

Wobbling Forth and Drifting Back: The Evolutionary History and Impact of Bacterial tRNA Modifications

Gaurav D. Diwan^{*,1,2} and Deepa Agashe^{*,1}

¹National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bengaluru, India

²SASTRA University, Thanjavur, India

*Corresponding authors: E-mails: gauravd@ncbs.res.in; dagashe@ncbs.res.in.

Associate editor: James McInerney

Abstract

Along with tRNAs, enzymes that modify anticodon bases are a key aspect of translation across the tree of life. tRNA modifications extend wobble pairing, allowing specific (“target”) tRNAs to recognize multiple codons and cover for other (“nontarget”) tRNAs, often improving translation efficiency and accuracy. However, the detailed evolutionary history and impact of tRNA modifying enzymes has not been analyzed. Using ancestral reconstruction of five tRNA modifications across 1093 bacteria, we show that most modifications were ancestral to eubacteria, but were repeatedly lost in many lineages. Most modification losses coincided with evolutionary shifts in nontarget tRNAs, often driven by increased bias in genomic GC and associated codon use, or by genome reduction. In turn, the loss of tRNA modifications stabilized otherwise highly dynamic tRNA gene repertoires. Our work thus traces the complex history of bacterial tRNA modifications, providing the first clear evidence for their role in the evolution of bacterial translation.

Key words: tRNA modifications, ancestral reconstruction, tRNA gene content, GC content.

Introduction

Compared with the total complement of 61 sense codons, most bacterial genomes contain only 25–46 unique tRNA species (Grosjean et al. 2010). This apparent shortfall in decoding ability is mitigated by GU wobble base pairing that allows a single tRNA to decode non-complementary codons (Crick 1966). Additionally, specific modifications to the first anticodon base of the tRNA also contribute to wobble base pairing, sometimes reducing translational errors (Yokoyama et al. 1985; Brierley et al. 1997; Marck and Grosjean 2002; Näsvalld et al. 2004, 2007; Agris et al. 2007; Björk and Hagervall 2014; Manickam et al. 2015). Base modifications are carried out by several tRNA modifying enzyme (ME) pathways, that chemically modify the wobble position of a large fraction (42%) of all tRNA species (Grosjean et al. 2010; Machnicka et al. 2015; Boccaletto et al. 2018). Given their potential impact on the effective tRNA pool and on translation dynamics, MEs should generally face strong selection, especially in species with a restricted tRNA repertoire. Indeed, in several bacteria, deleting MEs is lethal or otherwise deleterious for growth and translation (Wolf et al. 2002; Shippy et al. 2013; Björk and Hagervall 2014; Gao et al. 2016). However, in some cases MEs do not appear to be essential (Noguchi et al. 1982), suggesting strong yet variable evolutionary impacts of tRNA modifications. Recently, Novoa and colleagues showed that MEs unique to the three kingdoms of life have shaped kingdom-specific tRNA pools (Novoa et al. 2012). However, the detailed evolutionary history of diverse bacterial tRNA modifications and their impact on key components of bacterial translation remain unclear.

To address these gaps, we traced the evolutionary history of the five known bacterial tRNA modifications at the first wobble base of the anticodon: (c)mmn⁵(s²)U (5-carboxymethylaminomethyl 2-thiouridine and its variants), cmo⁵U (uridine-5-oxyacetic acid), k²C (2-lysyl-cytidine), Q (Queuosine), and I (Inosine) (Björk and Hagervall 2014). We first determined the occurrence of each modification pathway in 1093 sequenced bacteria that represent the diversity of eubacteria, and then used ancestral reconstruction to infer the major evolutionary gain or loss events in the bacterial phylogeny. To determine the causes and impacts of change in tRNA modification, we analyzed two other key components of translation: tRNA gene pools and codon bias, thought to coevolve under translational selection (Ikemura 1985; Kanaya et al. 1999). Finally, since both of these factors are in turn correlated with genomic GC content (Chen et al. 2004; Hershberg and Petrov 2009; Wald and Margalit 2014), we also mapped genomic GC content across the bacterial phylogeny. Thus, we aimed to generate a comprehensive model of the joint evolutionary history of important components of bacterial translation—tRNA genes and MEs—and their relationship with genomic GC content.

Each modified tRNA can recognize between 2 and 4 codons. Although there are no direct measurements of the effect of tRNA modification on translation rate, the presence of a multifunctional tRNA should amplify the effective tRNA pool, reducing ribosomal waiting times for the correct tRNA, and contributing to an increase in global translation speed. Modification also ensures that tRNA carrying amino acids coded by two synonymous codons do not recognize other

© The Author(s) 2018. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

nonsynonymous codons, potentially reducing mistranslation (Grosjean et al. 2010; Björk and Hagervall 2014). Thus, tRNA modifications may affect both translational speed and accuracy. Specifically, we hypothesized that the presence of a tRNA modification should selectively favor the tRNAs that it modifies (henceforth “target” RNA). Concomitantly, selection on tRNA that are not modified by the ME (henceforth “nontarget” tRNA; encoding the same amino acid) should be weakened, since their function would be redundant with the modified target tRNA. As such, we predicted that the evolutionary loss of nontarget tRNA should be permissible only in the presence of a modification that allows target tRNA to decode multiple sense codons. In contrast, the evolutionary loss of a modification should be associated with a generally irreversible expansion of the tRNA repertoire to include both target and nontarget tRNA. Finally, we expected that these changes in tRNA gene content should be correlated with changes in codon use as well as genomic GC content.

Our analyses indicate that the eubacterial ancestor had a large tRNA repertoire as well as most of the known bacterial tRNA modifications. However, some types of wobble were lost in multiple bacterial lineages. As predicted, ME loss was strongly associated with an expansion of the tRNA repertoire; either in conjunction with large shifts in genomic GC and associated changes in codon use, or in a GC independent manner. We suggest that the expanded tRNA set weakened selection for MEs, allowing their repeated loss via drift. Conversely, in rare cases, a shrinking of tRNA diversity may have favored the evolution of a novel modification; or, the chance evolution of a new modification may have relaxed selection on the tRNA pool. Our work represents the first systematic analysis of the joint evolutionary history of tRNA modifications, genome GC content and tRNA genes, and elucidates the role of tRNA modifications in the evolution of the bacterial translation machinery.

Results

The Evolutionary History of Bacterial tRNA Modifications Is Highly Dynamic

We determined the occurrence of tRNA modifications across 1093 bacteria using homology searches for complete protein sequences of previously described MEs (see Materials and Methods). We assume that detected homologs are functional, and have the same modification function as that of the previously described enzymes. Within each modification pathway, we focused on the following enzymes that directly modify the tRNA and complete the modification (see Materials and Methods), and whose absence should effectively prevent the modification: (1) MnmE-MnmG for the (c)mnm⁵(s²)U modification, (2) CmoA-CmoB for the cmo⁵U modification, (3) Tgt-QueA-QueG-QueH for the Q modification, (4) TadA for the I modification, and (5) TilS for the k²C modification. Note that in our analysis, we did not consider GU wobble pairing, which is not mediated by specific modification enzymes. We observed that except for TilS (found in 98.18% of the studied bacteria), most MEs are not found in many bacteria (fig. 1A), suggesting that different

modification pathways were gained or lost in specific lineages. The essential proteins RpoB and RplA (positive controls) were respectively present in 99.5% and 99.7% of analyzed bacteria, indicating that our homology detection method resulted in very few false negatives. When multiple enzymes are necessary for a modification, they tend to co-occur in bacterial lineages (fig. 1A). For instance, both MnmE and MnmG are absent only in Actinobacteria. Similarly, both CmoA and CmoB are primarily observed in γ -proteobacteria and ε -proteobacteria. However, this contradicts previous reports showing that the cmo⁵U modification is also active in *Bacillus subtilis* and *Mycobacterium bovis* str. BCG (Murao et al. 1976; Yamada et al. 2005; Chionh et al. 2016). Despite an extensive search, we could not reliably identify the reported enzymes in the sequences of any other organisms (see Materials and Methods for details). Hence, here we focus on the proteobacterial cmo⁵U modification. In contrast to the co-occurrence of CmoA and CmoB, for the Queuosine modification, we found that QueG is often missing even when Tgt and QueA are present. However, 244 of these species had the QueH enzyme (fig. 1B), which was recently shown to function as a nonorthologous replacement for QueG (Zallot, Ross, et al. 2017). The relative evolutionary instability of QueG corroborates a recent report suggesting that in some bacteria, Tgt alone may be enough to generate the Q modification (Zallot, Yuan, et al. 2017).

To determine the key evolutionary changes in major tRNA modification pathways across bacteria, we performed ancestral reconstruction with stochastic character mapping for each ME (fig. 1A; supplementary table S1, Supplementary Material online). We used a posterior probability value of ≥ 0.7 to assign the ME state at each internal node in the bacterial phylogeny. Four of the five modification systems—MnmE-MnmG, Tgt-QueA-QueG-QueH, TadA, and TilS—were predicted to occur at the root of the phylogenetic tree; that is, they were already present in the eubacterial ancestor (fig. 1B). In contrast, CmoA-CmoB were gained/evolved much later: once at the root of γ -proteobacteria, once in ε -proteobacteria, and once in some species of δ -proteobacteria (fig. 1B). However, given experimental evidence for functional cmo⁵U modification in species such as *B. subtilis* and *M. bovis* (Yamada et al. 2005; Chionh et al. 2016) that do not have well-supported homologs to the proteobacterial cmo⁵U genes, it is possible that the modification was ancient and the responsible genes diverged significantly in different bacterial lineages. Alternatively, these enzymes may have nonorthologous genes in the ancestor. However, we could not test these hypotheses because the relevant ME genes in these species are not known. Thus, in general, we suggest that most bacterial tRNA modifications are ancient. These ancient modifications were then independently lost multiple times, with a total of 12 major losses (on branches with at least 5 descendant bacterial species in our phylogeny; fig. 1B) and 19 minor loss events (in lineages with fewer than five descendants; supplementary fig. S1, Supplementary Material online). The only exception to this pattern of repeated ME gains/losses is the k²C modification (TilS), which appears to be evolutionarily stable across bacteria. Overall,

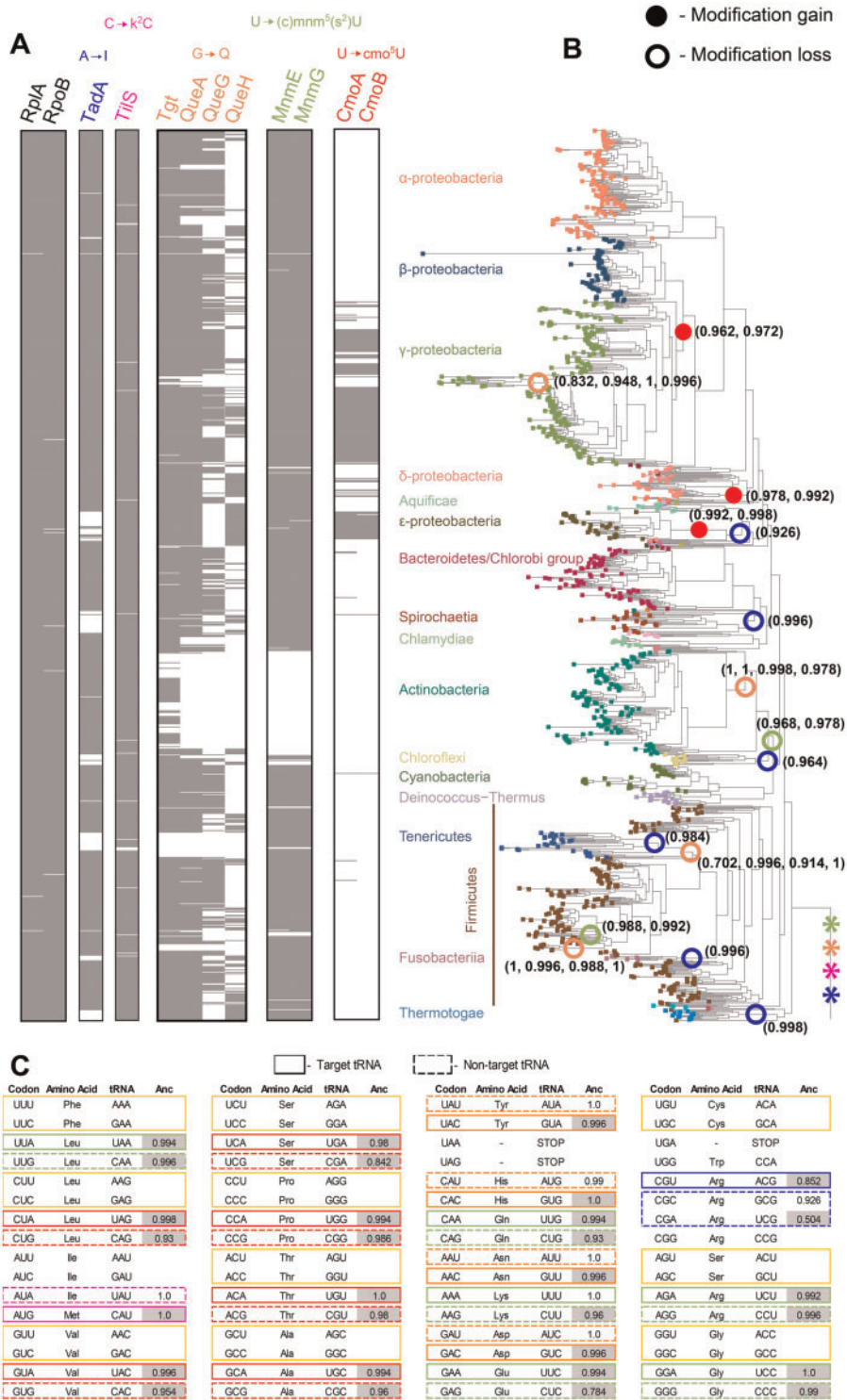


Fig. 1. The evolutionary history of bacterial tRNA modifications. (A) Columns indicate the presence (gray) or absence (white) of various tRNA MEs as noted above each column. The leftmost column shows two essential housekeeping proteins (*RpoB*- RNA polymerase beta subunit and *RplA*-50S Ribosomal protein L1) that serve as positive controls for homology detection. (B) Bacterial phylogenetic tree (Segata et al. 2013) pruned to show the evolutionary relationships between the 1093 bacterial species used in our study. Branch tips are colored by taxonomic group. Filled and open circles respectively mark the major gain and loss of tRNA modifications (with at least five descendent species in the relevant lineage), as inferred by ancestral reconstruction (see Materials and Methods). Circles are colored according to the modification, as indicated in (A). Values in parentheses indicate the posterior probability of each gain or loss event for each ME involved in the modification pathway. All other minor gain and loss events (with < 5 descendent species in the pruned phylogeny) are shown in [supplementary fig. S1, Supplementary Material](#) online. Asterisks at the root of the tree indicate the presence of the respective modification. (C) Genetic code table showing the target (solid boxes) and nontarget tRNA (dashed boxes) for each modification, colored as in (A). tRNA with a yellow border are those involved in GU wobble base pairing and were not considered in subsequent analyses. The column titled “Anc” indicates the presence (gray cells) or absence (white cells) of each tRNA in the eubacterial ancestor, inferred from ancestral reconstruction (see Materials and Methods). Value in each cell indicates the posterior probability of the inferred state for each tRNA.

our analysis demonstrates a highly dynamic evolutionary history of tRNA modifications that is dominated by repeated secondary losses and fewer gains (a total of 260 loss events and 105 gains across the phylogeny; [supplementary fig. S1, Supplementary Material](#) online).

tRNA Modification Loss Is Associated with More Nontarget tRNAs

The early evolution of nearly all tRNA modifications suggests that the eubacterial ancestor should have been able to decode most sense codons using only a subset of all possible tRNAs. Indeed, ancestral reconstruction of tRNA genes suggests that the eubacterial ancestor did have all target tRNAs (except tRNA_{Leu}^{UUU}), but did not possess nontarget tRNAs for the Q, k²C, and I modifications ([fig. 1C](#); note that the state of tRNA_{Arg}^{UCG} is ambiguous). Interestingly, for the (c)mnm⁵(s²)U modification, both target and nontarget tRNA were present in the eubacterial ancestor. Thus, a subsequent loss of any of these modifications should be strongly associated with the gain [or retention, in case of (c)mnm⁵(s²)U] of the respective nontarget tRNA. Both predictions were supported when we observed that the proportion of species that have nontarget tRNA is significantly different between bacteria with versus without a specific modification ($P < 0.05$, two-sample proportions test; [supplementary fig. S2, Supplementary Material](#) online). Overall, these patterns suggest that the loss of tRNA modifications was associated with the retention or secondary gain of nontarget tRNAs, and the gain of MEs was associated with the loss of nontarget tRNAs.

To account for phylogenetic relationships between analyzed bacteria, we compared the tRNA gene content of sister clades with contrasting modification status; that is, following a major modification gain or loss event (see Materials and Methods; [fig. 2](#)). For instance, since the (c)mnm⁵(s²)U modification was lost at the root of Actinobacteria (MnmE/MnmG enzymes, [fig. 1B](#)), we compared the tRNA gene content of species within the Actinobacteria clade with that of species in the sister group: the Cyanobacteria-Deinococcus-Thermus (CDT) clade. Similarly, we compared the target and nontarget tRNA gene copy numbers for descendent sister clades for 14 of the 15 major modification gain and loss events highlighted in [figure 1B](#) (we did not analyze Thermotogae, because its sister clade would consist of all other bacteria). Of these, four cases involving the Q modification are not informative with regard to nontarget tRNAs because these are absent in all bacteria ([fig. 2F–I](#)). Although the biochemical or evolutionary causes for the absence of these nontarget tRNA remain unknown, theoretically these tRNAs could exist, given the possible anticodon space. Therefore, we consider them in our analysis since they are useful for understanding the evolution of target tRNAs for the Q modification.

In most of the informative sister clade comparisons (8 out of 10 cases), species were more likely to have nontarget tRNAs if the modification was absent (compare bottom left vs. bottom right bar plots within each panel of [fig. 2](#)), supporting our hypothesis that modification loss (gain) is associated with the presence (absence) of nontarget tRNAs. For instance, in three out of the five comparisons made for the (c)mnm⁵(s²)U and

cmo⁵U modifications, at least five of the six nontarget tRNAs were present in a significantly higher proportion of species without the modification ([fig. 2A, C, and D](#); $P < 0.05$, two-sample proportions test). The two exceptions to this pattern occur in clades with relatively low sample sizes ([fig. 2B and E](#)), and may potentially reflect low statistical power. Alternatively, these may be biologically interesting exceptions that require further examination. Additionally, in all five comparisons made for the I modification, at least one nontarget tRNA (in most cases both) was present in significantly more species without the modification, compared with species with the modification ([fig. 2J–N](#); $P < 0.05$, two-sample proportions test).

In contrast to the wide variation in the occurrence of nontarget tRNAs, we found that target tRNA genes were largely stable irrespective of species' modification status (top row in each panel of [fig. 2](#)). Contrary to expectation, the gain or loss of the (c)mnm⁵(s²)U and cmo⁵U modifications does not appear to have any impact on target tRNAs ([fig. 2A–E](#); $P > 0.05$, two-sample proportions test); likely because the corresponding codons are still used in these species and cannot be decoded by nontarget tRNAs. In the case of the Q modification, target tRNAs were always present; but in two of four comparisons their copy number was lower in clades that had lost the ME ([fig. 2F–I](#); $P < 0.05$, two-sample proportions test). Inosine presents the only evidence supporting weakened selection on target tRNAs upon ME loss. In three of five comparisons, target tRNAs were absent in nearly all species without the modification ([fig. 2J–L](#)). In the remaining two comparisons, target tRNAs were present in fewer species without modifications ([fig. 2M and N](#); $P < 0.05$, two-sample proportions test). Together, these results suggest that except for the I modification, the evolution of modifications had no impact on the evolution of target tRNA genes.

Modification Status Is Associated with tRNA Diversity, GC Content, and Genome Size

To identify the cause and impact of evolutionary changes in tRNA modifications, we compared key translation-associated genomic features of sister clades with contrasting modification status. As shown above, we found that in 11 of 14 comparisons, the tRNA diversity of sister clades was significantly different ($P < 0.05$, Wilcoxon rank sum test; [fig. 3, supplementary fig. S3, Supplementary Material](#) online), indicating a strong association between the evolution of modifications and the tRNA gene pool. However, in four of these cases the direction of association opposed our initial expectation, and clades without the modification had lower tRNA diversity ([fig. 3G, I, M, and N](#)). In three of these cases involving Tenericutes and the endosymbiotic species from Enterobacterales ([fig. 3G, I, and M](#)), we also observed a significant reduction in the total number of tRNA genes ([supplementary fig. S3G, I, M, Supplementary Material](#) online), suggesting wide-ranging gene loss that is potentially a result of genome reduction. Indeed, previous reports have shown that reductive genome evolution in Mollicutes has led to the loss of several tRNA MEs (Yokobori et al. 2013; Grosjean et al. 2014). Thus, overall half the sister clade comparisons supported

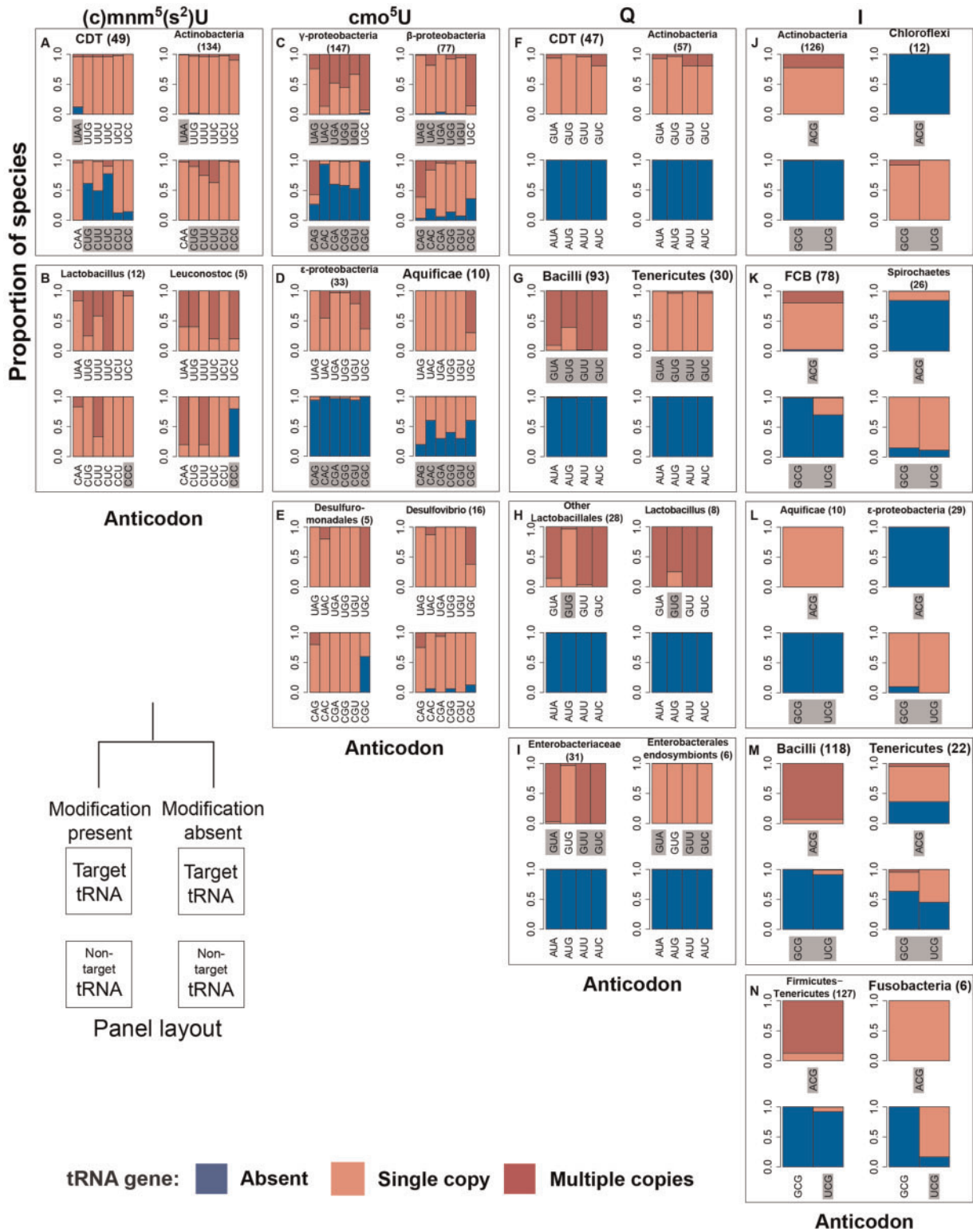


FIG. 2. tRNA gene content associated with modification gain or loss in sister clades. Gene copy number of target and nontarget tRNA within sister clades with contrasting modification status (marked in fig. 1B). Stacked bar plots show the proportion of species within each sister clade that lack a specific tRNA (blue bars), have a single copy of the gene (light red bars), or have multiple copies (dark red bars). The total number of species analyzed in each clade are given in parentheses above the bar plots. Each panel shows data for a specific sister clade comparison; panels by modification pathway. Gray shading represents anticodons (i.e., tRNA) for which the proportion of species without the tRNA is significantly different across groups, with respect to modification presence versus absence ($P < 0.05$; two-sample test for equality of proportions with continuity correction). CDT, Cyanobacteria-Deinococcus-Thermus; FCB, Fibrobacteres-Chlorobi-Bacteroidetes.

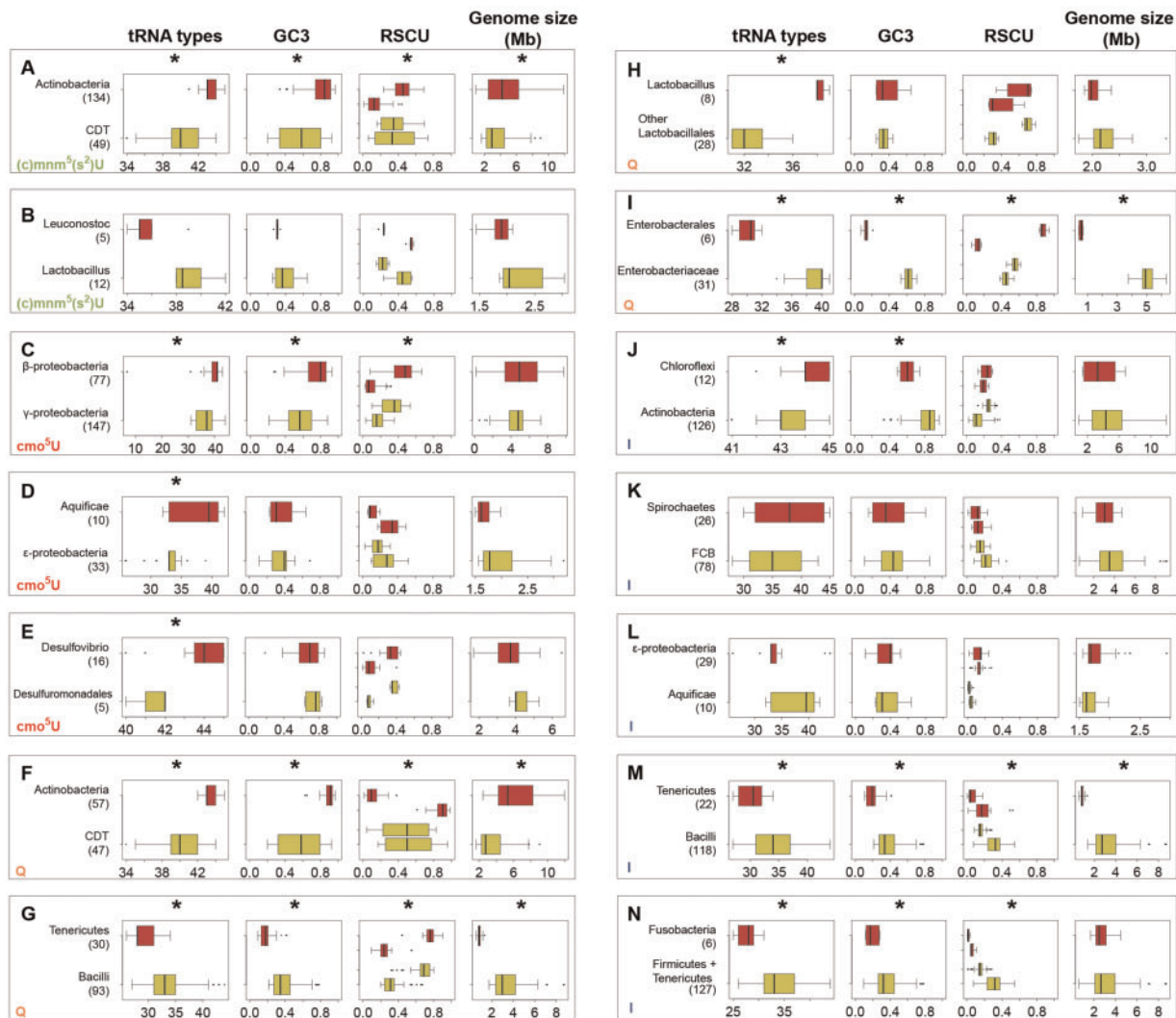


FIG. 3. Differences in key genomic features across sister clades with contrasting modification status. Box plots show the range of values of three genomic traits, across all species within a sister clade (red boxes = modification absent; green boxes = modification present). For the Relative Synonymous Codon Usage (RSCU) panel, asterisks indicate that the difference between the distributions represented by green box plots is significantly different than the difference between the distributions represented by the red box plots (i.e., codon usage is more biased in one case; $P < 0.05$; Wilcoxon rank sum test with P -values correction for multiple comparisons). For tRNA types, GC3 (GC content of the third codon base) and genome size, asterisks indicate a significant difference in distributions across the sister clades ($P < 0.05$; Wilcoxon rank sum test with P -values correction for multiple comparisons).

our prediction that the lack of modifications should be associated with greater tRNA diversity. Next, we tested whether changes in modification status and tRNA genes were also associated with changes in GC content and codon use.

In most cases showing altered tRNA diversity (8 out of 11), we found that the change in diversity was also associated with significant shifts in GC content, so that species in clades without the tRNA modification typically had more extreme GC3 values (i.e., GC content in the third base of codons, fig. 3; $P < 0.05$, Wilcoxon rank sum test; we observed similar results with genomic GC content, supplementary fig. S3, Supplementary Material online). In seven of these eight cases, species without the modification also had significantly more biased codon use (measured as relative synonymous codon use RSCU for all proteins; $P < 0.05$, Wilcoxon rank sum test;

we observed similar results for RSCU of highly expressed ribosomal proteins; supplementary fig. S3, Supplementary Material online). Finally, in five of the eight cases we observed significant differences in genome size across sister clades (fig. 3). Of these, in the two comparisons that involved Actinobacteria (fig. 3A and F) we observed that larger genomes were associated with the absence of the modification. As mentioned above, in the other three cases involving Tenericutes and Enterobacterales (fig. 3G, I, and M), the lack of modification may reflect overall genome reduction. Note that in every comparison where codon use and genome size were different, GC content was also skewed significantly. Thus, to test the overall effect of genome size and GC content on modification status, we carried out a phylogenetic regression analysis. We found that both GC content and genome size had a significant effect on the presence/absence of the Q

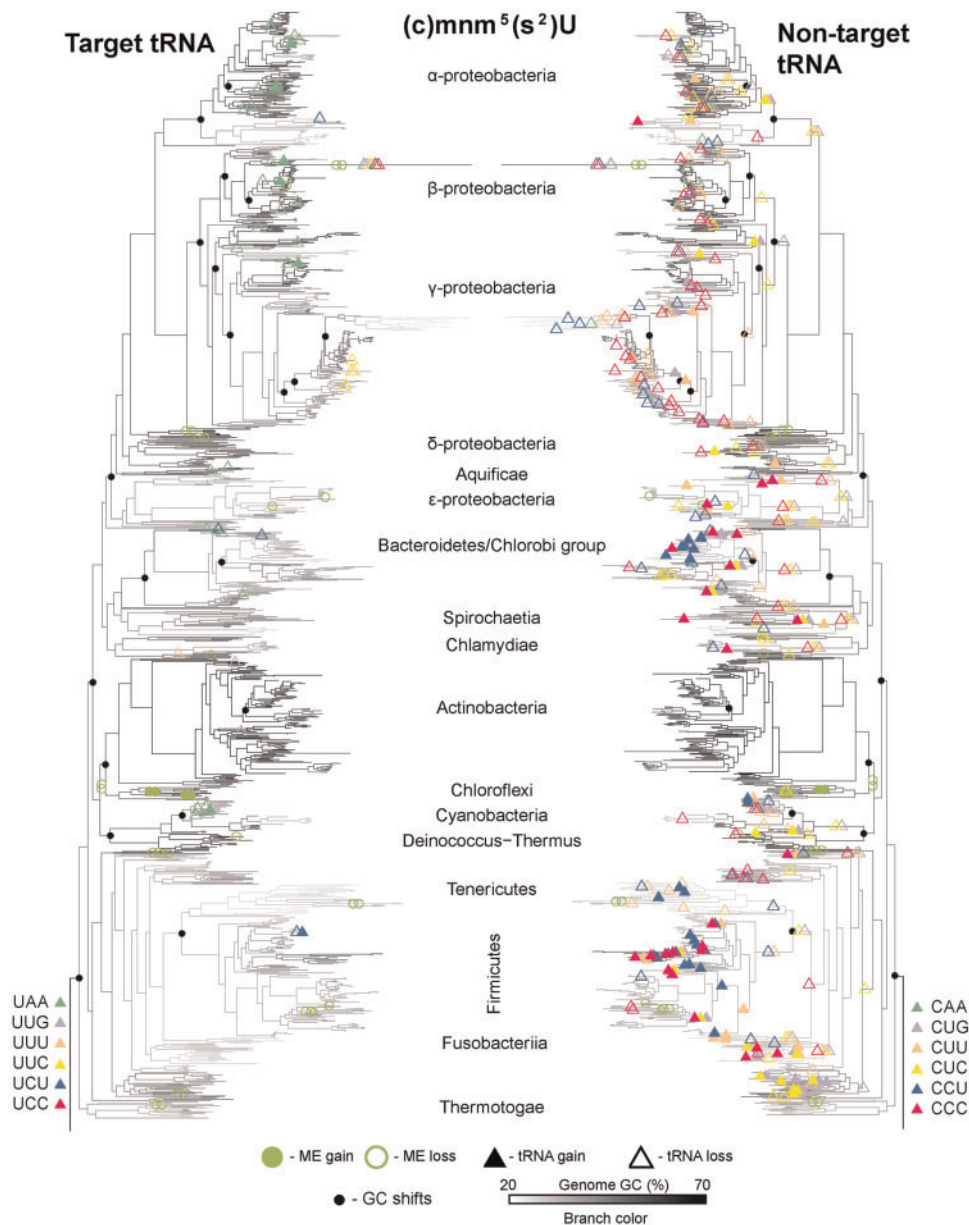


Fig. 4. Evolutionary dynamics of tRNA genes, MEs, and GC content for the (c)mnm⁵(s²)U modification. Branches are colored according to the inferred GC content for each branch, with white indicating low GC% and black indicating high GC%. Nodes that showed the largest GC shifts are indicated with black circles. Colored open and closed circles, respectively, indicate the loss and gain of each ME involved in the modification. Open and closed triangles indicate the loss and gain of tRNA genes relevant to the (c)mnm⁵(s²)U modification, with target tRNA indicated on the tree on the left and nontarget tRNA indicated on the mirror tree on the right.

modification alone (supplementary fig. S4, Supplementary Material online). A previous structural report shows that the Q base binds both C and U with a strength equivalent to GU pairing (Morris et al. 1999). This binding is weaker compared with the canonical GC base pair, and thus the Q modification may be lost in genomes with high or low GC content where canonical GC base pairs may be preferred. This hypothesis fits three out of four cases for the loss of the Q modification. Overall, these results suggest that except for the Q modification, modifications were lost in only a few instances of shifts in GC content. Thus, GC content may be the major ultimate driver of changes in tRNA gene content and the subsequent loss of modification may generate

selection to fix these changes. Altogether, 3 of the 14 major changes in modification status were not associated with significant shifts in either tRNA or GC content, and remain unexplained. These included one instance of the loss of the (c)mnm⁵(s²)U modification (fig. 3B) and two losses of the I modification (fig. 3K–L).

Finally, to quantify the evolutionary impact of tRNA modifications on tRNA gene pools, we simultaneously mapped changes in tRNA genes copies, MEs, and GC content on the bacterial phylogeny. As expected, following major GC shifts, nontarget tRNA showed several gain/loss events whereas target tRNA were more stable (figs. 4–6). As suggested by our sister clade comparison, in some cases modification gain or

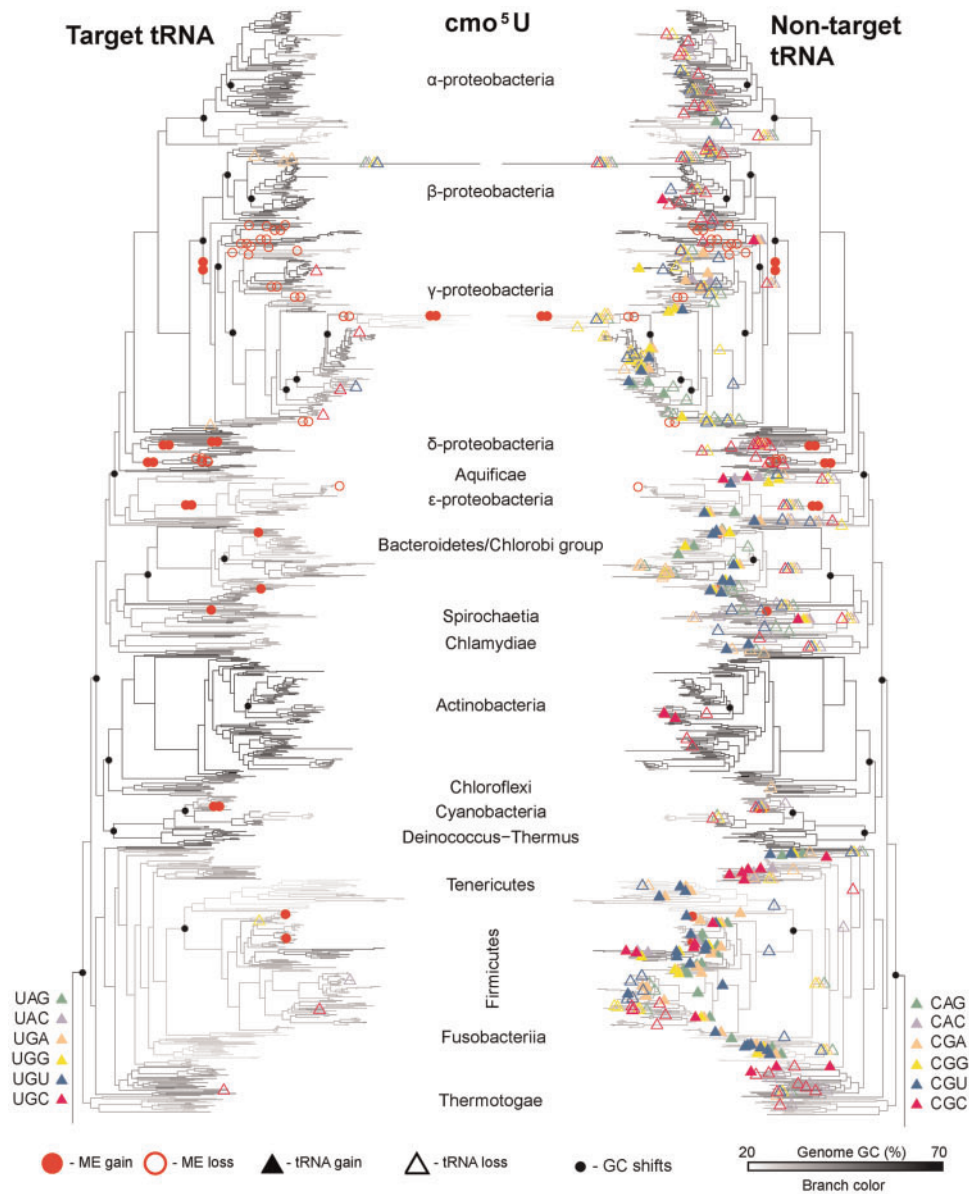


Fig. 5. Evolutionary dynamics of tRNA genes, MEs, and GC content for the cmo^5U modification: Branches are colored according to the inferred GC content for each branch, with white indicating low GC% and black indicating high GC%. Nodes that showed the largest GC shifts are indicated with black circles. Colored open and closed circles, respectively, indicate the loss and gain of each ME involved in the modification. Open and closed triangles indicate the loss and gain of tRNA genes relevant to the cmo^5U modification, with target tRNA indicated on the tree on the left and nontarget tRNA indicated on the mirror tree on the right.

loss was also preceded by a GC shift (8 out of 14 independent lineages). Moreover, a phylogenetic regression to test the impact of GC and MEs on tRNA genes revealed that GC content had the strongest impact on the dynamics of nontarget tRNA genes (fig. 7). However, as observed in the sister clade comparison, the phylogenetic regression and mapping also reveal clear instances of GC independent impacts of modification status on tRNA genes (figs. 4–7). The strongest impact of tRNA modification is observed the case of the Inosine modification, where the target tRNA was present in all species when the modification was present (fig. 6); and the nontarget tRNA were absent in most of these species (fig. 2D). On the other hand, the loss of target tRNAs and presence of nontarget tRNAs is strongly associated with the loss of the Inosine

modification (figs. 2D, 6, and 7). Thus, the bacterial tRNA repertoire is strongly affected by evolutionary changes in GC content as well as tRNA MEs.

Discussion

Over 50 years ago, Francis Crick proposed the phenomenon of wobble base pairing where one tRNA molecule can recognize multiple codons (Crick 1966). Since then, a large body of work has developed a detailed understanding of the enzymatic pathways that generate extended wobble pairing via base modifications of tRNA anticodons [reviewed in Björk and Hagervall (2014)]. Given their critical role in translation, MEs are expected to evolve under strong selection, and in



Fig. 6. Evolutionary dynamics of tRNA genes, MEs, and GC content for the I modification: Branches are colored according to the inferred GC content for each branch, with white indicating low GC% and black indicating high GC%. Nodes that showed the largest GC shifts are indicated with black circles. Colored open and closed circles, respectively, indicate the loss and gain of each ME involved in the modification. Open and closed triangles indicate the loss and gain of tRNA genes relevant to the I modification, with target tRNA indicated on the tree on the left and nontarget tRNA indicated on the mirror tree on the right.

turn influence the evolution of other components of the translation machinery (Grosjean et al. 2010). Here, we trace the evolutionary history of tRNA modifications across the bacterial phylogeny, showing that the eubacterial ancestor already had nearly all known modification pathways. Some modifications such as the k^2C modification were lost in only a few bacterial species. A previous report showed that the absence of the k^2C modification in such species was compensated by the presence of the nontarget tRNA^{lle}_{UAU} and altered specificity of the Isoleucine tRNA synthetase (Taniguchi et al. 2013). Additionally, we show that several major bacterial clades lost other modifications, often accompanied by increasing tRNA diversity and shifts in genome GC content and codon use. These results are consistent with the

hypothesis that an expanded tRNA pool gradually weakens selection on MEs, allowing their loss through drift. However, once the modification is lost, bacteria likely face strong selection to retain the expanded tRNA pool. In contrast, if GC shifts shrink the tRNA gene pool, selection would strongly favor the innovation or retention of tRNA modifications. We speculate that in Proteobacteria, such a process may have led to the evolution of the novel cmo^5U modification. Alternatively, Proteobacteria may have evolved the new modification by chance, secondarily weakening selection on the nontarget tRNAs. Interestingly, we also uncovered instances where changes in tRNA or GC content could not explain the evolution of tRNA MEs. In some of these cases, we suspect that overall genome reduction may have led to the loss of the

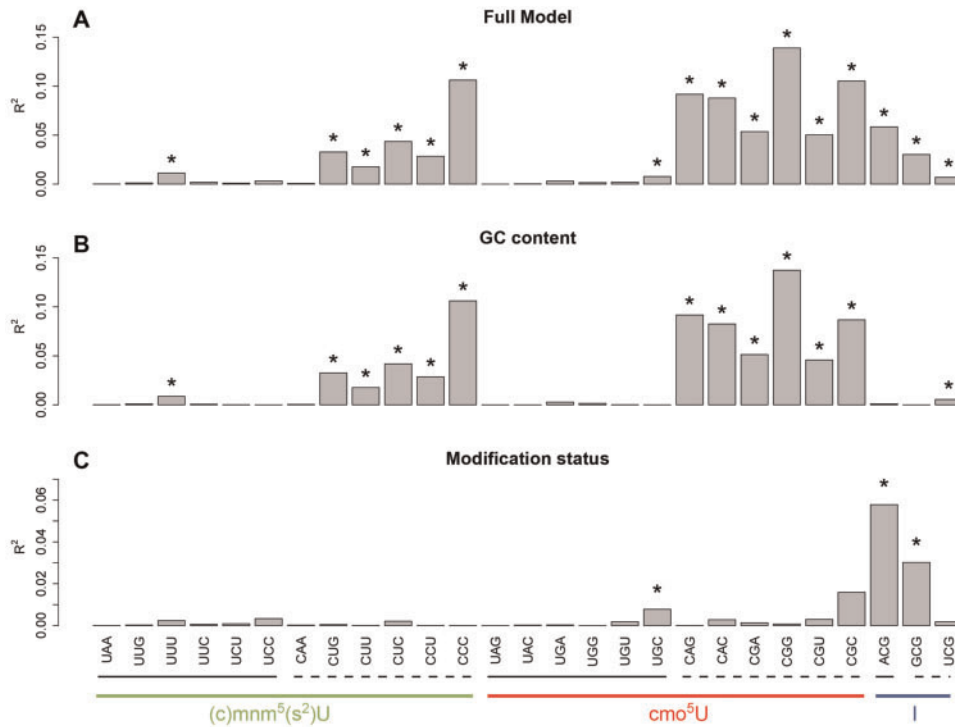


Fig. 7. Results of phylogenetic regressions for factors affecting tRNA gene presence or absence. Bar plots show the R^2 value for the phylogenetic regression for presence/absence of target and nontarget tRNAs for the (c)mm⁵(s²)U, cm⁵U, and I modifications. Results are shown for three different models: (A) Full model, tRNA presence/absence \sim Modification status (i.e., presence/absence) * GC content. (B) tRNA presence/absence \sim GC content (C) tRNA presence/absence \sim Modification status (i.e., presence/absence). Asterisks indicate a significant impact of the focal trait on the presence of the tRNA gene (see Materials and Methods). Target tRNAs are highlighted with a solid underline, and nontarget tRNAs with a dashed underline. Note the different γ -axis scale in (C).

modification; but three major cases of modification loss remain unexplained [loss of (c)mm⁵(s²)U in *Leuconostoc* and I in *Spirochaetes* and ϵ -proteobacteria]. Nonetheless, our results strongly support the conclusion that ME evolution was dominated by frequent losses driven by weakened selection. Note that reversing the direction of causality in our arguments requires implausible events: the loss of nontarget tRNAs cannot generate selection favoring the loss of tRNA modifications, and there is no reason to expect that the gain of a modification enzyme should selectively favor the retention of nontarget tRNAs. Similarly, it is difficult to imagine how evolutionary gains or losses of tRNA modifications could precipitate genome-scale changes in GC content or genome size. Thus, our systematic phylogenetic analysis clarifies the contrasting roles of strong versus weak selection acting on tRNA modifications, and on the evolutionary history of key components of bacterial translation.

Our results are generally consistent with the prediction that a lack of specific tRNA modifications should result in strong positive selection favoring the relevant nontarget tRNAs. However, we also observed some instructive exceptions to this pattern. For instance, comparing across sister clades, we found that \sim 33% of the species that do not have the cm⁵U modification also lack at least one nontarget tRNA. This should pose a problem because codons recognized by the missing nontarget tRNAs cannot be decoded efficiently. However, we observed that these codons were

significantly depleted in these genomes, compared with species that retained the nontarget tRNA (supplementary fig. S5, Supplementary Material online). Thus, these relatively rare codons may not pose a major problem even if they are inefficiently translated. It is also possible that these codons are decoded by the respective U-starting target tRNAs that can pair with A, G or U ending codons (Grosjean et al. 2010). This scenario is especially likely in species where the Q modification is absent. Recall that all bacteria lack the nontarget tRNA for this modification; hence species without the Q modification should be unable to translate many codons. We hypothesize that in species where both the Q modification as well as the nontarget tRNA are missing, there is an as yet undiscovered GU wobble base pairing between the target tRNA and codons complementary to the nontarget tRNA of the Q modification in these species.

Our analysis quantifies the role of specific tRNAs in the evolutionary dynamics of bacterial translation, both as drivers of selection on tRNA modifications and as a means to rapidly respond to shifts in genomic GC content. Specifically, we found that target tRNAs are evolutionarily largely stable whereas nontarget tRNAs are often gained or lost in lineages following a GC shift. Interestingly, the losses of the Inosine modification illustrate instances where tRNA genes were gained or lost independently of GC shifts. Overall, the strong phylogenetic correlation between nontarget tRNAs and genomic GC content echoes the previously reported association

between GC content and “auxiliary” tRNAs (tRNAs with a high number of gains and losses) defined by Wald and Margalit (Wald and Margalit 2014). The nearly ubiquitous presence of target tRNAs may reflect strong positive selection, since the function of target tRNA cannot be carried out by any other tRNA via wobble base pairing rules. Thus, target tRNA may be essential even if their complementary codons are used relatively rarely.

In summary, we propose that the evolution of diverse bacterial tRNA modification enzymes has been strongly influenced by changes in tRNA gene pools and genomic GC content, as well as by genome reduction in specific lineages. Our results support the hypothesis that these changes weakened selection on MEs and allowed their loss through drift. Conversely, we hypothesize that the loss of nontarget tRNA may have generated strong selection for the innovation of the cmo^5U modification in γ -proteobacteria and ϵ -proteobacteria. It is also possible that the tRNA loss succeeded only in the wake of the chance innovation of the cmo^5U modification. These patterns generate a number of testable predictions about the future evolutionary trajectory of translational components in various bacterial lineages. For instance, MEs may be lost in GC-rich bacteria without major consequences for translation; such losses should evolutionarily fix the expanded tRNA repertoire; and MEs should face strong positive selection in AT-rich clades with low tRNA diversity. Specifically, in AT rich bacteria, the combination of MEs and largely target tRNAs should confer higher fitness than a full tRNA set without MEs. We hope that future experimental work can test these predictions about the evolutionary dynamics of translation in specific bacterial lineages. Our analysis also shows that multiple key components of bacterial translation have evolved in close association with genomic GC content. Although the evolutionary and ecological forces driving GC shifts and GC independent changes in tRNA diversity remain unclear, our work highlights the rich evolutionary dynamics of wobble pairing, and more broadly, bacterial translation.

Materials and Methods

Phylogenetic Tree, Genomes, and tRNA Gene Data

All analyses were carried out in R (R Development Core Team 2015) using the phytools (Revell 2012) and seqinr (Charif et al. 2005) packages, unless otherwise specified. The R scripts for all analyses are available at https://github.com/gauravdiwan89/me_evolution_project. We used the bacterial phylogeny generated by Segata et al. (2013). This phylogeny was generated using an alignment of the 400 most conserved protein sequences across all prokaryotes. We downloaded the phylogeny and removed all archaea, as well as bacteria whose genomes have not been fully sequenced. We pruned the tree further using a distance based pruning function. Briefly, when closely related species had a pairwise phylogenetic distance ≤ 0.03 , we randomly sampled only one of these species. This allowed us to avoid excessive representation of some highly sequenced genomes, such as several *Escherichia coli* strains. This led to the final set of 1093 bacterial genomes that we used in our subsequent analyses.

We downloaded the tRNA gene copy numbers for each of these genomes from the Genomic tRNA database (Chan and Lowe 2009). For genomes where this information was not available, we determined the tRNA gene copy numbers by running the program tRNAscan-SE (Lowe and Eddy 1997) using the default parameters for bacterial genomes. The information for each of these genomes was represented using different identifiers in the various databases that we used. This information is consolidated in [supplementary table S1, Supplementary Material](#) online, consisting of: (1) species name according to NCBI taxonomy, (2) NCBI accession number, (3) Phylogenetic tree tip label in the tree, (4) IMG database identifier, and (5) Genomic tRNA database abbreviation.

Choice of Crucial tRNA MEs and Nontarget tRNA

The (c)mnm⁵(s²)U and Q modifications are created by elaborate ME pathways. For instance, our definition of the (c)mnm⁵(s²)U modification includes three different modifications: mnm5U, cmnm5U, and cmnm5s2U, each acting on different isoaccepting tRNA. These modifications have unique MEs that are involved in the last steps of the pathway. MnmE and MnmG are the common enzymes for each of these modifications and they carry out the first step of the modifying the Uridine base (Björk and Hagervall 2014 and references therein). Thus, we considered MnmE-MnmG to be crucial, and other enzymes such as MnmC, MnmA, MnmH, and trmL as auxiliary. The Q base is first biochemically synthesized as preQ₁ and then Q₁ in the bacterial cell using the enzymes FolE, QueC, QueD, QueE, and QueF. Subsequently, the Tgt enzyme replaces Guanosine in the anticodon stem loop with Q₁, and QueA and QueG convert it to Q (Björk and Hagervall 2014 and references within). A recent report shows that a nonorthologous enzyme QueH carries out the function of QueG in species that lack QueG but have the Q modification (Zallot, Ross, et al. 2017). Another report has also suggested the possibility of Tgt alone carrying out the Q modification in the absence of QueA and QueG, assuming the existence of a Q salvage pathway (Zallot, Yuan, et al. 2017). However, there is no evidence for this phenomenon across most bacteria in our analysis; hence, to be conservative, we considered all four enzymes (Tgt, QueA, QueG, and QueH) as crucial enzymes for the Q modification. For each modification, we only considered species that had all crucial enzymes as having the modification; and denoted species that lack any of these enzymes as lacking the modification.

The cmo^5U modification exists in tRNA molecules for four box codons (Leucine, Valine, Serine, Proline, Threonine, and Alanine) and therefore there were three possible nontarget tRNA in each case. However, tRNA with GNN anticodons (where N is any nucleotide) in each of these codon boxes can also recognize NNU codons via GU wobble pairing. Thus, we only focused on the nontarget tRNA that have CNN anticodons (indicated in [fig. 1C](#)).

Homology Search for tRNA MEs

We determined the presence/absence of all known tRNA MEs in the 1093 bacterial genomes using hidden Markov model (HMM) searches (Eddy 1998). First, we downloaded

all available reviewed protein sequences for each of the MEs from the UniProt-KB database. We created a multiple sequence alignment using the t-coffee program (Notredame et al. 2000) with default parameters. We used this alignment to build an HMM profile, and then used the “hmmsearch” command in the HMMER suite (Eddy 2009) to detect the presence/absence of each ME. We used the amino acid sequences of all proteins from the 1093 genomes (downloaded in November 2016 from the NCBI ftp site) as the sequence search space, and set all other parameters to default. Each search generated a table with accession numbers of potential homologous hits with a corresponding *e*-value and bit score. There is no consensus in the field with respect to setting an *e*-value cut-off to detect true homologs. Here, we present a slightly informative method of deciding *e*-value cut-offs for detecting true homologs. To infer presence/absence with precision, we implemented a dynamic *e*-value cut-off to eliminate spurious hits. We did this by plotting the number of species in which the ME would be detected as we increased the *e*-value cut-off (supplementary fig. S6, Supplementary Material online). We observed stable plateaus where the number of species in which an ME was detected did not change despite changing the *e*-value cut-off, suggesting that these represent robust *e*-value thresholds. Thus, for each enzyme we set an enzyme-specific *e*-value cut-off as the first *e*-value at the beginning of the most stable plateau (see supplementary fig. S6, Supplementary Material online). For each ME, homologous hits that had *e*-values lower than this cut-off were considered true homologues. The information on Refseq ID of the detected homologues, organism in which they were detected and their current annotation is shown in supplementary table S2, Supplementary Material online.

Previous reports have experimentally shown that the *cmo*^{5U} modification occurs in *B. subtilis* and *M. bovis* str. BCG (Murao et al. 1976; Yamada et al. 2005; Chionh et al. 2016). However, in our initial analysis we did not find CmoA and CmoB homologs in these or related species. A closer inspection of the *B. subtilis* genome also did not reveal these enzymes, as indicated in a previous report (Grosjean et al. 2014). Instead, there were a few entries for these enzymes in the UniProt database (each with a poor annotation score). We therefore tested whether these enzymes are true homologs of the proteobacterial CmoA and CmoB, using homology detection (using “jackhammer”) (Eddy 2009). We observed that only the query organisms showed the presence of these enzymes (data not shown). A survey of the SEED database (Overbeek et al. 2005) also revealed the absence of CmoA and CmoB in all but one *Bacillus* species. When we carried out a blastp analysis using CmoA and CmoB enzymes from *E. coli* in *M. bovis* str. BCG, we observed no significant hits for CmoA and no hits for CmoB with *e*-values $< 5 \times 10^{-3}$. Thus, although the modification may exist in several species, we could not reliably identify homologs of the proteobacterial CmoA and CmoB in other bacteria.

Ancestral Reconstruction of tRNA Modifications, Genomic GC Content, and tRNA Genes

We built a presence–absence matrix for each ME and used the binary state of each ME to infer gain or loss events along

the phylogeny. We used the stochastic character mapping function (Bollback 2006) in the phytools package in R, with the following parameters: transition rate matrix determination method—Bayesian MCMC; prior distribution on the root node of the tree—estimated; number of simulations—500. We fixed the number of simulations such that the posterior probability that an enzyme is present for 100 randomly sampled nodes did not change (supplementary fig. S7, Supplementary Material online). We used the enzyme with the most variable presence/absence state (queG) for this analysis, and assumed that calculations for all other enzymes would be more stable. The mapping resulted in 500 enzyme states at each node of the tree. We determined probabilities of each state at each node and assigned the state that had a posterior probability ≥ 0.7 (supplementary fig. S8, Supplementary Material online) as the state of the enzyme at that node. In case the probability was < 0.7 , we assumed the same state at the node as the preceding node. To determine gain and loss of tRNA modification pathways, we marked the branch that succeeded the gain or loss of all MEs for a given modification as the focal branch. Only these events are described in this study; however, all data for each enzyme are shown in supplementary fig. S1, Supplementary Material online.

We determined the GC content at each ancestral node in the bacterial phylogeny using the StableTraits program (Elliot and Mooers 2014) with 10 million iterations and all other parameters set to default. We then determined 20 nodes whose two immediate descendants had the largest difference in GC. Each of these nodes had at least 10 descendant species. These nodes representing major GC shifts are shown on the phylogeny along with the ancestral GC content (figs. 4–6).

We determined the evolutionary history of tRNA gains and losses using a similar method as described for MEs. However, in this case we carried out the ancestral reconstruction separately for each bacterial family, which allowed us to account for family-specific rates of evolution of tRNA genes. We used the ancestral state of each tRNA at the root of each bacterial family to infer the tRNA gene content of the eubacterial ancestor. Briefly, we converted tRNA gene copy numbers for all species into a binary vector by changing the state of tRNA genes with multiple copies to “1,” indicating presence. We then carried out ancestral reconstruction using stochastic character mapping and noted evolutionary transitions on the phylogenetic tree as before.

Identifying Sister Clades with Contrasting Modifying Enzyme Status

We identified clades where crucial enzymes for each tRNA modification were gained or lost, and determined the closest sister clade where the state of modification was opposite to that of the focal clade. We determined the sister clade agnostic of its taxonomic classification. We checked if there were at least five descendant extant species in this clade for each gain or loss event. If this criterion was not met, we picked the next closest sister clade neighboring the common ancestor of the above two clades. Within each clade, we excluded species that

had an opposite state for any of the MEs involved in each modification. We assigned the identity of the comparison group using the lowest level of taxonomic classification that encompassed the focal species.

Phylogenetic Regression

We carried out phylogenetic regression to test the impact of GC content and modifications on tRNA presence/absence, using the Maximum Likelihood Continuous Regression program in BayesTraits, similar to the method used by Wald and Margalit (Wald and Margalit 2014). We estimated the lambda parameter for each case and used default values for all other parameters. For each tRNA, we first tested the full model: tRNA presence/absence \sim GC content*tRNA modification status. To determine the impact of each explanatory variable, we compared the likelihood of the full regression model to the likelihood of the same regression when the R value was set to zero. For reduced models containing only GC content or modification status as the explanatory variable, we compared the likelihood of the regression of the reduced model with the likelihood of the regression of the full model. When these likelihoods were comparable ($P > 0.05$; likelihood ratio test), we inferred that the reduced model was sufficient to explain tRNA presence/absence, and hence had a significant impact on the evolution of the tRNA. We used the same method for testing the impact of GC content and genome size on modification status (supplementary fig S4, Supplementary Material online).

Supplementary Material

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

Acknowledgments

We thank Saurabh Mahajan for help with analyses; and Aparna Agarwal, Shyam Buddh, Vrinda Ravi Kumar, Saurabh Mahajan, and Laasya Samhita for constructive comments on the manuscript. This work was funded by the National Centre for Biological Sciences (NCBS-TIFR) and grants from the Department of Science and Technology, India (INSPIRE Faculty award, IFA-1 LSBM-64) and the Council for Scientific and Industrial Research, India [37(1629)/14/EMR-II]. G.D.D. acknowledges a Senior Research Fellowship from the University Grants Commission, India.

References

- Agris PF, Vendeix FAP, Graham WD. 2007. tRNA's wobble decoding of the genome: 40 years of modification. *J Mol Biol.* 366(1):1–13.
- Björk GR, Hagervall TG. 2014. Transfer RNA modification: presence, synthesis, and function. *EcoSal Plus* 6(1).
- Boccalletto P, Machnicka MA, Purta E, Piatkowski P, Baginski B, Wirecki TK, de Crécy-Lagard V, Ross R, Limbach PA, Kotter A, et al. 2018. MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* 46(D1):D303–D307.
- Bollback JP. 2006. SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* 7:88.
- Brierley I, Meredith MR, Bloys AJ, Hagervall TG. 1997. Expression of a coronavirus ribosomal frameshift signal in *Escherichia coli*: influence of tRNA anticodon modification on frameshifting. *J Mol Biol.* 270(3):360–373.
- Chan PP, Lowe TM. 2009. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* 37(Database):D93–D97.
- Charif D, Thioulouse J, Lobry JR, Perrière G. 2005. Online synonymous codon usage analyses with the ade4 and seqinR packages. *Bioinformatics* 21(4):545–547.
- 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 USA.* 101(10):3480–3485.
- Chionh YH, McBee M, Babu IR, Hia F, Lin W, Zhao W, Cao J, Dziergowska A, Malkiewicz A, Begley TJ, et al. 2016. tRNA-mediated codon-biased translation in mycobacterial hypoxic persistence. *Nat Commun.* 7:13302.
- Crick FH. 1966. Codon–anticodon pairing: the wobble hypothesis. *J Mol Biol.* 19(2):548–555.
- Eddy S. 1998. Profile hidden Markov models. *Bioinformatics* 14(9):755–763.
- Eddy SR. 2009. A new generation of homology search tools based on probabilistic inference. *Genome Inform.* 23(1):205–211.
- Elliot MG, Mooers AØ. 2014. Inferring ancestral states without assuming neutrality or gradualism using a stable model of continuous character evolution. *BMC Evol Biol.* 14:226.
- Gao T, Tan M, Liu W, Zhang C, Zhang T, Zheng L, Zhu J, Li L, Zhou R. 2016. GidA, a tRNA modification enzyme, contributes to the growth, and virulence of *Streptococcus suis* serotype 2. *Front Cell Infect Microbiol.* 6:1–11.
- Grosjean H, Breton M, Sirand-Pugnet P, Tardy F, Thiaucourt F, Citti C, Barré A, Yoshizawa S, Fourmy D, de Crécy-Lagard V, et al. 2014. Predicting the minimal translation apparatus: lessons from the reductive evolution of mollicutes. *PLoS Genet.* 10(5):e1004363.
- Grosjean H, de Crécy-Lagard V, Marck C. 2010. Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. *FEBS Lett.* 584(2):252–264.
- Hershberg R, Petrov DA. 2009. General rules for optimal codon choice. *PLoS Genet.* 5(7):e1000556.
- Ikemura T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol. Evol.* 2(1):13–34.
- 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.
- 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.
- Machnicka M, a, Olchowiak A, Grosjean H, Bujnicki JM. 2015. Distribution and frequencies of post-transcriptional modifications in transfer RNAs. *RNA Biol.* 6286:0.
- Manickam N, Joshi K, Bhatt MJ, Farabaugh PJ. 2015. Effects of tRNA modification on translational accuracy depend on intrinsic codon–anticodon strength. *Nucleic Acids Res.* 44(4):1871–1881.
- 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.
- Morris RC, Brown KG, Elliott MS. 1999. The effect of queuosine on tRNA structure and function. *J Biomol Struct Dyn.* 16(4):757–774.
- Murao K, Hasegawa T, Ishikura H. 1976. 5-Methoxyuridine: a new minor constituent located in the first position of the anticodon of tRNA^{Ala}, tRNA^{Thr}, and tRNA^{Val} from *Bacillus subtilis*. *Nucleic Acids Res.* 3(10):2851–2860.
- Näsvall SJ, Chen P, Björk GR. 2004. The modified wobble nucleoside uridine-5-oxyacetic acid in tRNA^{Pro}(cmo5UGG) promotes reading of all four proline codons in vivo. *RNA* 10(10):1662–1673.
- Näsvall SJ, Chen P, Björk GR. 2007. The wobble hypothesis revisited: uridine-5-oxyacetic acid is critical for reading of G-ending codons. *RNA* 13(12):2151–2164.

- Noguchi S, Nishimura Y, Hirota Y, Nishimura S. 1982. Isolation and characterization of an *Escherichia coli* mutant lacking tRNA-guanine transglycosylase. Function and biosynthesis of queuosine in tRNA. *J Biol Chem*. 257:6544–6550.
- Notredame C, Higgins D, Heringa J. 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol*. 302(1):205–217.
- Novoa EM, Pavon-Eternod M, Pan T, Ribas De Pouplana L. 2012. A role for tRNA modifications in genome structure and codon usage. *Cell* 149(1):202–213.
- Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang H-Y, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R, et al. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res*. 33(17):5691–5702.
- R Development Core Team. 2015. R: a language and environment for statistical computing.
- Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol*. 3(2):217–223.
- 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.
- Shippy DC, Eakley NM, Lauhon CT, Bochsler PN, Fadl AA. 2013. Virulence characteristics of *Salmonella* following deletion of genes encoding the tRNA modification enzymes GidA and MnmE. *Microb Pathog*. 57:1–9.
- Taniguchi T, Miyauchi K, Nakane D, Miyata M, Muto A, Nishimura S, Suzuki T. 2013. Decoding system for the AUA codon by tRNA Ile with the UAU anticodon in *Mycoplasma mobile*. *Nucleic Acids Res*. 41(4):2621–2631.
- Wald N, Margalit H. 2014. Auxiliary tRNAs: large-scale analysis of tRNA genes reveals patterns of tRNA repertoire dynamics. *Nucleic Acids Res*. 42(10):6552–6566.
- Wolf J, Gerber AP, Keller W. 2002. *tadA*, an essential tRNA-specific adenosine deaminase from *Escherichia coli*. *Embo J*. 21(14):3841–3851.
- Yamada Y, Matsugi J, Ishikura H, Murao K. 2005. *Bacillus subtilis* tRNA^{Pro} with the anticodon mo5UGG can recognize the codon CCC. *Biochim Biophys Acta Gene Struct Exp*. 1728(3):143–149.
- Yokobori SI, Kitamura A, Grosjean H, Bessho Y. 2013. Life without tRNA^{Arg}-adenosine deaminase TadA: evolutionary consequences of decoding the four CGN codons as arginine in Mycoplasmas and other Mollicutes. *Nucleic Acids Res*. 41(13):6531–6543.
- Yokoyama S, Watanabe T, Murao K, Ishikura H, Yamaizumi Z, Nishimura S, Miyazawa T. 1985. Molecular mechanism of codon recognition by tRNA species with modified uridine in the first position of the anticodon. *Proc Natl Acad Sci USA*. 82(15):4905–4909.
- Zallot R, Ross R, Chen WH, Bruner SD, Limbach PA, De Crécy-Lagard V. 2017. Identification of a novel epoxyqueuosine reductase family by comparative genomics. *ACS Chem Biol*. 12(3):844–851.
- Zallot R, Yuan Y, de Crécy-Lagard V. 2017. The *Escherichia coli* COG1738 member YhhQ is involved in 7-Cyanodeazaguanine (preQ0) transport. *Biomolecules* 7(4):12–13.