Translational Selection for Speed Is Not Sufficient to Explain Variation in Bacterial Codon Usage Bias

Saurabh Mahajan* and Deepa Agashe*

National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore, Karnataka, India

*Corresponding authors: E-mails: saurabh.mk@gmail.com; dagashe@ncbs.res.in.

Accepted: January 16, 2017

Abstract

Increasing growth rate across bacteria strengthens selection for faster translation, concomitantly increasing the total number of tRNA genes and codon usage bias (CUB: enrichment of specific synonymous codons in highly expressed genes). Typically, enriched codons are translated by tRNAs with higher gene copy numbers (GCN). A model of tRNA–CUB coevolution based on fast growth-associated selection on translational speed recapitulates these patterns. A key untested implication of the coevolution model is that translational selection should favor higher tRNA GCN for more frequently used amino acids, potentially weakening the effect of growth-associated selection on CUB. Surprisingly, we find that CUB saturates with increasing growth rate across γ-proteobacteria, even as the number of tRNA genes continues to increase. As predicted, amino acid-specific tRNA GCN is positively correlated with the usage of corresponding amino acids, but there is no correlation between growth rate associated changes in CUB and amino acid usage. Instead, we find that some amino acids—cysteine and those in the NNA/G codon family—show weak CUB that does not increase with growth rate, despite large variation in the corresponding tRNA GCN. We suggest that amino acid-specific variation in CUB is not explained by tRNA GCN because GCN does not influence the difference between translation times of synonymous codons as expected. Thus, selection on translational speed alone cannot fully explain quantitative variation in overall or amino acid-specific CUB, suggesting a significant role for other functional constraints and amino acid-specific codon features.

Key words: effective number of codons, tRNA gene copy number, growth rate, rRNA copy number, amino acid usage.

Introduction

Within bacterial genomes, the relative use of synonymous codons differs between highly expressed genes (HEGs) and most other genes. Codon use in most genes is governed by genome-specific nucleotide usage, such that GC-rich organisms typically use GC-rich codons and vice versa (Knight et al. 2001; Chen et al. 2004). However, HEGs are comparatively enriched for specific synonymous codons that are translated by more abundant tRNAs (Ikemura 1981; Kanaya et al. 1999). This enrichment is referred to as codon usage bias (CUB), with greater enrichment indicating stronger CUB. Enrichment of specific codons in HEGs may result from stronger selection on translation in two ways. First, the match between enriched codons and higher abundance of cognate tRNAs can result in faster elongation (Pedersen 1984; Spencer et al. 2012; Dana and Tuller 2014), making ribosome use more efficient (Andersson and Kurland 1990; Klumpp et al. 2012), and therefore providing a growth advantage (Andersson and Kurland 1990; Berg and Kurland 1997; Kudla et al. 2009). Alternatively, higher abundance of cognate tRNAs may also reduce missense errors by outcompeting near-cognate tRNAs (Kramer and Farabaugh 2006; but see Shah and Gilchrist 2010) or reduce nonsense errors (Gilchrist 2007; Stoletzki and Eyre-Walker 2007), thereby avoiding costs related to inaccurate or abortive translation (Gilchrist 2007; Stoletzki and Eyre-Walker 2007; Drummond and Wilke 2008). Such "translational selection" should be stronger for HEGs such as ribosomal proteins and elongation factors, because they constitute a large fraction of the protein mass in rapidly growing cells (Schaechter et al. 1958). At a macroevolutionary scale, growth rate itself varies widely across bacteria (Rocha 2004). With increasing growth rates, the fraction of total proteins contributed by ribosomal proteins and elongation factors also increases (Bremer and Dennis 2008; Scott et al. 2010). Therefore, translational selection on such HEGs is expected to become stronger.

Theoretical models based on the combined effect of CUB and tRNA gene copy number (GCN; a proxy for tRNA abundance) on translation elongation rate predict that these traits should coevolve (Bulmer 1987; Higgs and Ran 2008). Specifically, one detailed model (Higgs and Ran 2008) predicts that the coevolution between CUB and tRNA GCN should be influenced by multiple forces such as translational selection, bias in GC content, and amino acid usage. Comparisons across bacteria have confirmed two predictions of the model: 1) total tRNA GCN and CUB increase with growth associated translational selection (Rocha 2004; Higgs and Ran 2008), and 2) the identity of preferred codons and tRNA genes changes with increasing GC content (Higgs and Ran 2008; Perry 2015). However, the influence of amino acid usage on tRNA copy numbers and CUB has not been rigorously tested. The model predicts that frequently used amino acids should have higher tRNA GCN; and when a single tRNA type (i.e., anticodon) translates multiple codons, amino acids with higher tRNA GCN should have weaker CUB.

These seemingly counterintuitive predictions may be explained as follows. In coding sequences, amino acid use ranges from ∼1% (Cys) to ∼15% (Ala). For any amino acid, the overall benefit of an additional tRNA gene must be proportional to the usage of that amino acid, since gaining a tRNA gene will simultaneously improve the translation time (and/or accuracy) of a proportionate number of sites in coding sequences. Therefore, the number of tRNA gene copies dedicated to an amino acid must be proportional to the usage of that amino acid (Higgs and Ran 2008). How could this lead to a weaker CUB? The strength of selection at any codon site should be proportional to the time gained by using a fast codon instead of a slow codon. Following a simple model (Higgs and Ran 2008), translation time = 1/(translation rate), and translation rate = codon:anticodon pair-specific translation rate constant × tRNA abundance. In the simplest case of 2-fold degenerate amino acids encoded by NNU and NNC codons, both codons are decoded by GNN anticodons, but the NNC codon is decoded faster (Curran and Yarus 1989). Consider two such amino acids A and B, where the usage of B = 2 × usage of A, and tRNA copies for B = 2 × tRNA copies for A. If there are n copies of the GNN tRNA of A, the NNC codon is decoded at rate $k \times n$ and the wobble NNU codon at rate $(k/2) \times n$. Consequently, the time gained by using the faster NNC codon is $(2/(k \times n) - 1/(k \times n)) = 1/(k \times n)$ units. For amino acid B, the translation rate should double for both codons, that is, $k \times 2n$ for NNC codon and $(k/2) \times 2n = k \times n$ for NNU codon. In this case, the time gained is $(1/(k \times n) - 1/2 \times (k \times n)) = 1/2 \times (k \times n)$. As less time is gained in the second case, CUB for amino acid B should be weaker than the CUB for A. To generalize: if additional tRNA copies bear an anticodon that decodes multiple codons, all those codons should be translated faster; the benefit of favoring a particular codon should reduce; and hence CUB must weaken. Note that although increasing selection due to faster growth should apply similarly to all amino acids, this selection should be weaker for amino acids with more tRNA copies. These predictions can be extended to NNA/G codon family amino acids if the promiscuous UNN tRNA is more abundant, and the translation rate constants of the Watson–Crick and wobble pairs are similar to the NNU/C family. However, for 4- or 6-fold degenerate amino acids, it is difficult to predict quantitative trends in CUB, which will depend on the anticodon composition of tRNA copies and relative translation rates of different Watson–Crick and wobble pairs (Wald et al. 2012). Although the copy numbers of anticodon-specific tRNAs can be obtained from the genome, we are not aware of direct measurements of the relative translation rates of all possible codon: anticodon pairs. Therefore, for 4- or 6-fold degenerate amino acids, it is difficult to predict the strength of selection on CUB as a function of tRNA GCN.

Apart from tRNA abundance, other factors may also cause differences in CUB across amino acids. The general prediction for 2-fold amino acids of the NNU/C family assumes an identical translation rate constant (k above) across amino acids, because the codon: anticodon pairs involve the same bases at the wobble position. However, rate constants may differ due to the first two positions in the codon: anticodon pair (Bonekamp et al. 1989; Curran and Yarus 1989) or due to different tRNA modifications. Another possibility is that some codons may be preferred or avoided for peculiar reasons; for example, two of the four Glycine codons are strongly avoided to prevent translational pausing due to their similarity to Shine–Dalgarno like sequences (Li et al. 2012; Diwan and Agashe 2016). In such cases, amino acid-specific CUB may deviate from predictions based on tRNA abundances. Recent studies show that CUB varies across amino acids (Perry 2015), and that amino acid-specific tRNA GCN is positively correlated with amino acid usage (Du et al. 2017). However, the relationship of amino acid-specific CUB with growth rate, amino acid usage or tRNA GCN remains unclear. In addition, quantitative trends in anticodon-specific tRNA copy numbers are unclear, and there are no predictions about variation in amino acid-specific CUB, except in the simplest case of 2-fold degenerate amino acids. Understanding amino acid-specific tRNA copy numbers and CUB should deepen our understanding of translational selection and may also be important for designing optimal protein coding sequences.

Here, we assess the relationship of tRNA GCN, anticodon composition, and CUB with growth rate in a sample of 189 γ-proteobacteria for which genome data are available in public databases. We use rRNA copy number (rRNA CN) as a proxy for growth rate. Several studies suggest that faster growth is strongly associated with selection for higher rRNA CN: soil bacteria with higher rRNA CN form colonies faster on rich media (Klappenbach et al. 2000); *Escherichia coli* with lower rRNA CN have low fitness in nutrient rich conditions (Stevenson and Schmidt 2004; Gyorfy et al. 2015) but are favored in nutrient limited conditions (Gyorfy et al. 2015); and rRNA CN is positively correlated with maximal growth rate in the laboratory (Vieira-Silva and Rocha 2010; Roller et al. 2016) and with various other growth associated traits (Roller et al. 2016).

First, we analyzed average CUB and total tRNA gene copies (N_{tRNA}) as a function of rRNA CN and found discordant patterns for the two traits. Next, we analyzed across-species variation in tRNA GCN, anticodon composition, and CUB separately for each amino acid. Our comparative analysis across amino acids revealed that the impact of growth rate on tRNA GCN increases with amino acid usage, but the impact of growth rate on CUB is not correlated to amino acid usage. We found that some amino acids have weak CUB that does not change with increasing growth rate. Together, our results strongly indicate that factors other than translational speed modulate the strength of selection acting on CUB.

Results

Codon Usage Bias and Total tRNA Gene Numbers Show Distinct Patterns of Increase with rRNA Copy Number

We began by assessing the correlation between rRNA copy numbers (CN) and average magnitude of CUB, and between rRNA CN and total tRNA gene numbers (N_{tRNA}) for a large set of bacterial genomes (~1,000). These genomes from a published bacterial phylogeny (Segata et al. 2013) were chosen to reduce redundancy arising from numerous closely related strains (see Materials and Methods). We quantified CUB using two established metrics: 1) $\Delta ENC'$, the normalized difference in the effective number of codons (ENC′) between highly expressed ("HEG") and all other genes ("other"), accounting for nucleotide usage and averaged across amino acids (Wright 1990; Novembre 2002; Rocha 2004); and 2) S_{CUB}, a selection coefficient based on a population genetic model of CUB (Sharp et al. 2005). As done in previous studies, we defined HEGs as genes coding for ribosomal proteins, elongation factors, and RNA polymerase (Sharp et al. 2005; Higgs and Ran 2008; Vieira-Silva and Rocha 2010). These metrics are positively correlated (supplementary fig. S1*A*, Supplementary Material online) although they account for nucleotide usage in different ways, make different assumptions about selection on CUB, and use data for overlapping but distinct sets of amino acids. As established by previous studies, both CUB and N_{tRNA} are strongly positively correlated with rRNA CN (fig. 1). However, we noticed that the relationship of the two traits and rRNA CN differed in form. A saturating model fits the relationship of CUB and rRNA CN better than a linear model (see Materials and Methods; supplementary table S1, Supplementary Material online). In contrast, N_{tRNA} does not increase in bacteria with low rRNA CN (between 1 and 3), but increases substantially in bacteria with >3 rRNA copies (fig. 1*C*). A piecewise linear model with a breakpoint at rRNA CN = 3 fits this two-regime pattern better than a single linear model (supplementary table S1, Supplementary Material online). These observations suggest that the relative importance of CUB and N_{tRNA} for translation, or other constraints on these traits, may vary with growth rate. To account for the impact of phylogenetic relatedness between species on these correlations, we also assessed correlations among phylogenetically independent contrasts for each trait (see supplementary file, section 1.1, Supplementary Material online). We found weak but statistically significant positive correlations among all trait pairs tested in figure 1 (supplementary fig. S2, Supplementary Material online). In addition, we analyzed above relationships separately for six major bacterial clades (see supplementary file, section 1.2, Supplementary Material online). We found a strong

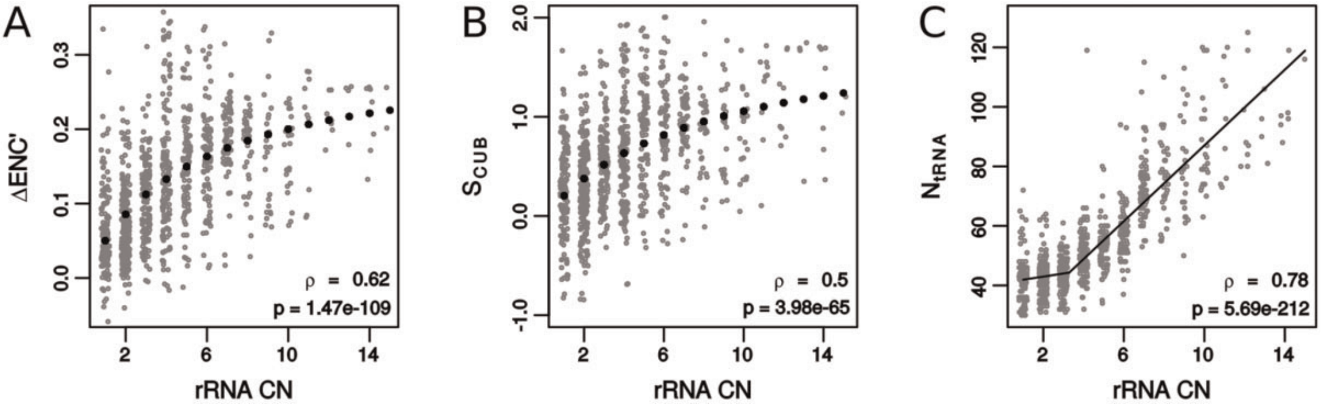

Fig. 1.—Correlations between codon usage bias (CUB), total tRNA gene copies (N_{tRNA}), and rRNA copy numbers (rRNA CN), across bacterial genomes. CUB was measured either as ΔENC′, the normalized difference in effective number of codons between highly expressed and other genes; or as S_{CUB}, a selection coefficient derived from a population genetics model. (*A*) ΔENC′ versus rRNA CN. (*B*) S_{CUB} versus rRNA CN. The fitted line represents a saturating model with three parameters. (*C*) N_{tRNA} versus rRNA CN. Each data point represents one bacterial genome ($n = 964$). We added a small jitter to rRNA CN since numerous data points have the same rRNA CN. ρ values represent Spearman's (nonparametric) correlation coefficients and *P*-values correspond to a one-way asymptotic permutation test for positive correlation. In panel C, lines represent piecewise linear models with two slopes. All fitted models were evaluated based on AIC differences reported in supplementary table S1, Supplementary Material online. Axes ranges were curtailed to magnify trends, causing up to eight data points to fall outside the axis range in each plot. Full data can be found in supplementary figure S1, Supplementary Material online.

correlation between N_{tRNA} and rRNA CN in all clades (supplementary fig. S3, Supplementary Material online), but only γ-proteobacteria showed a strong positive correlation and saturation in both ΔENC′ and S_{CUB}. Therefore, hereafter we focus on 189 genomes of γ-proteobacteria, representing one of the most widely sequenced clades that also exhibits the entire range of rRNA CN observed across bacteria.

The Impact of Growth Rate on tRNA Gene Copy Number Is Stronger for Frequently Used Amino Acids

Next, we turned to amino acid-specific differences in the impact of growth rate on tRNA GCN. Although tRNA GCN for all amino acids showed a strong positive correlation with rRNA CN, the extent and pattern of this association varied substantially (fig. 2). The range of tRNA GCN corresponding to an amino acid varied from 1–3 (His) to 3–20 (Arg). For multiple amino acids (Tyr, Lys, Cys, Val, Thr, Leu, Arg), tRNA GCN remains more or less constant until a threshold rRNA CN, and then increases at higher rRNA CN. This threshold appears to differ across amino acids. Such amino acid-specific early constancy-late increase patterns likely sum up to give rise to the pattern seen for total tRNA gene numbers (compare panels in fig. 2 with the γ-proteobacteria panel in supplementary fig. S3, Supplementary Material online). For Ile and Ala, we observed an exact linear relationship between tRNA GCN and rRNA CN in a subset of genomes. On more systematic investigation, we observed that up to seven tRNA copies of Ile, Ala, and Glu are part of rRNA operons in >50% of analyzed genomes, leading to a correlation between the respective tRNA GCN and rRNA CN. As with other amino acids, this correlation should still reflect growth-associated selection to increase tRNA copy numbers although it arises from a different molecular process. However, we cannot rule out the possibility that in these three cases, the correlation evolves for reasons other than translational selection. More generally, the patterns described here confirm that growth rate associated selection differentially impacts tRNA GCN depending on the amino acid.

As explained earlier, we expected that increasing growth rate should have a stronger impact on tRNA GCN of more frequently used amino acids. To test this, we fit linear regression models to the data shown in figure 2 and estimated two quantities for each amino acid: the tRNA GCN when rRNA CN = 1 (i.e., tRNA GCN in "slowest" growing bacteria), and the slope of tRNA GCN versus rRNA CN (i.e., the magnitude of increase in tRNA GCN per unit rRNA CN). As expected, both parameters were positively correlated ($P < 5\text{e}{-3}$) with amino acid usage in HEGs (fig. 3). This correlation could arise from the confounding effect of amino acid degeneracy since 4- and 6-fold degenerate amino acids are more frequently used and may need more tRNA genes to accommodate the larger number of synonymous codons. However, we also observed positive associations ($P < 0.05$) within the 2-fold degenerate amino acids (red data points in fig. 3). The correlation between tRNA GCN and amino acid usage persists ($P < 1\text{e}{-3}$) even if we ignore the amino acids whose tRNAs are part of rRNA operons. These results support the hypothesis that growth rate-associated translational selection on tRNA GCN increases with amino acid usage.

Specific Anticodons Contribute to Amino Acid-Specific Variation in tRNA GCN

Since amino acid-specific CUB will partly depend on the prevalence of the respective tRNA anticodons, we assessed how specific tRNA (with distinct anticodons) change as a function of increasing rRNA copy numbers. Apart from the well-known absence of ANN tRNAs (except Arg), we found five clear patterns (fig. 4). 1) As expected for 2-fold degenerate amino acids of the NNU/C codon family (and 3-fold degenerate Ile), GNN tRNAs increase in copy number. 2) For amino acids of the NNA/G codon family, both UNN and CNN tRNAs are present, but the UNN tRNAs are prevalent and increase in copy number with growth rate. The CNN tRNAs are absent in fast growing bacteria. 3) For 4-fold degenerate amino acids (except Gly), UNN tRNAs are prevalent and their gene copies increase strongly with growth rate. The GNN tRNAs are also present but their gene copies increase to a smaller extent. The opposite is true for Gly. For all 4-fold degenerate amino acids, CNN tRNAs are either absent or are present in single copies. 4) For 6-fold degenerate Leu, anticodons can be divided into a four-box (NAG) and a two-box (YAA). Among the four-box tRNAs, UAG or CAG anticodons are prevalent in different sets of genomes, while the UAA anticodon is prevalent in the two-box tRNAs. 5) 6-fold degenerate Arg can also be divided into a four-box (NCG) and a two-box (YCU). The ACG anticodon is prevalent among the four-box tRNAs, while the UCU anticodon is prevalent among the two-box tRNAs.

Thus, the NNA/G codon family amino acids are similar to the NNU/C codon family because the tRNAs (UNN) that can decode both codons are prevalent and increase with growth rate. These patterns set up the expectation for an inverse relationship between tRNA GCN and CUB for all these amino acids, tested as described in the following section. In the case of 4-fold amino acids, the UNN and GNN tRNAs can decode multiple overlapping codons (Marck and Grosjean 2002; Weixlbaumer et al. 2007), but their relative efficiencies at decoding different codons have not been comprehensively measured. Hence, unlike 2-fold degenerate amino acids, we cannot predict quantitative differences in CUB across 4- or 6-fold degenerate amino acids. Nonetheless, we analyzed variation in CUB of these amino acids and discuss the potential sources of this variation (see Discussion section).

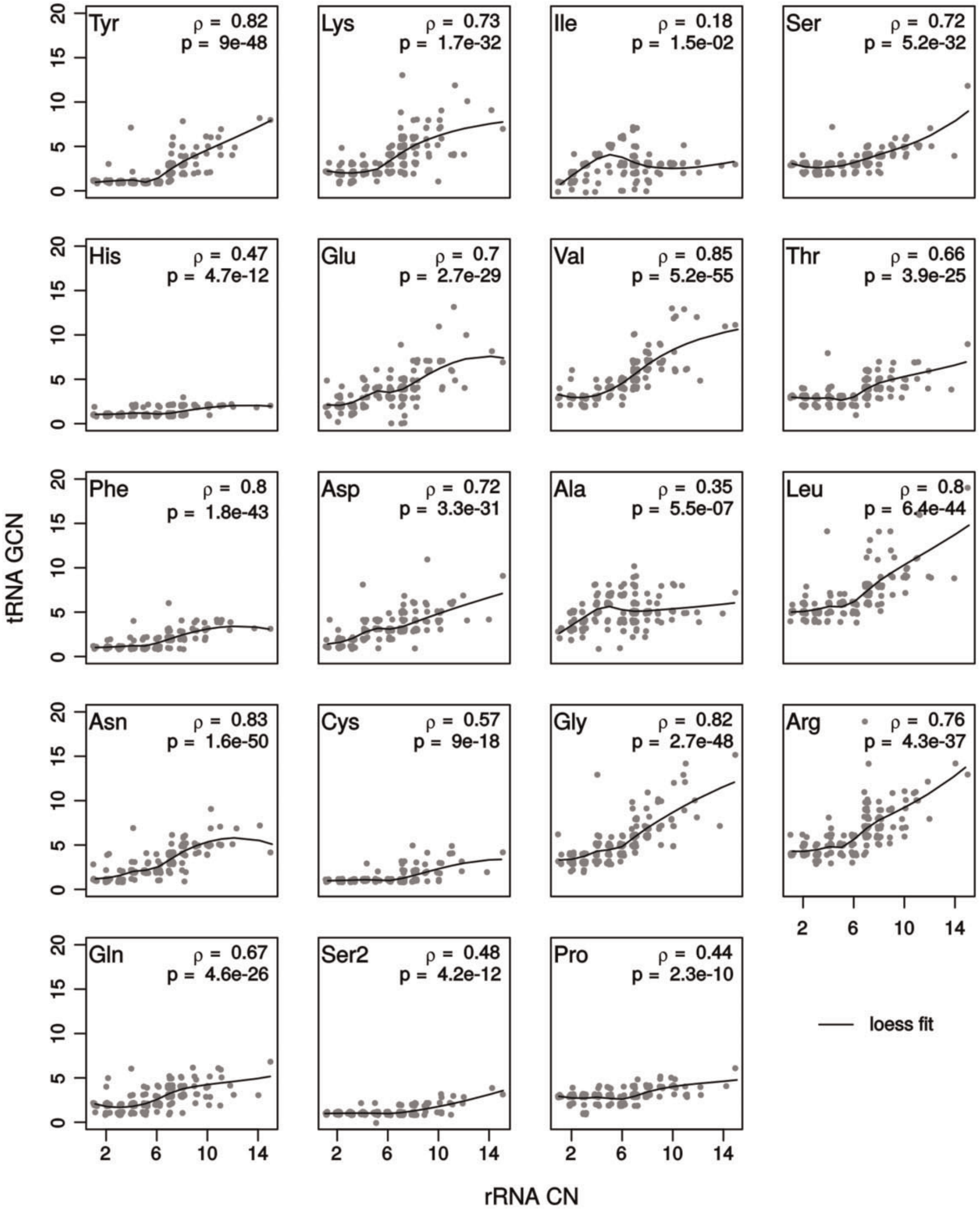

Fig. 2.—Relationship between amino acid specific tRNA gene copy number (tRNA GCN) and rRNA copy numbers (rRNA CN) in γ-proteobacteria. For all amino acids, tRNA GCN increased with rRNA CN. ρ represents the Spearman's correlation coefficient and *P*-values correspond to a one-way asymptotic permutation test for positive correlation. Smoothened Loess fits are shown to highlight trends. Each data point represents one genome (n=189). Since multiple genomes have identical rRNA CN and tRNA GCN, we added a small jitter to both variables. As a result, the size of grey clusters approximates the number of genomes at the same *x*–*y* values.

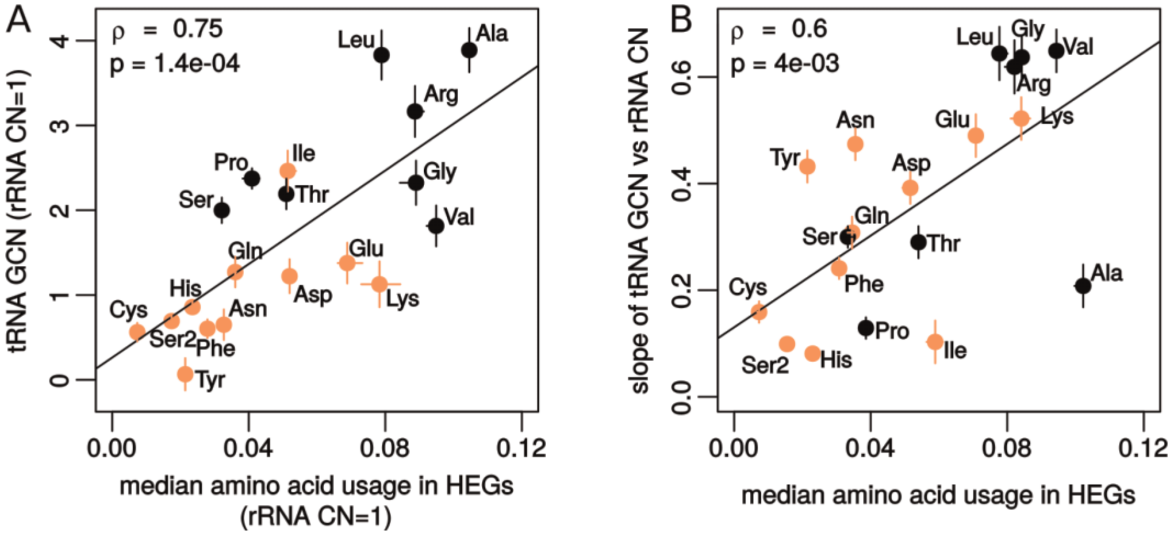

Fig. 3.—Amino acid usage and tRNA gene copy number (tRNA GCN) in γ-proteobacteria. We calculated amino acid usage as the median value (across genomes) of the fraction of coding sites corresponding to a particular amino acid in highly expressed genes (HEGs). Predicted values of tRNA GCN at rRNA CN=1 represent tRNA GCN in the slowest growing bacteria. Slopes represent the increase in tRNA GCN per unit rRNA CN. (*A*) Median amino acid usage and predicted tRNA GCN when rRNA CN=1. (*B*) Median amino acid usage and slope of tRNA GCN versus rRNA CN. Red circles show data for 2- or 3-fold degenerate amino acids and black circles indicate 4- or 6-fold degenerate amino acids. ρ represents Spearman's correlation coefficient and *P*-values correspond to one-way asymptotic permutation test for positive correlation. Vertical bars are standard errors from the linear regression fit, and horizontal bars are interquartile ranges for amino acid usage. In most cases, IQR is smaller than the width of the circles.

The Impact of Growth Rate on CUB Varies across Amino Acids, but Is Not Correlated with Amino Acid Usage or tRNA GCN

To quantify amino acid-specific CUB, we calculated an amino acid-specific version of ΔENC′ (see Materials and Methods and supplementary text, section 1.3, Supplementary Material online) for all amino acids, and S_{CUB} for 2-fold degenerate amino acids (including Ile) as in previous studies (Sharp et al. 2005; Higgs and Ran 2008). The motivation for and limitations of these metrics are discussed in the supplementary text, section 1.3, Supplementary Material online. We first describe general patterns of change in amino acid-specific CUB with growth rate, before addressing the correlation of CUB with amino acid usage or tRNA GCN.

We found substantial variation in the relationship between amino acid-specific ΔENC′ and rRNA CN (fig. 5). There was no correlation between rRNA CN and ΔENC′ for Gln, Lys, Glu, Asp, or Cys ($P > 0.076$); but we observed a moderate positive correlation for Tyr, His, Phe, Gly, Pro, Leu ($0.4 < \rho < 0.6$; $P < 1e{-}8$), and a strong positive correlation for Asn, Ser2, Ile, Val, Ser ($\rho > 0.6$; $P < 1e{-}18$). To avoid potential biases due to the peculiarities of ΔENC′ (supplementary text, section 1.3, Supplementary Material online), we also analyzed the relationship of S_{CUB} with rRNA CN for 2-fold degenerate amino acids. Consistent with patterns in ΔENC′, we observed a strong positive correlation between S_{CUB} and rRNA CN for Tyr, Phe, and Asn; and no correlation for Gln, Lys, and Cys (supplementary fig. S5, Supplementary Material online). Although we observed a positive correlation for Glu and Asp (unlike ΔENC′), the actual S_{CUB} values for these amino acids were consistently smaller than other amino acids across the entire range of rRNA CN. Overall, our results indicate no effect of growth rate on CUB for Gln, Lys, or Cys, and a weak effect for Glu and Asp. Note that these five amino acids with no or weak effect of growth include all 2-fold degenerate amino acids encoded by NNA/G codons (three of three). On the other hand, growth rate had a strong impact on CUB of five of seven amino acids encoded by NNU/C codons.

To test whether the (weak) impact of growth rate on CUB was associated with frequently used amino acids and high tRNA GCN, we quantified the impact of growth rate on CUB as the increase in CUB per unit increase in rRNA CN. We then analyzed two sets of correlations: one between amino acid usage and the impact of growth rate on CUB; and another between impact of growth rate on tRNA GCN and the impact of growth rate on CUB. As explained above, in both cases, we expected to see a negative correlation for 2-fold degenerate amino acids. Contrary to our expectation, we found that the impact of growth rate on CUB was negatively correlated ($P = 0.043$) with amino acid usage only if CUB was quantified as ΔENC′ (fig. 6*A*), but not when it was quantified as S_{CUB} (fig. 6*B*). Moreover, the impact of growth rate on CUB was not correlated with the impact on tRNA GCN in either case (fig. 6*C* and *D*). Hence, the differences in the impact of growth rate on CUB across amino acids were explained neither by amino acid usage nor by tRNA GCN.

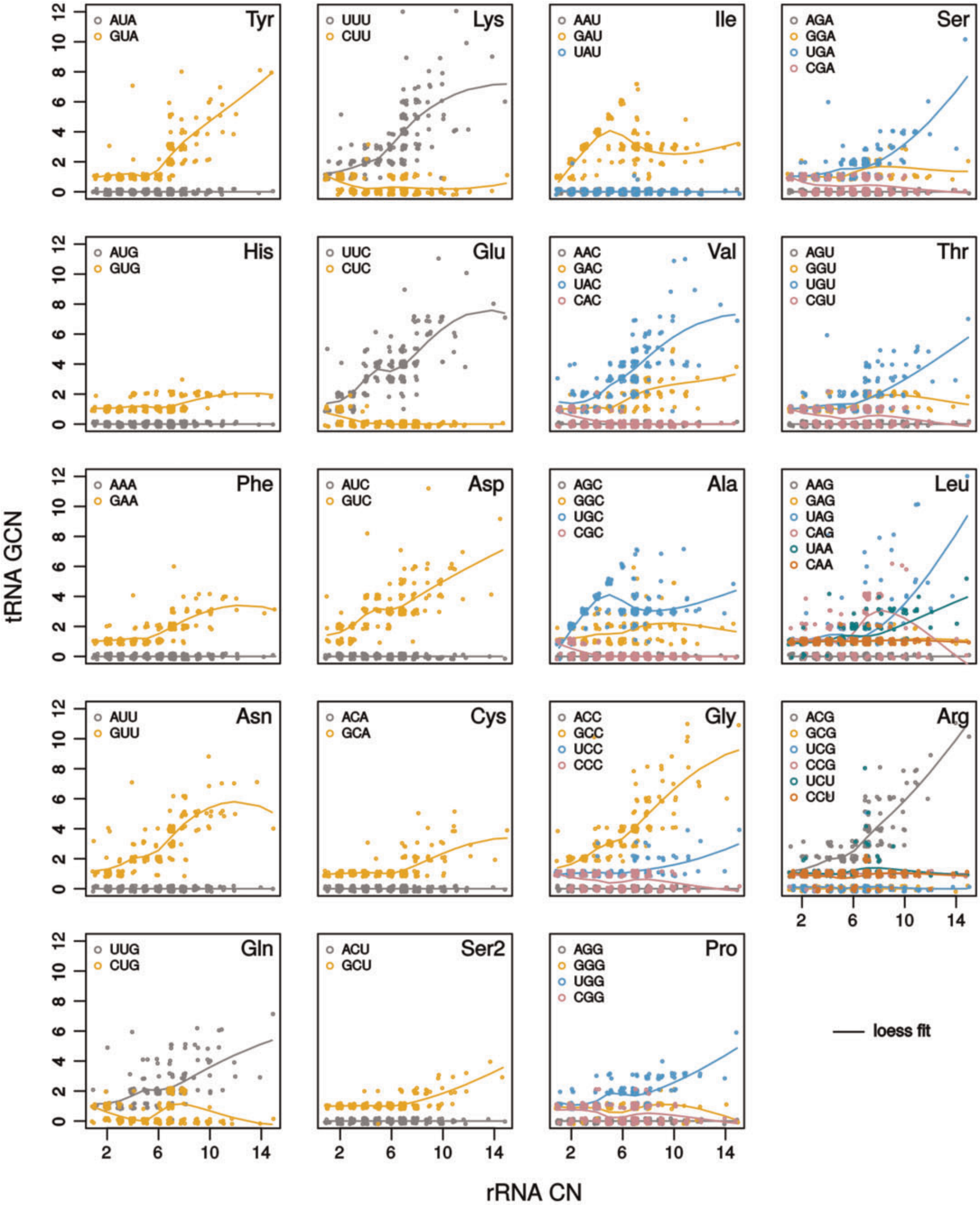

FIG. 4.—Anticodon specific tRNA gene copy number (tRNA GCN) in γ-proteobacteria. GCN of tRNAs bearing each anticodon are plotted against rRNA copy numbers (rRNA CN) of 189 genomes. Amino acids are arranged column-wise in increasing order of degeneracy, and the anticodon identities appear in the legend at top left. For 2-fold degenerate amino acids, the GNN anticodons (orange) are prevalent in the NNU/C codon family; while the UNN anticodons (grey) are prevalent in the NNA/G codon family. For most 4-fold degenerate amino acids, the UNN anticodons (blue) are most prevalent, followed by the GNN anticodons (orange). Glycine (Gly) is an exception where the GNN anticodon (orange) prevails over UNN anticodons (blue). For Leucine (Leu), the CNN (magenta) or UNN (blue) anticodons are prevalent in different set of genomes. For Arginine (Arg), the ANN anticodon is prevalent over others. Since multiple genomes have identical rRNA CN and tRNA GCN, we added a small jitter to both variables. As a result, the size of clusters approximates the number of genomes at the same *x*–*y* values. Smoothened Loess fits are shown to aid visualization.

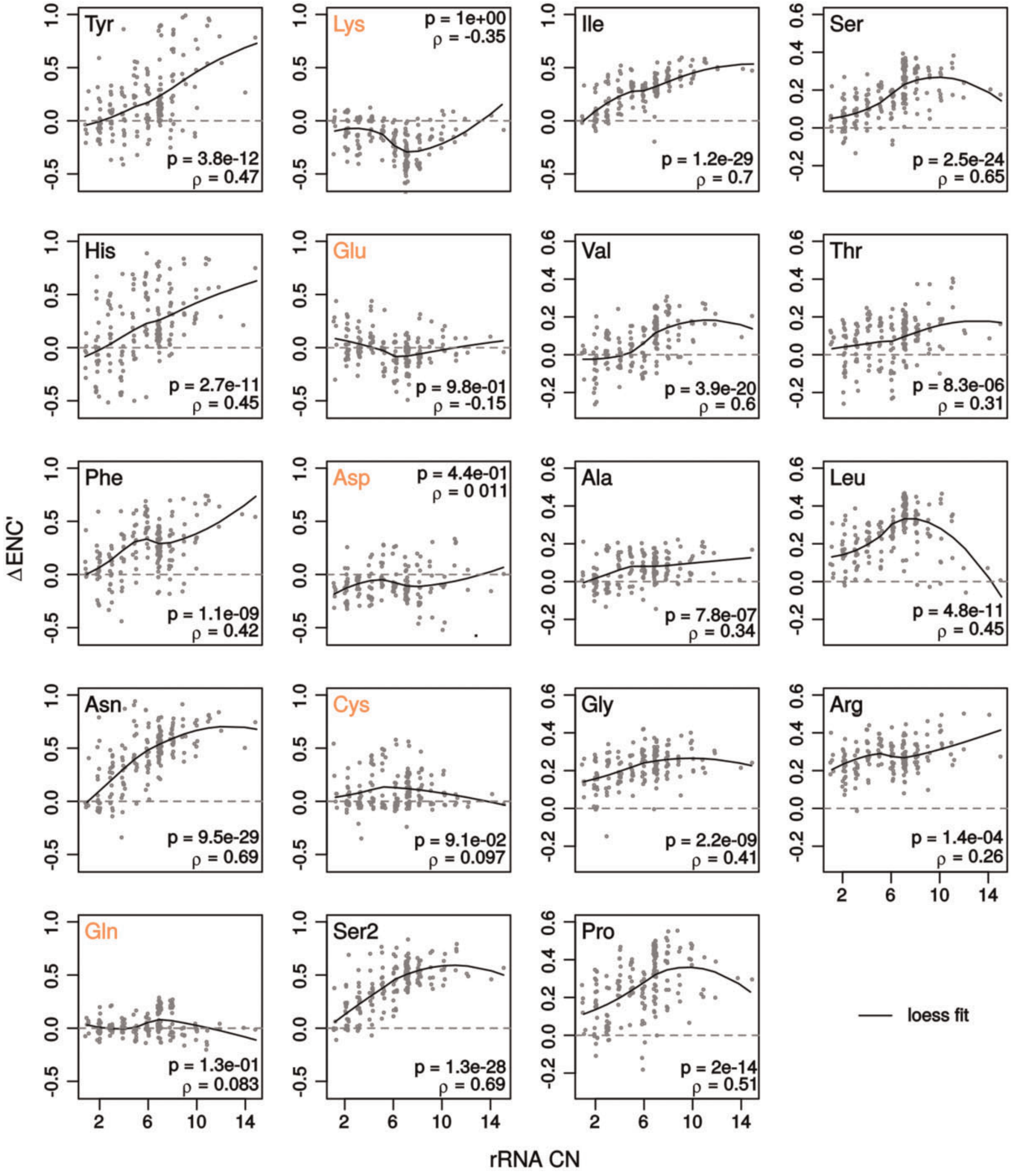

Fig. 5.—Correlations between amino acid specific codon usage bias (CUB) and rRNA copy numbers (rRNA CN) in γ-proteobacteria. CUB is represented by amino acid-specific ΔENC′, the normalized difference in effective number of codons between HEGs and all other genes. Red labels indicate amino acids with no positive correlation between ΔENC′ and rRNA CN. Each data point represents one genome (n=189). The dashed grey line indicates an absence of CUB. ρ is the Spearman's correlation coefficient and *P*-values correspond to a one-way asymptotic permutation test for positive correlation. Smoothened Loess fits are shown to aid visualization. *Y*-axes were set to identical scales within 2- or 3-fold, and 4- or 6-fold degenerate amino acid sets, for ease of visual comparison.

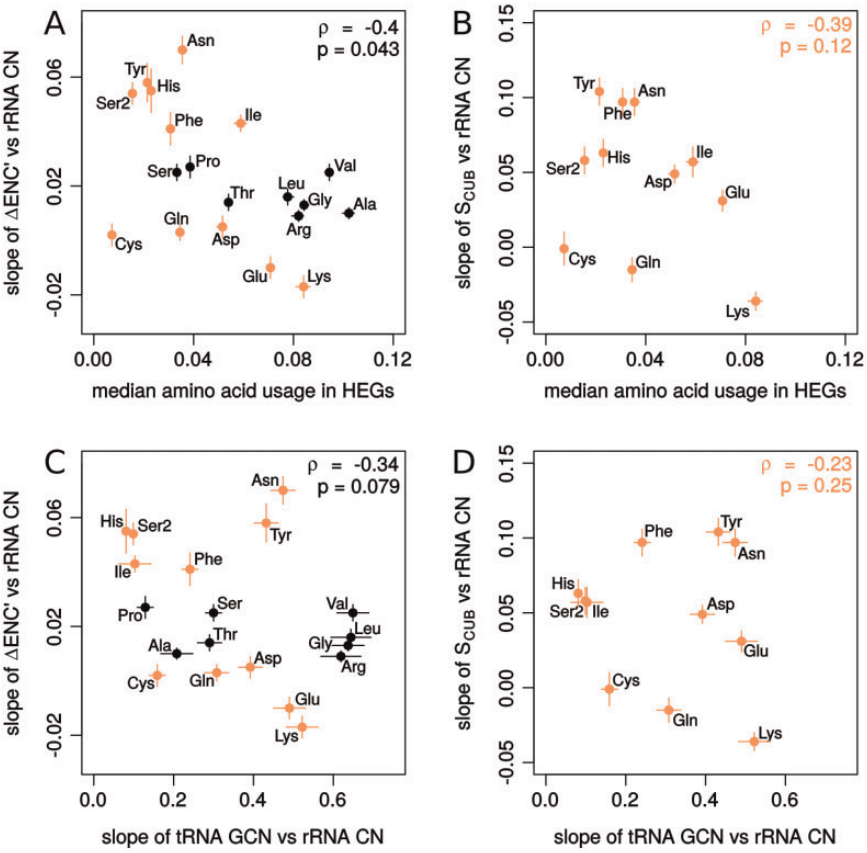

FIG. 6.—Association between the impact of growth rate on CUB and amino acid usage or tRNA GCN in γ-proteobacteria. Amino acid usage was calculated as the median values (across genomes) of the fraction of coding sites in HEGs that belong to a particular amino acid. The impact of growth rate on CUB is represented by the increase in ΔENC′ or S_{CUB} per unit change in rRNA CN, estimated by fitting linear regression models. (*A*) Median amino acid usage and slope of ΔENC′ versus rRNA CN. (*B*) Median amino acid usage and slope of S_{CUB} versus rRNA CN. (*C*) Slope of tRNA GCN versus rRNA CN and slope of ΔENC′ versus rRNA CN. (*D*) Slope of tRNA GCN versus rRNA CN and slope of S_{CUB} versus rRNA CN. Red circles show data for 2- or 3-fold amino acids and black circles indicate 4- or 6-fold degenerate amino acids. ρ is the Spearman's correlation coefficient and *P*-values correspond to a one-way asymptotic permutation test for a positive correlation. In panels A and B, horizontal bars are interquartile ranges of amino acid usage. Vertical bars in all plots are standard errors from the linear regression fits. In most cases IQR of amino acid usage is smaller than the width of the circles.

In the case of 4- and 6-fold degenerate amino acids, we also noticed that the impact of growth rate on CUB was more narrowly distributed compared with the 2-fold degenerate amino acids (black vs. red circles in fig. 6*A* and *C*). For 4-fold degenerate amino acids, this may arise from a consistent enrichment of NNU (and/or NNA) codons and avoidance of NNG (and/or NNC) codons (supplementary fig. S6, Supplementary Material online), irrespective of varying tRNA GCN (figs. 2 and 4). This may suggest that the strength of selection on codon usage in these cases is not influenced by the observed variation in tRNA abundances.

Some previous studies have raised the possibility that ribosomal protein genes may have peculiar codon usage distinct from other HEGs (Hershberg and Petrov 2012). To test whether our results depend on inclusion of ribosomal protein genes as HEG, we repeated the above tests after redefining HEG as genes with lowest 1% ENC′ in each genome, excluding any genes in our original HEG set. We find similar results even after redefining HEG. The impact of growth rate remains similar for all amino acids, except Cys, where a positive correlation is observed between ΔENC′ and rRNA CN. Particularly, the impact of growth rate on CUB of 2-fold degenerate amino acids is negatively correlated with amino acid usage ($P = 0.0015$) or tRNA GCN ($P = 0.028$) only if CUB is quantified as ΔENC′, but not when it is quantified as S_{CUB} (supplementary fig. S7, Supplementary Material online).

tRNA GCN, CUB, and Codon-Specific Translation Times in *E. coli*

As described above, we found that the amino acid-specific impact of growth rate on CUB is not correlated with the impact on tRNA GCN (fig. 6*C* and *D*), whereas we expected a negative correlation. This expectation was based on the following assumptions: 1) higher tRNA GCN for frequently used amino acids should reduce the difference in translation times

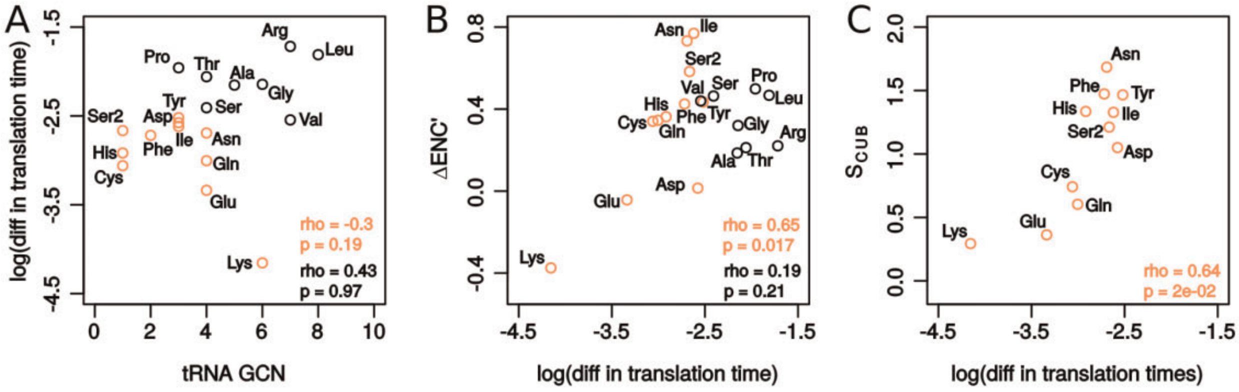

Fig. 7.—Amino acid specific translation time differences, tRNA gene copy number (tRNA GCN) and codon usage bias (CUB) in *E. coli*. Mean typical translation time for each codon for *E. coli* was obtained from Dana and Tuller (2014), and the difference in translation times, that is, |maximum – minimum| of the codon set of each amino acid were calculated. CUB (ΔENC′ and S_{CUB}) was calculated as previously described. S_{CUB} could be calculated only for two-box amino acids. (*A*) Difference in translation time and tRNA GCN. (*B*) ΔENC′ and difference in translation times. (*C*) S_{CUB} and difference in translation times. Red circles indicate 2- or 3-fold and black circles 4- or 6-fold degenerate amino acids. ρ is Spearman's correlation coefficient and *P*-values correspond to a one-way asymptotic permutation test for positive correlation. In the case of A and B correlations were also separately assessed for 2-fold degenerate amino acids, these are indicated in red text.

between synonymous codons, and 2) CUB is proportional to translation time differences between synonymous codons. However, these assumptions have not been explicitly verified and codon-specific translation times have been comprehensively estimated *in vivo* only in *E. coli* (a γ-proteobacterium) and *Bacillus subtilis* (a Firmicute). Using published translation time estimates based on ribosome profiling data (Dana and Tuller 2014), we found that the difference in translation times of the fastest and slowest synonymous codons of amino acids was not correlated with corresponding tRNA GCN (fig. 7*A* and supplementary fig. S8*A*, Supplementary Material online). This may explain why the impact of growth rate on amino acid specific CUB does not correlate with the impact on tRNA GCN. CUB was positively correlated ($P < 0.02$) with the difference in translation times between synonymous codons, but only for 2-fold degenerate amino acids in *E. coli* (fig. 7*B* and *C*). Interestingly, in *E. coli*, the smallest differences in translation times between synonymous codons were observed for the NNA/G codon sets of Lys, Glu, Gln, and NNU/C codons of Cys. These observations are consistent with the idea that a smaller increase in translation speed results in weaker CUB. However, these correlations disappear if Lys and/or Glu are excluded, and do not hold true for 4- or 6-fold degenerate amino acids or in *B. subtilis* (supplementary fig. S8*B* and *C*, Supplementary Material online). Overall, we thus find little evidence for the assumption that tRNA GCN mediated changes in translation times contribute to amino acid-specific variation in CUB.

Discussion

Bacterial growth rate appears to be under strong ecological selection, with faster growth preferred in nutrient rich conditions (Roller et al. 2016). Such selection is expected to impact the evolution of key factors affecting translation rate, including ribosomal RNA genes, tRNA pools, and codon use. Although the influence of fast growth-associated selection on tRNA GCN and CUB has been recognized for many years (Rocha 2004; Sharp et al. 2005; Higgs and Ran 2008), the quantitative nature of this "imprint of growth" (Vieira-Silva and Rocha 2010) is less clear. By investigating genome level tRNA GCN and CUB across ~200 γ-proteobacterial genomes, and using rRNA copy number as a proxy for growth rate, we gained new insights about the impact of growth rate on these traits. First, we found that total tRNA gene numbers (N_{tRNA}) and average CUB respond differently to growth associated selection, and covary strongly only within a narrow range of moderate growth rates. N_{tRNA} showed an early constancy-late increase pattern, while CUB showed an early increase-late saturation pattern with increasing rRNA CN. Similar patterns were also observed individually for multiple amino acids. Our observations cannot be explained if the costs and benefits of translational components scale linearly with growth rate. We suggest that factors other than growth associated translational selection may constrain trait values at very low and very high growth rates.

At low growth rates, the simplest explanation for constancy of N_{tRNA} may be that tRNA abundance is not a limiting factor for translation, or that single copies of genes are sufficient in this range of growth rates. Alternatively, constant N_{tRNA} may also result from negative selection against investment in additional tRNA gene copies. On the other hand, at higher growth rates, CUB may saturate because any further enrichment of codons preferred for speed may have detrimental effects on translational accuracy, or on other sequence features such as mRNA structure and stability. Fast codons are less accurately translated in at least some cases (Dix and Thompson 1989), and many sites in mRNAs may be under selection to maintain features other than elongation rate, including short- and long-range RNA structure (Kudla et al. 2009; Tuller et al. 2010; Kelsic et al. 2016) or cotranslational protein folding (Chaney et al. 2017). Although it is clear that using only "good codons" may actually impair protein expression and fitness (Agashe et al. 2013), the effect of increasing CUB on specific sequence features has not been comprehensively explored. Such effects can be investigated by simulating sequences with increasing CUB and quantifying other sequence features like folding energy, base pairing propensities, or occurrence of regulatory motifs. Apart from these possibilities, we acknowledge that our inferences may be influenced by the relationship between rRNA CN and growth rate. For instance, the saturation in CUB could be explained if rRNA CN continued to increase even when growth rate and associated selection on CUB ceased to change (i.e., if rRNA CN increased under independent selection unrelated to growth rate). At present, it is difficult to reliably assess this possibility due to the limited number of organisms with very high growth rates. But we must also highlight that CUB and rRNA CN are both likely to be affected by shared long-term selection on translation, whereas growth rates are typically measured only in the laboratory. Hence, rRNA CN may be a more realistic predictor of selection on other translational traits such as CUB.

We also show for the first time that the imprint of growth on tRNA GCN and CUB varies across amino acids, and—contrary to expectation—variation in tRNA GCN does not explain variation in CUB for 2-fold degenerate amino acids. A quantitative match between tRNA GCN and CUB across amino acids critically depends on tRNA GCN affecting the difference in translation time of synonymous codons. We tested this assumption using codon-specific translation time and tRNA GCN of *E. coli* and *B. subtilis*, and found that tRNA GCN did not explain amino acid-specific differences in the translation times of synonymous codons. This supports the possibility that translation times are influenced by other factors that differ across amino acids,for example, the first two base pairs in the codon:anticodon interaction. Nevertheless, CUB should reflect the time gained by using the fast codon for each amino acid. After directly testing the correlation between amino acid-specific translation time differences and CUB in these bacteria, we found some evidence for a connection between translation times and CUB. We note that the translation time data for this analysis were obtained from single experiments in specific conditions (Li et al. 2012; Dana and Tuller 2014) and may not represent typical values for all bacteria in various growth conditions. Combined measurements of tRNA concentrations and codon-specific translation time under different conditions for different bacteria should clarify this issue further. As with 2-fold degenerate amino acids, we also found discordant impacts of growth rate on tRNA and CUB of 4-fold degenerate amino acids. In this case, CUB can be mostly attributed to consistent enrichment in NNU (and/or NNA) codons versus the consistent avoidance of NNG (and/or NNC) codons. Only two tRNA types—with UNN and GNN anticodons—in turn explain most of the variation in tRNA GCN across 4-fold degenerate amino acids. However, there is little variation in the impact of growth rate on CUB. We suggest that these discordant patterns may arise because the observed changes in copy numbers of UNN and GNN tRNAs are not sufficient to change the translation rate of various codons. This hypothesis can be tested only by measuring the translation rate of various codon:anticodon pairs and the effect of tRNA abundance on these rates.

The mismatch between tRNA GCN, translation time differences of synonymous codons, and CUB raise the possibility that CUB is influenced by selection other than for translational speed. We also observed that CUB for Cys, Lys, and Gln is weak, that is, there is little difference in codon usage of highly expressed versus other genes. In addition, CUB for these amino acids does not increase with growth rate across bacteria; and we found only a weak increase in CUB for Glu and Asp. This suggests that codon use for these amino acids is not under selection for translational speed. This is surprising because the typical tRNA GCN for these amino acids ranges from 2 (Cys) to 6 (Lys) in fast growing bacteria, and growth rate should have thus impacted their CUB. Altogether, our results suggest that the selection on CUB for some amino acids is constrained by factors other than translation speed.

Another interesting aspect of our results is the differential behavior of 2-fold degenerate amino acids encoded by NNA/G versus NNU/C codons. Out of the ten 2-fold degenerate amino acids, three are encoded by NNA/G codons and seven by NNU/C codons. Across γ-proteobacteria, there was no or little impact of growth rate on the CUB of any NNA/G codon family amino acids. In *E. coli*, differences in translation times within NNA/G codon families were also among the lowest. On the other hand, CUB of amino acids encoded by most NNU/C codon families increased robustly (except Asp and Cys). It is possible that this is just a chance occurrence given the small number of amino acids with NNA/G codons. However, NNA/G codon family tRNAs share a distinct modification (mnm^{5}s^{2}U^{34}) in their UNN tRNAs (Yokoyama et al. 1985) and this tRNA type is often the only tRNA corresponding to these amino acids in γ-proteobacteria. It is possible that mnm^{5}s^2-modifications of UNN tRNAs increases the translation speed of both NNA and NNG codons, reducing selection for favoring either codon, ultimately resulting in a lack of significant CUB. This still leaves open the question of why there is no increase in the CUB for Cys, and a weaker change for Asp compared with other amino acids encoded by NNU/C codons.

Regardless of the reasons for amino acid-specific differences in CUB, we expect the degree of CUB to be consistent with the actual fitness consequences of altering the codon use. Few experimental studies have systematically varied codon use for every amino acid and measured its effect on growth. A recent study in *E. coli* (Kelsic et al. 2016) measured the fitness consequences of replacing every native codon of *infA* (a highly expressed translation initiation factor) with every possible alternate codon to uncover the fitness consequences of using a specific synonymous codon for each amino acid. This data set offered us the opportunity to compare amino acid-specific differences in CUB with their fitness consequences. In agreement with a strong preference for NNC codons of 2-fold degenerate amino acids in *E. coli* (and other bacteria), the fitness of strains carrying NNC codons for Tyr, His, Phe, Asn, Ser2, and Ile was higher than corresponding NNU codons (fig. 2B in Kelsic et al. 2016). On the contrary, even though CUB for Lys, Gln, and Cys is weak in *E. coli* and does not increase with growth rates across bacteria, specific synonymous codons for each of these amino acids (AAA–Lys, CAG–Gln, UGC–Cys) conferred higher fitness in this experiment. In Glu, both synonymous codons had similar fitness impacts, consistent with our observation of weak CUB. Overall, the strength of amino acid-specific CUB in *E. coli* appears to be inconsistent with the fitness impact of codons in this experiment. The mismatch between weak CUB and strong fitness effect in the above experiment is puzzling, but to test its generality, we need experiments that manipulate codon use of multiple HEGs in many bacteria. If these effects are more general, it will support the hypothesis that codons of some amino acids are favored for reasons independent of gene expression and growth rate, and therefore do not manifest as biased codon use in HEG. Further investigations of the effect of synonymous codon choice of these amino acids on other aspects of translation such as translational accuracy and mRNA structure may be required to resolve this issue.

In summary, this study reveals several riders to the predicted coevolution of tRNA GCN and CUB. By studying both traits in the context of growth rate-associated selection, we found patterns that were unexpected, given current assumptions about the mechanistic basis of tRNA–CUB coevolution. These patterns suggest the presence of other constraints that shape the evolution of these traits. We suggest that some constraints may act across many amino acids (tRNA costs, conflicting selection between codon use and other sequence features), while others may be specific to particular amino acids (specific codon:anticodon interactions and their effect on translation rates or other features). Although identifying the exact nature of the constraints is beyond the scope of our study, we suggest several hypotheses, and computational and laboratory experiments that can be used to test them. We hope that such experiments will further our understanding of the interrelation between growth rate and key features of translation in bacteria.

Materials and Methods

Data sets

We curated a data set of about ~1,000 bacterial genomes based on a phylogeny from Segata et al. (2013). The curation involved removing very closely related taxa and known endosymbionts. To remove closely related taxa, we used an in-house script to traverse all internal nodes starting from the root of the phylogeny, identified terminal branch lengths leading to all descendant taxa, and chose only one representative taxon from all taxa that were within a specific branch length threshold. Endosymbiont taxa were identified based on literature and removed manually. We obtained genomes of selected taxa by identifying the closest genome found in the NCBI genomes database, and manually selected a genome when multiple genomes for the same taxonomic id were available. In some cases this led to the selection of genomes of a closely related strain (instead of the one included in the original phylogeny). Only completely sequenced genomes (not contigs and scaffolds) were retained. We obtained rRNA copy number data from the rrnDB v5.1 (Stoddard et al. 2015), or from the IMG database (Markowitz et al. 2012) (downloaded on May 5, 2016). We obtained tRNA GCN directly from GtRNAdb (Chan and Lowe 2009) (downloaded August 13, 2016). For some genomes that were missing in the GtRNAdb, we detected tRNA genes via tRNAscan-SE (Lowe and Eddy 1997) using default parameters for bacteria. We excluded Ile and Ala tRNA genes from the calculation of total tRNA gene numbers (N_{tRNA}) because they are often found within rRNA operons, thus introducing a superfluous correlation between N_{tRNA} and rRNA copy numbers (rRNA CN). We calculated codon usage statistics such as codon counts using ENCprime software (Novembre 2002), and then nucleotide and amino acid usage using the codon count data. We used in-house scripts to calculate actual metrics of CUB, that is, ΔENC′ and S_{CUB}. All data sets and scripts used for calculations and analysis can be accessed at https://github.com/saurabh-mk/tRNA_CUB_aa.

As in previous studies, we separated the six codons of Serine into two families of two and four codons each, and treated them as separate amino acids (Ser2 and Ser). This separation is warranted because these codon families are different at two nucleotides, and thus will be decoded by independent sets of tRNAs. Moreover, mutational changes are much less likely to change codons from one Serine family to other, making the evolutionary dynamics of these two codon families independent.

Calculation of Average ΔENC′ and S_{CUB}

Both metrics of CUB (ΔENC′ and S_{CUB}) are based on a comparison of codon usage in HEGs and remaining genes in the genome. We defined HEGs as genes encoding ribosomal proteins, RNA polymerase subunits, and EF–Tu subunits, identified based on gene annotations from the coding sequence files provided in the NCBI genomes database. All remaining genes were classified as "other." We concatenated all genes within each set and calculated a single codon count per set. We calculated position specific (1st/2nd/3rd position in coding frame) nucleotide usage frequencies using the codon counts.

We calculated S_{CUB} as suggested by Sharp et al. (2005). In brief, for each of the four amino acids—Phe, Tyr, Ile, Asn, encoded by NNU and NNC codons, we calculated S_{CUB} as

$$S_{\mathrm{CUB}} = \ln(n_C^{\mathrm{HEG}} n_U^{\mathrm{other}} / n_U^{\mathrm{HEG}} n_C^{\mathrm{other}}), \qquad (1)$$

where n is the number of codons, subscript indicates the codon and superscript indicates the gene set (HEG vs. other). Finally, we calculated average S_{CUB} by summing amino-acid specific S_{CUB} weighted by amino acid usage in the HEGs.

Next, we calculated the second CUB metric based on the differences in the effective number of codons (ENC) as

$$\Delta\mathrm{ENC}' = (\mathrm{ENC}'_{\mathrm{other}} - \mathrm{ENC}'_{\mathrm{HEG}})/\mathrm{ENC}'_{\mathrm{other}}. \qquad (2)$$

ENC′ for each set (HEGs and other) was calculated separately. All the following calculations are based on the method developed by Novembre (2002). First, for each amino acid a χ^2 statistic that represents the deviation of codon usage from expected codon usage was calculated as

$$\chi^2_{\mathrm{aa}} = \sum_{i=1}^{k} \left\{ \frac{n_{\mathrm{aa}} (f_i - e_i)^2}{e_i} \right\}, \qquad (3)$$

where aa is the amino acid, k is redundancy of the amino acid, n_{aa} is the total number of sites at which the amino acid is used, f_i is the actual codon use frequency, and e_i is the expected codon use frequency. Expected codon usage was defined as a product of position specific nucleotide frequencies. For example, the expected frequency of the AAT codon would be $f_1^A \times f_2^A \times f_3^T$, where f is the frequency of the nucleotide in the superscript at the coding position in the subscript. Position-specific nucleotide frequencies were calculated separately for each gene-set from the respective codon counts. We normalized the expected frequencies by the sum of expected frequencies of all sense codons. Next, we calculated codon usage heterogeneity ($\mathrm{F}'_{\mathrm{aa}}$) as

$$F'_{\mathrm{aa}} = \frac{(\chi^2_{\mathrm{aa}} + n_{\mathrm{aa}} - k)}{(k n_{\mathrm{aa}} - k)}, \qquad (4)$$

where the symbols are as defined above. This normalization has the effect of making the inverse of F'_{aa} a number between 1 and k, which can be interpreted as the effective number of codons, ENC′. After this, we obtained the average ENC′ by first calculating the average F'_{aa} within a redundancy class (i.e., separately for 2-, 3-, 4-, 6-fold degenerate amino acids) and then calculating ENC′ as

$$\mathrm{ENC}' = \sum_r \frac{n_r}{F'_r}, \qquad (5)$$

where r is the redundancy class, n_r is the number of amino acids belonging to that redundancy class (9, 1, 5, 3, respectively), and F'_r is the average heterogeneity in the redundancy class. Finally, average ΔENC′ was calculated as

$$\Delta\mathrm{ENC}' = \frac{(\mathrm{ENC}'_{\mathrm{other}} - \mathrm{ENC}'_{\mathrm{HEG}})}{\mathrm{ENC}'_{\mathrm{other}}}. \qquad (6)$$

Calculation of amino acid specific ΔENC′ and S_{CUB}

$$\mathrm{ENC}'^{\mathrm{aa}} = \frac{1}{F'^{\mathrm{aa}}}. \qquad (7)$$

At this point, we defined the amino acid specific ΔENC′ as

$$\Delta\mathrm{ENC}'^{\mathrm{aa}} = \frac{(\mathrm{ENC}'^{\mathrm{aa}}_{\mathrm{other}} - \mathrm{ENC}'^{\mathrm{aa}}_{\mathrm{HEG}})}{(k-1)}. \qquad (8)$$

This metric represents CUB and usually takes values between 0 (when codon usage in HEGs and other genes is identical) and 1 (codon usage in other genes is unbiased and HEGs is completely biased).

To calculate amino acid-specific S_{CUB}, we only considered the magnitude of CUB, ignoring the identity of the preferred synonymous codon. For 2-fold degenerate amino acids encoded by NNU/C codons and Ile, we calculated S_{CUB} as

$$S_{\mathrm{CUB}} = \left| \ln\left(\frac{n_C^{\mathrm{HEG}} n_U^{\mathrm{other}}}{n_U^{\mathrm{HEG}} n_C^{\mathrm{other}}} \right) \right|.$$

For 2-fold degenerate amino acids encoded by NNA/G codons, we calculated S_{CUB} as

$$S_{\mathrm{CUB}} = \left| \ln\left(\frac{n_A^{\mathrm{HEG}} n_G^{\mathrm{other}}}{n_A^{\mathrm{HEG}} n_G^{\mathrm{other}}} \right) \right|,$$

where n is the number of codons, subscript indicates the codon and superscript indicates the gene set (HEG vs. other).

Statistical Analysis

We always tested correlations using Spearman's rank correlation (nonparametric) and assessed significance using the asymptotic permutation test included in the cor.test function of stats package in base R (R Core Team 2015). We fit a saturating model with the form

$$\Delta\mathrm{ENC}' \text{ or } S_{\mathrm{CUB}} = \frac{a + b \times (\mathrm{rRNA\ CN})}{(c + \mathrm{rRNA\ CN})}$$

using nonlinear least squares in R using the nls function. Using the segmented.lm() function in the package "segmented"

(Muggeo 2008), we fit a piecewise linear model with one breakpoint to the data set with tRNA gene numbers vs. rRNA CN. We evaluated alternate models based on AIC values without correction, because sample sizes were much larger (at least 10-fold) than the square of the number of parameters in the models. We fit smoothened Loess lines to data by choosing an arbitrary smoothening parameter value that removed kinks in the resulting fit. These were only used for the purpose of visualizing trends.

We characterized the impact of growth rate on amino acid specific CUB or tRNA GCN by fitting a linear regression model to the relationship between trait value and rRNA CN. From this regression equation, we calculated or obtained the estimated trait value when rRNA CN = 1, and the slope of the relationship between the trait versus rRNA CN.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We thank an anonymous reviewer for constructive suggestions that improved the paper. We thank Gaurav Diwan for compiling the tRNA gene copy number data and contributing to the script for curating the phylogeny. This study was supported by the Council of Scientific and Industrial Research of the Government of India (an SPM graduate fellowship to S.M. and a research grant to D.A., 37(1629)/14/EMR-II); the Department of Science and Technology of the Government of India (an INSPIRE faculty award to D.A., IFA-13 LSBM-64); and the National Centre for Biological Sciences of India.

Literature Cited

Agashe D, Martinez-Gomez NC, Drummond DA, Marx CJ. 2013. Good codons, bad transcript: large reductions in gene expression and fitness arising from synonymous mutations in a key enzyme. Mol Biol Evol. 30(3):549–560.

Andersson SG, Kurland CG. 1990. Codon preferences in free-living microorganisms. Microbiol Rev. 54(2):198–210.

Berg OG, Kurland CG. 1997. Growth rate-optimised tRNA abundance and codon usage. J Mol Biol. 270(4):544–550.

Bonekamp F, Dalbøge H, Christensen T, Jensen KF. 1989. Translation rates of individual codons are not correlated with tRNA abundances or with frequencies of utilization in *Escherichia coli*. J Bacteriol. 171(11):5812–5816.

Bremer H, Dennis PP. 2008. Modulation of chemical composition and other parameters of the cell at different exponential growth rates. EcoSal Plus 3(1): .

Bulmer M. 1987. Coevolution of codon usage and transfer RNA abundance. Nature 325(6106):728–730.

Chan PP, Lowe TM. 2009. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. Nucleic Acids Res. 37(Database issue):D93–D97.

Chaney JL, et al. 2017. Widespread position-specific conservation of synonymous rare codons within coding sequences. PLoS Comput Biol. 13(5):e1005531.

Chen SL, Lee W, Hottes AK, Shapiro L, McAdams HH. 2004. Codon usage between genomes is constrained by genome-wide mutational processes. Proc Natl Acad Sci U S A. 101(10):3480–3485.

Curran JF, Yarus M. 1989. Rates of aminoacyl-tRNA selection at 29 sense codons in vivo. J Mol Biol. 209(1):65–77.

Dana A, Tuller T. 2014. The effect of tRNA levels on decoding times of mRNA codons. Nucleic Acids Res. 42(14):9171–9181.

Diwan GD, Agashe D. 2016. The frequency of internal Shine–Dalgarno-like motifs in prokaryotes. Genome Biol Evol. 8(6):1722–1733.

Dix DB, Thompson RC. 1989. Codon choice and gene expression: synonymous codons differ in translational accuracy. Proc Natl Acad Sci U S A. 86(18):6888–6892.

Drummond DA, Wilke CO. 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell 134(2):341–352.

Du M-Z, et al. 2017. Co-adaption of tRNA gene copy number and amino acid usage influences translation rates in three life domains. DNA Res. 24:623–633. .

Gilchrist MA. 2007. Combining models of protein translation and population genetics to predict protein production rates from codon usage patterns. Mol Biol Evol. 24(11):2362–2372.

Gyorfy Z, et al. 2015. Engineered ribosomal RNA operon copy-number variants of *E. coli* reveal the evolutionary trade-offs shaping rRNA operon number. Nucleic Acids Res. 43(3):1783–1794.

Hershberg R, Petrov DA. 2012. On the limitations of using ribosomal genes as references for the study of codon usage: a rebuttal. PLoS ONE 7(12):e49060.

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(11):2279–2291.

Ikemura T. 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. J Mol Biol. 151:389–409

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(1):143–155.

Kelsic ED, et al. 2016. RNA structural determinants of optimal codons revealed by MAGE-seq. Cell Syst. 3(6):563–571.e6.

Klappenbach JA, Dunbar JM, Schmidt TM. 2000. rRNA operon copy number reflects ecological strategies of bacteria. Appl Environ Microbiol. 66(4):1328–1333.

Klumpp S, Dong J, Hwa T. 2012. On ribosome load, codon bias and protein abundance. PLoS ONE 7(11):e48542.

Knight RD, Freeland SJ, Landweber LF. 2001. A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. Genome Biol. 2:research0010.1–research0010.13.

Kramer EB, Farabaugh PJ. 2006. The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. RNA 13(1):87–96. .

Kudla G, Murray AW, Tollervey D, Plotkin JB. 2009. Coding-sequence determinants of gene expression in *Escherichia coli*. Science 324(5924):255–258.

Li G-W, Oh E, Weissman JS. 2012. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature 484(7395):538–541.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25(5):955–964.

Marck C, Grosjean H. 2002. tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. RNA 8(10):1189–1232.

Markowitz VM, et al. 2012. IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Res. 40(D1):D115–D122.

Muggeo VMR. 2008. Segmented: an R package to fit regression models with broken-line relationships. R News 8:20–25.

Novembre JA. 2002. Accounting for background nucleotide composition when measuring codon usage bias. Mol Biol Evol. 19(8):1390–1394.

Pedersen S. 1984. *Escherichia coli* ribosomes translate in vivo with variable rate. EMBO J. 3(12):2895–2898.

Perry RHJ. 2015. The evolution of codon usage and base composition [Ph.D. thesis]. University of Edinburgh. http://hdl.handle.net/1842/17870, last accessed January 26 2018.

R Core Team. 2015. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/, last accessed January 26 2018.

Rocha EPC. 2004. Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. Genome Res. 14(11):2279–2286.

Roller BRK, Stoddard SF, Schmidt TM. 2016. Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. Nat Microbiol. 1(11):16160.

Schaechter M, Maaloe O, Kjeldgaard NO. 1958. Dependency on medium and temperature of cell size and chemical composition during balanced grown of *Salmonella typhimurium*. J Gen Microbiol. 19(3):592–606.

Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T. 2010. Interdependence of cell growth and gene expression: origins and consequences. Science 330(6007):1099–1102.

Segata N, Börnigen D, Morgan XC, Huttenhower C. 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nat Commun. 4:2304.

Shah P, Gilchrist MA. 2010. Effect of correlated tRNA abundances on translation errors and evolution of codon usage bias. PLoS Genet. 6(9):e1001128.

Sharp PM, Bailes E, Grocock RJ, Peden JF, Sockett RE. 2005. Variation in the strength of selected codon usage bias among bacteria. Nucleic Acids Res. 33(4):1141–1153.

Spencer PS, Siller E, Anderson JF, Barral JM. 2012. Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. J Mol Biol. 422(3):328–335.

Stevenson BS, Schmidt TM. 2004. Life history implications of rRNA gene copy number in *Escherichia coli*. Appl Environ Microbiol. 70(11):6670–6677.

Stoddard SF, Smith BJ, Hein R, Roller BRK, Schmidt TM. 2015. rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. Nucleic Acids Res. 43(D1):D593–D598.

Stoletzki N, Eyre-Walker A. 2007. Synonymous codon usage in *Escherichia coli*: selection for translational accuracy. Mol Biol Evol. 24(2):374–381.

Tuller T, Waldman YY, Kupiec M, Ruppin E. 2010. Translation efficiency is determined by both codon bias and folding energy. Proc Natl Acad Sci U S A. 107(8):3645–3650.

Vieira-Silva S, Rocha EPC. 2010. The systemic imprint of growth and its uses in ecological (meta)genomics. PLoS Genet. 6(1):e1000808.

Wald N, Alroy M, Botzman M, Margalit H. 2012. Codon usage bias in prokaryotic pyrimidine-ending codons is associated with the degeneracy of the encoded amino acids. Nucleic Acids Res. 40(15):7074–7083.

Weixlbaumer A, et al. 2007. Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines. Nat Struct Mol Biol. 14(6):498.

Wright F. 1990. The 'effective number of codons' used in a gene. Gene 87(1):23–29.

Yokoyama S, et al. 1985. Molecular mechanism of codon recognition by tRNA species with modified uridine in the first position of the anticodon. Proc Natl Acad Sci U S A. 82(15):4905–4909.

Associate editor: Ruth Hershberg