The evolution and functional significance of the programmed ribosomal frameshift in *prfB*

Authors: Cassidy R. Prince[1], Isabella N. Lin[1], Heather A. Feaga[1]*

Affiliations:

[1] Department of Microbiology, Cornell University, Ithaca, NY 14853

* Corresponding author haf54@cornell.edu

`

Keywords: translation termination; RF-2; *prfB*; frameshifting; ribosome; *Bacillus subtilis*

Abstract

Release Factor 2 (RF2) is one of two peptide release factors that terminate translation in bacteria. In *Escherichia coli*, the gene encoding RF2, *prfB*, contains an in-frame premature RF2-specific stop codon. Therefore, a programmed ribosomal frameshift is required to translate full-length RF2. Here, we investigate the diversity of *prfB* frameshifting through bioinformatic analyses of >12,000 genomes. We present evidence that *prfB* frameshifting autoregulates RF2 levels throughout the bacterial domain since (i) the *prfB* in-frame stop codon is always TGA or TAA, both of which are recognized by RF2, and never the RF1-specific TAG stop codon, and (ii) species that lack the autoregulatory programmed frameshift likely need higher RF2 levels since, on average, they have significantly higher RF2-specific stop codon usage. Overexpression of *prfB* without the autoregulatory frameshift motif is toxic to *Bacillus subtilis*, an organism with intermediate RF2-specific stop codon usage. We did not detect the programmed frameshift in any Actinobacteriota. Consistent with this finding, we observed very low frameshift efficiency at the *prfB* frameshift motif in the Actinobacterium *Mycobacterium smegmatis*. Our work provides a more complete picture of the evolution of the RF2 programmed frameshifting motif, and its usage to prevent toxic overexpression of RF2.

Introduction

Ribosomes decode mRNA using a triplet code, with each codon specifying a single amino acid. In most bacteria, the open reading frame (ORF) is set by the base-pairing of the Shine-Dalgarno (SD) sequence within the mRNA with the anti-Shine-Dalgarno sequence in the 16S rRNA of the small ribosomal subunit (1, 2). This interaction ensures that the start codon is positioned within the P-site during translation initiation. The start codon is recognized by an *N*-formyl-methionyl-tRNA, the large ribosomal subunit joins, and elongation proceeds. Each codon is sequentially decoded by an aminoacylated tRNA molecule until a stop codon is reached. Bacteria have three stop codons: UAG, UGA, and UAA. Translation termination occurs when one of two bacterial release factors (RF1 or RF2) recognizes the stop codon and catalyzes hydrolysis of the bond between the nascent peptide and the P-site tRNA. RF1 can recognize UAG and UAA stop codons, and RF2 can recognize UGA and UAA stop codons (3). After hydrolysis of the peptide from the P-site tRNA, the ribosomal subunits are recycled and can initiate on a new transcript.

Translation occurs with high fidelity (4). However, ribosomes can frameshift from the original ORF, altering the identity of downstreamcodons (5). Frameshifting events are often regarded as dangerous for the cell, as many frameshifting events are unintentional and wasteful, producing off-target proteins that are likely to be degraded. In some genes, ribosomal frameshifting has been evolutionarily selected, or “programmed,” as a unique mechanism of gene regulation or as a mechanism of making two distinct proteins from the same gene (6–8). In cases of programmed ribosomal frameshifting, a frameshift is required to create a full-length protein. Programmed ribosomal frameshifts often have two key components: a pausing site, and a “slippery” sequence that assists with ribosomal frameshifting (9). Slippery sequences commonly include repeated bases (e.g. “AAAAA” or “CUUU”) that allow tRNA recognition in multiple frames (10). Sequence elements like internal Shine-Dalgarno sequences or difficult-to-translate

sequences around the slippery sequence can also encourage ribosome pausing, and thus frameshifting (11–15).

One example of programmed ribosomal frameshifting is found in *prfB*, the gene that encodes the essential protein Release Factor 2 (RF2) (16). Foundational studies have examined the conservation of the *prfB* programmed ribosomal frameshift motif across ~300 bacterial species (16–18). These studies demonstrate that the programmed frameshift motif typically contains an internal Shine-Dalgarno sequence, slippery sequence, and premature in-frame UGA stop codon that can be recognized by RF2 (8, 19, 20) (Fig. 1). When concentrations of RF2 are high, RF2 will terminate translation at the premature stop codon, creating a truncated peptide (21). However, when RF2 concentrations are low, the ribosome will pause at the internal Shine-Dalgarno sequence, then slip into the +1 frame, thereby bypassing the internal stop codon to create full-length, functional RF2 (8). Since it is thought that RF2 levels directly regulate the translation of the *prfB* transcript, this frameshifting mechanism is considered autoregulatory. Outstanding questions include whether the *prfB* frameshift motif is autoregulatory outside of well-studied species like *E. coli*, why some species lack the programmed frameshift motif, and how this unique regulatory mechanism evolved.

Here, we report a survey of 12,751 bacterial genomes across 21 phyla to determine the prevalence and conserved features of programmed ribosomal frameshifting in *prfB*. We find that most genomes encode a *prfB* that requires a ribosomal frameshift to produce full-length RF2. Among these genomes, the frameshift sequence motif is extremely well-conserved, including the identity of the premature stop codon, which is nearly always an RF2-specific UGA stop codon. We did not find any instance of the RF1-specific stop codon within the motif. These data solidify the autoregulatory role of the programmed ribosomal frameshift in *prfB* across the bacterial domain. Consistent with this finding, overexpression of *prfB* lacking the frameshift motif is toxic to *Bacillus subtilis*, which further suggests that autoregulating RF2 is important for fitness. Next, we examined the species that lack the autoregulatory frameshift motif and found that they have

significantly higher RF2 stop codon usage, which may explain why RF2 does not need to be autoregulated in these organisms. Finally, we determined that genomes of the phylum Actinobacteriota completely lack the *prfB* programmed frameshift motif and show that the Actinobacterium *Mycobacterium smegmatis* exhibits an inability to efficiently frameshift at the motif. Cumulatively, our results support the autoregulatory function of the *prfB* frameshift across the bacterial domain, and identify key phyla, including Actinobacteriota, that do not require *prfB* autoregulation.

Results

Sequence elements of the programmed frameshift within *prfB* are hyperconserved

To determine the prevalence of the *prfB* programmed frameshift motif, we analyzed the *prfB* sequences of 12,751 representative bacterial species genomes from the NCBI RefSeq database as annotated by NCBI (22). We identified the *prfB* sequence in each of these genomes and determined whether they contained a premature, in-frame stop codon within the *prfB* reading frame. Of the 12,751 genomes surveyed, 8160 (64.0%) contain a premature stop codon in *prfB*, indicating that these organisms require a programmed frameshift to produce functional RF2. We then aligned these *prfB* sequences to generate a nucleotide sequence logo of the programmed frameshift motif (Fig. 1). Within this motif, the purine-rich internal Shine-Dalgarno sequence is hyper-conserved, with the consensus sequence approaching “GGGGG” (Fig. 1B). An even more highly conserved slippery sequence of “CTTT” occurs 5 nucleotides downstream of the internal Shine-Dalgarno sequence. The last T of the slippery sequence is the first T of the premature stop codon. The cytosine following the premature stop codon is also highly conserved, likely because “TGAC” is a known poor translation terminator, and poor termination would permit more frequent frameshifting (23). This canonical “CTTTGA” sequence is present in nearly every *prfB* frameshift motif. The exceptions included 21 genomes with a long poly-thymine tract in the slippery sequence, in which $tRNA_{Phe}$ would decode the codon before the stop codon instead of $tRNA_{Leu}$ (i.e. “TTTTTGA”). These genomes belong to predominantly low-GC organisms (mean GC of 34.6%) in Aquificota and Gammaproteobacteria. In these genomes, the *prfB* sequence retains the internal Shine-Dalgarno sequence as well as the spacing between the Shine-Dalgarno sequence and the slippery sequence, and the identity of the premature stop codon is TGA.

We also aligned the *prfB* sequences of species that do not contain a premature stop codon to determine whether these sequences retained any elements of the programmed

frameshift motif. In these *prfB* sequences the internal Shine-Dalgarno and slippery sequence are not found in the analogous region of the sequence, suggesting that these elements must be lost in organisms that encode a fully in-frame RF2 (Fig. 1B).

The RF1-specific TAG stop codon is not detected as the premature stop codon in the programmed frameshift motif

Bacteria terminate translation using one of three stop codons: TAA, TAG, or TGA. TAA is recognized by either RF1 or RF2, whereas TAG is RF1-specific and TGA is RF2-specific. To assess the conservation of the RF2-specific stop codon within the programmed frameshift motif, we determined the identity of the premature in-frame stop codon in the 8160 genomes that contain the motif. We found that 98.6% of genomes with the motif contain the RF2-specific TGA stop codon as the premature stop codon ("CTTTGA") (Fig. 2). 1.4% of genomes contain a TAA premature stop codon ("CTTTAA"). Genomes encoding TAA as the premature stop codon are found randomly amongst phyla and retained among strains of a species (Fig. S1). These findings suggest that the TAA codon is poorly tolerated since it does not become fixed within particular clades and that the premature TGA is preferred for RF2 autoregulation (Fig. 2). More importantly, none of the 8160 genomes encode an in-frame RF1-specific TAG stop codon. The extreme prevalence of TGA as the premature stop codon in the motif and the total absence of TAG suggests that the purpose of the programmed frameshift motif is indeed RF2 autoregulation.

Overexpression of RF2 is toxic to *Bacillus subtilis*

The extreme conservation of the RF2-specific premature stop codon in the *prfB* programmed frameshift motif suggests that autoregulation of RF2 levels imparts a strong selective advantage. In *E. coli*, even a minor three-fold increase in RF2 is sufficient to show a modest growth defect (24). *In vitro* and *in vivo E. coli* studies show that increased release factor

concentration leads to increased premature termination at sense codons (24, 25). To test whether RF2 overexpression has a deleterious effect in a Gram-positive organism, we expressed a variant of *prfB* lacking the premature in-frame stop codon under the control of a xylose-inducible promoter in *Bacillus subtilis*. Cells expressing this construct fail to form colonies on plates at 30°C (Fig. 3), whereas cells expressing wild-type *prfB* containing the programmed frameshift motif grow similarly to cells containing empty vector (Fig. 3). These results indicate that abolishing translational regulation of *prfB* has a negative effect on fitness, and that the autoregulation imparted by the frameshift motif is sufficient to control RF2 expression. We also overexpressed RF1 from the same promoter. We found that RF1 overexpression also reduces fitness, but to a far lesser extent than RF2 overexpression.

We next tested whether removing the frameshift motif in *prfB* at its native chromosomal locus in *B. subtilis* would impact fitness. Cells missing the frameshift motif in *prfB* grow comparably to cells with wild-type *prfB* in LB at 30°C and 37°C (Fig. S2), suggesting that *prfB* autoregulation is not important for *B. subtilis* fitness under these conditions. Nevertheless, our overexpression data indicate that there may be conditions or a concentration threshold where uncontrolled expression of RF2 is detrimental and that the programmed frameshift is sufficient to autoregulate RF2 even at high levels of overexpression.

The programmed frameshift in *prfB* is broadly conserved in bacteria, but completely absent from Actinobacteriota

Next, we determined the phylogenetic relationship between species that do not use the programmed frameshift motif in *prfB* to autoregulate RF2 expression. The frameshifting motif is distributed widely across bacterial phyla (Fig. 4). In 14 out of 19 phyla with >10 available genomes, >50% of species contain the premature stop and frameshifting motif (Fig. 5A), demonstrating strong conservation of the programmed frameshift. A major exception is

Actinobacteriota. We surveyed 2658 genomes in this phylum and did not detect a premature stop codon or frameshifting motif in *prfB* in any of these genomes. These data suggest that the common ancestor of Actinobacteriota lacked the motif.

Other phyla in which less than 20% of genomes contained the frameshift motif include Mycoplasmatota, Thermotogota, and Aquificota + Campylobacterota + Deferribacterota (Fig. 5A). Interestingly, species in Mycoplasmatota, a phylum closely related to Firmicutes, utilize nearly zero TGA stop codons (3), and an early report indicates that *prfB* is missing from the *Mycoplasmoides genitalium* genome (26). Instead, several *Mycoplasmoides* species decode TGA codons as tryptophan (27). Our genome database contained 199 Mycoplasmatota genomes, but only 21 genomes that met CheckM contamination standards contained *prfB*. Therefore, a lack of *prfB* is common among Mycoplasmatota genomes. None of these *prfB* sequences contain the frameshift motif. Species that do not encode *prfB* include important pathogens like *Mycoplasmoides genitalium*, *Mycoplasmoides pneumoniae*, and *Metamycoplasma hominis*. Mycoplasmatota species with *prfB* are predominantly found within the Acholeplasmataceae family.

The taxonomic rank at which organisms have the frameshifting motif varies from phylum to species. In phyla in which the majority of genomes contain the motif (e.g. Betaproteobacteria), it is likely and parsimonious that the common ancestor of the phylum contained the motif, and that the motif was lost in various recent lineages. Altogether, these data suggest that RF2 autoregulation was present in the common ancestor of bacteria.

Genomes lacking the programmed frameshifting motif in *prfB* have higher GC content and more TGA stop codon usage

We next explored genome characteristics of organisms that do not utilize the programmed frameshift in *prfB*. We found that genomes that lack the programmed frameshift motif have significantly higher GC content than genomes with the motif (62% average GC, $p <$

2.2e-16) (Fig. 5C). Our finding remains significant even with the removal of the well-represented high-GC Actinobacteriota genomes ($p < 1.1e-07$) (Fig. S3). GC content positively correlates with RF2-specific TGA codon usage (3) (Fig. S4). Therefore, we hypothesized that organisms lacking RF2 autoregulation would also encode more RF2-specific stop codons. To test this, we compared terminal stop codon usage between genomes with and without the programmed frameshift for a random subset of 1000 genomes. Genomes that lost the *prfB* frameshift motif have significantly higher RF2-specific TGA terminal stop codon usage than genomes that retained the motif ($p < 2.2e-16$). Again, our findings are significant even when Actinobacterial genomes are excluded ($p < 2.2$ e-06) (Fig. S3). Therefore, a higher demand for RF2 due to increased RF2-specific TGA stop codon usage may explain the loss of the RF2 autoregulation in some species.

Ribosomal frameshifting is inefficient at the *prfB* programmed frameshift motif in the Actinobacterium *Mycobacterium smegmatis*

No Actinobacterial genomes that we surveyed (n = 2658) contain the *prfB* frameshift. We hypothesized that Actinobacteriota may be unable to frameshift at the motif. To assay frameshifting efficiency, we compared the frameshifting efficiencies of an organism that natively contains the frameshift motif, *B. subtilis*, and an Actinobacterium that lacks the motif, *Mycobacterium smegmatis*. We designed analogous constructs for *B. subtilis* and *M. smegmatis* that contain two fluorescent proteins separated by the *prfB* frameshift motif from *B. subtilis*, including ~60 bp upstream and downstream of the motif (*prfB* FS motif) (Fig. 6A). In *M. smegmatis*, the fluorescent proteins and regions upstream and downstream of the *prfB* motif were codon optimized for *M. smegmatis* to avoid ribosomes stalling at rare codons. As a control for production of full-length protein we used a construct that was simply a fusion of the two fluorescent proteins without any inserted sequence (no insert). As a control for production of the truncated protein we used a construct identical to the *prfB* FS motif construct including the

premature TGA stop codon but lacking the frameshifting motif elements (*prfB* no FS motif) (Fig. 6A). *B. subtilis* frameshifts at the *prfB* motif with an efficiency of 52.7% ± 2.8% (Fig. 6B). Conversely, *M. smegmatis* frameshifts inefficiently at the motif with an efficiency of 6.6% ± 4.4% (Fig. 6B). As expected, neither organism exhibits strong frameshifting at the *prfB* variant without the frameshifting motif (Fig. 6B).

Interestingly, the no-insert reporter protein levels are ~15 times higher than the other two reporters in *M. smegmatis* (Fig. 6B). The truncated protein product that results from termination at the premature stop codon in *prfB* is thought to be degraded (21). Both the "*prfB* FS motif" and "*prfB* no FS motif" constructs encode the same amino acid sequence up to the premature stop codon. Therefore, we attribute the difference in *M. smegmatis* reporter protein levels to degradation of the protein product that terminates at the premature stop codon, in accordance with the expected fate of truncated *prfB*. More work is needed to determine the precise reason for the low-level expression of the truncated reporter protein. However, regardless of the mechanism, these results suggest that encoding the frameshift motif in *prfB* would be highly detrimental in *M. smegmatis* because it would likely result in a severe decrease in RF2 levels.

Discussion

The mechanism of the programmed ribosomal frameshift in *prfB* is highly characterized as autoregulatory in *E. coli* (8, 19, 20, 28, 29), but few other studies extend to additional organisms (16, 30, 31). In this work, we performed a large-scale bioinformatics study followed by targeted wet-lab characterization to explore the nature and conservation of *prfB* programmed ribosomal frameshifting in diverse bacteria. We expand upon foundational studies (16–18) to show the high sequence conservation and broad phylogenetic distribution of the *prfB* programmed frameshift motif across >12,000 diverse genomes.

Our *prfB* sequence analysis revealed the extreme conservation of the RF2-specific TGA stop codon as the *prfB* premature stop codon, suggesting that the *prfB* programmed ribosomal

frameshifting mechanism is autoregulatory in all bacteria. We note that while the RF1-specific stop codon was never detected in the motif, the stop codon that can be recognized by both RF1 and RF2 (TAA) was used as the *prfB* premature stop codon for a small proportion of genomes (Fig. 2). These genomes were not restricted to one phylum, but rather were present in small clades of bacterial species scattered across the phylogeny, suggesting that in these clades the stop codon mutated from TGA to TAA in a recent ancestor. Crucially, RF2 can also recognize TAA, and therefore even in these species the autoregulatory mechanism would not be lost.

The conservation of the *prfB* programmed frameshift suggests that it imparts a strong selective advantage. Consistent with this prediction, we found that overexpressing RF2 from *prfB* without the programmed frameshift was toxic to *B. subtilis* at 30°C whereas overexpressing RF2 from *prfB* encoding the programmed frameshift caused no noticeable growth defect compared to wild-type cells (Fig. 3). Notably, when expressed from its native locus, the frameshift motif is dispensable in the standard lab conditions we tested. More work is needed to identify conditions tha cause increased transcription and necessitate autoregulation

Why is RF-2 overexpression toxic? Toxicity may be due to premature termination at sense codons, leading to wasteful and potentially toxic truncated peptides. Codon recognition by bacterial release factors is governed purely by kinetics. Thus, increasing the relative concentration of release factor to tRNAs increases the likelihood of a release factor incorrectly recognizing a sense codon and terminating translation (24, 25). Moreover, release factor methylation by PrmC at the conserved GGQ motif increases RF2 stop codon specificity (32, 33). Therefore, uncontrolled expression of RF2 could also result in a greater proportion of unmethylated RF2 in the cell, further increasing promiscuous termination (32). RF2, but not RF1, also serves additional roles in quality control. For example, RF2 binds ArfA on the ribosome to rescue ribosomes at nonstop mRNAs (34). In vitro, RF2 also aids in post peptidyl transfer quality control via premature termination at misincorporated amino acids (35). Therefore, RF2 can terminate translation in contexts outside of canonical translation, indicating potential inherent promiscuity.

Consistent with previous reports (16, 17), we did not identify any Actinobacterial genomes that require a frameshift within *prfB* to make full-length RF2. Persson and Atkins hypothesized that this was due to a major loss event in the Actinobacterial common ancestor (17). Our results support their hypothesis, as it is a parsimonious explanation for the lack of frameshifting in all 2658 Actinobacterial genomes surveyed alongside broad conservation in other phyla. This conclusion is further supported by recent attempts to root the bacterial tree of life, which indicate that Actinobacteriota are not the closest phylum to the proposed root and, therefore, are not the most genetically similar to the last bacterial common ancestor (36, 37). Thus, it is more likely that Actinobacteriota lost the motif than the occurrence of multiple ancient gain events within the tree.

We found that bacteria that have lost the programmed frameshift in *prfB* had significantly higher RF2-specific stop codon usage ($p < 2.2e^{-16}$) (Fig. 5). Even when we excluded Actinobacteriota, which make up a large proportion of genomes that do not encode the motif, this finding was still highly significant ($p = 2.2e^{-06}$)(Fig. S3). Bacterial release factor concentrations correlate with cognate stop codon usage (3, 38). The direction of causality is unknown, but a recent in silico study proposed that release factor concentrations adapted to stop codon usage (39). High GC content strongly correlates with high RF2-specific TGA stop codon usage but not with TAG stop codon usage (Fig S3) (3). Therefore, a likely explanation for RF2 autoregulation loss is that high GC content and high RF2-specific stop codon usage increased the demand for RF2, and RF2 autoregulation was subsequently lost to satisfy this demand.

High levels of RF2 are also predicted to decrease programmed frameshifting efficiency within *prfB* because more RF2 would be available to terminate translation at the in-frame TGA stop codon. We tested the *prfB* frameshifting efficiency in the model Actinobacterium, *Mycobacterium smegmatis,* and found that *M. smegmatis* exhibits low frameshift efficiency at the canonical *prfB* frameshift motif (7% frameshifting efficiency). At present, we cannot determine whether *M. smegmatis* ribosomes are less prone to frameshifting at this motif or whether high RF2 concentrations terminate translation before the frameshift can take place. The only other

organism reported to have similarly low frameshifting efficiency at the *prfB* motif is *Flavobacterium johnsoniae,* which rarely uses Shine-Dalgarno sequences during translation initiation (30, 40). Surprisingly, *F. johnsoniae* retains the *prfB* frameshift motif despite its low frameshifting efficiency. The low efficiency may be tolerated because the *F. johnsoniae* genome has an unusually low proportion of RF2-specific stop codons (7% of stop codons) (30), and therefore a low demand for RF2. Consistent with this hypothesis, organisms that have lost the programmed frameshift in *prfB* use significantly more RF2-specific stop codons than organisms that have retained it (Fig. 5C).

Our work supports a model in which the programmed ribosomal frameshift motif in *prfB* was present in the last common ancestor of bacteria and autoregulates RF2 expression in nearly all bacterial species. In many species that have lost the motif, it is likely that high TGA stop codon usage increased demand for RF2 and so RF2 autoregulation is no longer required. While our work offers a comprehensive survey of the evolution and purpose of RF2 autoregulation, future molecular studies are essential to determine the precise mechanism underlying RF2 mediated toxicity. Moreover, structural studies are needed to yield crucial insights into ribosomal differences that affect frameshifting efficiency at this motif in diverse bacteria.

Data Availability Statement

Data for all genomes surveyed can be found in Table S1. Scripts for data acquisition and analyses are available on GitHub at https://github.com/cassprince/prfB_evolution.

Materials and Methods

Strains and media. All strains were derived from *B. subtilis* 168 *trpC2* and *M. smegmatis* MC^2 155. *B. subtilis* was grown shaking in Lysogeny Broth media at 37˚C, and *M. smegmatis* was grown shaking in 7H9 Middlebrook media at 37°C as indicated. Antibiotics were used at final concentrations of 1x MLS (1 µg/mL erythromycin and 25 µg/mL lincomycin), 5 µg/mL chloramphenicol, and 20 µg/mL kanamycin. Plasmids used in this study are described in Table 1 and novel plasmid sequences are available on the project GitHub at https://github.com/cassprince/prfB_evolution/blob/main/data/plasmid_sequences.fasta. All experiments were performed in biological triplicate.

***prfB* mutant growth curves.** Chromosomal *prfB* mutants were confirmed by whole-genome sequencing (SeqCenter). Cultures grown to log-phase were normalized to an OD600 of 0.005, deposited into Thermo Scientific 96-well flat bottom plates (Cat.No. 167008), and shaken at 2mm amplitude in a BioTek Synergy H1 microplate reader, Gen5 3.11, at 30°C or 37°C. OD600 values were obtained every 15 minutes for 24 hours.

***prfB* overexpression.** Overnight cultures were normalized to an OD600 of 0.05. For spot plates, tenfold serial dilutions were spotted onto LB containing 1x MLS and 0, 1, or 5% xylose and incubated at 30°C for 24 hours. For measurement of overexpression levels, cultures were grown at 30°C to an OD of 1 and induced with 0, 1, or 5% xylose. Cells were harvested and pelleted after 2.5 hours of induction.

Frameshift reporter lysates, western blots, and Coomassie gels. *M. smegmatis* overnight cultures were normalized to an OD600 of 0.2. Cells were then harvested and pelleted after 12 hours. *B. subtilis* overnight cultures were normalized to an OD600 of 0.05. Reporter expression was induced with 1mM IPTG when cultures reached an OD600 of 1. Cells were then harvested and pelleted after 30 minutes of induction. *B. subtilis* and *M. smegmatis* cell pellets were treated with lysis buffer (10 mM Tris pH 8, 50 mM EDTA, 1 mg/mL lysozyme) for 10 minutes at 37°C. *M. smegmatis* cells were further lysed using bead beating for five cycles of 20 seconds at 4350 rpm with 3 minutes on ice between cycles. All lysates were mixed with SDS loading dye, boiled at 90°C for 5 minutes, and cooled on ice. For *M. smegmatis*, the protein levels of the full-length reporter were approximately 15x greater than those of the experimental *prfB* frameshifting reporters based on band intensity. Therefore, the lysates for the full-length reporter strain were diluted 15x to normalize the protein levels between reporters.

Proteins were run on a 12% SDS-PAGE gel for 70 minutes at 150 V. To measure total protein levels, SDS-PAGE gels were stained with Coomassie blue dye for 30 minutes and destained 5 times for 30 minutes. To measure *prfB* overexpression levels and reporter frameshifting levels, proteins were transferred from SDS-PAGE gels to a PVDF membrane (BioRad) for 100 minutes at 300 mAmps. The membrane was blocked in 3% bovine serum albumin (BSA) overnight at 4°C. Anti-FLAG antibody conjugated to horseradish peroxidase (Sigma SAB4200119) was added to the BSA for 1.5 hours at room temperature. The membrane was washed with PBS-T three times for 5 minutes at room temperature and developed with ECL substrate and enhancer (Biorad 170-5060). Band intensities were quantified using ImageJ v1.53k (41). P-values for differences in band intensity were calculated with the R stats v4.2.2 package using a Welch two-sampled t-test.

***prfB* sequence acquisition and analyses.** *prfB* nucleotide sequences were downloaded from representative prokaryotic genomes in NCBI RefSeq as annotated by the NCBI Prokaryotic

Genome Annotation Pipeline (22). The NCBI accession numbers, species names, and taxids for all genomes used can be found in Table S1. *prfB* sequences were aligned using MAFFT v7.453 (42). Python scripts were used to identify premature stop codons by searching for any stop codon that was in-frame but not found in the final three nucleotides of the sequence. The surrounding region was then extracted, and the identity of the stop codon was recorded. If no premature stop codon was found, the expected region of the frameshifting motif (based on multiple sequence alignment) was extracted. The scripts utilized the biopython v1.78 (43) package for sequence manipulation. The extracted regions were then converted to sequence logos using the logomaker v0.8 package (44).

Phylogenetic analyses. 16S rRNA sequences were identified and acquired using BLAST v2.13.0. Sequences were aligned using MAFFT v7.453 (42). The alignments were applied to FastTree v2.1.11 (45) to infer a maximum likelihood tree. Trees were visualized using the ggtree v3.6.2 package (46). FastTree produces unrooted phylogenies, so trees were midpoint rooted using the phangorn v2.11.1 package (47). The simplified tree (Fig. 5A) was produced by randomly selecting a representative genome for each phylum and subsetting the large 16S tree (Fig. 4). The identities of the randomly selected genomes can be found in Table 1. Taxonomic classification was assigned to genomes using the NCBI Taxonomy database (48) and taxonkit v0.14.1 (49). GC content for each genome was downloaded from NCBI. To determine terminal stop codon usage, the coding sequences were downloaded as annotated by NCBI PGAP for a random subset of 1000 genomes. The list of genomes in the subset can be found in Table S2. A novel Python script recorded the last three nucleotides of each coding sequence per genome and utilized the biopython v1.78 package for sequence manipulation. P-values for differences in GC content and terminal stop codon usage between “frameshift” and “no frameshift” genomes were calculated with the R stats v4.2.2 package using a Welch two-sample t-test.

Acknowledgements

HAF and CRP were supported by NIH R35GM147049. CRP was supported by a Graduate Research Fellowship from the National Science Foundation.

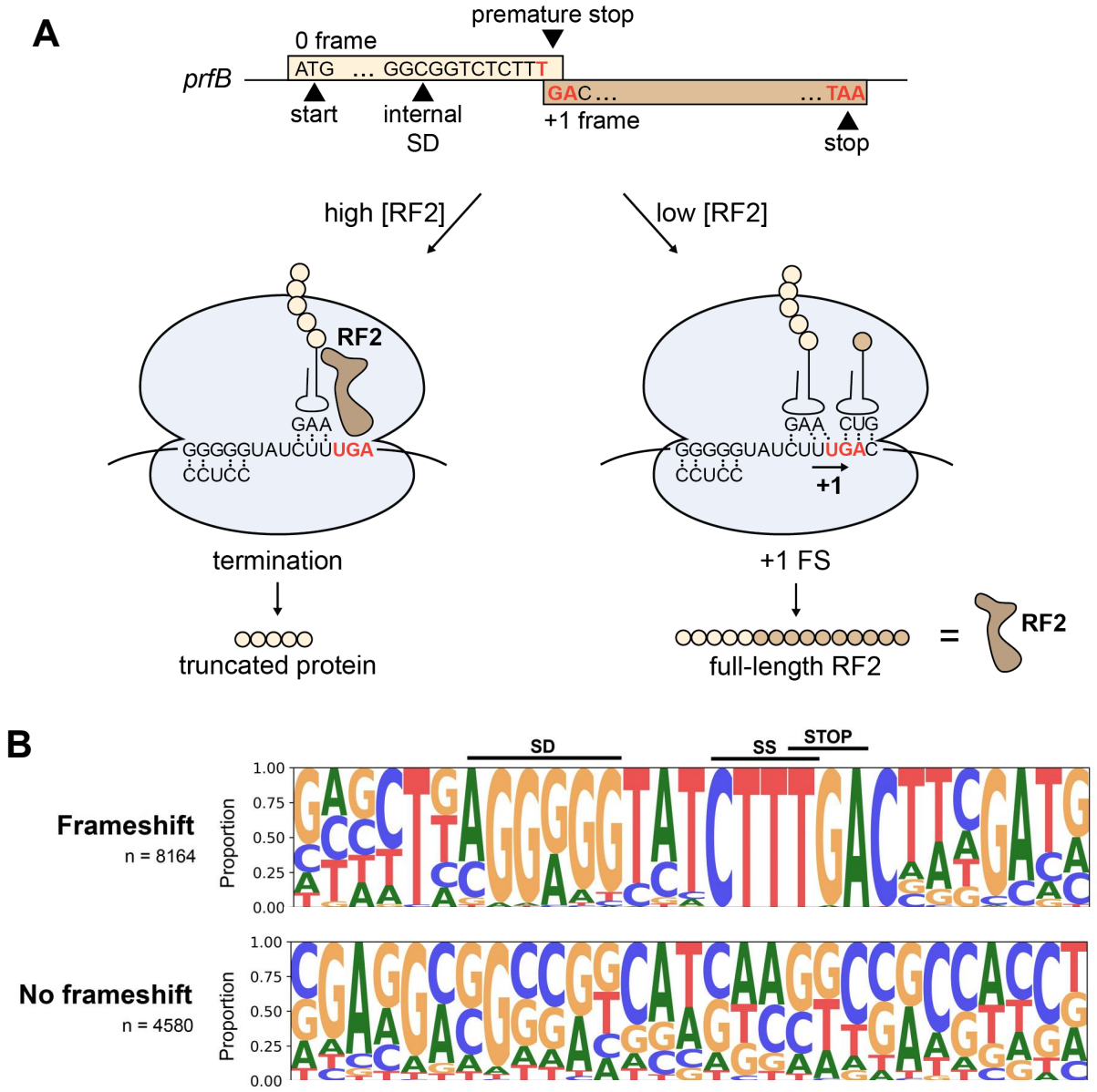

Figure 1: The internal *prfB* Shine-Dalgarno, slippery sequence, and stop codon sequences are highly conserved. (A) Schematic of *prfB* programmed ribosomal frameshift mechanism. (B) Sequence logos of *prfB* frameshifting motif regions for genomes with and without a premature stop codon in *prfB*. The "No frameshift" logo captures the region in which the frameshifting motif is expected based on a large-scale alignment of all *prfB* sequences. Labels above the black lines correspond to the following frameshifting motif components: SD, Shine-Dalgarno; SS, slippery sequence; STOP, internal stop codon.

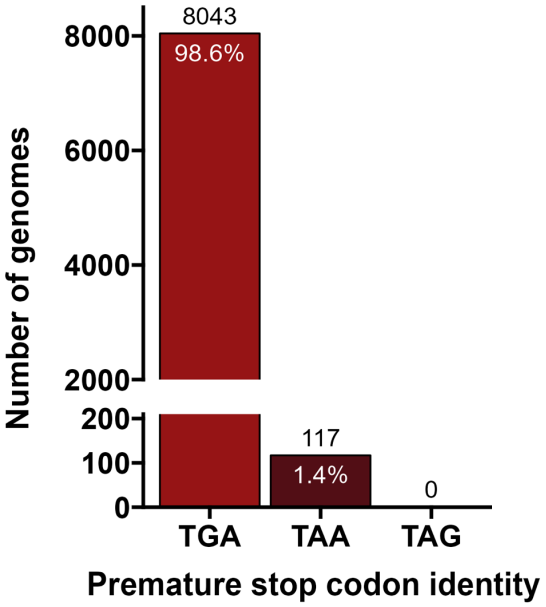

Figure 2: The premature stop codon in the *prfB* frameshifting motif is always recognizable by RF2. Identity of the premature stop codon within the *prfB* frameshifting motif.

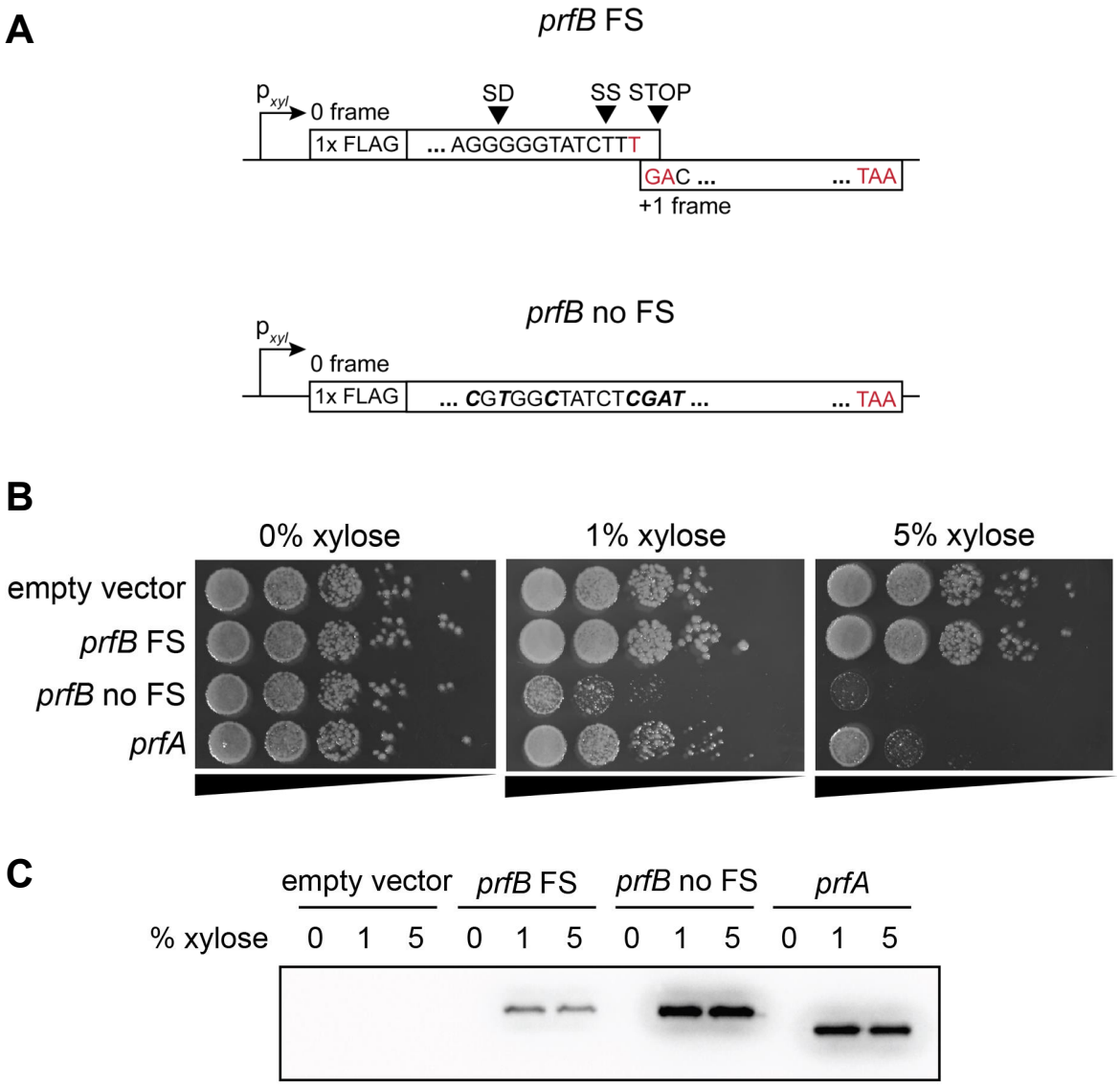

Figure 3: Overexpression of RF2 without the autoregulatory programmed frameshifting motif is toxic to *B. subtilis*. (A) Schematic of *prfB* overexpression vectors in *B. subtilis*. For the *prfB* construct without frameshift (*prfB* no FS) the SD, slippery sequence, and in-frame stop codon have been mutated to eliminate all key elements of the motif. The exact base changes that were made are bold and italicized. (B) *B. subtilis* cells overexpressing *prfB* were serially diluted and plated on varying levels of xylose for induction of *prfB* variant overexpression. Spot plates are representative of three independent biological replicates. (C) Western blot of full-length RF2 variant and RF1 levels during overexpression.

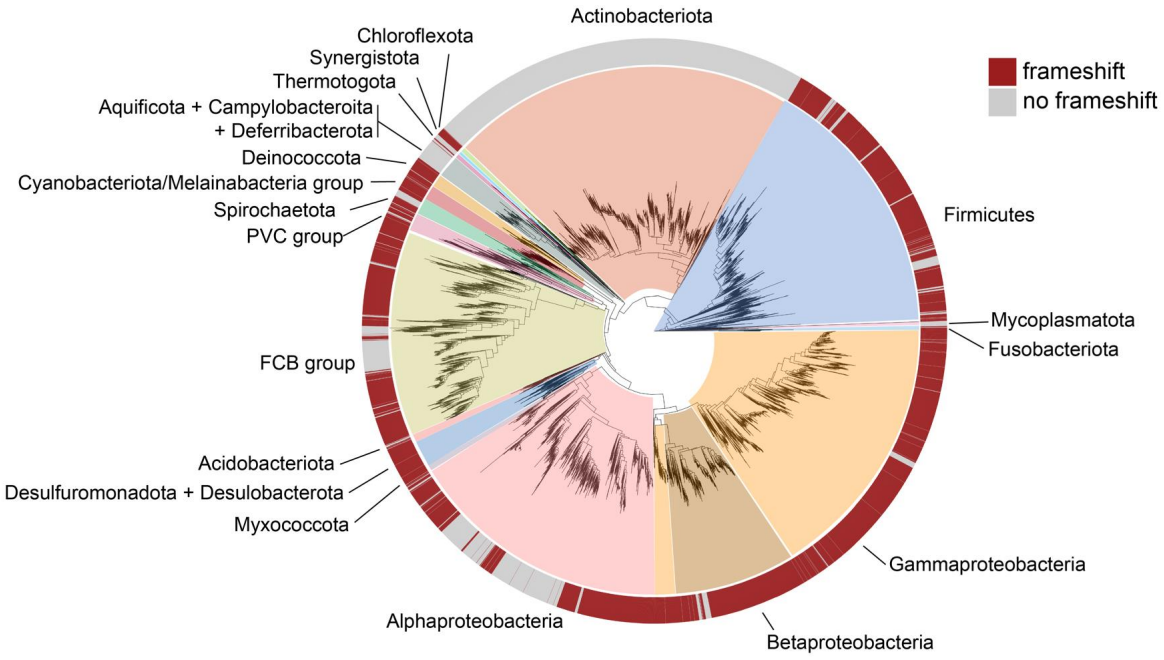

Figure 4: The *prfB* frameshifting motif is broadly distributed across bacterial phyla. A 16S maximum-likelihood phylogenetic tree showing distribution of genomes that encode the programmed frameshift within *prfB*. Phyla with more than 10 available and high-quality reference genomes are shown.

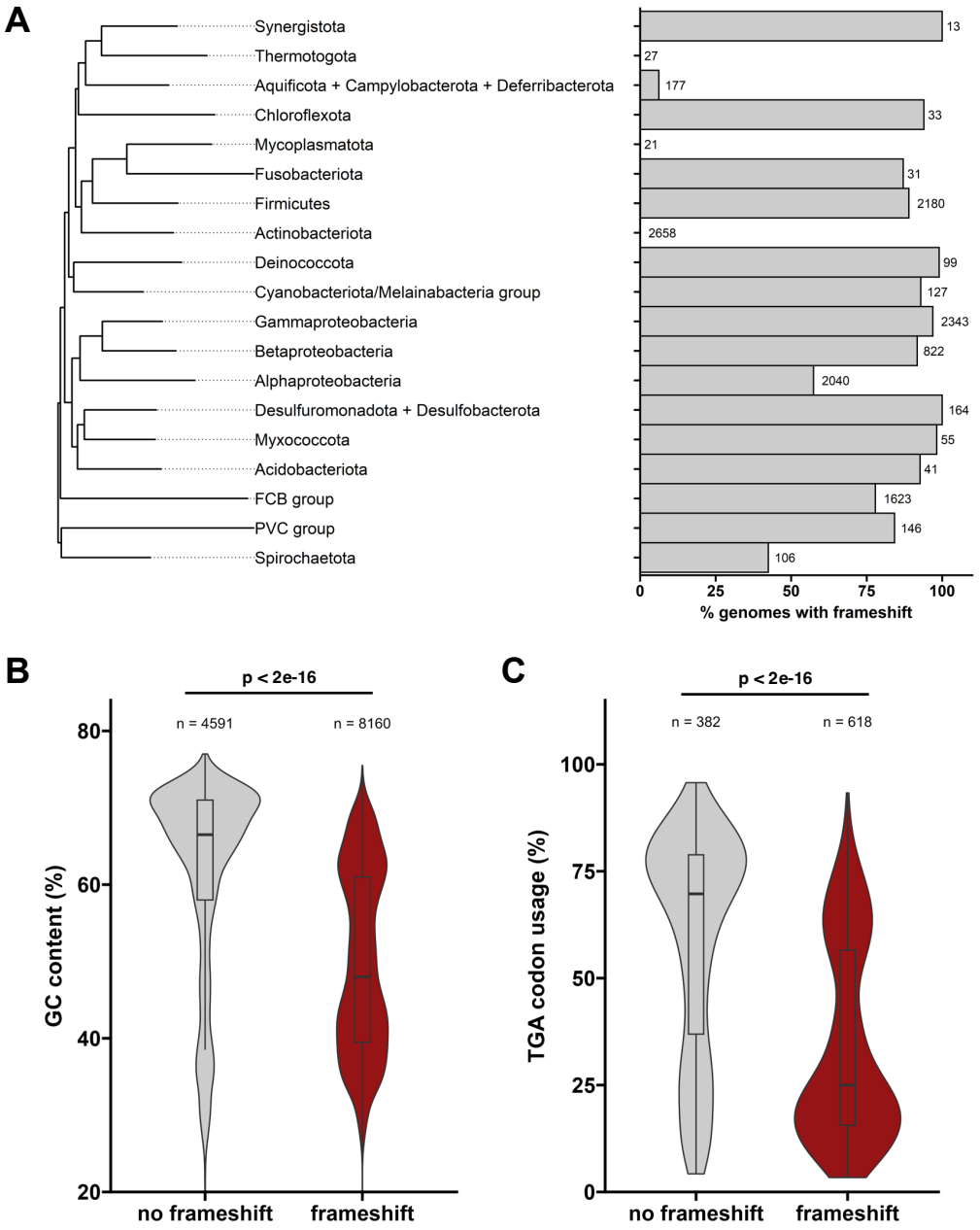

Figure 5: Genomes that encode the programmed frameshift motif within *prfB* have significantly higher GC content and TGA stop codon usage. (A) Percent of genomes containing the *prfB* frameshifting motif per taxon. A subtree of the large 16S tree was created using a random single representative genome for each phylum. The number of analyzed genomes per phylum is located to the right of each bar. (B) GC content of genomes separated by genomes with and without the *prfB* frameshift motif. P-values in B and C indicate the results of a Welch two-sample *t* test. (C) RF2-specific TGA stop codon usage in a random subset of 1000 genomes, separated by genomes with and without the *prfB* frameshift motif.

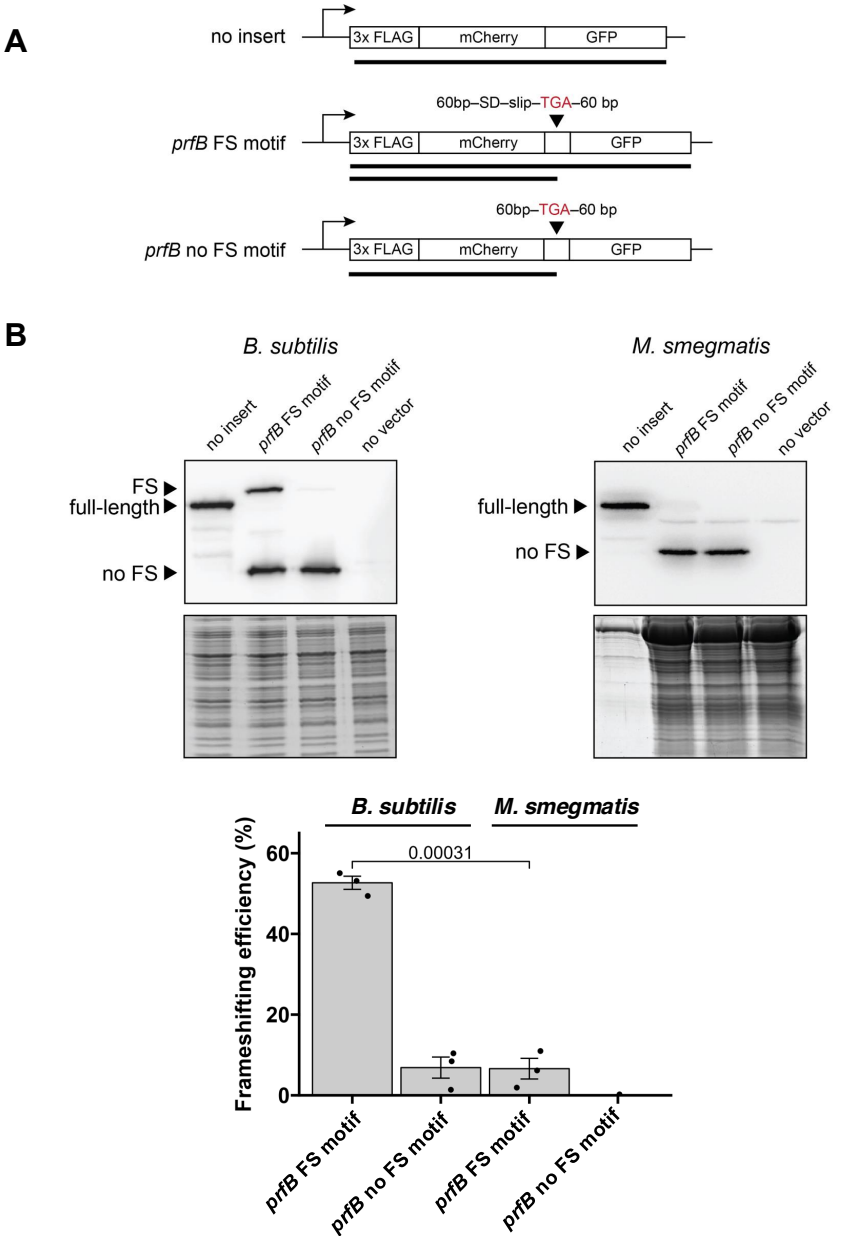

Figure 6. *Mycobacterium smegmatis* cannot efficiently frameshift at the canonical *prfB* frameshifting motif. (A) Schematics for analogous *prfB* frameshifting reporters in *B. subtilis* and *M. smegmatis*. (B) Western blots showing both frameshifted and termination products resulting from expression of the constructs pictured in the schematic. (C) Quantification of western blots. Frameshifting efficiency = frameshifted protein produced/(frameshifted + non-frameshifted protein produced). P-values indicate the results of a Welch two-sample *t* test.

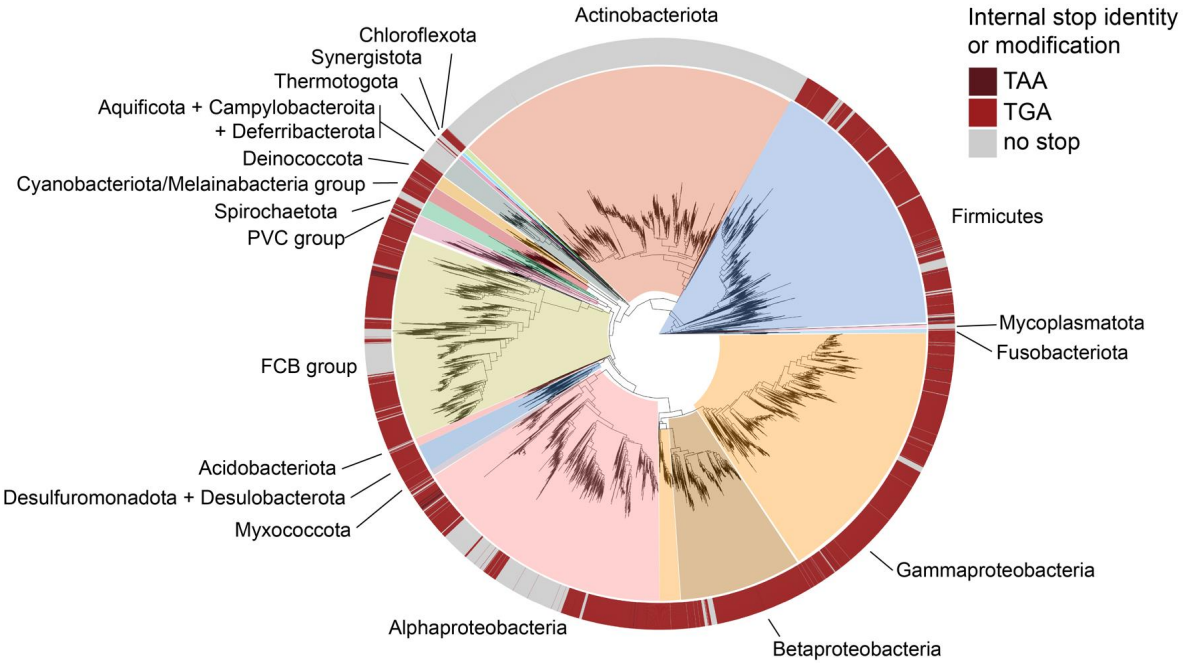

Figure S1. TAA is poorly represented and randomly distributed as the premature stop codon in the frameshifting motif. 16S maximum-likelihood phylogenetic tree from Figure 4 showing the distribution of premature stop codons in *prfB*. Phyla with more than 10 available and high-quality reference genomes are shown.

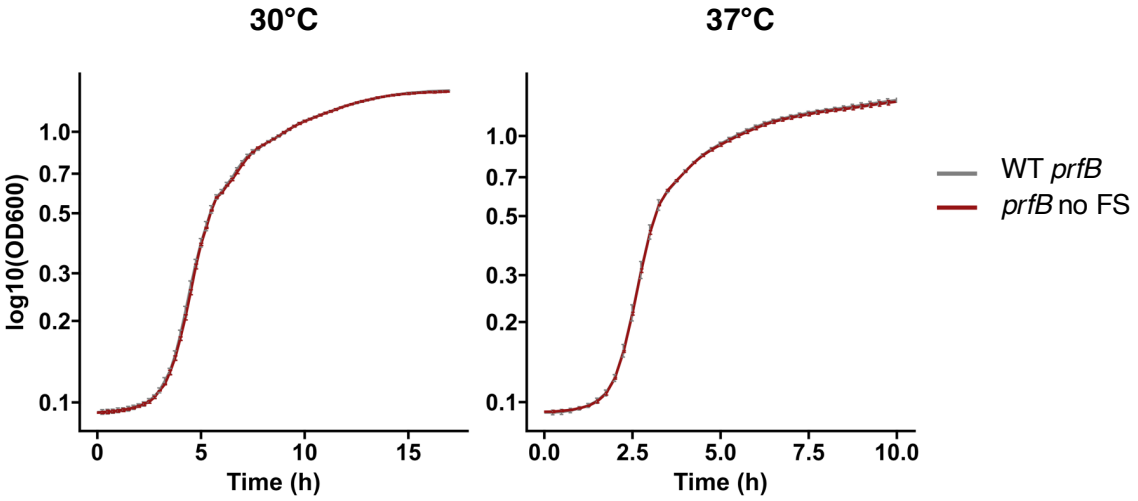

Figure S2. Removal of the *prfB* frameshift motif at the chromosomal locus does not affect fitness. Growth curves of *B. subtilis* encoding *prfB* with and without the frameshifting motif at the chromosomal locus. Growth curves were performed in LB at 30˚C and 37˚C in a plate reader. Error bars indicate standard deviation of biological triplicate.

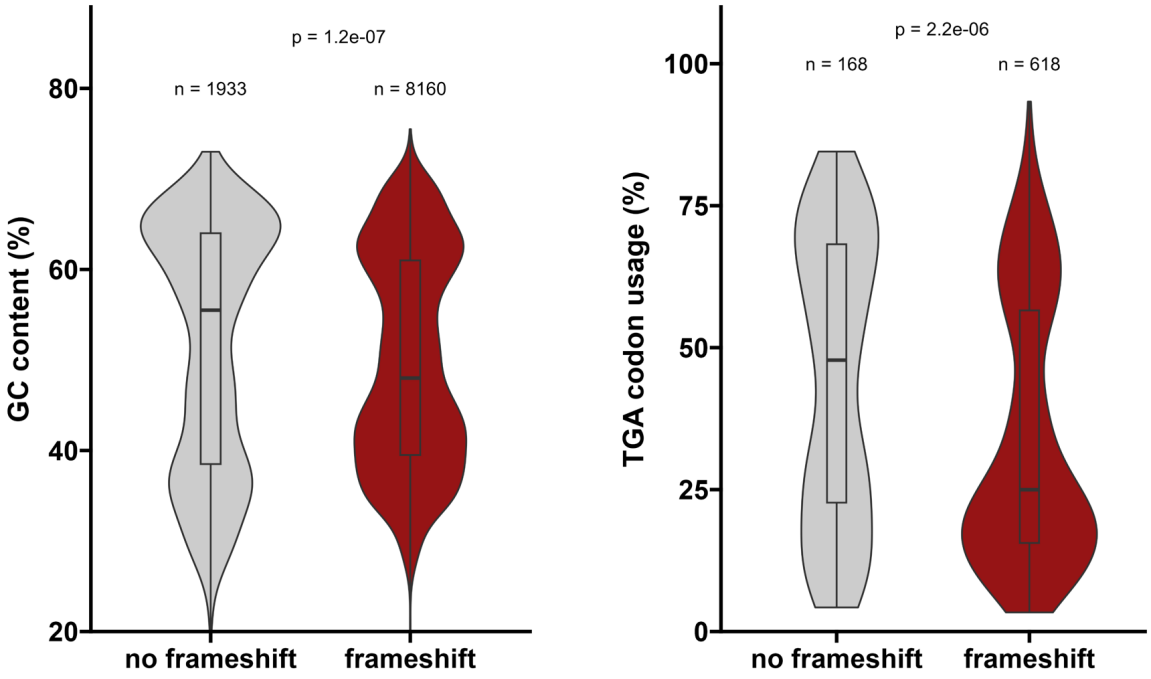

Figure S3. RF2-specific stop codon usage is significantly higher in species that lack the *prfB* frameshifting motif even when Actinobacterial genomes are excluded. Violin plots of genomic GC content and TGA stop codon usage as in Figure 5 but without Actinobacterial genomes. P-values indicate the results of a Welch two-sample *t* test.

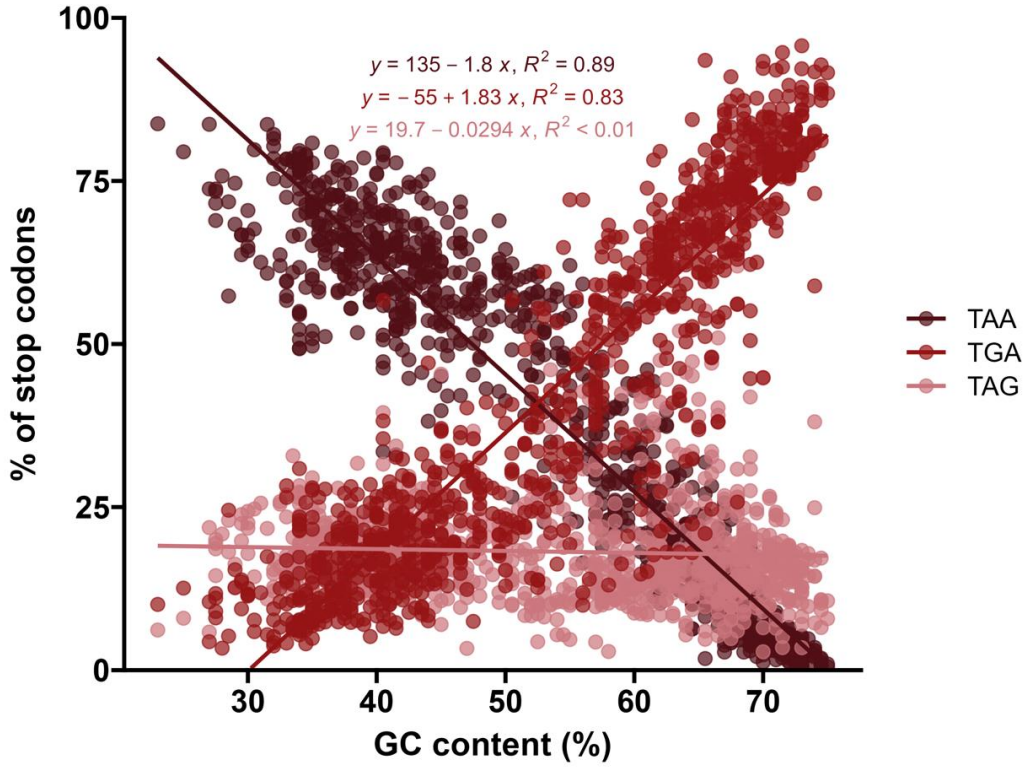

Figure S4. TGA stop codon usage increases with GC content.
Terminal stop codon usage in all coding sequences for a subset of 1000 random reference genomes plotted verses GC content. Each genome has one point for each of the three stop codons.

Table 1

Strain	Description	Source
HAF1	*B. subtilis* wild type 168 trpC2	(50)
HAF477	*M. smegmatis* wild type MC^2 155	Kenneth Keiler
CP203	168 trpC2 ECE743 empty vector	
CP73	168 trpC2 ECE743 P_{xylA}*-1xFLAG-B. subtilis prfB*	
CP215	168 trpC2 ECE743 P_{xylA}*-1xFLAG-B. subtilis prfB with recoded frameshifting motif*	
CP266	168 trpC2 ECE743 P_{xylA}*-1xFLAG-B. subtilis prfA*	
CP201	168 trpC2 *prfB::3xFLAG-B. subtilis prfB*	
CP202	168 trpC2 *prfB::3xFLAG-B. subtilis prfB with recoded frameshift motif*	
CP109	168 trpC2 pHF328 *sacA::*$P_{hyperspank}$*-3xFLAG-mcherry-gfp*	
CP127	168 trpC2 pHF328 *sacA::*$P_{hyperspank}$*-3xFLAG-mcherry-B.subtilis prfB frameshifting motif-gfp*	
CP271	168 trpC2 pHF328 *sacA::*$P_{hyperspank}$*-3xFLAG-mcherry-B.subtilis prfB recoded frameshifting motif-TGA-gfp*	
CP252	MC^2 155 pMV306hsp P_{hsp60}*-3xFLAG-mcherry-gfp* codon optimized for *M. smegmatis*	

CP253	MC^2 155 pMV306hsp P_{hsp60}-*3xFLAG-mcherry-B. subtilis prfB frameshifting motif-gfp* codon optimized for *M. smegmatis*
CP267	MC^2 155 pMV306hsp P_{hsp60}-*3xFLAG-mcherry-B. subtilis prfB recoded frameshifting motif-TGA-gfp* codon optimized for *M. smegmatis*

Plasmid	Description	Source
ECE743	empty vector, ori1030, XylR-P_{xylA} upstream of MCS, amp^r, mls^r (replicative)	(51)
pHF328	pDR111 $P_{hyperspank}$ and MCS cloned into ECE174 backbone at the BamHI and EcoRI sites. Integration at *sacA*.	This study
pRP1028	empty *Bacillus* shuttle vector, I-SceI site, $spec_r$	(52)
pRP1099	facilitator plasmid, *I-SceI*, kan^r (replicative)	(52)
pCP66	ECE743 P_{xylA}-*1xFLAG-B. subtilis prfB* FS (replicative)	This study
pCP209	ECE743 P_{xylA}-*1xFLAG-B. subtilis prfB* no FS (replicative)	This study
pCP254	ECE743 P_{xylA}-*1xFLAG-B. subtilis prfA* (replicative)	This study
pCP103	pRP1028 *3xFLAG-B. subtilis prfB* no FS	This study
pCP104	pRP1028 *3xFLAG-B. subtilis prfB*	This study

pCP105	pHF328 *sacA::*$P_{hyperspank}$*-3xFLAG-mcherry-gfp*	This study
pCP125	pHF328 *sacA::*$P_{hyperspank}$*-3xFLAG-mcherry-B.subtilis prfB FS-gfp*	This study
pCP269	pHF328 *sacA::*$P_{hyperspank}$*-3xFLAG-mcherry-B.subtilis prfB no FS motif-gfp*	This study
pCP234	pMV306hsp P_{hsp60}*-3xFLAG-mcherry-gfp* codon optimized for *M. smegmatis* (replicative)	This study
pCP238 (replicative)	pMV306hsp P_{hsp60}*-3xFLAG-mcherry-B. subtilis prfB FS-gfp* codon optimized for *M. smegmatis*	This study
pCP258 (replicative)	pMV306hsp P_{hsp60}*-3xFLAG-mcherry-B. subtilis prfB no FS-gfp* codon optimized for *M. smegmatis*	This study

Primer	Sequence
CP70-F	5’-GGTGATGTACTTACTATATGAAATAAAATGCATCTGTAGAATTC-3’
CP71-R	5’-GGGCCTCCTTTGATTCGAGGTCAAAGAGA-3’
CP72-F	5’-TCTCTTTGACCTCGAATCAAAGGAGGCCC-3’
CP73-R	5’-CATGATTACGCCAAGCTTGCATGCTTATGAAAGC-3’
CP76-F	5’-ATAACAATTAAGCTTGGAGGAAAAAAAATGGATTATAAAGACGACGACG-3’
CP77-F	5’-ACCTTTAGACAGACCTGAATTCGAGCTCGGTACCC-3’

61	CP78-F	5’-CCGAGCTCGAATTCAGGTCTGTCTAAAGGTGAAGAACTG-3’
62	CP79-R	5’-TTGCATGCGGCTAGCTTATTTGTAGAGCTCATCCATGCCG-3’
63	CP97-F	5’-GATTATAAAGATGATGATGATAAAGAACTTAGCGAGATACGGGC-3’
64	CP98-R	5’-ATTACGCCAAGCTTGCATGCTTATGAAAGCTTAGAACGCAGGTAGG-3’
65	CP95-R	5’-TTTATCATCATCATCTTTATAATCCATTTTTTTTCCTCC-3’
66	CP96-F	5’-GCATGCAAGCTTGGCGTAATC-3’
67	CP120-F	5’-ATGATGATGATAAAGTCGACGTGTTAGACCGTTTAAAATCAATTGAAGAACG-3’
68	CP121-R	5’-ACCATGATTACGCCAAGCTTTTAACCTTCCGACTGCTGAAGCTTGC-3’

69

70 <u>Accession numbers of representative genomes for Figure 5A</u>

71	Acidobacteriota	GCF_003131205.1	Firmicutes	GCF_002243665.1
72	Actinobacteriota	GCF_021183725.1	Fusobacteriota	GCF_000023905.1
73	Alphaproteobacteria	GCF_001458195.1	Gammaproteobacteria	GCF_003852045.1
74	Aquificota + Campylobacterota + Deferribacterota	GCF_005843985.1	Mycoplasmatota	GCF_000397185.1
75	Betaproteobacteria	GCF_007830455.1	Myxococcota	GCF_000280925.3
76	Chloroflexota	GCF_002532075.1	PVC group	GCF_901538355.1
77	Cyanobacteriota/Melainabacteria group	GCF_022848905.1	Spirochaetota	GCF_000758165.1
78	Deinococcota	GCF_000309885.1	Synergistota	GCF_000025885.1
79	Desulfuromonadota + Desulfobacterota	GCF_020886695.1	Thermotogota	GCF_000504105.1
80	FCB group	GCF_003339505.1		

82 References

83

84 1. Shine J, Dalgarno L. 1974. The 3'-terminal sequence of Escherichia coli 16S
85 ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites.
86 Proc Natl Acad Sci U S A 71:1342–1346.

87 2. Steitz JA, Jakes K. 1975. How ribosomes select initiator regions in mRNA: base pair
88 formation between the 3' terminus of 16S rRNA and the mRNA during initiation of
89 protein synthesis in Escherichia coli. Proc Natl Acad Sci U S A 72:4734–4738.

90 3. Korkmaz G, Holm M, Wiens T, Sanyal S. 2014. Comprehensive Analysis of Stop
91 Codon Usage in Bacteria and Its Correlation with Release Factor Abundance *. 44.
92 Journal of Biological Chemistry 289:30334–30342.

93 4. Zaher HS, Green R. 2009. Fidelity at the molecular level: lessons from protein
94 synthesis. Cell 136:746–762.

95 5. Mao Y, Qian S-B. 2024. Making sense of mRNA translational "noise." Semin Cell
96 Dev Biol 154:114–122.

97 6. Flower AM, McHenry CS. 1990. The gamma subunit of DNA polymerase III
98 holoenzyme of Escherichia coli is produced by ribosomal frameshifting. Proceedings
99 of the National Academy of Sciences 87:3713–3717.

100 7. Meydan S, Klepacki D, Karthikeyan S, Margus T, Thomas P, Jones JE, Khan Y,
101 Briggs J, Dinman JD, Vázquez-Laslop N, Mankin AS. 2017. Programmed Ribosomal

102 Frameshifting Generates a Copper Transporter and a Copper Chaperone from the
103 Same Gene. 2. Molecular Cell 65:207–219.

104 8. Craigen WJ, Caskey CT. 1986. Expression of peptide chain release factor 2
105 requires high-efficiency frameshift. 6076. Nature 322:273–275.

106 9. Atkins JF, Loughran G, Bhatt PR, Firth AE, Baranov PV. 2016. Ribosomal
107 frameshifting and transcriptional slippage: From genetic steganography and
108 cryptography to adventitious use. 15. Nucleic Acids Research 44:7007–7078.

109 10. Sharma V, Prère M-F, Canal I, Firth AE, Atkins JF, Baranov PV, Fayet O. 2014.
110 Analysis of tetra- and hepta-nucleotides motifs promoting -1 ribosomal frameshifting
111 in Escherichia coli. 11. Nucleic Acids Res 42:7210–7225.

112 11. Weiss RB, Dunn DM, Atkins JF, Gesteland RF. 1987. Slippery runs, shifty stops,
113 backward steps, and forward hops: -2, -1, +1, +2, +5, and +6 ribosomal
114 frameshifting. Cold Spring Harb Symp Quant Biol 52:687–693.

115 12. Devaraj A, Fredrick K. 2010. Short spacing between the Shine–Dalgarno
116 sequence and P codon destabilizes codon–anticodon pairing in the P site to
117 promote +1 programmed frameshifting. 6. Molecular Microbiology 78:1500–1509.

118 13. Larsen B, Wills NM, Gesteland RF, Atkins JF. 1994. rRNA-mRNA base pairing
119 stimulates a programmed -1 ribosomal frameshift. 22. Journal of Bacteriology
120 176:6842–6851.

121 14. Larsen B, Gesteland RF, Atkins JF. 1997. Structural probing and mutagenic
122 analysis of the stem-loop required for Escherichia coli dnaX ribosomal frameshifting:

programmed efficiency of 50%11Edited By D. E. Draper. 1. Journal of Molecular Biology 271:47–60.

15. Gamper HB, Masuda I, Frenkel-Morgenstern M, Hou Y-M. 2015. Maintenance of protein synthesis reading frame by EF-P and m(1)G37-tRNA. Nat Commun 6:7226.

16. Baranov PV, Gesteland RF, Atkins JF. 2002. Release factor 2 frameshifting sites in different bacteria. 4. EMBO Rep 3:373–377.

17. Persson BC, Atkins JF. 1998. Does disparate occurrence of autoregulatory programmed frameshifting in decoding the release factor 2 gene reflect an ancient origin with loss in independent lineages? 13. J Bacteriol 180:3462–3466.

18. Bekaert M, Atkins JF, Baranov PV. 2006. ARFA: a program for annotating bacterial release factor genes, including prediction of programmed ribosomal frameshifting. 20. Bioinformatics 22:2463–2465.

19. Weiss RB, Dunn DM, Dahlberg AE, Atkins JF, Gesteland RF. 1988. Reading frame switch caused by base-pair formation between the 3′ end of 16S rRNA and the mRNA during elongation of protein synthesis in Escherichia coli. The EMBO Journal 7:1503–1507.

20. Craigen WJ, Cook RG, Tate WP, Caskey CT. 1985. Bacterial peptide chain release factors: conserved primary structure and possible frameshift regulation of release factor 2. Proc Natl Acad Sci U S A 82:3616–3620.

142 21. Márquez V, Wilson DN, Tate WP, Triana-Alonso F, Nierhaus KH. 2004.
143 Maintaining the ribosomal reading frame: the influence of the E site during
144 translational regulation of release factor 2. Cell 118:45–55.

145 22. Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretdin A, Coulouris G,
146 Chitsaz F, Derbyshire MK, Durkin AS, Gonzales NR, Gwadz M, Lanczycki CJ, Song
147 JS, Thanki N, Wang J, Yamashita RA, Yang M, Zheng C, Marchler-Bauer A,
148 Thibaud-Nissen F. 2020. RefSeq: expanding the Prokaryotic Genome Annotation
149 Pipeline reach with protein family model curation. D1. Nucleic Acids Res 49:D1020–
150 D1028.

151 23. Poole ES, Brown CM, Tate WP. 1995. The identity of the base following the stop
152 codon determines the efficiency of in vivo translational termination in Escherichia
153 coli. 1. EMBO J 14:151–158.

154 24. Jørgensen F, Adamski FM, Tate WP, Kurland CG. 1993. Release Factor-
155 dependent False Stops are Infrequent in *Escherichia coli*. Journal of Molecular
156 Biology 230:41–50.

157 25. Freistroffer DV, Kwiatkowski M, Buckingham RH, Ehrenberg M. 2000. The
158 accuracy of codon recognition by polypeptide release factors. Proceedings of the
159 National Academy of Sciences 97:2046–2051.

160 26. Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD,
161 Bult CJ, Kerlavage AR, Sutton G, Kelley JM, Fritchman RD, Weidman JF, Small KV,
162 Sandusky M, Fuhrmann J, Nguyen D, Utterback TR, Saudek DM, Phillips CA,
163 Merrick JM, Tomb JF, Dougherty BA, Bott KF, Hu PC, Lucier TS, Peterson SN,

164 Smith HO, Hutchison CA, Venter JC. 1995. The minimal gene complement of
165 Mycoplasma genitalium. Science 270:397–403.

166 27. Inamine JM, Ho KC, Loechel S, Hu PC. 1990. Evidence that UGA is read as a
167 tryptophan codon rather than as a stop codon by Mycoplasma pneumoniae,
168 Mycoplasma genitalium, and Mycoplasma gallisepticum. J Bacteriol 172:504–506.

169 28. Donly BC, Edgar CD, Adamski FM, Tate WP. 1990. Frameshift autoregulation in
170 the gene for Escherichia coli release factor 2: partly functional mutants result in
171 frameshift enhancement. 22. Nucleic Acids Res 18:6517–6522.

172 29. Weiss RB, Dunn DM, Atkins JF, Gesteland RF. 1987. Slippery runs, shifty stops,
173 backward steps, and forward hops: -2, -1, +1, +2, +5, and +6 ribosomal
174 frameshifting. Cold Spring Harb Symp Quant Biol 52:687–693.

175 30. Naeem FM, Gemler BT, McNutt ZA, Bundschuh R, Fredrick K. 2024. Analysis of
176 programmed frameshifting during translation of prfB in Flavobacterium johnsoniae.
177 2. RNA 30:136–148.

178 31. Lalanne J-B, Parker DJ, Li G-W. 2021. Spurious regulatory connections dictate
179 the expression-fitness landscape of translation factors. Mol Syst Biol 17:e10302.

180 32. Pundir S, Ge X, Sanyal S. 2021. GGQ methylation enhances both speed and
181 accuracy of stop codon recognition by bacterial class-I release factors. Journal of
182 Biological Chemistry 296.

183 33. Dinçbas-Renqvist V, Engström Å, Mora L, Heurgué-Hamard V, Buckingham R,
184 Ehrenberg M. 2000. A post-translational modification in the GGQ motif of RF2 from
185 Escherichia coli stimulates termination of translation. EMBO J 19:6900–6907.

186 34. Demo G, Svidritskiy E, Madireddy R, Diaz-Avalos R, Grant T, Grigorieff N, Sousa
187 D, Korostelev AA. Mechanism of ribosome rescue by ArfA and RF2. eLife 6:e23687.

188 35. Petropoulos AD, McDonald ME, Green R, Zaher HS. 2014. Distinct Roles for
189 Release Factor 1 and Release Factor 2 in Translational Quality Control. 25. J Biol
190 Chem 289:17589–17596.

191 36. Coleman GA, Davín AA, Mahendrarajah TA, Szánthó LL, Spang A, Hugenholtz
192 P, Szöllősi GJ, Williams TA. 2021. A rooted phylogeny resolves early bacterial
193 evolution. 6542. Science 372:eabe0511.

194 37. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ,
195 Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA,
196 Finstad KM, Amundson R, Thomas BC, Banfield JF. 2016. A new view of the tree of
197 life. Nat Microbiol 1:1–6.

198 38. Wei Y, Wang J, Xia X. 2016. Coevolution between Stop Codon Usage and
199 Release Factors in Bacterial Species. Molecular Biology and Evolution 33:2357–
200 2367.

201 39. Ho AT, Hurst LD. 2022. Variation in Release Factor Abundance Is Not Needed to
202 Explain Trends in Bacterial Stop Codon Usage. Molecular Biology and Evolution
203 39:msab326.

204 40. McNutt ZA, Gandhi MD, Shatoff EA, Roy B, Devaraj A, Bundschuh R, Fredrick K.
205 2021. Comparative Analysis of anti-Shine- Dalgarno Function in Flavobacterium
206 johnsoniae and Escherichia coli. Front Mol Biosci 8:787388.

207 41. Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years
208 of image analysis. Nat Methods 9:671–675.

209 42. Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software
210 Version 7: Improvements in Performance and Usability. 4. Molecular Biology and
211 Evolution 30:772–780.

212 43. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I,
213 Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL. 2009. Biopython: freely available
214 Python tools for computational molecular biology and bioinformatics. 11.
215 Bioinformatics 25:1422–1423.

216 44. Tareen A, Kinney JB. 2020. Logomaker: beautiful sequence logos in Python. 7.
217 Bioinformatics 36:2272–2274.

218 45. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – Approximately Maximum-
219 Likelihood Trees for Large Alignments. 3. PLOS ONE 5:e9490.

220 46. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. ggtree: an r package for
221 visualization and annotation of phylogenetic trees with their covariates and other
222 associated data. 1. Methods in Ecology and Evolution 8:28–36.

223 47. Schliep KP. 2011. phangorn: phylogenetic analysis in R. 4. Bioinformatics
224 27:592–593.

48. Schoch CL, Ciufo S, Domrachev M, Hotton CL, Kannan S, Khovanskaya R, Leipe D, Mcveigh R, O'Neill K, Robbertse B, Sharma S, Soussov V, Sullivan JP, Sun L, Turner S, Karsch-Mizrachi I. 2020. NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database (Oxford) 2020:baaa062.

49. Shen W, Ren H. 2021. TaxonKit: A practical and efficient NCBI taxonomy toolkit. Journal of Genetics and Genomics 48:844–850.

50. Gaidenko TA, Kim T-J, Price CW. 2002. The PrpC Serine-Threonine Phosphatase and PrkC Kinase Have Opposing Physiological Roles in Stationary-Phase Bacillus subtilis Cells. J Bacteriol 184:6109–6114.

51. Popp PF, Dotzler M, Radeck J, Bartels J, Mascher T. 2017. The Bacillus BioBrick Box 2.0: expanding the genetic toolbox for the standardized work with Bacillus subtilis. Sci Rep 7:15058.

52. Improvements to a Markerless Allelic Exchange System for Bacillus anthracis | PLOS ONE. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0142758#sec002. Retrieved 7 August 2024.