ for each minimal shifting context (NNN|N for +1 shifts, N|NNN for −1 shifts) across genomes, we find that the majority of contexts exhibits a lower propensity for frameshifting with increasing GC content (negative tau in fig. 3). The altered relationship between GC content and

cost is therefore not simply a function of different codon or dicodon usage, that is, less shifting-prone motifs being used more frequently at high GC content; systematic GC-linked changes in tRNA profiles must be a contributing factor. Conspicuously, GC-rich genomes typically sport a more diverse repertoire of anticodons (fig. 4, Kanaya et al. 1999; Rocha 2004; Higgs and Ran 2008; Ran and Higgs 2010). In particular, tRNAs with C or G in the first anticodon position, which we would expect to bind most stably to G- and C-ending codons, respectively, are typically present in high-GC genomes where G/C-ending codons are common but frequently spared in medium- or low-GC genomes (fig. 5) where these codons are read via wobble pairing with U in the first anticodon position. This is in line with theoretical expectations about the diversity of tRNAs required for efficient translation (Higgs and Ran 2008). We suggest that, in addition, larger anticodon repertoires in high-GC genomes will be selectively favorable as they reduce the burden of frameshifting error in genomes vulnerable to incurring large per-error costs.

What these results highlight, above all, is that comparing translational cost estimates between genomes will be misleading when sequence features are considered in isolation because other critical parameters ($p_i$) can and do differ between genomes. In this context, we realize that our empirical evaluation falls short of giving a comprehensive comparative costing because we cannot at present incorporate translation levels ($t_i$). We are keenly aware that, especially in fast-growing organisms, a large proportion of realized cost might be incurred by a relatively small number of highly expressed genes so that taking average cost across all sites (or even genes) might not adequately reflect genomic fitness burden. Once comprehensive quantitative transcriptome data becomes available for an extremely high-GC genome, it will be interesting to incorporate this information into our model to derive genuinely comparative genome-wide cost estimates.

## A Critical Test for a Role of Frameshifting in Shaping Gene Anatomy: GC-Rich Genomes Show Weaker 5′-3′ Gradients in Translational Robustness

Above we hypothesized that, in addition to selection on translational efficiency (Higgs and Ran 2008), increased richness of the tRNA repertoire in GC-rich genomes might be at least in part an adaptation to the comparatively larger per-error cost of frameshifting in these genomes. Is there, however, any evidence consistent with frameshifting as an important force in molecular evolution? Selection against premature termination errors predicts a gradient in codon adaptation toward greater decoding accuracy at the 3′ end of mRNAs, predicated on $n_{pre}$ as the principal process cost. But $n_{pre}$ also represents an important component of
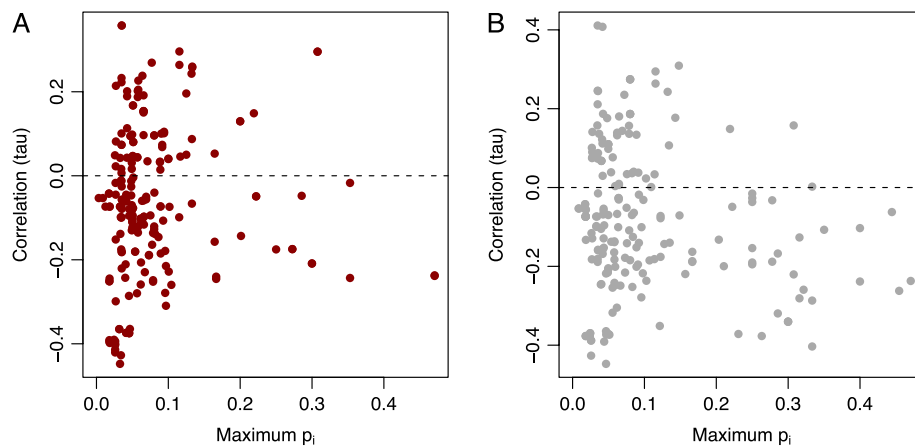
**FIG. 3.**—Frameshifting probabilities decline with GC content for the majority of frameshifting contexts. We computed nonparametric correlations (Kendall's tau) between genomic GC3 content and $p_i$ for all possible frameshifting contexts (NNN|N or N|NNN for +1 and −1 shifts, respectively) across genomes. Each data point represents one such context. The majority of (A) +1 and (B) −1 contexts exhibit negative correlation coefficients, indicating that, for the respective context, the probability of frameshifting decreases with increasing GC3. In particular, there is typically a negative correlation for frameshifting contexts that can be especially prone to shifting (high maximum $p_i$).

the process cost of frameshifting errors. Does selection against frameshifting errors contribute to intragenic gradients in codon adaptation?

The unique process cost dynamic of frameshifting errors, namely the existence of a post-error cost ($n_{post}$), allows us to test for frameshifting involvement as follows: Consider an mRNA with very high GC content. At the extreme, even slipping up right at the start of the message leads to exactly the same cost as slipping up at the 3′ end because the ribosome will never encounter an off-frame stop and therefore keep



**FIG. 4.**—Anticodon repertoire (the number of different anticodons among all tRNA genes in the genome) increases with GC3 content.

on translating until the mRNA terminates. By implication, GC-rich genomes should benefit relatively less from greater robustness ($1 - p_i$) against frameshifting errors toward the 3′ end of genes. We therefore predict that, if frameshifting avoidance is a relevant force determining heterogeneity in codon composition along the mRNA, the difference in frameshifting robustness between 5′ and 3′ ends will decline with increasing GC content. In contrast, selection against premature termination errors does not predict 5′-3′ differential robustness to ameliorate with rising GC content.

Replicating Huang et al.'s (2009) approach, we computed pairwise differentials in average frameshifting robustness across the terminal 5′ and 3′ 100 codons. Note that this analysis is internally controlled so that we do not expect differences in expression across genes and genomes to affect results. We observe a clear-cut tendency toward less pronounced 5′-3′ differences with increasing GC content (fig. 6), supporting a role for selection against frameshifting errors. Results are virtually identical when we exclude the first and last 30 codons, which are likely under selection for translational regulation (Tuller et al. 2010; 30-codon cutoff conservatively estimated from prokaryotic data in their fig. 2E and F).

May this trend simply be a consequence of codon choice becoming less flexible at more extreme GC content? This would predict that differences in terminal robustness should also decline toward the AT-biased end of the spectrum. This we do not observe: We split the data into genomes with >50% GC and <50% GC and confined analysis to genomes where the most AT-biased genome was as far away from the 50% threshold as the most GC-biased genome (range 11–89% GC). We found significant positive relationships between GC3 and differential robustness for
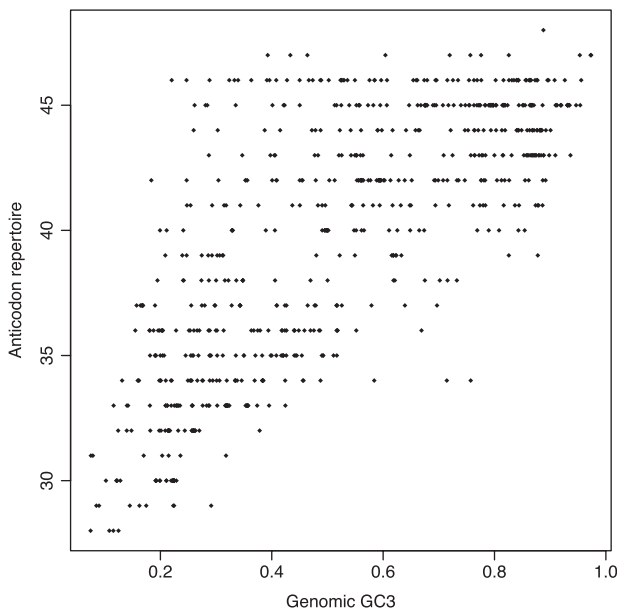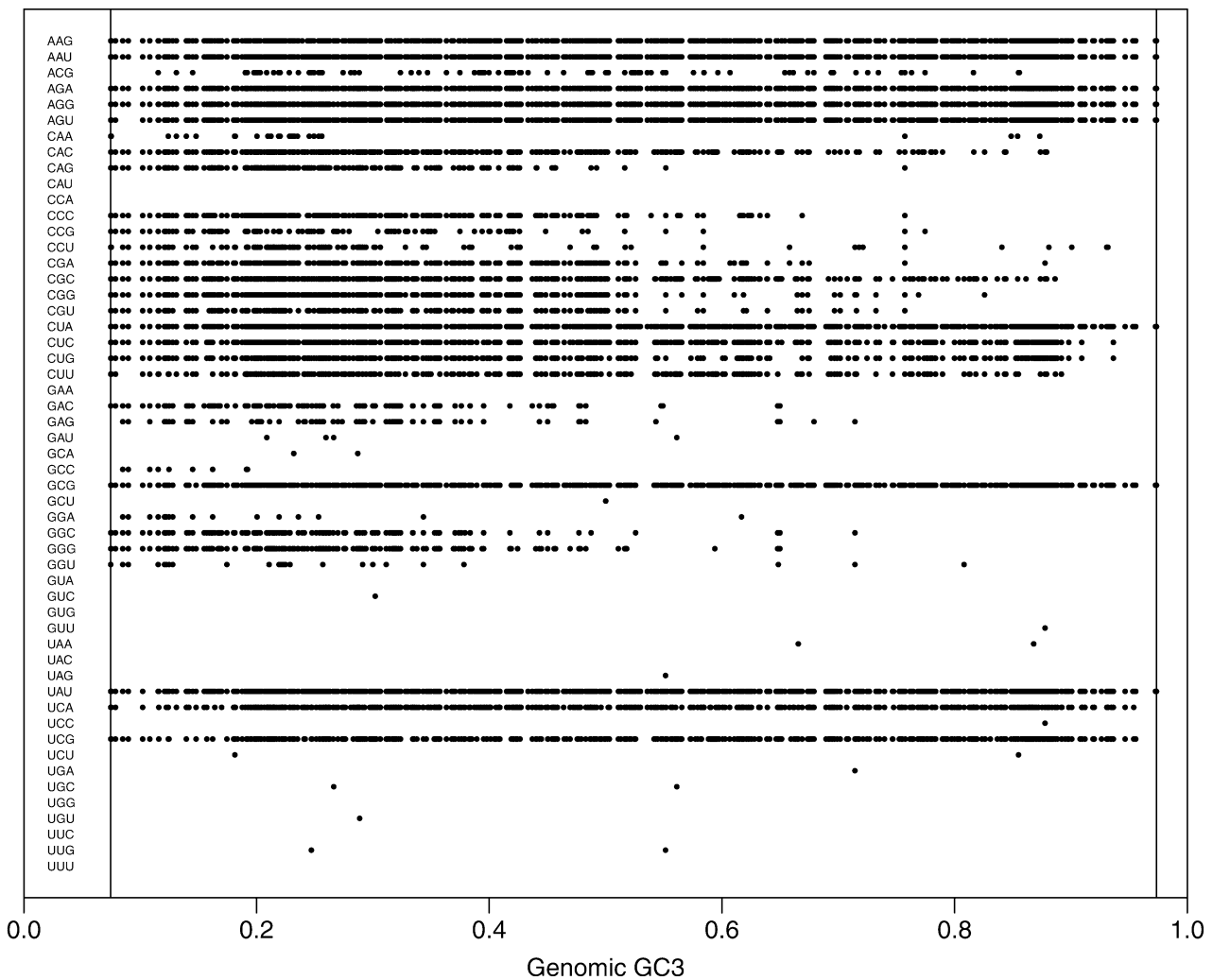
FIG. 5.—Anticodon sparing strategies as a function of GC content. The "absence" of a particular anticodon (rows) from a particular genome (columns, ordered by genomic GC3 content) is indicated to highlight that tRNAs with C or G in the first position of the anticodon are more frequently spared in genomes toward the lower end of the GC spectrum.

genomes with >50% GC (+1: rho = 0.32, $P = 1.91 \times 10^{-08}$; −1: rho = 0.19, $P = 0.0008$, $N = 302$), yet no significant negative trends for genomes with <50% GC (+1: rho = −0.067, $P = 0.24$; −1: rho = 0.014, $P = 0.81$, $N = 312$). Moreover, an exponential fit outperforms a quadratic fit (Akaike's Information Criterion: −8,635 vs. –8,995) suggesting that a model that lacks an increase toward the AT-biased end of the spectrum provides a better description of the data.

## Discussion

The simple model of frameshifting process cost presented above illustrates a number of key issues relevant to assessing the role of frameshifting errors in shaping gene anatomy.

First, the notion that $n_{post}$ is typically short is misleading for genomes with high GC content.

Second, depending on the evolutionary question under consideration, arguments concerning the likely costliness of frameshifting have been focused on either $n_{pre}$ or $n_{post}$. But it is important to acknowledge that frameshifting incurs a compound cost ($n_{pre} + n_{post}$), which distinguishes this particular translational error from, for example, premature termination or drop-off errors, which only incur $n_{pre}$. Such differences in cost dynamics can be exploited to attribute signatures of selection for translational accuracy to specific error classes. We explore these differences in the context of process costs because the link between the length of an erroneous polypeptide and its fitness cost should be linearly proportional. This does not imply, however, that product cost is unrelated to length. In fact, it seems likely that longer frameshifted tracts will on average also be less likely to be soluble and, consequently, have a greater potential to be
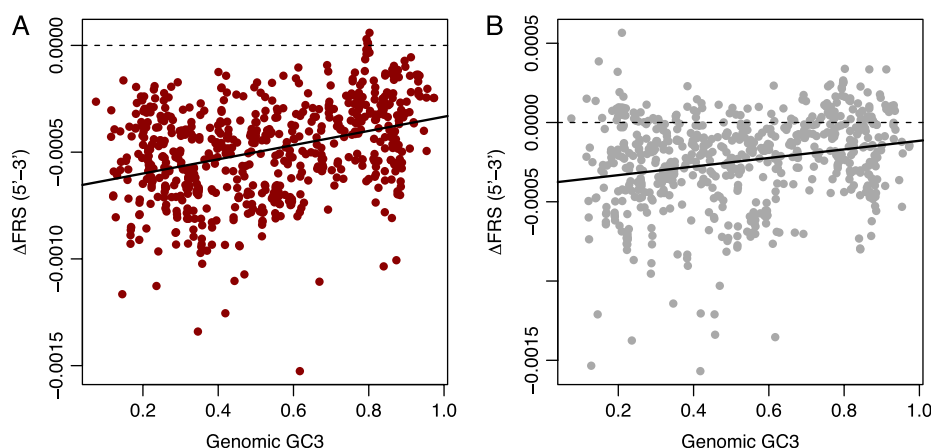
**Fig. 6.**—Differences in frameshifting robustness become less pronounced with increasing GC3. Frameshifting robustness scores (FRSs) were computed as $1 - p_i$ for each of the 5′ and 3′ terminal 100 codons and averaged across codons. Averages were then subtracted (5′ − 3′; pairwise by gene) to determine differential robustness across gene ends. Both for (*A*) +1 and (*B*) −1 frameshifts, mean differential robustness (computed across all sites and all genes) approaches zero, consistent with decreasing differences in process cost in high-GC genomes (see main text).

disruptive, although—in contrast to process costs—the specific amino acid context will be critically important in this regard. Thus, high-GC genomes are likely faced with higher per-error product costs as well as process costs.

Third, comparative genomic analysis of frameshifting costs reveals that considering mRNA sequences in isolation and ignoring vital differences in translational machineries between genomes will produce a deceptive guide to fitness burden. In order to arrive at a genuine comparative estimate of the selective leverage of translational error, it will be imperative to incorporate differences in translation levels between genes and genomes, but our results already highlight the importance of differences in tRNA repertoire for relative susceptibility to translational error. It is intriguing that systematic changes in tRNA repertoire with GC content correlate with a reduction in the expected fitness burden related to frameshifting. But does this imply that differences in tRNA repertoires represent selected adaptations to reduce frameshifting costs or is anticodon diversity under selection for other reasons, for example translational efficiency (Higgs and Ran 2008), and reduction in error rates constitutes a fortuitous side effect? These two explanations are by no means mutually exclusive and might assume different relative importance depending on the lifestyle of the organism under consideration. For example, one would expect translational efficiency to be relatively more important in r-selected species where fast growth is critical for fitness. Fundamentally, the answer to this question will hinge on accurate quantitative determination of fitness costs of erroneous versus slow protein production.

Although our results clearly demonstrate that the link between process costs and GC content is readily transformed by differences in the translational apparatus, more concrete quantitative aspects of the current model should be interpreted with caution. For example, is the higher cost of −1 frameshifting in high-GC genomes real or rather an indication that the model does not incorporate an important determinant of frameshifting dynamics? Although it is conceivable that GC genomes find it intrinsically hard to reduce the cost of frameshifting errors and therefore genuinely shoulder a greater fitness burden in relation to frameshifting, it remains a distinct possibility that this cost is not actually incurred because high-GC genomes exhibit certain (adaptive) features in *cis* or *trans* which our model fails to capture. Notably, we adhere to prokaryotic consensus rules for anticodon:codon interactions proposed by Grosjean et al. (2010) to model binding stabilities and therefore propensities for frameshifting (see supplementary methods, Supplementary Material online), principally because this allows us to compare cost estimates across genomes. These rules are inevitably generalizations because decoding capacities cannot be perfectly predicted from sequence information alone. Importantly, anticodon residues themselves as well as tRNA nucleotides outside the anticodon loop can be posttranscriptionally modified in a variety of ways, with marked effects on decoding capacity (Cochella and Green 2005; Daviter et al. 2006; Grosjean et al. 2010) and/or translational fidelity (reviewed in Saks and Conery 2007), explicitly including reading frame maintenance (Qian and Björk 1997; Björk et al. 1999; Herr et al. 1999; Urbonavicius et al. 2003). Decoding accuracy is further affected by variation in other components of the translation machinery. This includes nucleotide substitutions or modifications in ribosomal RNA, which can cause more or less accurate decoding (Rodnina and Wintermeyer 2001; Baxter-Roshek et al. 2007). In addition, differences in cellular environment, notably $Mg^{2+}$ ion concentrations (Gromadski and Rodnina 2004), can affect translation kinetics with implications for

accuracy, proofreading behavior, and anticodon:codon affinities. Finally, we characterize accidental frameshifting as a local error, solely dependent on interactions at the focal codon and its immediate upstream or downstream neighbor. However, it is apparent from the analysis of programmed frameshifts that downstream secondary structure (hairpins, pseudoknots, etc.) in particular can dramatically affect the rates of shifting, probably at least in part by affecting ribosomal progression and thus residency at a given site (Farabaugh 1996).

Despite these various simplifications and uncertainties, however, the results presented here reinforce the notion that translational errors have been an important force in shaping mRNA anatomy and further suggest that selection might have shaped tRNA repertoires to reduce frameshifting errors.

## Supplementary Material

Supplementary methods, figures 1–2, and table are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

## Acknowledgments

## Literature Cited

Abe T, et al. 2009. tRNADB-CE: tRNA gene database curated manually by experts. Nucleic Acids Res. 37:D163–D168.

Atkins JF, Gesteland RF, Reid BR, Anderson CW. 1979. Normal tRNAs promote ribosomal frameshifting. Cell. 18:1119–1131.

Baxter-Roshek JL, Petrov AN, Dinman JD. 2007. Optimization of ribosome structure and function by rRNA base modification. PLoS One. 2:e174.

Björk GR, et al. 1999. Transfer RNA modification: influence on translational frameshifting and metabolism. FEBS Lett. 452:47–51.

Cochella L, Green R. 2005. An active role for tRNA in decoding beyond codon:anticodon pairing. Science. 308:1178–1180.

Curran JF, Yarus M. 1989. Rates of aminoacyl-tRNA selection at 29 sense codons in vivo. J Mol Biol. 209:65–77.

Daviter T, Gromadski KB, Rodnina MV. 2006. The ribosome's response to codon-anticodon mismatches. Biochimie. 88:1001–1011.

Dong H, Nilsson L, Kurland CG. 1996. Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. J Mol Biol. 260:649–663.

Drummond DA, Wilke CO. 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell. 134:341–352.

Drummond DA, Wilke CO. 2009. The evolutionary consequences of erroneous protein synthesis. Nat Rev Genet. 10:715–724.

Eyre-Walker A. 1996. Synonymous codon bias is related to gene length in Escherichia coli: selection for translational accuracy? Mol Biol Evol. 13:864–872.

Farabaugh PJ. 1996. Programmed translational frameshifting. Annu Rev Genet. 30:507–528.

Farabaugh PJ, Björk GR. 1999. How translational accuracy influences reading frame maintenance. EMBO J. 18:1427–1434.

Fluitt A, Pienaar E, Viljoen H. 2007. Ribosome kinetics and aa-tRNA competition determine rate and fidelity of peptide synthesis. Comput Biol Chem. 31:335–346.

Gallant J, Lindsley D. 1993. Ribosome frameshifting at hungry codons: sequence rules, directional specificity and possible relationship to mobile element behaviour. Biochem Soc Trans. 21:817–821.

Gallant JA, Lindsley D. 1992. Leftward ribosome frameshifting at a hungry codon. J Mol Biol. 223:31–40.

Gromadski KB, Rodnina MV. 2004. Kinetic determinants of high-fidelity tRNA discrimination on the ribosome. Mol Cell. 13:191–200.

Grosjean H, de Crecy-Lagard V, Marck C. 2010. Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. FEBS Lett. 584:252–264.

Herr AJ, Atkins JF, Gesteland RF. 1999. Mutations which alter the elbow region of tRNA2Gly reduce T4 gene 60 translational bypassing efficiency. EMBO J. 18:2886–2896.

Higgs PG, Ran W. 2008. Coevolution of codon usage and tRNA genes leads to alternative stable states of biased codon usage. Mol Biol Evol. 25:2279–2291.

Huang Y, Koonin EV, Lipman DJ, Przytycka TM. 2009. Selection for minimization of translational frameshifting errors as a factor in the evolution of codon usage. Nucleic Acids Res. 37:6799–6810.

Itzkovitz S, Alon U. 2007. The genetic code is nearly optimal for allowing additional information within protein-coding sequences. Genome Res. 17:405–412.

Kanaya S, Yamada Y, Kudo Y, Ikemura T. 1999. Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of Bacillus subtilis tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. Gene. 238:143–155.

Kolor K, Lindsley D, Gallant JA. 1993. On the role of the P-site in leftward ribosome frameshifting at a hungry codon. J Mol Biol. 230:1–5.

Muto A, Osawa S. 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. Proc Natl Acad Sci U S A. 84:166–169.

Parker J. 1989. Errors and alternatives in reading the universal genetic code. Microbiol Rev. 53:273–298.

Qian Q, Björk GR. 1997. Structural alterations far from the anticodon of the tRNAProGGG of Salmonella typhimurium induce +1 frameshifting at the peptidyl-site. J Mol Biol. 273:978–992.

Qin H, Wu WB, Comeron JM, Kreitman M, Li WH. 2004. Intragenic spatial patterns of codon usage bias in prokaryotic and eukaryotic genomes. Genetics. 168:2245–2260.

Ran W, Higgs PG. 2010. The influence of anticodon-codon interactions and modified bases on codon usage bias in bacteria. Mol Biol Evol. 27(9):2129–2140.

Rocha EP. 2004. Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. Genome Res. 14:2279–2286.

Rodnina MV, Wintermeyer W. 2001. Ribosome fidelity: tRNA discrimination, proofreading and induced fit. Trends Biochem Sci. 26:124–130.

Saks ME, Conery JS. 2007. Anticodon-dependent conservation of bacterial tRNA gene sequences. RNA. 13:651–660.

Shachrai I, Zaslaver A, Alon U, Dekel E. 2010. Cost of unneeded proteins in E. coli is reduced after several generations in exponential growth. Mol Cell. 38:1–10.

Sipley J, Goldman E. 1993. Increased ribosomal accuracy increases a programmed translational frameshift in Escherichia coli. Proc Natl Acad Sci U S A. 90:2315–2319.

Sorek R, Cossart P. 2010. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. Nat Rev Genet. 11:9–16.

Stoebel DM, Dean AM, Dykhuizen DE. 2008. The cost of expression of Escherichia coli lac operon proteins is in the process, not in the products. Genetics. 178:1653–1660.

Stoletzki N, Eyre-Walker A. 2007. Synonymous codon usage in Escherichia coli: selection for translational accuracy. Mol Biol Evol. 24:374–381.

Tuller T, et al. 2010. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell. 141:344–354.

Urbonavicius J, et al. 2003. Transfer RNA modifications that alter +1 frameshifting in general fail to affect -1 frameshifting. RNA. 9:760–768.

Wagner A. 2005. Energy constraints on the evolution of gene expression. Mol Biol Evol. 22:1365–1374.

Wagner A. 2007. Energy costs constrain the evolution of gene expression. J Exp Zool B Mol Dev Evol. 308:322–324.

**Associate editor:** Bill Martin