

1 THE EVOLUTIONARY HISTORY AND IMPACT OF BACTERIAL tRNA 2 MODIFICATIONS

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

4

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

6 ²SASTRA University, Thanjavur, India

7

8 *Correspondence:

9 dagashe@ncbs.res.in

10

11 ABSTRACT

12 Along with tRNAs, enzymes that modify anticodon bases are a key aspect of translation across the
13 tree of life. tRNA modifications extend wobble pairing, allowing specific (“target”) tRNAs to
14 recognize multiple codons and cover for other (“non-target”) tRNAs, often improving translation
15 efficiency and accuracy. However, the detailed evolutionary history and impact of tRNA modifying
16 enzymes has not been analyzed. Using ancestral reconstruction of five tRNA modifications across
17 1093 bacteria, we show that most modifications were ancestral to eubacteria, but were repeatedly
18 lost in many lineages. Most modification losses coincided with evolutionary shifts in non-target
19 tRNAs, often driven by increased bias in genomic GC and associated codon use, or by genome
20 reduction. In turn, the loss of tRNA modifications stabilized otherwise highly dynamic tRNA gene
21 repertoires. Our work thus traces the complex history of bacterial tRNA modifications, providing
22 the first clear evidence for their role in the evolution of bacterial translation.

23

24 **KEYWORDS:** tRNA modifications, ancestral reconstruction, tRNA gene content, GC content

1 INTRODUCTION

2 Compared to the total complement of 61 sense codons, most bacterial genomes contain only 25 – 46
3 unique tRNA species (Grosjean, de Crécy-Lagard, & Marck, 2010). This apparent shortfall in
4 decoding ability is mitigated by G::U wobble base pairing that allows a single tRNA to decode non-
5 complementary codons (Crick, 1966). Additionally, specific modifications to the first anticodon
6 base of the tRNA also contribute to wobble base pairing, sometimes reducing translational errors
7 (Björk & Hagervall, 2014; Brierley, Meredith, Bloys, & Hagervall, 1997; Manickam, Joshi, Bhatt,
8 & Farabaugh, 2015; Marck & Grosjean, 2002; Nasvall, Chen, & Bjork, 2004; Näsval, Chen, &
9 Björk, 2007; Yokoyama et al., 1985). Base modifications are carried out by several tRNA modifying
10 enzyme (ME) pathways, modifying the wobble position of a large fraction (42%) of all tRNA
11 species (Boccaletto et al., 2018; Grosjean et al., 2010; Machnicka, Olchowik, Grosjean, & Bujnicki,
12 2015). Given their potential impact on the effective tRNA pool and on translation dynamics, MEs
13 should generally face strong selection, especially in species with a restricted tRNA repertoire.
14 Indeed, in several bacteria, deleting modifying enzymes is lethal or otherwise deleterious for growth
15 and translation (Björk & Hagervall, 2014; Gao et al., 2016; Shippy, Eakley, Lauhon, Bochsler, &
16 Fadl, 2013; Wolf, Gerber, & Keller, 2002). However, in some cases MEs do not appear to be
17 essential (Noguchi, Nishimura, Hirota, & Nishimura, 1982), suggesting strong yet variable
18 evolutionary impact of tRNA modifications. Recently, Novoa and colleagues showed that MEs
19 unique to the three kingdoms of life have shaped kingdom-specific tRNA pools (Novoa, Pavon-
20 Eternod, Pan, & Ribas De Pouplana, 2012). However, the detailed evolutionary history of diverse
21 bacterial tRNA modifications and their impact on key components of bacterial translation remain
22 unclear.

23 To address these gaps, we traced the evolutionary history of the five known bacterial tRNA
24 modifications at the first wobble base of the anticodon: (c)mm⁵(s²)U (5-
25 carboxymethylaminomethyl 2-thiouridine and its variants), cmo⁵U (uridine-5-oxyacetic acid), k²C
26 (2-lysyl-cytidine), Q (Queuosine) and I (Inosine) (Björk & Hagervall, 2014). We first determined
27 the occurrence of each modification pathway in 1093 sequenced bacteria that represent the diversity
28 of eubacteria, and then used ancestral reconstruction to infer the major evolutionary gain or loss
29 events in the bacterial phylogeny. To determine the causes and impacts of change in tRNA
30 modification, we analyzed two other key components of translation: tRNA gene pools and codon
31 bias, thought to coevolve under translational selection (Ikemura, 1985; Kanaya, Yamada, Kudo, &
32 Ikemura, 1999). Finally, since both of these factors are in turn correlated with genomic GC content
33 (Chen, Lee, Hottes, Shapiro, & McAdams, 2004; Hershberg & Petrov, 2009; Wald & Margalit,
34 2014), we also mapped genomic GC content across the bacterial phylogeny. Thus, we aimed to

1 generate a comprehensive model of the joint evolutionary history of important components of
2 bacterial translation – tRNA genes and modifying enzymes – and their relationship with genomic
3 GC content.

4 Each modified tRNA can recognize between 2 to 4 codons. Although there are no direct
5 measurements of the effect of tRNA modification on translation rate, the presence of a
6 multifunctional tRNA should amplify the effective tRNA pool, reducing ribosomal waiting times
7 for the correct tRNA and contributing to an increase in global translation speed. Modification also
8 prevents tRNA carrying amino acids coded by two synonymous codons from recognizing other
9 non-synonymous codons, potentially reducing mistranslation (Björk & Hagervall, 2014; Grosjean et
10 al., 2010). Thus, tRNA modifications may affect both translational speed and accuracy. Specifically,
11 we hypothesized that the presence of a tRNA modification should selectively favor the tRNAs that
12 it modifies (henceforth “target” RNA). Concomitantly, selection on tRNA that are not modified by
13 the ME (henceforth “non-target” tRNA; encoding the same amino acid) should be weakened, since
14 their function would be redundant with the modified target tRNA. As such, we predicted that the
15 evolutionary loss of non-target tRNA should be permissible only in the presence of a modification
16 that allows target tRNA to decode multiple sense codons. In contrast, the evolutionary loss of a
17 modification should be associated with a generally irreversible expansion of the tRNA repertoire to
18 include both target and non-target tRNA. Finally, we expected that these changes in tRNA gene
19 content should be correlated with changes in codon use as well as genomic GC content.

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

31

32 **RESULTS**

1 **The evolutionary history of bacterial tRNA modifications is highly dynamic**

2 We determined the occurrence of tRNA modifications across 1093 bacteria using homology
3 searches for complete protein sequences of previously described MEs (see Methods). We assume
4 that detected homologs are functional, and have the same modification function as that of the
5 previously described enzymes. Within each modification pathway, we focused on the following
6 enzymes that directly modify the tRNA and complete the modification (see Methods), and whose
7 absence should effectively prevent the modification. i) MnmE-MnmG for the (c)mnm5(s2)U
8 modification ii) CmoA-CmoB for the cmo5U modification iii) Tgt-QueA-QueG for the Q
9 modification iv) TadA for the I modification and v) TilS for the k2C modification. Note that in our
10 analysis, we did not consider G::U wobble pairing, which is not mediated by specific modification
11 enzymes. We observed that except for TilS, most MEs are not found in all bacteria (Fig. 1A),
12 suggesting that different modification pathways were gained or lost in specific lineages. The
13 essential genes RpoB and RplA (positive controls) were respectively present in 99.5% and 99.7% of
14 analyzed bacteria, indicating that our homology detection method resulted in very few false
15 negatives. When multiple enzymes are necessary for a modification, they tend to co-occur in
16 bacterial lineages (Fig. 1A). For instance, both MnmE and MnmG are absent only in Actinobacteria.
17 Similarly, both CmoA and CmoB are primarily observed in γ -proteobacteria and ϵ -proteobacteria.
18 However, this contradicts previous reports showing that the cmo5U modification is also active in
19 *Bacillus subtilis* and *Mycobacterium bovis* str. BCG (Chionh et al., 2016; Murao, Hasegawa, &
20 Ishikura, 1976; Yamada, Matsugi, Ishikura, & Murao, 2005). Despite an extensive search, we could
21 not reliably identify the reported enzymes in the sequences of any other organisms (see Methods for
22 details). Hence, here we focus on the proteobacterial cmo5U modification. In contrast to the co-
23 occurrence of CmoA and CmoB, for the Queosine modification, we found that QueG is often
24 missing even when Tgt and QueA are present. This relative evolutionary instability of QueG
25 corroborates a recent report suggesting that Tgt alone may be enough to generate the Q
26 modification (Zallot, Yuan, & de Crécy-Lagard, 2017).

27 To determine the key evolutionary changes in major tRNA modification pathways across bacteria,
28 we performed ancestral reconstruction with stochastic character mapping for each ME (Fig. 1A;
29 Table S1). We used a posterior probability value of ≥ 0.7 to assign the ME state at each internal
30 node in the bacterial phylogeny. Four of the five modification systems – MnmE-MnmG, Tgt-QueA-
31 QueG, TadA and TilS – were predicted to already occur at the root of the phylogenetic tree, i.e.
32 were present in the eubacterial ancestor (Fig. 1B). In contrast, CmoA-CmoB were gained/evolved
33 much later: once at the root of γ -proteobacteria, once in ϵ -proteobacteria, and once in some species
34 of δ -proteobacteria (Fig. 1B). Thus, except for the cmo5U modification, most bacterial tRNA

1 modifications are ancient. These ancient modifications were then independently lost multiple times,
2 with a total of 12 major losses (on branches with at least 5 descendant bacterial species in our
3 phylogeny; Fig. 1B) and 19 minor loss events (in lineages with fewer than five descendants; Fig.
4 S1). The only exception to this pattern of repeated ME gains/losses is the k2C modification (TilS),
5 which appears to be evolutionarily stable across bacteria. Overall, our analysis demonstrates a
6 highly dynamic evolutionary history of tRNA modifications that is dominated by repeated
7 secondary losses and only a few gains (a total of 44 loss events and 13 gains across the phylogeny;
8 Fig. S1).

9

10 **tRNA modification loss is associated with more non-target tRNAs**

11 The early evolution of nearly all tRNA modifications suggests that the eubacterial ancestor should
12 have been able to decode most sense codons using only a subset of all possible tRNAs. Indeed,
13 ancestral reconstruction of tRNA genes suggests that the eubacterial ancestor did have all target
14 tRNAs (except $\text{tRNA}_{\text{Leu}}^{\text{UUU}}$), but did not possess non-target tRNAs for the Q, k2C and I modifications
15 (Fig. 1C; note that the state of $\text{tRNA}_{\text{Arg}}^{\text{UCG}}$ is ambiguous). Interestingly, for the
16 (c)mnm5(s2)U modification, both target and non-target tRNA were present in the eubacterial
17 ancestor. Thus, a subsequent loss of any of these modifications should be strongly associated with
18 the gain (or retention, in case of (c)mnm5(s2)U) of the respective non-target tRNA. Both
19 predictions were supported when we observed that the proportion of species that have non-target
20 tRNA is significantly different between bacteria with vs. without a specific modification ($p < 0.05$,
21 2-sample proportions test; Fig. S2). Overall, these patterns suggest that the loss of tRNA
22 modifications was associated with the retention or secondary gain of non-target tRNAs, and the
23 gain of MEs was associated with the loss of non-target tRNAs.

24 To account for phylogenetic relationships between analysed bacteria, we compared the tRNA gene
25 content of sister clades with contrasting modification status; i.e. following a major modification
26 gain or loss event (see Methods; Fig. 2). For instance, since the (c)mnm5(s2)U modification was
27 lost at the root of Actinobacteria (MnmE/MnmG enzymes, Fig. 1B), we compared the tRNA gene
28 content of species within the Actinobacteria clade with that of species in the sister group: the
29 Cyanobacteria-Deinococcus-Thermus (CDT) clade. Similarly, we compared the target and non-
30 target tRNA gene copy numbers for descendent sister clades for 14 of the 15 major modification
31 gain and loss events highlighted in Fig. 1B (we did not analyze Thermotogae, because its sister
32 clade would consist of all other bacteria). Of these, 4 cases involving the Q modification are not
33 informative with regard to non-target tRNAs because these are absent in all bacteria (Fig. 2F–2I).

1 However, we describe these comparisons below since they are useful for understanding the
2 evolution of target tRNAs.

3 In most of the informative sister clade comparisons (8 out of 10 cases), species were more likely to
4 have non-target tRNAs if the modification was absent (compare bottom left vs. bottom right bar
5 plots within each panel of Fig. 2), supporting our hypothesis that modification loss (gain) is
6 associated with the presence (absence) of non-target tRNAs. For instance, in three out of the five
7 comparisons made for the (c)mm5(s2)U and cmo5U modifications, at least 5 of the 6 non-target
8 tRNAs were present in a significantly higher proportion of species without the modification (Fig.
9 2A, 2C and 2D; $p < 0.05$, 2-sample proportions test). The two exceptions to this pattern occur in
10 clades with relatively low sample sizes (Fig. 2B and 2E), and may potentially reflect low statistical
11 power. Alternatively, these may be biologically interesting exceptions that require further
12 examination. Additionally, in all five comparisons made for the I modification, at least one non-
13 target tRNA (in most cases both) was present in significantly more species without the
14 modification, compared to species with the modification (Fig. 2J – 2N; $p < 0.05$, 2-sample
15 proportions test).

16 In contrast to the wide variation in the occurrence of non-target tRNAs, we found that target tRNA
17 genes were largely stable irrespective of species' modification status (top row in each panel of Fig.
18 2). Contrary to expectation, the gain or loss of the (c)mm5(s2)U and cmo5U modifications does
19 not appear to have any impact on target tRNAs (Fig. 2A–2E; $p > 0.05$, 2-sample proportions test);
20 likely because the corresponding codons are still used in these species and cannot be decoded by
21 non-target tRNAs. In the case of the Q modification, target tRNAs were always present; but in 2 of
22 4 comparisons their copy number was lower in clades that had lost the ME (Fig. 2F–2I; $p < 0.05$, 2-
23 sample proportions test). Inosine presents the only evidence supporting weakened selection on
24 target tRNAs upon ME loss. In 3 of 5 comparisons, target tRNAs were absent in nearly all species
25 without the modification (Fig. 2J–2L). In the remaining 2 comparisons, target tRNAs were present
26 in fewer species without modifications (Fig. 2M and 2N; $p < 0.05$, 2-sample proportions test).
27 Together, these results suggest that except for the I modification, the evolution of modifications had
28 no impact on the evolution of target tRNA genes.

29

30 **Modification status is associated with tRNA diversity, GC content and genome size**

31 To identify the cause and impact of evolutionary changes in tRNA modifications, we compared key
32 translation-associated genomic features of sister clades with contrasting modification status. As

1 shown above, we found that in 11 of 14 comparisons, the tRNA diversity of sister clades was
2 significantly different ($p < 0.05$, Wilcoxon rank sum test; Fig. 3 Fig. S3), indicating a strong
3 association between the evolution of modifications and the tRNA gene pool. However, in four of
4 these cases the direction of association opposed our initial expectation, and clades without the
5 modification had lower tRNA diversity (Figs. 3G, 3I, 3M and 3N). In three of these cases involving
6 Tenericutes and the endosymbiotic species from Enterobacterales (Figs. 3G, 3I and 3M), we also
7 observed a significant reduction in the total number of tRNA genes (Fig. S3G, S3I, S3M),
8 suggesting wide-ranging gene loss that is potentially a result of genome reduction in these clades.
9 Indeed, previous reports have shown that reductive genome evolution in Mollicutes has led to the
10 loss of several tRNA modifying enzymes (Grosjean et al., 2014; Yokobori, Kitamura, Grosjean, &
11 Bessho, 2013). Thus, overall half the sister clade comparisons supported our initial prediction that
12 the lack of modifications should be associated with greater tRNA diversity. Next, we tested whether
13 changes in modification status and tRNA genes were also associated with changes in GC content
14 and codon use.

15 In most cases showing altered tRNA diversity (8 out of 11), we found that the change in diversity
16 was also associated with significant shifts in GC content, so that species in clades without the tRNA
17 modification typically had more extreme GC3 values (i.e. GC content in the third base of codons,
18 Fig. 3; $p < 0.05$, Wilcoxon rank sum test; we observed similar results with genomic GC content,
19 Fig. S3). In seven of these eight cases, species without the modification also had significantly more
20 biased codon use (measured as relative synonymous codon use RSCU for all proteins; $p < 0.05$,
21 Wilcoxon rank sum test; we observed similar results for RSCU of highly expressed ribosomal
22 proteins; Fig. S3). Finally, in five of the eight cases we observed significant differences in genome
23 size across sister clades (Fig. 3). Of these, in the two comparisons that involved Actinobacteria
24 (Figs. 3A and 3F) we observed that larger genomes were associated with the absence of the
25 modification. As mentioned above, in the other three cases involving Tenericutes and
26 Enterobacterales (Figs. 3G, 3I and 3M), the lack of modification may reflect overall genome
27 reduction. Note that in every comparison where codon use and genome size were different, GC
28 content was also skewed significantly. Thus, to test the overall effect of genome size and GC
29 content on modification status, we carried out a phylogenetic regression. We found that both GC
30 content and genome size had a significant effect on the presence/absence of the Q modification
31 alone (Fig S4). These results suggest that except for the Q modification, modifications were lost in
32 only a few instances of shifts in GC content. Thus, GC content may be the major ultimate driver of
33 changes in tRNA gene content and the subsequent loss of modification may generate selection to fix
34 these changes. Altogether, three of the 14 major changes in modification status were not associated
35 with significant shifts in either tRNA or GC content, and remain unexplained. These included one

1 instance of the loss of the (c)mm5(s2)U modification (Fig. 3B) and two losses of the I
2 modification (Figs. 3K–L).

3 Finally, to quantify the evolutionary impact of tRNA modifications on tRNA gene pools, we
4 simultaneously mapped changes in tRNA genes copies, MEs, and GC content on the bacterial
5 phylogeny. As expected, following major GC shifts, non-target tRNA showed several gain/loss
6 events whereas target tRNA were more stable (Figs. 4–6). As suggested by our sister clade
7 comparison, in some cases modification gain or loss was also preceded by a GC shift (8 out of 14
8 independent lineages). Moreover, a phylogenetic regression to test the impact of GC and MEs on
9 tRNA genes revealed that GC content had the strongest impact on the dynamics of non-target tRNA
10 genes (Fig. 7). However, as observed in the sister clade comparison, the phylogenetic regression
11 and mapping also reveal clear instances of GC-independent impacts of modification status on tRNA
12 genes (Figs. 4–7). The strongest impact of tRNA modification is observed the case of the Inosine,
13 where the target tRNA was present in all species when the modification was present (Fig. 6); and
14 the non-target tRNA were absent in most of these species (Fig. 2D). On the other hand, the loss of
15 target tRNAs and presence of non-target tRNAs is strongly associated with the loss of the Inosine
16 modification (Fig. 2D, Fig. 6, Fig. 7). Thus, the bacterial tRNA repertoire is strongly affected by
17 evolutionary changes in GC content as well as tRNA modifying enzymes.

18

19 **DISCUSSION**

20 Over 50 years ago, Francis Crick described the phenomenon of wobble base pairing where one
21 tRNA molecule can recognize multiple codons (Crick, 1966). Since then, a large body of work has
22 developed a detailed understanding of the enzymatic pathways that generate extended wobble
23 pairing via base modifications of tRNA anticodons (reviewed in (Björk & Hagervall, 2014)). Given
24 their critical role in translation, modifying enzymes are expected to evolve under strong selection,
25 and in turn influence the evolution of other components of the translation machinery (Grosjean et
26 al., 2010). Here, we trace the evolutionary history of tRNA modifications across the bacterial
27 phylogeny, showing that the eubacterial ancestor already had nearly all known modification
28 pathways. Several major bacterial clades subsequently lost the modifications, often accompanied by
29 increasing tRNA diversity and shifts in genome GC content and codon use. These results are
30 consistent with the hypothesis that an expanded tRNA pool gradually weakens selection on MEs,
31 allowing their loss through drift. However, once the modification is lost, bacteria likely face strong
32 selection to retain the expanded tRNA pool. In contrast, if GC shifts shrink the tRNA gene pool,
33 selection would strongly favor the innovation or retention of tRNA modifications. We speculate that

1 in Proteobacteria, such a process may have led to the evolution of the novel cmo5U modification.
2 Alternatively, Proteobacteria may have evolved the new modification by chance, secondarily
3 weakening selection on the non-target tRNAs. Interestingly, we also uncovered instances where
4 changes in tRNA or GC content could not explain the evolution of tRNA modifying enzymes. In
5 some of these cases, we suspect that overall genome reduction may have led to the loss of the
6 modification; but three major cases of modification loss remain unexplained (loss of (c)mnm5(s2)U
7 in *Leuconostoc* and I in *Spirochaetes* and ϵ -proteobacteria). Nonetheless, our results strongly
8 support the conclusion that ME evolution was dominated by frequent losses driven by weakened
9 selection. Note that reversing the direction of causality in our arguments would not make logical
10 sense: the loss of non-target tRNAs cannot generate selection favoring the loss of tRNA
11 modifications, and there is no reason to expect that the gain of a modification enzyme should
12 selectively favor the retention on non-target tRNAs. Similarly, it is difficult to imagine how
13 evolutionary gains or losses of tRNA modifications could precipitate genome-scale changes in GC
14 content or genome size. Thus, our systematic phylogenetic analysis clarifies the contrasting roles of
15 strong vs. weak selection acting on tRNA modifications, and on the evolutionary history of key
16 components of bacterial translation.

17 Our results were generally consistent with the prediction that a lack of specific tRNA modifications
18 should result in strong positive selection favoring the relevant non-target tRNAs. However, we also
19 observed some instructive exceptions to this pattern. For instance, comparing across sister clades,
20 we found that ~33% of the species that do not have the cmo5U modification also lack at least one
21 non-target tRNA. This should pose a problem because codons recognized by the missing non-target
22 tRNAs cannot be decoded efficiently. However, we observed that these codons were significantly
23 depleted in these genomes, compared to species that retained the non-target tRNA (Fig. S5). Thus,
24 these relatively rare codons may not pose a major problem even if they are inefficiently translated.
25 It is also possible that these codons are decoded by the respective U-starting target tRNAs that can
26 pair with A, G or U ending codons (Grosjean et al., 2010). This scenario is especially likely in
27 species where the Q modification is absent. Recall that all bacteria lack the non-target tRNA for this
28 modification; hence species without the Q modification should be unable to translate many codons.
29 We hypothesize that in species where both the Q modification as well as the non-target tRNA are
30 missing, there is an as yet undiscovered G::U wobble base pairing between the target tRNA and
31 codons complementary to the non-target tRNA in these species.

32 Our analysis quantifies the role of specific tRNAs in the evolutionary dynamics of bacterial
33 translation, both as drivers of selection on tRNA modifications and as a means to rapidly respond to
34 shifts in genomic GC content. Specifically, we found that target tRNAs are evolutionarily largely

1 stable (except for those modified by Inosine), whereas non-target tRNAs are often gained or lost in
2 lineages following a GC shift. The strong phylogenetic correlation between non-target tRNAs and
3 genomic GC content echoes the previously reported association between GC content and
4 “auxiliary” tRNAs (tRNAs with a high number of gains and losses) defined by Wald and Margalit
5 (Wald & Margalit, 2014). The nearly ubiquitous presence of target tRNAs may reflect strong
6 positive selection, since the function of target tRNA cannot be carried out by any other tRNA via
7 wobble base pairing rules. Thus, target tRNA may be essential even if their complementary codons
8 are used relatively rarely.

9 In summary, we propose that the evolution of diverse bacterial tRNA modification enzymes has
10 been strongly influenced by changes in tRNA gene pools and genomic GC content, as well as by
11 genome reduction in specific lineages. Our results support the hypothesis that these changes
12 weakened selection on MEs and allowed their loss through drift. Conversely, we hypothesize that
13 the loss of non-target tRNA may have generated strong selection for the innovation of the cmo5U
14 modification in γ -proteobacteria and ϵ -proteobacteria. It is also possible that the tRNA loss
15 succeeded only in the wake of the chance innovation of the cmo5U modification. These patterns
16 generate a number of broad, testable predictions about the future evolutionary trajectory of
17 translational components in various bacterial lineages. For instance, MEs may be lost in GC-rich
18 bacteria without major consequences for translation; such losses should evolutionarily fix the
19 expanded tRNA repertoire; and MEs should face strong positive selection in AT-rich clades with
20 low tRNA diversity. Specifically, in AT rich bacteria, the combination of MEs and largely target
21 tRNAs should confer higher fitness than a full tRNA set without MEs. Our analysis also shows that
22 multiple key components of bacterial translation have evolved in close association with genomic
23 GC content. Although the evolutionary and ecological forces driving GC shifts and GC-independent
24 changes in tRNA diversity remain unclear, our analyses highlight the rich evolutionary dynamics of
25 wobble pairing, and more broadly, bacterial translation.

26

27 **METHODS**

28 **Phylogenetic tree, genomes and tRNA gene data**

29 All analyses were carried out in R (R Development Core Team, 2015) using the phytools (Revell,
30 2012) and seqinr (Charif, Thioulouse, Lobry, & Perrière, 2005) packages, unless otherwise
31 specified. The R scripts for all analyses are available at
32 https://github.com/gauravdiwan89/me_evolution_project. We used the bacterial phylogeny

1 generated by Segata and co-workers (Segata, Börnigen, Morgan, & Huttenhower, 2013). This
2 phylogeny was generated using an alignment of the 400 most conserved protein sequences across
3 all prokaryotes. We downloaded the phylogeny
4 (https://bitbucket.org/nsegata/phylophlan/wiki/bs_tree.reroot.nwk) and removed all archaea, as well
5 as bacteria whose genomes have not been fully sequenced. We pruned the tree further using a
6 distance based pruning function. Briefly, when closely related species had a pairwise phylogenetic
7 distance ≤ 0.03 , we randomly sampled only one of these species. This allowed us to remove the
8 excessive representation of some highly sequenced genomes, such as several *Escherichia coli*
9 strains. This led to the final set of 1093 bacterial genomes that we used in our subsequent analyses.

10 We downloaded the tRNA gene copy numbers for each of these genomes from the Genomic tRNA
11 database (Chan & Lowe, 2009). For genomes where this information was not available, we
12 determined the tRNA gene copy numbers by running the program tRNAscan-SE (Lowe & Eddy,
13 1996) using the default parameters for bacterial genomes. The information for each of these
14 genomes was represented using different identifiers in the various databases that we used. This
15 information is consolidated in Table S1, consisting of: i) species name according to NCBI
16 taxonomy, ii) NCBI accession number, iii) Phylogenetic tree tip label in the tree, iv) IMG database
17 identifier and v) Genomic tRNA database abbreviation.

18 **Choice of crucial tRNA MEs and non-target tRNA**

19 The (c)mm5(s2)U and Q modifications are created by elaborate ME pathways. For instance, our
20 definition of the (c)mm5(s2)U modification includes three different modifications: mnm5U,
21 cmnm5U and cmnm5s2U, each of them acting on different isoaccepting tRNA. These modifications
22 have unique MEs that are involved in the last steps of the pathway. MnmE and MnmG are the
23 common enzymes for each of these modifications and they carry out the first step of the modifying
24 the Uridine base (Björk & Hagervall, 2014) and references therein. Thus, we considered MnmE-
25 MnmG to be crucial, and other enzymes such as MnmC, MnmA, MnmH and trmL as auxiliary. The
26 Q base is first biochemically synthesized as Q₁ in the bacterial cell using the enzymes FolE, QueC,
27 QueD, QueE and QueF. Subsequently, the Tgt enzyme replaces Guanosine in the anticodon stem
28 loop with Q₁ and QueA and QueG convert it to Q ((Björk & Hagervall, 2014) and references
29 within). A recent report has suggested the possibility of Tgt alone carrying out the Q modification in
30 the absence of QueA and QueG, assuming the existence of a Q salvage pathway (Zallot et al.,
31 2017). However, there is no evidence for this phenomenon across most bacteria from our analysis;
32 hence, we considered Tgt, QueA and QueG as the crucial enzymes for the Q modification.

1 The cmo5U modification exists in tRNA molecules for four box codons (Leucine, Valine, Serine,
2 Proline, Threonine and Alanine) and therefore there were three possible non-target tRNA in each
3 case. However, tRNA with GNN anticodons (where N is any nucleotide) in each of these codon
4 boxes can also recognize NNU codons via G::U wobble pairing. Thus, we only focused on the non-
5 target tRNA that have CNN anticodons (indicated in Fig. 1C).

6 **Homology search for tRNA MEs**

7 We determined the presence/absence of all known tRNA MEs in the 1093 bacterial genomes using
8 Hidden Markov Model (HMM) searches (S. Eddy, 1998). First, we downloaded all available
9 reviewed protein sequences for each of the MEs from the UniProt-KB database. We created a
10 multiple sequence alignment using the t-coffee program (Notredame, Higgins, & Heringa, 2000)
11 with default parameters. We used this alignment to build an HMM profile, and then used the
12 “hmmsearch” command in the HMMER suite (S. R. Eddy, 2009) to detect the presence/absence of
13 each ME. We used the amino acid sequences of all proteins from the 1093 genomes (downloaded in
14 November 2016 from the NCBI ftp site) as the sequence search space, and set all other parameters
15 to default. Each search generated a table with accession numbers of potential homologous hits with
16 a corresponding e-value and bit-score. There is no consensus in the field with respect to setting an
17 e-value cut-off to detect true homologs. Here, we present a slightly informative method of deciding
18 e-value cut-offs for detecting true homologs. To infer presence/absence with precision, we
19 implemented a dynamic e-value cut-off to eliminate spurious hits. We did this by plotting the
20 number of species in which the ME would be detected as we increased the e-value cut-off (Fig. S6).
21 We observed stable plateaus where the number of species in which an ME was detected did not
22 change despite changing the e-value cut-off, suggesting that these represent robust e-value
23 thresholds. Thus, for each enzyme we set an enzyme-specific e-value cut-off as the first e-value at
24 the beginning of the most stable plateau (see Fig. S6). For each ME, homologous hits that had e-
25 values lower than this cut-off were considered true homologues.

26 Previous reports have experimentally shown that the cmo5U modification occurs in *Bacillus subtilis*
27 and *Mycobacterium bovis* str. BCG (Chionh et al., 2016; Murao et al., 1976; Yamada et al., 2005).
28 However, in our initial analysis we did not find CmoA and CmoB homologs in these or related
29 species. A closer inspection of the *Bacillus subtilis* genome also did not reveal these enzymes, as
30 indicated in a previous report (Grosjean et al., 2014). Instead, there were a few entries for these
31 enzymes in the UniProt database (each with a poor annotation score). We therefore tested whether
32 these enzymes are true homologs of the proteobacterial CmoA and CmoB, using homology
33 detection (using ‘jackhmmmer’). We observed that only the query organisms showed the presence of

1 these enzymes (data not shown). A survey of the SEED database (Overbeek et al., 2005) also
2 revealed the absence of CmoA and CmoB in all but one *Bacillus* species. When we carried out a
3 blastp analysis using CmoA and CmoB enzymes from *Escherichia coli* in *Mycobacterium bovis* str.
4 BCG, we observed no significant hits for CmoA and no hits for CmoB with e-values $< 5e-3$. Thus,
5 although the modification may exist in several species, we could not reliably identify homologs of
6 the proteobacterial CmoA and CmoB in other bacteria.

7 **Ancestral reconstruction of tRNA modifications, genomic GC content, and tRNA genes**

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

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

27 We determined the evolutionary history of tRNA gains and losses using a similar method as
28 described for MEs. However, in this case we carried out the ancestral reconstruction separately for
29 each bacterial family, which allowed us to account for family-specific rates of evolution of tRNA
30 genes. We used the ancestral state of each tRNA at the root of each bacterial family to infer the
31 tRNA gene content of the eubacterial ancestor. Briefly, we converted tRNA gene copy numbers for
32 all species into a binary vector by changing the state of tRNA genes with multiple copies to “1” i.e.

1 presence. We then carried out ancestral reconstruction using stochastic character mapping and noted
2 evolutionary transitions on the phylogenetic tree as before.

3 **Identifying sister clades with contrasting modifying enzyme status**

4 We identified clades where crucial enzymes for each tRNA modification were gained or lost, and
5 determined the closest sister clade where the state of modification was opposite to that of the focal
6 clade. We determined the sister clade agnostic of its taxonomic classification. We checked if there
7 were at least 5 descendant extant species in this clade for each gain or loss event. If this criterion
8 was not met, we picked the next closest sister clade neighboring the common ancestor of the above
9 two clades. Within each clade, we excluded species that had an opposite state for any of the MEs
10 involved in each modification. We assigned the identity of the comparison group using the lowest
11 level of taxonomic classification that encompassed the focal species.

12 **Phylogenetic regression**

13 We carried out phylogenetic regression to test the impact of GC content and modifications on tRNA
14 presence/absence, using the Maximum Likelihood Continuous Regression program in BayesTraits,
15 similar to the method used by Wald and Margalit (Wald & Margalit, 2014). We estimated the
16 lambda parameter for each case and used default values for all other parameters. For each tRNA, we
17 first tested the full model: tRNA presence/absence ~ GC content*tRNA modification status. To
18 determine the impact of each explanatory variable, we compared the likelihood of the full
19 regression model to the likelihood of the same regression when the R value was set to zero. For
20 reduced models containing only GC content or modification status as the explanatory variable, we
21 compared the likelihood of the regression of the reduced model with the likelihood of the regression
22 of the full model. When these likelihoods were comparable ($p > 0.05$; Likelihood Ratio test), we
23 inferred that the reduced model was sufficient to explain tRNA presence/absence, and hence had a
24 significant impact on the evolution of the tRNA. We used the same method for testing the impact of
25 GC content and genome size on modification status (Fig. S4).

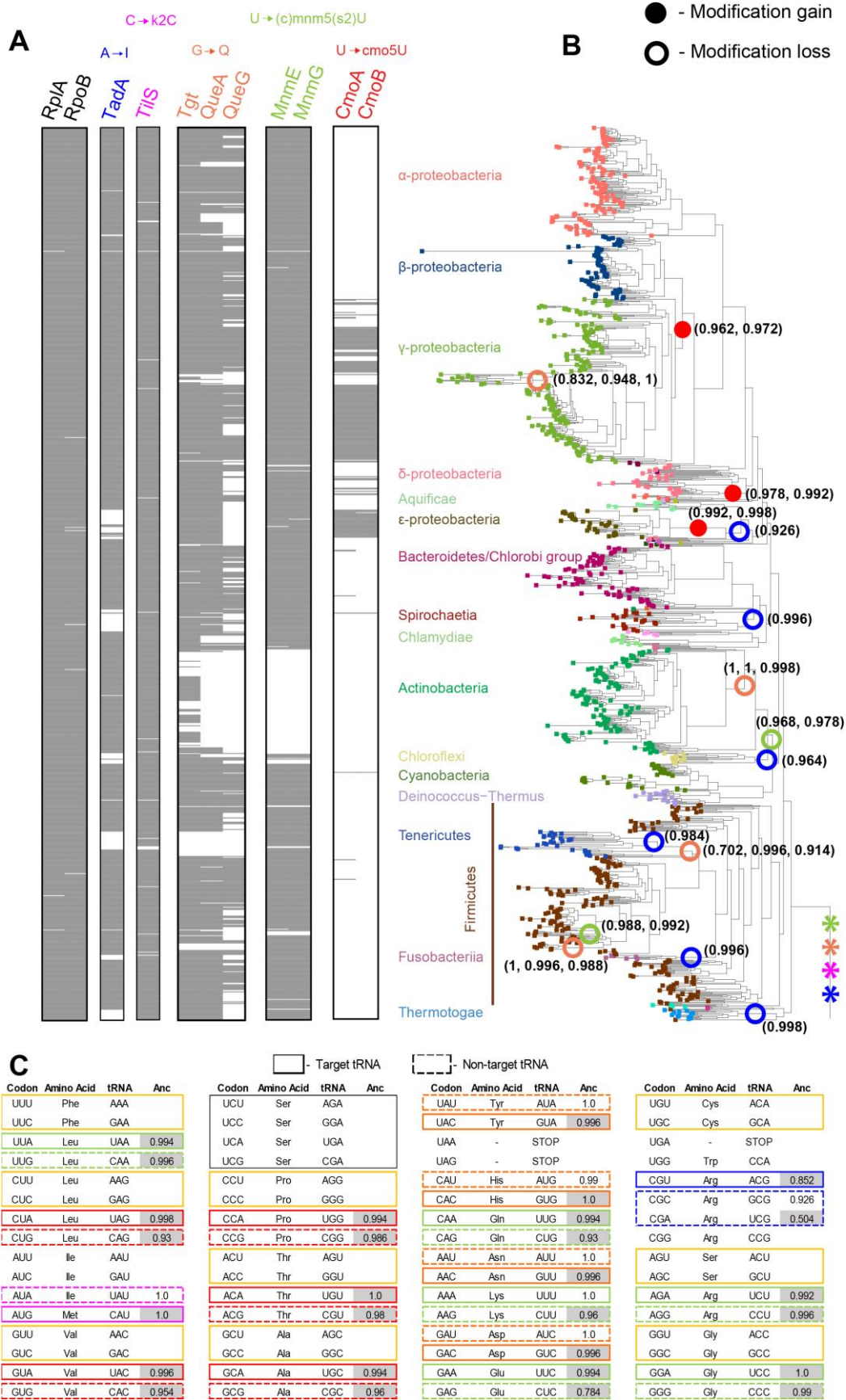
26

27 **ACKNOWLEDGEMENTS**

28 We thank Saurabh Mahajan for help with analyses; and Aparna Agarwal, Shyam Buddh, Vrinda
29 Ravi Kumar, Saurabh Mahajan and Laasya Samhita for constructive comments on the manuscript.
30 This work was funded by the National Centre for Biological Sciences (NCBS-TIFR) and grants
31 from the Department of Science and Technology, India (INSPIRE Faculty award, IFA-1 LSBM-64)

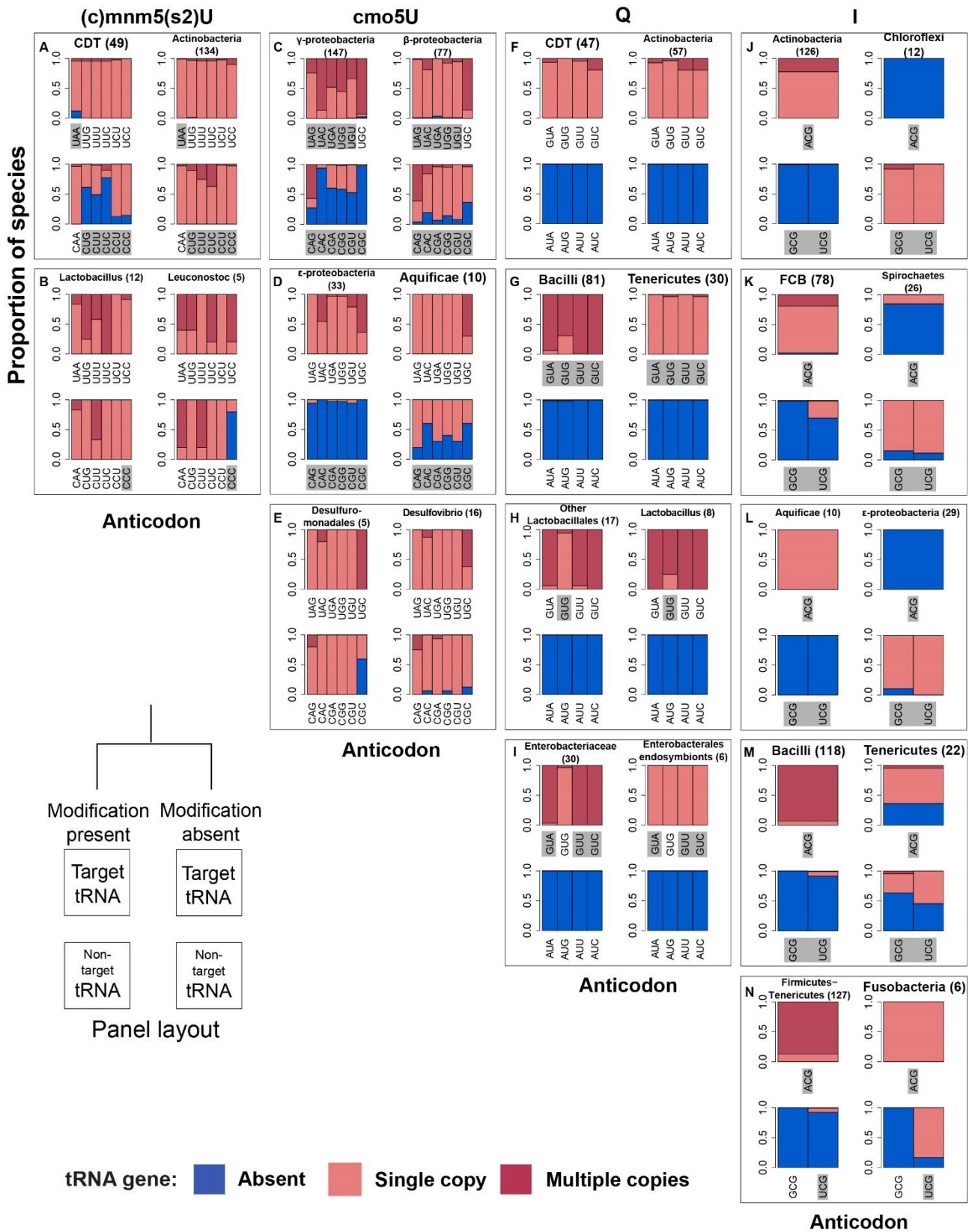
- 1 and the Council for Scientific and Industrial Research, India (37(1629)/14/EMR-II). GDD
- 2 acknowledges a Senior Research Fellowship from the University Grants Commission, India.

1 FIGURES



2

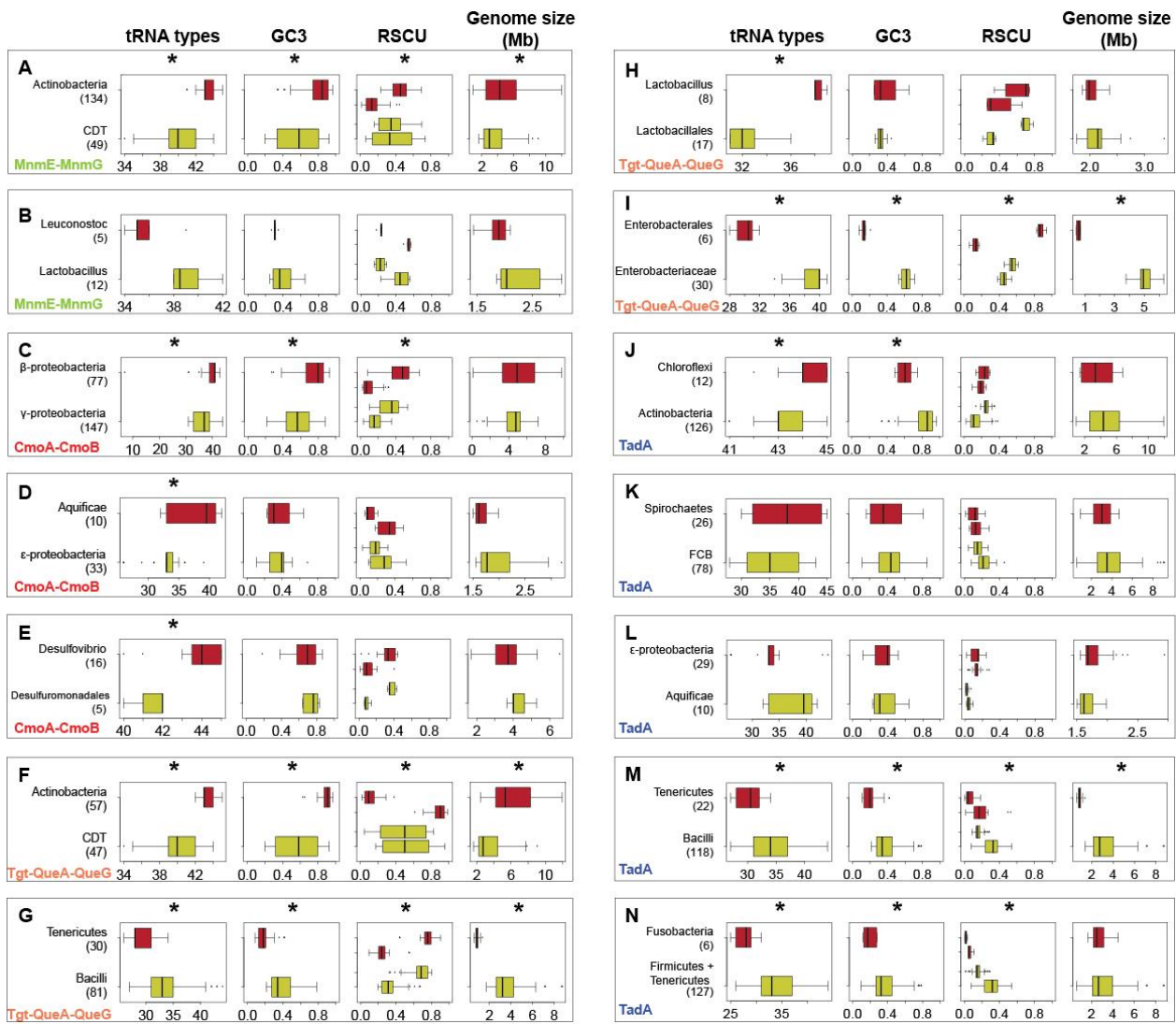
1 **Fig. 1: The evolutionary history of bacterial tRNA modifications.** (A) Columns indicate the
2 presence (gray) or absence (white) of various tRNA modifying enzymes as noted above each
3 column. The leftmost column shows two essential housekeeping proteins (RpoB- RNA polymerase
4 beta subunit and RplA- 50S Ribosomal protein L1) that serve as positive controls for homology
5 detection. (B) Bacterial phylogenetic tree (Segata et al., 2013) pruned to show the evolutionary
6 relationships between the 1093 bacterial species used in our study. Branch tips are colored by
7 taxonomic group. Filled and open circles respectively mark the major gain and loss of tRNA
8 modifications (with at least 5 descendent species in the relevant lineage), as inferred by ancestral
9 reconstruction (see Methods). Circles are colored according to the modification, as indicated in
10 panel A. Values in parentheses indicate the posterior probability of each gain or loss event for each
11 ME involved in the modification pathway. All other minor gain and loss events (with < 5
12 descendent species in the pruned phylogeny) are shown in Fig. S1. Asterisks at the root of the tree
13 indicate the presence of the respective modification. (C) Genetic code table showing the target
14 (solid boxes) and non-target tRNA (dashed boxes) for each modification, colored as in panel A.
15 tRNA with a yellow border show G::U wobble base pairing and were not considered in subsequent
16 analyses. Serine tRNA are highlighted with a black box to indicate the ability of tRNA_{Ser}^{GGA} to decode
17 all Serine codons. The column titled “Anc” indicates the presence (grey cells) or absence (white
18 cells) of each tRNA in the eubacterial ancestor, inferred from ancestral reconstruction (see
19 Methods). Value in each cell indicates the posterior probability of the inferred state for each tRNA.



1

2 **Fig. 2: tRNA gene content associated with modification gain or loss in sister clades.** Gene copy
 3 number of target and non-target tRNA within sister clades with contrasting modification status
 4 (marked in Fig. 1B). Stacked bar plots show the proportion of species within each sister clade that
 5 lack a specific tRNA (blue bars), have a single copy of the gene (light red bars), or have multiple

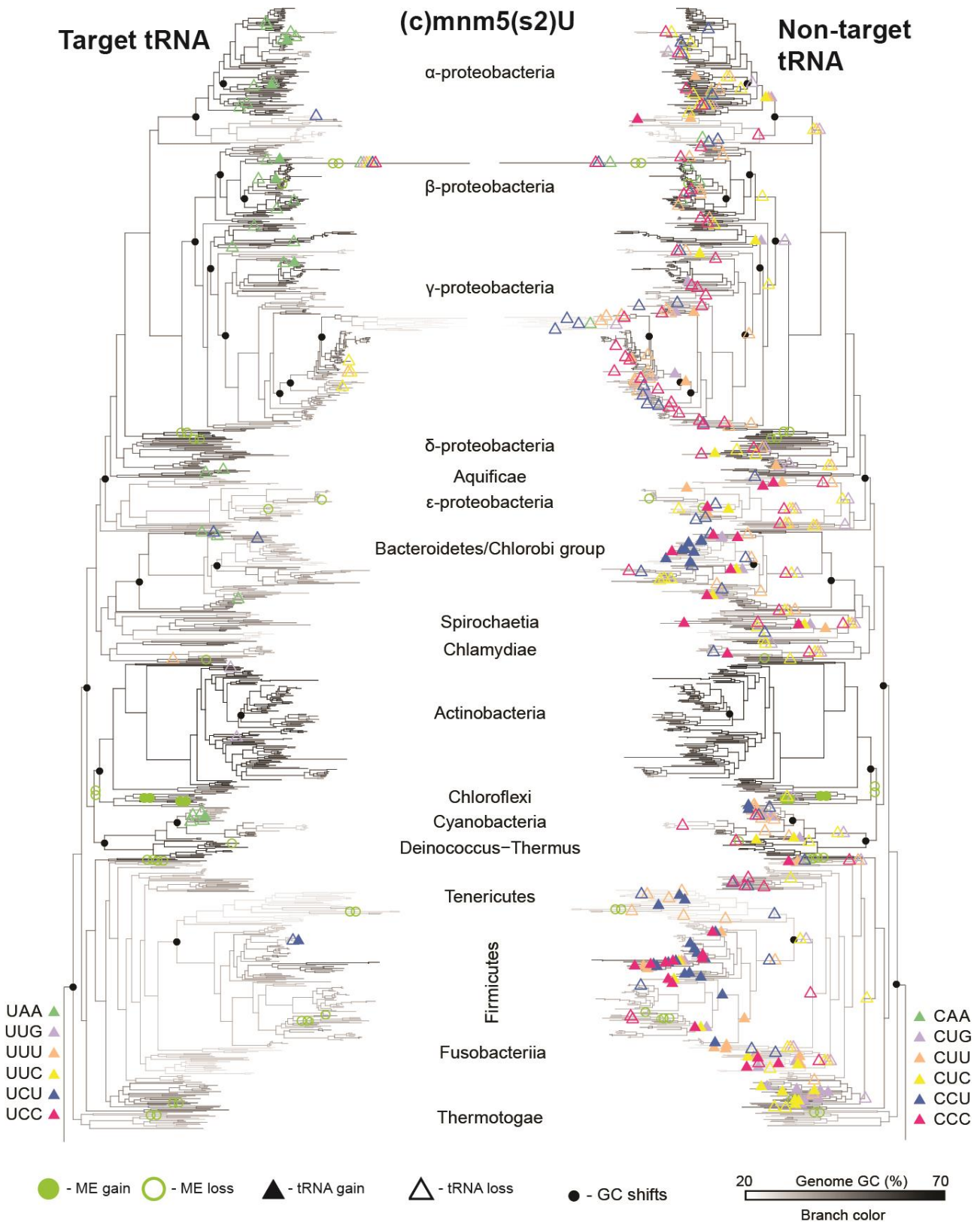
1 copies (dark red bars). The total number of species analyzed in each clade are given in parentheses
2 above the bar plots. Each panel shows data for a specific sister clade comparison; panels are
3 grouped by modification pathway. Gray shading represents anticodons (i.e. tRNA) for which the
4 proportion of species without the tRNA is significantly different across groups, with respect to
5 modification presence vs. absence ($p < 0.05$; 2-sample test for equality of proportions with
6 continuity correction). CDT – Cyanobacteria-Deinococcus-Thermus, FCB – Fibrobacteres-
7 Chlorobi-Bacteroidetes.



1

2 **Fig. 3: Differences in key genomic features across sister clades with contrasting modification**
 3 **status.** Box plots show the range of values of three genomic traits, across all species within a sister
 4 clade (see Fig. 3; red boxes = modification absent; green boxes = modification present). For the
 5 Relative Synonymous Codon Usage (RSCU) panel, asterisks indicate that the difference between
 6 the distributions represented by green boxplots is significantly different than the difference between
 7 the distributions represented by the red boxplots (i.e., codon usage is more biased in one case; $p <$
 8 0.05 ; Wilcoxon rank sum test with p -values correction for multiple comparisons). For tRNA types,
 9 GC3 (GC content of the third codon base) and genome size, asterisks indicate a significant
 10 difference in distributions across the sister clades ($p < 0.05$; Wilcoxon rank sum test with p -values
 11 correction for multiple comparisons).

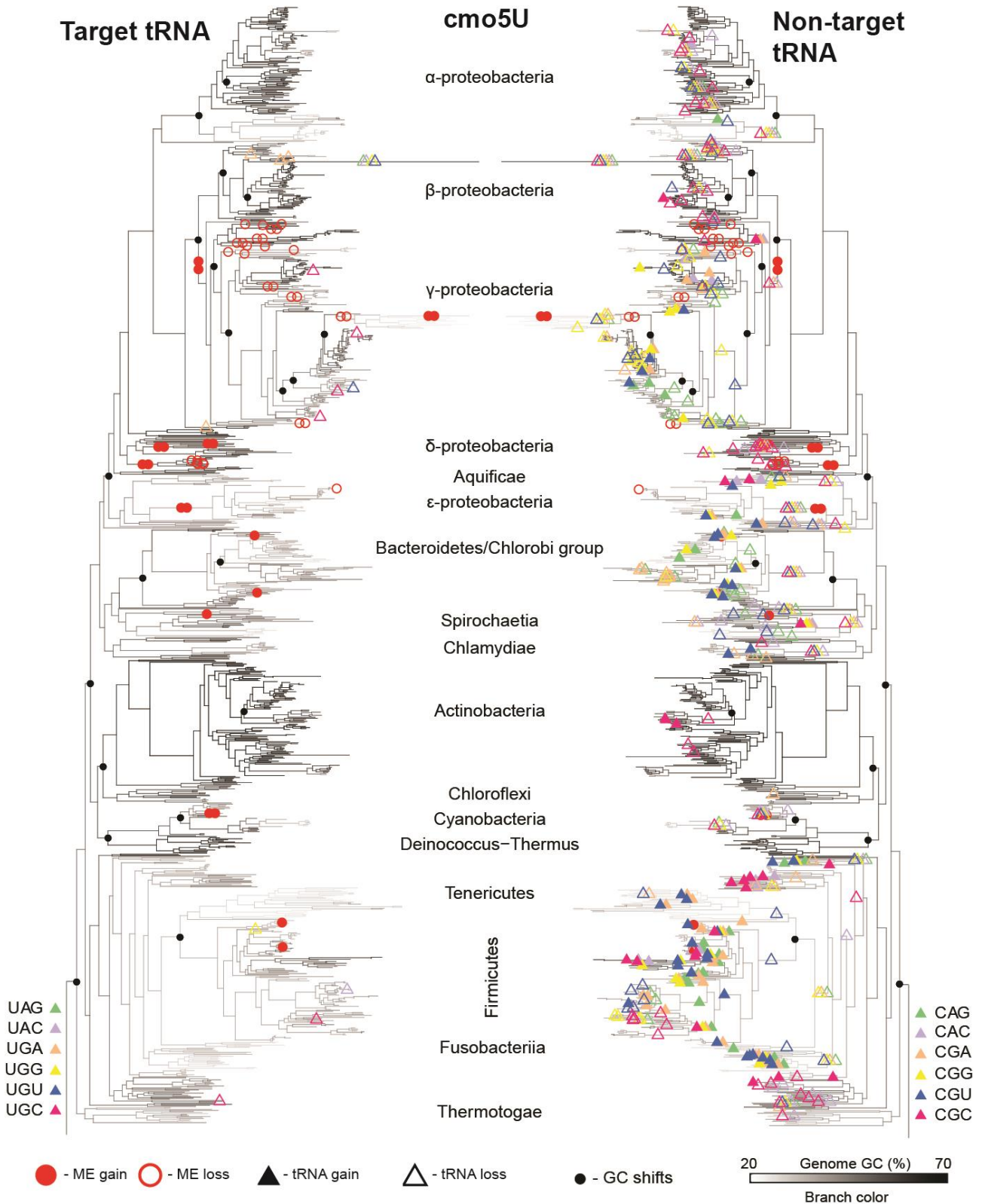
12



1
2 **Fig. 4: Evolutionary dynamics of tRNA genes, MEs and GC content for the (c)mnm5(s2)U**
3 **modification.** Branches are colored according to the inferred GC content for each branch, with
4 white indicating low GC% and black indicating high GC%. Nodes that showed the largest GC shifts
5 are indicated with black circles. Colored open and closed circles respectively indicate the loss and
6 gain of each ME involved in the modification. Open and closed triangles indicate the loss and gain

- 1 of tRNA genes relevant to the (c)mm5(s2)U modification, with target tRNA indicated on the tree
- 2 on the left and non-target tRNA indicated on the mirror tree on the right.
- 3

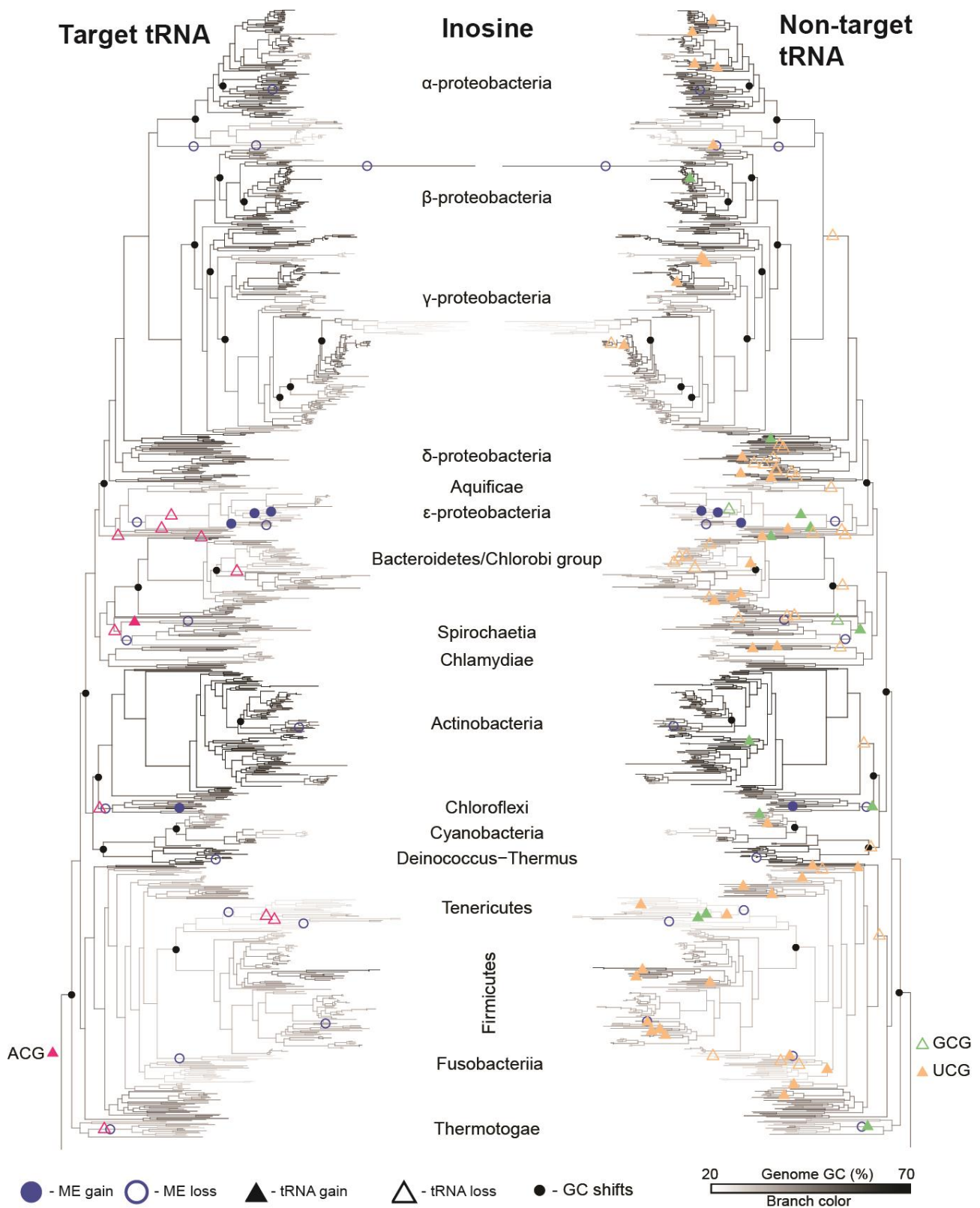
1



2

3 **Fig. 5: Evolutionary dynamics of tRNA genes, MEs and GC content for the cmo5U**
 4 **modification:** Branches are colored according to the inferred GC content for each branch, with
 5 white indicating low GC% and black indicating high GC%. Nodes that showed the largest GC shifts
 6 are indicated with black circles. Colored open and closed circles respectively indicate the loss and

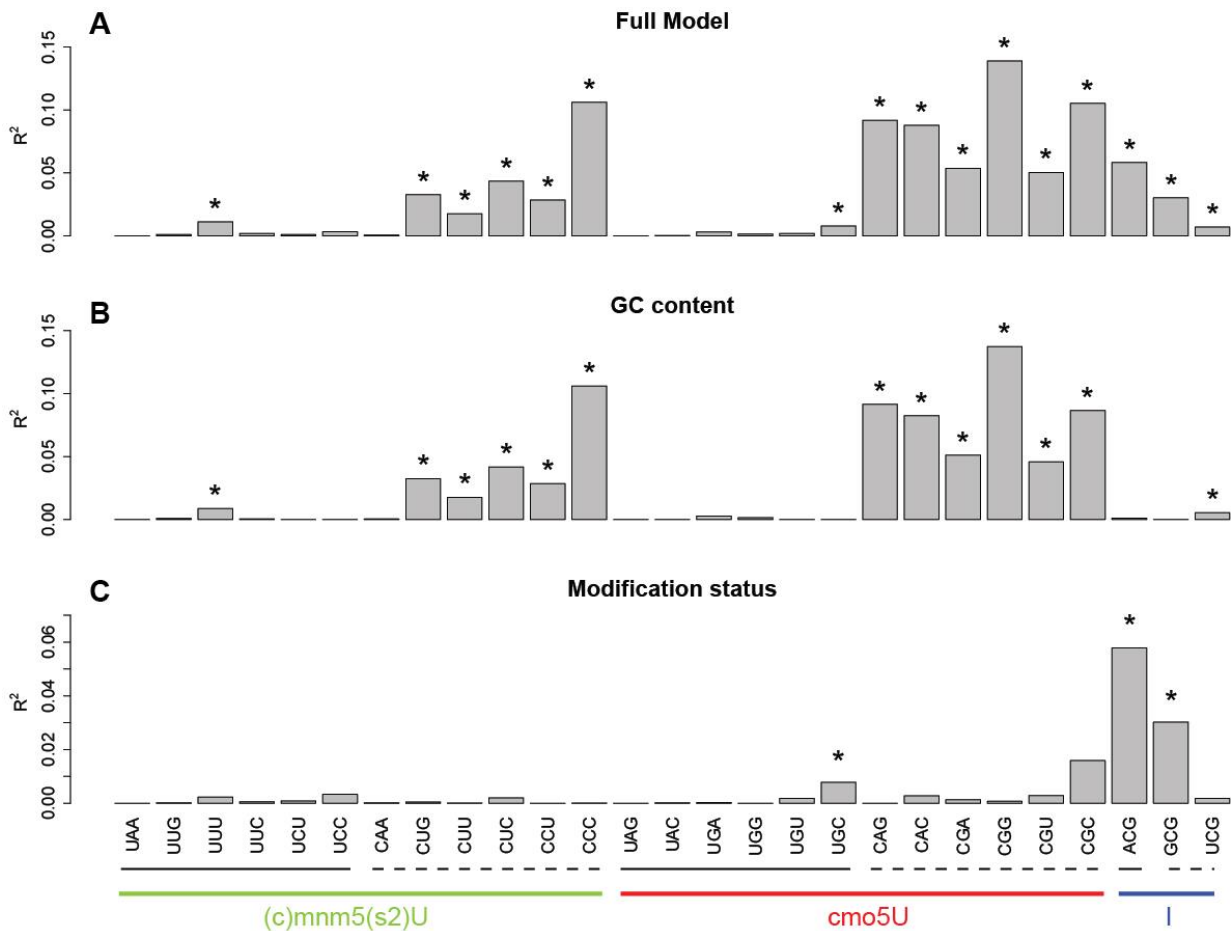
- 1 gain of each ME involved in the modification. Open and closed triangles indicate the loss and gain
- 2 of tRNA genes relevant to the cmo5U modification, with target tRNA indicated on the tree on the
- 3 left and non-target tRNA indicated on the mirror tree on the right.
- 4
- 5



1
2 **Fig. 6: Evolutionary dynamics of tRNA genes, MEs and GC content for the I modification:**
3 Branches are colored according to the inferred GC content for each branch, with white indicating
4 low GC% and black indicating high GC%. Nodes that showed the largest GC shifts are indicated
5 with black circles. Colored open and closed circles respectively indicate the loss and gain of each
6 ME involved in the modification. Open and closed triangles indicate the loss and gain of tRNA

- 1 genes relevant to the I modification, with target tRNA indicated on the tree on the left and non-
- 2 target tRNA indicated on the mirror tree on the right.
- 3

1



2

3 **Fig. 7: Results of phylogenetic regressions for factors affecting tRNA gene presence or**
 4 **absence.** Bar plots show the R² value for the phylogenetic regression for presence/absence of target
 5 and non-target tRNAs for the (c)mnm5(s2)U, cmo5U and I modifications. Results are shown for
 6 three different models: (A) Full model, tRNA presence/absence ~ Modification status (i.e.
 7 presence/absence) * GC content. (B) tRNA presence/absence ~ GC content (C) tRNA
 8 presence/absence ~ Modification status (i.e. presence/absence). Asterisks indicate a significant
 9 impact of the focal trait on the presence of the tRNA gene (see Methods). Target tRNAs are
 10 highlighted with a solid underline, and non-target tRNAs with a dashed underline. Note the
 11 different y-axis scale in panel C.

12

1 REFERENCES

- 2 Björk, G. R., & Hagervall, T. G. (2014). Transfer RNA Modification: Presence, Synthesis, and
3 Function. *EcoSal Plus*, 6(1). <http://doi.org/10.1128/ecosalplus.ESP-0007-2013>
- 4 Boccaletto, P., Machnicka, M. A., Purta, E., Piątkowski, P., Bagiński, B., Wirecki, T. K., ...
5 Bujnicki, J. M. (2018). MODOMICS: a database of RNA modification pathways. 2017 update.
6 *Nucleic Acids Research*, 46(D1), D303–D307. <http://doi.org/10.1093/nar/gkx1030>
- 7 Bollback, J. P. (2006). SIMMAP: stochastic character mapping of discrete traits on phylogenies.
8 *BMC Bioinformatics*, 7(1), 88. <http://doi.org/10.1186/1471-2105-7-88>
- 9 Brierley, I., Meredith, M. R., Bloys, A. J., & Hagervall, T. G. (1997). Expression of a coronavirus
10 ribosomal frameshift signal in *Escherichia coli*: influence of tRNA anticodon modification on
11 frameshifting. *Journal of Molecular Biology*, 270(3), 360–373.
12 <http://doi.org/10.1006/jmbi.1997.1134>
- 13 Chan, P. P., & Lowe, T. M. (2009). GtRNAdb: A database of transfer RNA genes detected in
14 genomic sequence. *Nucleic Acids Research*, 37(SUPPL. 1), 93–97.
15 <http://doi.org/10.1093/nar/gkn787>
- 16 Charif, D., Thioulouse, J., Lobry, J. R., & Perrière, G. (2005). Online synonymous codon usage
17 analyses with the ade4 and seqinR packages. *Bioinformatics*, 21(4), 545–547.
18 <http://doi.org/10.1093/bioinformatics/bti037>
- 19 Chen, S. L., Lee, W., Hottes, A. K., Shapiro, L., & McAdams, H. H. (2004). Codon usage between
20 genomes is constrained by genome-wide mutational processes. *Proceedings of the National
21 Academy of Sciences of the United States of America*, 101(10), 3480–3485.
22 <http://doi.org/10.1073/pnas.0307827100>
- 23 Chionh, Y. H., McBee, M., Babu, I. R., Hia, F., Lin, W., Zhao, W., ... Dedon, P. C. (2016). tRNA-
24 mediated codon-biased translation in mycobacterial hypoxic persistence. *Nature
25 Communications*, 7, 13302. <http://doi.org/10.1038/ncomms13302>
- 26 Crick, F. H. (1966). Codon--anticodon pairing: the wobble hypothesis. *Journal of Molecular
27 Biology*, 19(2), 548–55. [http://doi.org/10.1016/S0022-2836\(66\)80022-0](http://doi.org/10.1016/S0022-2836(66)80022-0)
- 28 Eddy, S. (1998). Profile hidden Markov models. *Bioinformatics*, 14(9), 755–763.
29 <http://doi.org/btb114> [pii]

- 1 Eddy, S. R. (2009). A New Generation of Homology Search Tools Based on Probabilistic Inference.
2 *Genome Informatics*, 23(1), 205–211. http://doi.org/10.1142/9781848165632_0019
- 3 Elliot, M. G., & Mooers, A. Ø. (2014). Inferring ancestral states without assuming neutrality or
4 gradualism using a stable model of continuous character evolution. *BMC Evolutionary*
5 *Biology*, 14(1), 226. <http://doi.org/10.1186/s12862-014-0226-8>
- 6 Gao, T., Tan, M., Liu, W., Zhang, C., Zhang, T., Zheng, L., ... Zhou, R. (2016). GidA, a tRNA
7 Modification Enzyme, Contributes to the Growth, and Virulence of *Streptococcus suis*
8 Serotype 2. *Frontiers in Cellular and Infection Microbiology*, 6(April), 1–11.
9 <http://doi.org/10.3389/fcimb.2016.00044>
- 10 Grosjean, H., Breton, M., Sirand-Pugnet, P., Tardy, F., Thiaucourt, F., Citti, C., ... Blanchard, A.
11 (2014). Predicting the minimal translation apparatus: lessons from the reductive evolution of
12 Mollicutes. *PLoS Genetics*, 10(5), e1004363. <http://doi.org/10.1371/journal.pgen.1004363>
- 13 Grosjean, H., de Crécy-Lagard, V., & Marck, C. (2010). Deciphering synonymous codons in the
14 three domains of life: Co-evolution with specific tRNA modification enzymes. *FEBS Letters*,
15 584(2), 252–264. <http://doi.org/10.1016/j.febslet.2009.11.052>
- 16 Hershberg, R., & Petrov, D. A. (2009). General Rules for Optimal Codon Choice. *PLoS Genetics*,
17 5(7), e1000556. <http://doi.org/10.1371/journal.pgen.1000556>
- 18 Ikemura, T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms.
19 *Molecular Biology and Evolution*, 2(1), 13–34.
- 20 Kanaya, S., Yamada, Y., Kudo, Y., & Ikemura, T. (1999). Studies of codon usage and tRNA genes of
21 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: Gene expression level
22 and species-specific diversity of codon usage based on multivariate analysis. *Gene*, 238(1),
23 143–155. [http://doi.org/10.1016/S0378-1119\(99\)00225-5](http://doi.org/10.1016/S0378-1119(99)00225-5)
- 24 Lowe, T. M., & Eddy, S. R. (1996). TRNAscan-SE: A program for improved detection of transfer
25 RNA genes in genomic sequence. *Nucleic Acids Research*, 25(5), 955–964.
26 <http://doi.org/10.1093/nar/25.5.0955>
- 27 Machnicka, M. a, Olchowik, A., Grosjean, H., & Bujnicki, J. M. (2015). Distribution and
28 frequencies of post-transcriptional modifications in transfer RNAs. *RNA Biology*,
29 6286(December), 0. <http://doi.org/10.4161/15476286.2014.992273>
- 30 Manickam, N., Joshi, K., Bhatt, M. J., & Farabaugh, P. J. (2015). Effects of tRNA modification on

- 1 translational accuracy depend on intrinsic codon–anticodon strength. *Nucleic Acids Research*,
2 gkv1506. <http://doi.org/10.1093/nar/gkv1506>
- 3 Marck, C., & Grosjean, H. (2002). tRNomics: analysis of tRNA genes from 50 genomes of
4 Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific
5 features. *RNA (New York, N.Y.)*, 8(10), 1189–232. <http://doi.org/10.1017/S1355838202022021>
- 6 Murao, K., Hasegawa, T., & Ishikura, H. (1976). 5-Methoxyuridine: A New Minor Constituent
7 Located in the First Position of the Anticodon of tRNA^{Ala}, tRNA^{Thr}, and tRNA^{Val} from
8 *Bacillus Subtilis*. *Nucleic Acids Research*, 3(10), 2851–2860.
9 <http://doi.org/10.1093/nar/3.10.2851>
- 10 Nasvall, S. J., Chen, P., & Bjork, G. R. (2004). The modified wobble nucleoside uridine-5-oxyacetic
11 acid in tRNA^{Pro}(cmo5UGG) promotes reading of all four proline codons in vivo. *RNA (New*
12 *York, N.Y.)*, 10(10), 1662–73. <http://doi.org/10.1261/rna.7106404>
- 13 Näsval, S. J., Chen, P., & Björk, G. R. (2007). The wobble hypothesis revisited: uridine-5-
14 oxyacetic acid is critical for reading of G-ending codons. *RNA (New York, N.Y.)*, 13(Iii), 2151–
15 2164. <http://doi.org/10.1261/rna.731007>
- 16 Noguchi, S., Nishimura, Y., Hirota, Y., & Nishimura, S. (1982). Isolation and characterization of an
17 *Escherichia coli* mutant lacking tRNA-guanine transglycosylase. Function and biosynthesis of
18 queuosine in tRNA. *The Journal of Biological Chemistry*, 257(11), 6544–50.
- 19 Notredame, C., Higgins, D., & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate
20 multiple sequence alignment. *Journal of Molecular Biology*, 302(1), 205–217.
21 <http://doi.org/10.1006/jmbi.2000.4042>
- 22 Novoa, E. M., Pavon-Eternod, M., Pan, T., & Ribas De Pouplana, L. (2012). A role for tRNA
23 modifications in genome structure and codon usage. *Cell*, 149(1), 202–213.
24 <http://doi.org/10.1016/j.cell.2012.01.050>
- 25 Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V, Chuang, H.-Y., Cohoon, M., ... Vonstein,
26 V. (2005). The subsystems approach to genome annotation and its use in the project to annotate
27 1000 genomes. *Nucleic Acids Research*, 33(17), 5691–702. <http://doi.org/10.1093/nar/gki866>
- 28 R Development Core Team. (2015). R: A Language and Environment for Statistical Computing.
29 <http://doi.org/10.1038/sj.hdy.6800737>
- 30 Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other

- 1 things). *Methods in Ecology and Evolution*, 3(2), 217–223. <http://doi.org/10.1111/j.2041->
2 210X.2011.00169.x
- 3 Segata, N., Börnigen, D., Morgan, X. C., & Huttenhower, C. (2013). PhyloPhlAn is a new method
4 for improved phylogenetic and taxonomic placement of microbes. *Nature Communications*, 4.
5 <http://doi.org/10.1038/ncomms3304>
- 6 Shippy, D. C., Eakley, N. M., Lauhon, C. T., Bochsler, P. N., & Fadl, A. A. (2013). Virulence
7 characteristics of Salmonella following deletion of genes encoding the tRNA modification
8 enzymes GidA and MnmE. *Microbial Pathogenesis*, 57, 1–9.
9 <http://doi.org/10.1016/j.micpath.2013.01.004>
- 10 Wald, N., & Margalit, H. (2014). Auxiliary tRNAs: Large-scale analysis of tRNA genes reveals
11 patterns of tRNA repertoire dynamics. *Nucleic Acids Research*, 42(10), 6552–6566.
12 <http://doi.org/10.1093/nar/gku245>
- 13 Wolf, J., Gerber, A. P., & Keller, W. (2002). tadA, an essential tRNA-specific adenosine deaminase
14 from Escherichia coli. *The EMBO Journal*, 21(14), 3841–51.
15 <http://doi.org/10.1093/emboj/cdf362>
- 16 Yamada, Y., Matsugi, J., Ishikura, H., & Murao, K. (2005). Bacillus subtilis tRNA^{Pro} with the
17 anticodon mo⁵UGG can recognize the codon CCC. *Biochimica et Biophysica Acta - Gene*
18 *Structure and Expression*, 1728(3), 143–149. <http://doi.org/10.1016/j.bbaexp.2005.02.011>
- 19 Yokobori, S. I., Kitamura, A., Grosjean, H., & Bessho, Y. (2013). Life without tRNA^{Arg}-adenosine
20 deaminase TadA: Evolutionary consequences of decoding the four CGN codons as arginine in
21 Mycoplasmas and other Mollicutes. *Nucleic Acids Research*, 41(13), 6531–6543.
22 <http://doi.org/10.1093/nar/gkt356>
- 23 Yokoyama, S., Watanabe, T., Murao, K., Ishikura, H., Yamaizumi, Z., Nishimura, S., & Miyazawa,
24 T. (1985). Molecular mechanism of codon recognition by tRNA species with modified uridine
25 in the first position of the anticodon. *Proceedings of the National Academy of Sciences*, 82(15),
26 4905–4909. <http://doi.org/10.1073/pnas.82.15.4905>
- 27 Zallot, R., Yuan, Y., & de Crécy-Lagard, V. (2017). The Escherichia coli COG1738 Member YhhQ
28 Is Involved in 7-Cyanodeazaguanine (preQ₀) Transport. *Biomolecules*, 7(1), 1–13.
29 <http://doi.org/10.3390/biom7010012>