Initiator tRNA Genes Template the 3′ CCA End at High Frequencies in Bacteria

David H. Ardell [1, 2, *] and Ya-Ming Hou [3]

[1] Program in Quantitative and Systems Biology, University of California, Merced, 5200 North Lake Road, Merced, CA 95343

[2] Molecular and Cell Biology Unit, School of Natural Sciences, University of California, Merced, 5200 North Lake Road, Merced, CA 95343

[3] Department of Biochemistry and Molecular Biology, Thomas Jefferson University, 233 South 10th Street, BLSB 220, Philadelphia, PA 19107, U.S.A.

* To whom correspondence should be addressed. +1 209 228 2953. Email: dardell@ucmerced.edu

December 22, 2015

ABSTRACT

While the CCA sequence at the mature 3′ end of tRNAs is conserved and critical for translational function, a genetic template for this sequence is not always contained in tRNA genes. In eukaryotes and archaea, the CCA ends of tRNAs are synthesized post-transcriptionally by CCA-adding enzymes. In bacteria, tRNA genes template CCA sporadically. In order to understand variation in how prokaryotic tRNA genes template CCA, we re-annotated tRNA genes in the tRNAdb-CE database. Among 132,129 prokaryotic tRNA genes, initiator tRNA genes template CCA at the highest average frequency (74.1%) over all functional classes except selenocysteine and pyrrolysine tRNA genes (88.1% and 100% respectively). Across bacterial phyla and a wide range of genome sizes, many lineages exist in which predominantly initiator tRNA genes template CCA. Preferential retention of CCA in initiator tRNA genes evolved multiple times during reductive genome evolution in Bacteria. Also, in a majority of cyanobacterial and actinobacterial genera, predominantly initiator tRNA genes template CCA. We suggest that cotranscriptional synthesis of initiator tRNA CCA 3′ ends can complement inefficient processing of initiator tRNA precursors, "bootstrap" rapid initiation of protein synthesis from a non-growing state, or contribute to an increase in cellular growth rates by reducing overheads of mass and energy to maintain nonfunctional tRNA precursors. More generally, CCA templating in structurally non-conforming tRNA genes can afford cells robustness and greater plasticity to respond rapidly to environmental changes and stimuli.

Running Head: CCA-templating in bacterial initiator tRNA genes

INTRODUCTION

All active tRNA molecules must contain a CCA sequence at the 3′-end as the site for amino acid attachment and for interaction with the ribosome during protein synthesis (Betat, Rammelt et al. 2010, Vortler and Morl 2010, Betat and Morl 2015). While essential for tRNA activities, the CCA sequence is generally not encoded in tRNA genes but is added post-transcriptionally. Exceptions are found in bacteria, where some tRNA genes contain a template of the CCA sequence for direct synthesis at the time of transcription. However, CCA-templating is not necessarily conserved among tRNA genes with different functional identities or among bacterial species across different phyla. To explore whether there is potential selective pressure for tRNA genes to template CCA in bacteria, we undertook a reannotation of publicly available tRNA gene data.

One source of error in the annotation of tRNA genes concerns the functional classification of genes for tRNAs with CAU anticodons. These include genes for both the initiator and elongator tRNA^{Met} and specific elongator $\text{tRNA}^{\text{Ile}}_{\text{CAU}}$ isoacceptors in bacteria and archaea. In the latter case, transcribed CAU anticodons are post-transcriptionally modified to distinguish them from the unmodified CAU anticodons of cytosolic tRNA^{Met} (Suzuki and Miyauchi 2010). However, currently available tRNA gene-finders annotate all three classes as elongator tRNA^{Met} genes (Ardell 2010). The TFAM tRNA functional classifier, which uses profile-based models of whole tRNA sequences (Ardell and Andersson 2006, Tåquist, Cui et al. 2007), can differentiate all three tRNA functional classes with generally high specificity and sensitivity (Silva, Belda et al. 2006). However, the $\text{tRNA}^{\text{Ile}}_{\text{CAU}}$ class evolves more rapidly than other classes, so that even though the TFAM 1.4 Proteobacterial-specific model generalizes well to some other Bacterial phyla, this model does not generalize well to all (Freyhult, Cui et al. 2007). An alternative TFAM model (Amrine, Swingley et al. 2014), for just genes for tRNAs with CAU anticodons, is based on a custom annotation of such genes in a wide sampling of bacterial taxa (Silva, Belda et al. 2006). Although this alternative model is imperfect in its sensitivity and specificity (Silva, Belda et al. 2006), as discussed further below, its performance is satisfactory and suitable for the present study.

Here we apply the alternative “Silva TFAM model” to improve the functional annotation of tRNA genes with CAU anticodons in the high quality public database tRNAdb-CE (Abe, Ikemura et al. 2009, Abe, Inokuchi et al. 2014). In our analysis, we found that genes for the initiator class of tRNAs across the bacterial domain consistenly template CCA with significantly higher frequencies than elongator tRNA genes. This CCA-templating can provide unique advantages to initiator tRNA for rapid maturation, aminoacylation, and initiation of protein synthesis.

RESULTS

Functional Reannotation of Bacterial Genes in tRNAdb-CE v.8

The tRNAdb-CE v0.8 database uses TFAM 1.4 for functional classification of bacterial tRNA genes (as described in http://trna.ie.niigata-u.ac.jp/trnadb/method.html). However, the Proteobacterial model for the $tRNA^{Ile}{}_{CAU}$ elongator class that comes with TFAM 1.4 does not generalize well to all bacterial phyla (Freyhult, Cui et al. 2007). Therefore, we reannotated 9,914 bacterial genes for tRNAs with CAU anticodons in tRNAdb-CE v0.8 using the more general Silva TFAM model derived from the analysis. This model is also provided as supplementary data in the present work. By applying the Silva TFAM model, we revised the functional classification of 4,362 of 9,914 genes (≈ 43.9%). Reclassification frequencies are presented in Table 1, showing that most of the changes involve reclassification of genes from $tRNA^{Met}$ to initiator $tRNA^{fMet}$ or to $tRNA^{Ile}{}_{CAU}$. Reannotated data are provided in supplementary materials.

Structural Reannotation of Bacterial Genes in tRNAdb-CE v.8

A well-designed feature of the tRNAdb-CE v0.8 data model lies in that its gene records contain not only annotated gene sequences but also ten bases of genomic context both up- and downstream. Inspection of tRNAdb-CE v0.8 data revealed multiple genes with an annotated 3´-end sequence other than CCA, followed by 3´-trailer sequences that begin with the sequence CCA. To confidently assess whether these genes might template a 3´-CCA-end for their gene products, we assigned Sprinzl coordinates

(Sprinzl, Horn et al. 1998) to these bases for each gene sequence. These coordinates were not provided in tRNAdb-CE v0.8. We did this by implementing a dynamic programming algorithm to optimize base-pairing of the acceptor 3′-end region against the database-annotated 5′-end. Although our acceptor-end annotations were almost always identical with those annotated in the database, they enabled us to confidently and consistently assign Sprinzl coordinates to the 3′-end region of each gene. Using this technique, we annotated an additional 2,866 bacterial tRNA genes out of 129,989 (or 2.2%) records as containing the CCA template at the 3′-end in the sequence framework of Sprinzl coordinates 74-76.

To clarify why we could identify an additional 2,866 tRNA genes in tRNAdb-CE v0.8 that template CCA, we ran tRNA gene-finding programs on the database records. We used ARAGORN v1.0 (Laslett and Canback 2004) and tRNAscan-SE v.1.23 (Lowe and Eddy 1997) in default eukaryotic tRNA gene-finding mode, and tRNAscan-SE v.1.23 in Bacterial mode (with the -B option). We found that tRNAscan-SE v.1.23, when run in its default eukaryotic gene-finding mode, never annotates nucleotides at positions 74-76 irrespective of sequence.

An exception to this rule was with selenocysteine tRNA genes, for which tRNAscan-SE in eukaryotic mode does annotate positions 74-76 if they contain the CCA sequence. From this observation we conclude that a likely cause of misannotations in tRNAdb-CE is user error in genome annotation pipelines. This is particularly notable when users of tRNAscan-SE use its default eukaryotic gene-finding mode on prokaryotic genomes. Such errors may then be incessently propagated in public and private databases.

Frequencies of CCA-templating in Bacterial tRNA Genes

With our reannotated tRNAdb-CE data in hand, we calculated frequencies of CCA-templating in tRNA genes across different tRNA functional classes and taxonomic groupings as defined by NCBI Taxonomy. Figure 1 visualizes our data summarized by prokaryotic genus. Prokaryotic clades exhibit all four possible patterns: 1) all tRNAs genes template CCA, 2) few or no tRNA genes template CCA, 3) primarily initiator genes template CCA, or — most rarely — 4) primarily elongator tRNA genes template CCA.

The five best-sampled phyla in our dataset, as defined by number of distinct genera with at least one genome sequenced, are Proteobacteria, Bacillus/Clostridium, Actinobacteria, Bacteroidetes/Clorobi, and Cyanobacteria. These five phyla exhibit three of four patterns described above in a strikingly consistent pattern by phylum. Practically all tRNA genes template CCA in Proteobacteria and Bacillus/Clostridium, except in certain reduced genomes, most of which template CCA only in initiator tRNA genes, or in no tRNA genes at all. In Cyanobacteria and Actinobacteria, on the other hand, primarily only the initiator tRNA genes template CCA, with certain exceptions. For example, a clade of Actinobacteria with relatively small genomes exists in which both initiator and elongator tRNA genes template CCA at high frequencies. In the Bacteroidetes/Chlorobi group, most tRNA genes do not template CCA, except for one lineage, the Solitalea, in which only initiator tRNA genes template CCA. In all five of the most-sampled phyla, there exist both small and moderately-sized genomes in which only initiator tRNA genes template CCA, or no tRNA genes at all template CCA. Certain Myxococcales, among the Deltaproteobacteria, are exceptional in having among the largest genomes that we observed and yet no tRNA genes or only initiator tRNA genes template CCA.

Less-sampled phyla are also quite heterogeneous in our dataset. In the Thermotogae, Deinococcus/Thermus and Tenericutes, all tRNA genes template CCA. Spirochaetae do not template CCA in any genes, while in Deferribacteres, only initiator tRNA genes template CCA. In the most rare pattern we observed, in only a few archaeal or bacterial genera, primarily elongator tRNA genes and not initiator tRNA genes template CCA.

In order to better visualize these data down to individual genomes and separating different elongator classes, we created an interactive javascript-based taxonomic navigator for our results visualized with heatmaps in any ordinary web browser. The full interactive data navigator is available as supplementary materials to this work. A static view on these data is also provided as a searchable PDF in supplementary materials. Figure 2 presents a snapshot from this browser with some notable detailed results for Bacterial tRNA genes. The figure shows columns of frequency data, corresponding to functional classes of tRNA genes, sorted left to right by decreasing average frequency at which tRNA genes template CCA over all prokaryotic genomic sequences in our sample. This analysis reveals that initiator tRNA genes (labeled as Ini) template CCA at the highest frequency (74.1%), versus 66.2% for elongator tRNA genes generally in prokaryotes. Bacterial initiator tRNA genes template CCA at a frequency of 74.7%, second only to selenocysteinyl tRNA genes ($tRNA^{Sec}$, Sec = selenocysteine), with a frequency of 89.2%. We also found that genes for $tRNA^{Asp}$ and $tRNA^{Asn}$ template CCA at the highest frequencies among all canonical elongator tRNA genes. Below we describe some of the notable results shown in Figure 2.

Cyanobacteria. Among bacterial phyla we observed, Cyanobacteria have the most striking and consistent pattern in which specifically initiator and not elongator tRNA genes template CCA. The overall rates are 64.1% for initiator tRNA genes versus 25.7% for the next highest gene class, which are elongator $tRNA^{Tyr}$ genes. But different cyanobacterial lineages exhibit considerable variation in this trait. For example, among Prochlorales and Nostocales genomes — comprising both the smallest and largest average genome sizes, respectively — the frequencies at which initiator tRNA genes template CCA are 91.7% and 88.9%, while elongator tRNA genes template CCA at only 3.4% and 8.2% respectively. In 11 out of 12 of *Prochlorococcus* genomes and 9 out of 13 *Synechococcus* genomes, only initiator tRNA genes template CCA. Initiator tRNA genes template CCA at very different rates in sister orders Oscillatoriales and Chroococcales within subclass Oscillatoriophycideae: 87.5 and 43.8% respectively.

The Cyanobacteria are also unusual in that different strains and groups feature specific elongator tRNA gene classes that also template CCA at intermediate rates (above 10%) while other elongator classes template CCA at lower rates (below 10%). Usually, if in any one genome the initiator tRNA gene or genes template CCA, at least one elongator tRNA gene class will also template CCA at an intermediate rate. The elongator gene class that templates CCA most consistently across the phylum is the tRNA^{Tyr} gene class. In *Nostocales*, tRNA^{Tyr} genes template CCA at a frequency of (41.7%), while tRNA^{Asn} and tRNA^{Gln} elongator genes also template CCA at a high relative rate (36.8% and 23.7%).

Proteobacteria. All proteobacterial tRNA genes generally template CCA at consistently high rates: 96.1% overall (Figure 2). Yet proteobacterial initiator tRNA genes template CCA at 98.0%, significantly higher than proteobacterial elongators ($\chi 2 = 23.625$, d.f. = 1, $p < 10^{-4}$ by Fisher's Exact Test with a Yates correction). Closer examination of the proteobacterial variation (supplementary materials) reveals that while many free-living proteobacteria template CCA at high rates, endosymbiotic γ-proteobacteria and α-proteobacteria with reduced genomes show similar patterns to those described above for cyanobacteria with reduced genomes. In these cases, initiator tRNA genes appear to be the only class to consistently template CCA, while several elongator classes also template CCA. For example, in most *Buchnera aphidicola* genomes, about eight or nine additional elongator tRNA classes template CCA at intermediate to high rates while other classes do not template CCA, as previously reported (Hansen and Moran 2012). However, not previously reported is that in all *Buchnera* strain genomes except one, initiator tRNA genes always template CCA. Furthermore, like in the Cyanobacteria, in the smallest of the *Buchnera* genomes, only initiator tRNA genes template CCA. This same pattern holds in other endosymbiotic γ-proteobacteria genomes such as Ca. *Blochmannia, Wigglesworthia, Glossina*, Ca. *Baumannia,* Ca. *Carsonella,* Ca. *Portiera,* as well as α-proteobacteria endosymbionts such as *Wolbachia*. In contrast, among the smallest γ-proteobacterial genomes like Ca. *Hodgkinia,* none of the tRNA genes template CCA.

Other bacterial phyla. Many diverse genera and classes of bacteria preferentially template CCA in their initiator tRNA genes (Supplementary files). Examples include *Geobacillus*, *Thermoaerobacter, Ruminococcus, Thermomicrobiales, Deferribacter,* Thermodesulfobacteria*, Mycobacterium, Propionibacterium, Frankia,* and *Bifidobacterium*. As shown in Figure 2, within the Bacillus/Clostridium phylum, frequency variation in this genomic trait also extensive. An unusual pattern is found in the pathogenic Staphylococcaeceae and Listeraceae families, and also the Lactobacillales, which contain both pathogens and non-pathogens, in which initiator tRNA genes never template CCA, even while elongator tRNA genes do template CCA at intermediate rates. For example, in Staphylococcaceae about 60% of elongator tRNA genes template CCA and in Listeraceae about 24% of elongator tRNA genes template CCA, while in Lactobacillales, 1.4% of elongator genes template CCA. Yet among the 257 genome representatives of these three families in our dataset, not one initiator tRNA gene templates CCA.

Archaea. We found no need for structural or functional reannotation of archaeal tRNA genes in tRNAdb-CE v.8. Figure 3 presents a snapshot from this browser with some of our most notable results for archaeal tRNA genes. While there are fairly high frequencies of CCA-templating in archaeal tRNA genes overall, at 30.4%, we found that initiator tRNA genes in Archaea do not template CCA at any especially high frequency among tRNA genes, which presents a major difference from bacteria. Other than this, we observed extensive phyletic variation in this trait across Archaea. Crenarchaeota tRNA genes template CCA at a rate of 50.5%, while Euryarchaeota tRNA genes template CCA at about half of that rate. Within Crenarchaeota, tRNA genes in the Sulfolobales template CCA at 3.8%, but in the Desulforococcales this rate is 84.8%. All four tRNA^{Pyl} (Pyl = pyrrolysine) genes template CCA in the Methanomicrobia. Contrary to the generalization that Archaea and Eukarya do not template CCA, there exist lineages in both the Crenarchaeota and Euryarchaeota in which all or nearly all tRNA genes template CCA, for example in the Desulfurococcales, Protoarchaea, and Methanopyri. Although variation exists across tRNA functional classes in a phyletic pattern, no obvious overall pattern emerges.

DISCUSSION

We observed widespread phyletic variation in the frequencies and patterns at which tRNA genes template CCA across functional classes in prokaryotic genomes. Across diverse bacterial and archaeal clades, frequencies range between 0 to 100%. The key finding is that initiator tRNA genes have the greatest class-specific frequency of CCA-templating in bacteria after tRNA^{Sec} genes. Furthermore, in diverse bacterial lineages, especially among the reduced genomes of free-living Cyanobacteria and host-associated endosymbiotic Proteobacteria, initiator tRNA genes template CCA at uniquely high frequencies. In Proteobacteria, all tRNA genes template CCA at high rates, but initiator tRNA genes have the highest overall rate, second only to tRNA^{Sec} genes.

We believe that the tRNA gene reannotations that led to our results are accurate. The most important source of reannotation errors would be from our reclassification of tRNA gene function (Table 1). Note that no previously annotated initiator tRNA genes were reclassified in our analysis, but rather a substantial fraction of genes annotated as elongator tRNA^{Met} were reclassified as tRNA^{iMet} or tRNA^{Ile}. Of these reclassifications, detection of initiator tRNA genes by TFAM has very high sensitivity and specificity (Ardell and Andersson 2006, Silva, Belda et al. 2006). This is because initiator tRNA sequence and structure is highly conserved over the three domains of life (Marck and Grosjean 2002).

For example, one of the Cyanobacterial lineages shown in Figure 2 — *Gloeobacter* — is annotated as not having any initiator tRNA genes in tRNAdb-CE v.8. In our analysis of the tRNA gene complements of 2323 prokaryotic genomes in tRNAdb-CE v.8, initiator tRNA genes were not annotated in only 15 genomes (0.6%). We spot-checked several of these aberrant tRNA gene complements by examining their score distributions with TFAM 0.4 to verify that there were no viable candidates for initiator tRNA genes in these gene complements. No tRNA genes in any of the gene complements that we checked scored outside of the normal background distribution for the initiator tFAM model. We believe that initiator tRNA genes may simply be missing from the genome annotations that were aggregated in tRNAdb-CE v.8. Moreover, statistically, our results are robust to these missing data.

The initiator tRNA of protein synthesis in Bacteria is known as tRNA^{fMet} , because its charged methionine moiety contains a formyl group attached to the α-amino group. By templating CCA in tRNA^{fMet} genes, Bacteria can directly synthesize tRNA^{fMet} with the CCA sequence at the 3′-end. Below we hypothesize five non-mutually exclusive potential advantages for tRNA genes to template CCA.

Our first hypothesis is that certain tRNA classes, particularly initiator tRNAs, may have relatively non-conforming structures that lead to inefficient processing in shared tRNA maturation pathways. For example, tRNA^{fMet} in Bacteria is exceptional in that it contains a mismatched C-A pair at the 1-72 position of the acceptor end, providing a C1-A72 motif for recognition by initiation factors to initiate protein synthesis ((Lee and RajBhandary 1991). All elongator tRNAs contain a Watson-Crick (W-C) base pair at the 1-72 position and therefore are discriminated against by initiation factors. However, the C1-A72 motif of $\text{tRNA}^{\text{fMet}}$ compromises the efficiency of processing at the 5´-end (Meinnel and Blanquet 1995), so direct transcription of the CCA sequence can be a critical component to mitigate this reduced efficiency (Wegscheid and Hartmann 2006, Wegscheid and Hartmann 2007). In contrast, initiator tRNAs in Archaea have Watson-Crick base-pairs in the 1-72 position (Marck and Grosjean 2002) so we conjecture that they do not share this “Achilles heel” problem with Bacterial initiator tRNAs, but instead are efficiently processed at the 5´-end without any requirement for a 3´-end CCA sequence. This is consistent with our observation that initiator tRNA genes in Archaea do not have a particularly high frequency of CCA-templating.

Second, direct templating of the CCA sequence in tRNAs can potentially increase the maximal growth rate of cells. Under conditions of rapid growth, the co-transcriptional synthesis of 3´-terminal CCA in tRNAs can increase the allocation of cellular resources directly to the synthesis of new proteomic biomass and growth in two ways: first, by reducing or eliminating steady-state cellular pools of species of nonfunctional tRNA precursors, which reduces the mass and energy overhead of the translational machinery itself, and second, by reducing the steady-state fraction of ribosomes devoted to synthesizing tRNA-affiliated proteins such as CCA-adding enzyme (Ehrenberg and Kurland 1984, Klumpp, Scott et al. 2013).

Third, given that translational initiation is rate-limiting in protein synthesis (Vind, Sorensen et al. 1993), and therefore a key determinant of maximal growth rate (Ehrenberg and Kurland 1984, Hersch, Elgamal et al. 2014, Pop, Rouskin et al. 2014), cells selected for a high maximum growth rate may need to efficiently maintain high concentrations of initiator tRNA^{fMet} for rapid growth. The costs of maturation of a tRNA to a growing cell should increase proportionally with the concentration of that tRNA, and initiator tRNA concentration increases more with growth rate in *E. coli* than elongator tRNAs (Dong, Nilsson et al. 1996), so the fitness impact of templating CCA in initiator tRNAs should be greater than in elongator tRNAs in rapidly growing cells. For example, the record-high growth rates reported among *Vibrio* species (Aiyar, Gaal et al. 2002) is associated with very high initiator tRNA gene copy numbers in *Vibrio* genomes (Ardell and Andersson 2006). Consistent with the above, all initiator tRNA genes in *Vibrio* template CCA in the present analysis.

Fourth, rapid synthesis of initiator tRNAs through co-transcriptional synthesis of CCA could reduce the lag phase associated with the transition to growth by reducing the waiting time to increase initiator tRNA concentration. Importantly, this “bootstrapping” trait may be important for all cells, including free-living and endosymbiotic bacteria under reductive genome evolution, and not just for cells capable of rapid growth.

Many such cells could have an advantage in the rapid initiation of protein synthesis from a quiescent state in response to environmental change. Indeed, we have shown that each nucleotide addition for post-transcriptional synthesis of CCA requires the CCA enzyme to proofread tRNA integrity (Dupasquier, Kim et al. 2008, Hou 2010), which likely delays maturation of newly transcribed tRNAs.

Fifth, for elongator tRNA genes, direct templating of CCA can facilitate more rapid synthesis of corresponding tRNA elongators to help cells avoid transient depletion of specific ternary complexes and the detrimental consequences that such shortages may have on the accuracy of protein synthesis and proteomic integrity. The supply-demand theory of tRNA charging dynamics (Elf, Nilsson et al. 2003) predicts wide variability in sensitivity of charging levels of tRNA species to perturbations, such as amino acid starvation, affecting specific elongator tRNAs for both proteomically abundant and rare amino acids such as Leucine, Tyrosine and Phenylalanine. Stalled ribosomes caused by shortages of specific ternary complexes increase translational misreading at corresponding “hungry” codons (O'Farrell 1978, Gamper, Masuda et al. 2015), including frame-shift errors (Gallant and Lindsley 1998), all of which can cause protein misfolding, aggregation, and damage (Drummond and Wilke 2009).

While many cyanobacteria with reduced genomes are not fast-growing, they may generally be subject to multiple constraints of chronic nutrient limitation and a heavy burden of a large fraction of proteome dedicated to autotrophic functions (Burnap

2015). When combined, these factors may lead to “proteomic constraints” from small cell sizes, an exacerbation of macromolecular crowding, and increased sensitivity to mistranslation of the most abundant parts of the proteome (Burnap 2015). We suggest that the relatively high frequency at which tRNA^{Tyr} genes template CCA in Cyanobacteria (Fig. 2) is associated with a unique biological sensitivity to depletion of charged tRNA^{Tyr}. Tyrosine residues are critically important for both catalysis and stability of RuBisCo (Esquivel, Pinto et al. 2006), one of the most abundant proteins in Cyanobacteria (Wegener, Singh et al. 2010). In light of this hypothesis it is remarkable that there exists a d-Tyr-tRNA^{Tyr} deacylase that is conserved and apparently unique to Cyanobacteria (Wydau, van der Rest et al. 2009), which helps maintain the accuracy of tRNA^{Tyr} charging. Competition experiments that model biologically relevant conditions with Cyanobacterial strains with or without CCA-templating for tRNA^{Tyr}, as well as biochemical assays, could test this hypothesis.

We further suggest that the advantages of avoiding supply shortages and streamlining tRNA biogenesis pathways may extend to other elongator tRNAs that we found to template CCA in an often lineage-specific manner. Selenocysteine and Pyrrolysine tRNAs both have complex biosynthetic/maturation pathways and both template CCA at high frequencies in our analysis. Similarly, biosynthesis of Asn-tRNA^{Asn} involves two steps, first by synthesizing a mispaired Asp-tRNA^{Asn}, followed by conversion of Asp to Asn (Curnow, Ibba et al. 1996, Becker and Kern 1998, Bailly, Blaise et al. 2007). Indeed, genes for both tRNA^{Asp} and tRNA^{Asn} template CCA at high frequencies.

Although the synthesis of Gln-tRNA^{Gln} also relies on a two-step pathway involving transamidation of Glu on Glu-tRNA^{Gln} (Gagnon, Lacoste et al. 1996), the frequencies for tRNA^{Glu} and tRNA^{Gln} are among the lowest we observed in Bacteria overall. Further analysis and experiments will be necessary to fully understand the patterns reported in this paper.

Re-annotation of tRNA gene sequences was essential to our discovery that CCA-templating is a major feature of initiator tRNA genes. This shows the importance for genome annotation projects of using tRNA gene-finders with taxonomically correct models. More generally, this work demonstrates the importance of using bioinformatic assets carefully to maximize scientific returns.

MATERIALS AND METHODS

Data. Version 8 (October, 2014) of the tRNAdb-CE database (Abe, Inokuchi et al. 2014) was downloaded on November 4, 2014. NCBI Taxonomy data (NCBI Resource Coordinators 2014) was downloaded on November 13, 2014.

Functional Reannotation of CAU-anticodon tRNAs. We classified bacterial CAU-anticodon-templating tRNA genes as templating methionine elongators, lysidinylated isoleucine elongators or initiators using TFAM version 1.4 (Ardell and Andersson 2006, Tåquist, Cui et al. 2007) with a general bacterial model for this purpose based on a previously published analysis (Silva, Belda et al. 2006).

Structural Annotation of 3´- ends. To annotate Sprinzl coordinates to the 3´-end of each tRNAdb-CE sequence record, we implemented a dynamic programming algorithm to optimize base-pairing of the annotated 3´-end of the mature tRNA in each record against its own annotated 5´-end and trailer sequence.

For each sequence record we obtained the 5′-most seven bases of the annotated acceptor stem sequence and reversed it to obtain sequence *x*. Given sequence *x*, we computed its optimal pairing against a second sequence *y* defined by the last 12 bases of the annotated 3′-end and the first five bases of the annotated 3′-trailer using the simple dynamic programming algorithm described here.

Let x and y be finite sequences over the alphabet $\Sigma = \{A, C, G, U\}$, with lengths m and n respectively. We compute a matrix H whose elements are specified as follows:

$H(i , 1) = 0$ for all i, such that $1 \leq i \leq m;$

$H(1 , j) = 0$ for all i, such that $1 \leq j \leq n;$

$H(I , j) = \max (D, U, L)$ for all i and j, such that $2 \leq i \leq m$ and $2 \leq j \leq n,$ and

$D = H(i - 1 , j - 1) + s(x_i , y_j) + a(x_i , y_j);$

$U = H(i - 1 , j) + g;$ and

$L = H(i , j - 1) + g,$

where x_i is the ith base in sequence x of length $m = 7$, y_j is the jth base in sequence y of length $n = 17$, $s(x_i , y_j) = 4$, for $(x_i , y_j) \in \{ (A,U), (U,A), (C,G), (G,C), (G,U), (U,G)\}$ and $s(x_i , y_j) = 1$ otherwise, $a(x_i , y_j)$ is an annotation bonus if x_i and y_j were annotated as paired in tRNAdb-CE, $g = -5$ is a linear gap penalty, and $H(i , j)$ is the maximum base-pairing score obtained on sequence prefixes $x[1,i]$ and $y[1,j]$. We compared results both with and without an annotation bonus, *i.e.* we recomputed $H(m , n)$ for every record using the bonus $a(x_i , y_j) = 1$ or no bonus $a(x_i , y_j) = 0$.

Statistics and Visualization of Genome Size and CCA-templating Data. After reannotation, we considered a tRNA gene to template CCA if Sprinzl bases 74 through 76 contained the sequence CCA. We used genome size data downloaded as a "genome report" from NCBI Genome on October 26, 2015 (NCBI Resource Coordinators 2014) and visualized data using the Interactive Tree of Life (Letunic and Bork 2007, Letunic and Bork 2011)

ACKNOWLEDGEMENTS

We thank the support of NSF grant 1344279 to DHA and NIH grants, 1R01 GM114343, 5U01 GM108972, and 1R01 GM068561 to YMH.

REFERENCES

Abe, T., T. Ikemura, Y. Ohara, H. Uehara, M. Kinouchi, S. Kanaya, Y. Yamada, A. Muto and H. Inokuchi (2009). "tRNADB-CE: tRNA gene database curated manually by experts." Nucleic Acids Res **37**(Database issue): D163--168.

Abe, T., H. Inokuchi, Y. Yamada, A. Muto, Y. Iwasaki and T. Ikemura (2014). "tRNADB-CE: tRNA gene database well-timed in the era of big sequence data." Frontiers in Genetics **5**: 114.

Aiyar, S. E., T. Gaal and R. L. Gourse (2002). "rRNA promoter activity in the fast-growing bacterium Vibrio natriegens." J Bacteriol **184**(5): 1349-1358.

Amrine, K. C. H., W. D. Swingley and D. H. Ardell (2014). "tRNA signatures reveal a polyphyletic origin of SAR11 strains among alphaproteobacteria." PLoS computational biology **10**(2): e1003454.

Ardell, D. H. (2010). "Computational analysis of tRNA identity." FEBS Lett **584**(2): 325--333.

Ardell, D. H. and S. G. E. Andersson (2006). "TFAM detects co-evolution of tRNA identity rules with lateral transfer of histidyl-tRNA synthetase." Nucleic Acids Res **34**(3): 893--904.

Bailly, M., M. Blaise, B. Lorber, H. D. Becker and D. Kern (2007). "The transamidosome: a dynamic ribonucleoprotein particle dedicated to prokaryotic tRNA-dependent asparagine biosynthesis." Mol Cell **28**(2): 228-239.

Becker, H. D. and D. Kern (1998). "Thermus thermophilus: a link in evolution of the tRNA-dependent amino acid amidation pathways." Proc Natl Acad Sci U S A **95**(22): 12832-12837.

Betat, H. and M. Morl (2015). "The CCA-adding enzyme: A central scrutinizer in tRNA quality control." Bioessays **37**(9): 975-982.

Betat, H., C. Rammelt and M. Morl (2010). "tRNA nucleotidyltransferases: ancient catalysts with an unusual mechanism of polymerization." Cell Mol Life Sci **67**(9): 1447-1463.

Burnap, R. L. (2015). "Systems and Photosystems: Cellular Limits of Autotrophic Productivity in Cyanobacteria." Frontiers in Bioengineering and Biotechnology **3**(1).

Curnow, A. W., M. Ibba and D. Soll (1996). "tRNA-dependent asparagine formation." Nature **382**(6592): 589-590.

Dong, H. J., L. Nilsson and C. G. Kurland (1996). "Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates." J Mol Biol **260**(5): 649-663.

Drummond, D. A. and C. O. Wilke (2009). "The evolutionary consequences of erroneous protein synthesis." Nat Rev Genet **10**(10): 715-724.

Dupasquier, M., S. Kim, K. Halkidis, H. Gamper and Y. M. Hou (2008). "tRNA integrity is a prerequisite for rapid CCA addition: implication for quality control." J Mol Biol **379**(3): 579-588.

Ehrenberg, M. and C. G. Kurland (1984). "Costs of accuracy determined by a maximal growth rate constraint." Q Rev Biophys **17**(1): 45-82.

Elf, J., D. Nilsson, T. Tenson and M. Ehrenberg (2003). "Selective charging of tRNA isoacceptors explains patterns of codon usage." Science **300**(5626): 1718-1722.

Esquivel, M. G., T. S. Pinto, J. Marin-Navarro and J. Moreno (2006). "Substitution of tyrosine residues at the aromatic cluster around the betaA-betaB loop of rubisco small subunit affects the structural stability of the enzyme and the in vivo degradation under stress conditions." Biochemistry **45**(18): 5745-5753.

Freyhult, E., Y. Cui, O. Nilsson and D. H. Ardell (2007). "New computational methods reveal tRNA identity element divergence between Proteobacteria and Cyanobacteria." Biochimie **89**(10): 1276--1288.

Gagnon, Y., L. Lacoste, N. Champagne and J. Lapointe (1996). "Widespread use of the glu-tRNAGln transamidation pathway among bacteria. A member of the alpha purple bacteria lacks glutaminyl-trna synthetase." J Biol Chem **271**(25): 14856-14863.

Gallant, J. A. and D. Lindsley (1998). "Ribosomes can slide over and beyond "hungry" codons, resuming protein chain elongation many nucleotides downstream." Proc Natl Acad Sci U S A **95**(23): 13771-13776.

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

Hansen, A. K. and N. A. Moran (2012). "Altered tRNA characteristics and 3' maturation in bacterial symbionts with reduced genomes." Nucleic Acids Res **40**(16): 7870--7884.

Hersch, S. J., S. Elgamal, A. Katz, M. Ibba and W. W. Navarre (2014). "Translation initiation rate determines the impact of ribosome stalling on bacterial protein synthesis." J Biol Chem **289**(41): 28160-28171.

Hou, Y. M. (2010). "CCA addition to tRNA: implications for tRNA quality control." IUBMB Life **62**(4): 251-260.

Klumpp, S., M. Scott, S. Pedersen and T. Hwa (2013). "Molecular crowding limits translation and cell growth." Proc Natl Acad Sci U S A **110**(42): 16754-16759.

Laslett, D. and B. Canback (2004). "ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences." Nucleic Acids Res **32**(1): 11-16.

Lee, C. P. and U. L. RajBhandary (1991). "Mutants of Escherichia coli initiator tRNA that suppress amber codons in Saccharomyces cerevisiae and are aminoacylated with tyrosine by yeast extracts." Proc Natl Acad Sci U S A **88**(24): 11378-11382.

Letunic, I. and P. Bork (2007). "Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation." Bioinformatics **23**(1): 127-128.

Letunic, I. and P. Bork (2011). "Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy." Nucleic Acids Res **39**(Web Server issue): W475-478.

Lowe, T. M. and S. R. Eddy (1997). "tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence." Nucleic Acids Res **25**(5): 955--964.

Marck, C. and H. Grosjean (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.

Meinnel, T. and S. Blanquet (1995). "Maturation of pre-tRNA(fMet) by Escherichia coli RNase P is specified by a guanosine of the 5'-flanking sequence." J Biol Chem **270**(26): 15908-15914.

NCBI Resource Coordinators (2014). "Database resources of the National Center for Biotechnology Information." Nucleic Acids Research **42**(Database issue): D7--17.

O'Farrell, P. H. (1978). "The suppression of defective translation by ppGpp and its role in the stringent response." Cell **14**(3): 545-557.

Pop, C., S. Rouskin, N. T. Ingolia, L. Han, E. M. Phizicky, J. S. Weissman and D. Koller (2014). "Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation." Mol Syst Biol **10**: 770.

Silva, F. J., E. Belda and S. E. Talens (2006). "Differential annotation of tRNA genes with anticodon CAT in bacterial genomes." Nucleic Acids Res **34**(20): 6015--6022.

Sprinzl, M., C. Horn, M. Brown, A. Ioudovitch and S. Steinberg (1998). "Compilation of tRNA sequences and sequences of tRNA genes." Nucleic Acids Research **26**(1): 148--153.

Suzuki, T. and K. Miyauchi (2010). "Discovery and characterization of tRNAIle lysidine synthetase (TilS)." FEBS letters **584**(2): 272--277.

Tåquist, H., Y. Cui and D. H. Ardell (2007). "TFAM 1.0: an online tRNA function classifier." Nucleic Acids Research **35**(Web Server issue): W350--353.

Vind, J., M. A. Sorensen, M. D. Rasmussen and S. Pedersen (1993). "Synthesis of proteins in Escherichia coli is limited by the concentration of free ribosomes. Expression from reporter genes does not always reflect functional mRNA levels." J Mol Biol **231**(3): 678-688.

Vortler, S. and M. Morl (2010). "tRNA-nucleotidyltransferases: highly unusual RNA polymerases with vital functions." FEBS Lett **584**(2): 297-302.

Wegener, K. M., A. K. Singh, J. M. Jacobs, T. Elvitigala, E. A. Welsh, N. Keren, M. A. Gritsenko, B. K. Ghosh, D. G. Camp, R. D. Smith and H. B. Pakrasi (2010). "Global proteomics reveal an atypical strategy for carbon/nitrogen assimilation by a cyanobacterium under diverse environmental perturbations." Molecular \& cellular proteomics: MCP **9**(12): 2678--2689.

Wegscheid, B. and R. K. Hartmann (2006). "The precursor tRNA 3'-CCA interaction with Escherichia coli RNase P RNA is essential for catalysis by RNase P in vivo." RNA **12**(12): 2135-2148.

Wegscheid, B. and R. K. Hartmann (2007). "In vivo and in vitro investigation of bacterial type B RNase P interaction with tRNA 3'-CCA." Nucleic Acids Res **35**(6): 2060-2073.

Wydau, S., G. van der Rest, C. Aubard, P. Plateau and S. Blanquet (2009). "Widespread Distribution of Cell Defense against d-Aminoacyl-tRNAs." The Journal of Biological Chemistry **284**(21): 14096--14104.

TABLES

		Silva TFAM Model			
		Met	Ini	kIle	Sum
Proteobacterial TFAM	Met	2751	261	173	710
	Ini	0	182	0	182
	kIle	11	0	977	988
	Sum	2762	443	271	991

Table 1: Reannotation of bacterial tRNAs with CAU anticodons in tRNAdb-CE v.8 using a custom model based on the analysis of (Silva et al., 2006)

FIGURE LEGENDS

Figure 1. Summarized frequencies at which elongator tRNA genes and initiator tRNA genes template 3′-CCA against average genome size in different prokaryotic genera. NCBI-Taxonomy based cladogram of prokaryotic genera in tRNAdb-CE v.8 showing average genome size (radial light blue bars) and average fractions at which elongator tRNA genes template 3′-CCA (blue circles) and initiator tRNA genes template 3′-CCA (red circles).

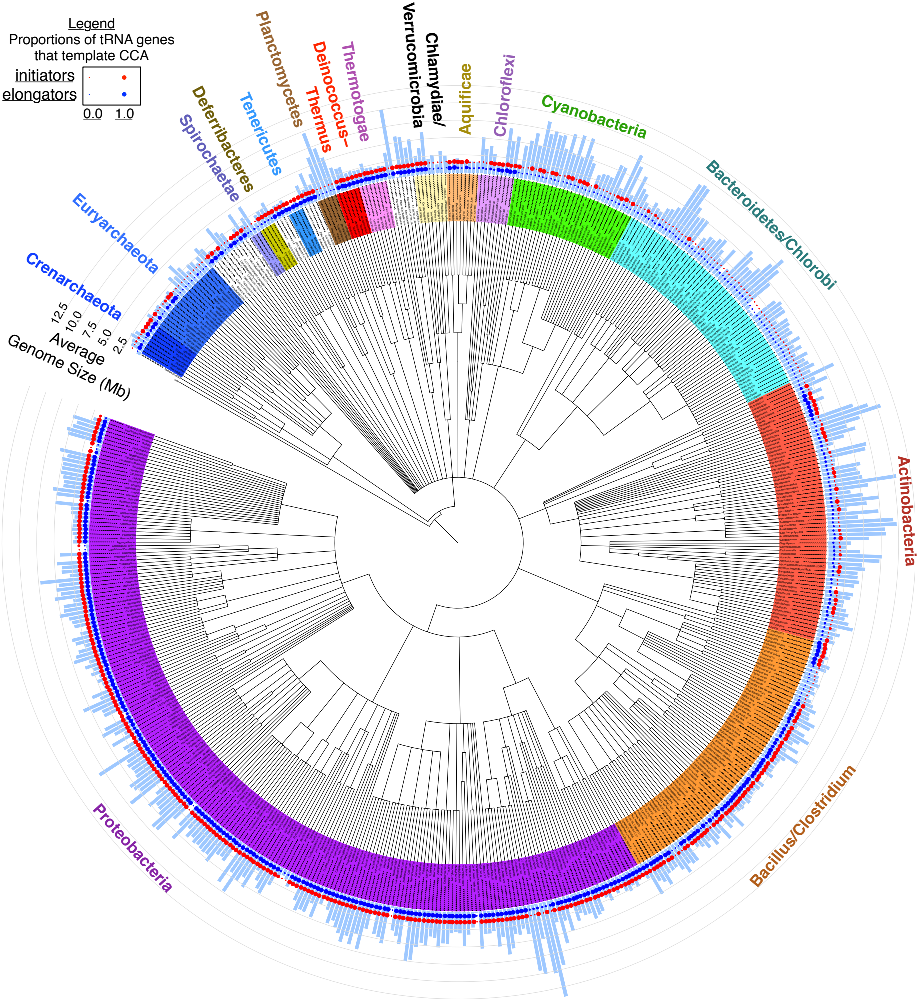
Figure 2. Summarized genome size and CCA frequency data in Bacterial clades broken out by tRNA functional class. Except for columns labelled "All," "SeC," and "Pyl," all columns of frequency data are sorted in decreasing order from left to right in frequency at which tRNA genes template CCA over all prokaryotic genomes that we sampled. Clades are defined as in NCBI Taxonomy. Column labels correspond to IUPAC three-letter amino acid charging identity except for "Ini" (initiators) and

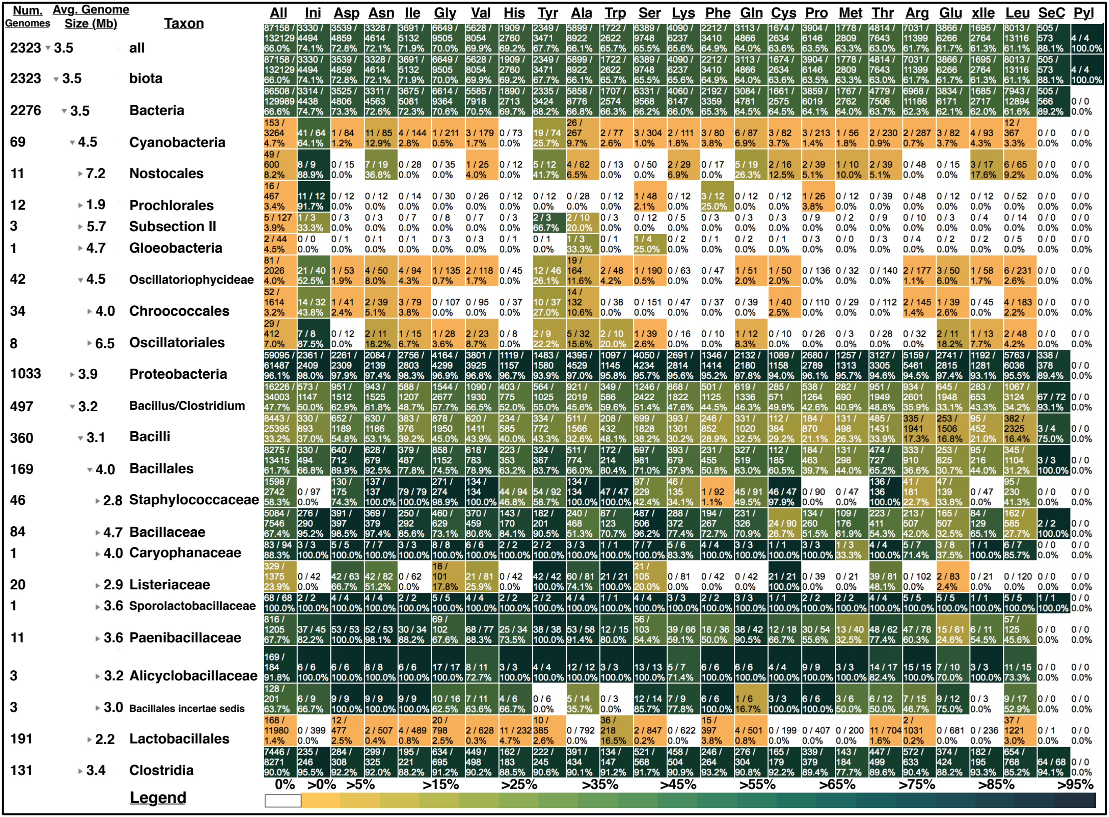
“xIle” (AUA-codon-reading isoleucine isoacceptors). “All” summarizes frequency data over all tRNA classes.

Figure 3. Summarized genome size and CCA frequency data in Archaeal clades broken out by tRNA functional class. Annotations are the same as in Figure 2.

SUPPLEMENTARY MATERIALS

1. **README.txt —** description of each file.
2. **tCE-Nov5-2014.3.fas —** reannotated tRNAdb-CE v.8 data
3. **prokaryotes_NCBI_102615.txt —** genome size data from NCBI.
4. **addresses_v4.txt.gz —** NCBI taxonomy data required by compute_heatmap.pl
5. **silva.coveam** — TFAM model used to reannotated tRNA functions.
6. **model_cca.pl —** Perl script for structural reannotation of 3′ends.
7. **compute_heatmap.pl —** Perl script to generate full data browser and iTol input.
8. **iTol_color_definitions.txt —** Color definitions for clades to generate Figure 1.
9. **phyloT.tre —** Phylogenetic tree of NCBI taxon IDs from NCBI Taxonomy.
10. **CCA-HEATMAP-TAX-BROWSER/Ardell_Hou.html** — web-browser based navigator of full dataset.
11. **Ardell_Hou_all_data.pdf —** static and searchable PDF with full dataset.

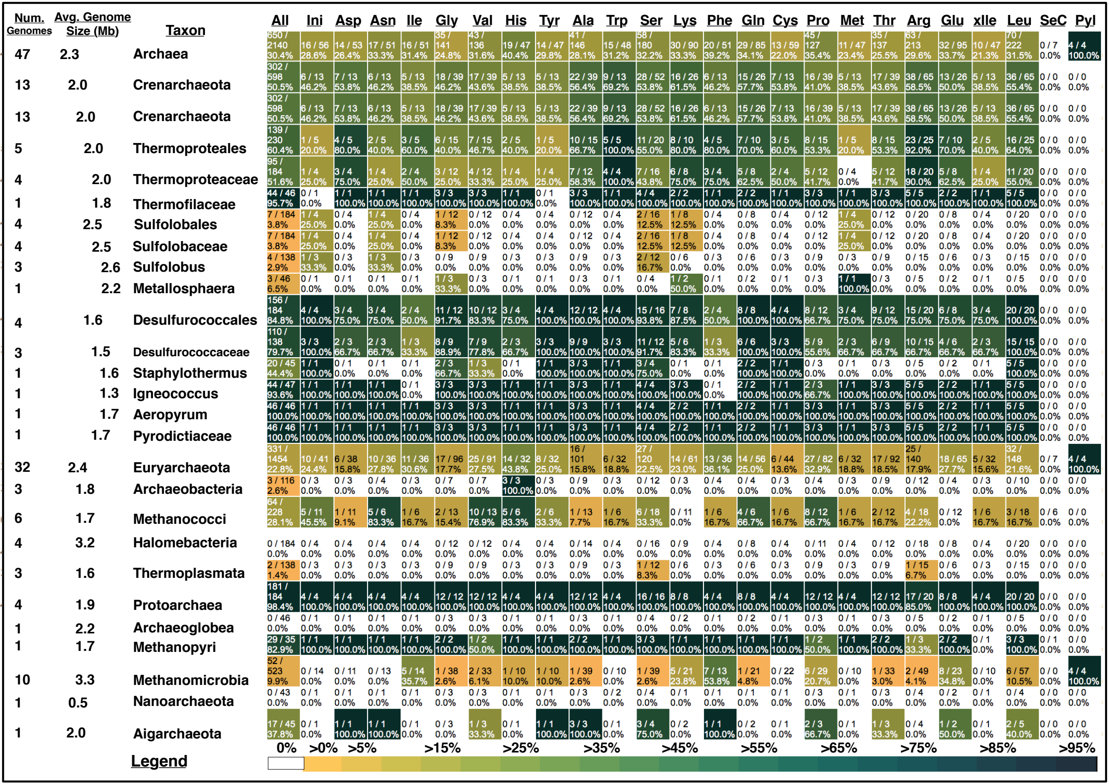

Legend
Proportions of tRNA genes that template CCA
initiators
elongators
0.0
1.0
Average Genome Size (Mb)
2.5
5.0
7.5
10.0
12.5
Crenarchaeota
Euryarchaeota
Spirochaetae
Deferribacteres
Tenericutes
Planctomycetes
Deinococcus–Thermus
Thermotogae
Verrucomicrobia
Chlamydiae/
Aquificae
Chloroflexi
Cyanobacteria
Bacteroidetes/Chlorobi
Actinobacteria
Bacillus/Clostridium
Proteobacteria



Num. Genomes	Avg. Genome Size (Mb)	Taxon	All	Ini	Asp	Asn	Ile	Gly	Val	His	Tyr	Ala	Trp	Ser	Lys	Phe	Gln	Cys	Pro	Met	Thr	Arg	Glu	xIle	Leu	SeC	Pyl
2323	▾ 3.5	all	87158 / 132129 66.0%	3330 / 4494 74.1%	3539 / 4859 72.8%	3328 / 4614 72.1%	3691 / 5132 71.9%	6649 / 9505 70.0%	5628 / 8054 69.9%	1909 / 2760 69.2%	2349 / 3471 67.7%	5899 / 8922 66.1%	1722 / 2622 65.7%	6389 / 9748 65.5%	4090 / 6237 65.6%	2212 / 3410 64.9%	3113 / 4866 64.0%	1674 / 2634 63.6%	3904 / 6146 63.5%	1778 / 2809 63.3%	4814 / 7643 63.0%	7031 / 11399 61.7%	3866 / 6266 61.7%	1695 / 2764 61.3%	8013 / 13116 61.1%	505 / 573 88.1%	4 / 4 100.0%
2323	▾ 3.5	biota	87158 / 132129 66.0%	3330 / 4494 74.1%	3539 / 4859 72.8%	3328 / 4614 72.1%	3691 / 5132 71.9%	6649 / 9505 70.0%	5628 / 8054 69.9%	1909 / 2760 69.2%	2349 / 3471 67.7%	5899 / 8922 66.1%	1722 / 2622 65.7%	6389 / 9748 65.5%	4090 / 6237 65.6%	2212 / 3410 64.9%	3113 / 4866 64.0%	1674 / 2634 63.6%	3904 / 6146 63.5%	1778 / 2809 63.3%	4814 / 7643 63.0%	7031 / 11399 61.7%	3866 / 6266 61.7%	1695 / 2764 61.3%	8013 / 13116 61.1%	505 / 573 88.1%	4 / 4 100.0%
2276	▾ 3.5	Bacteria	86508 / 129989 66.6%	3314 / 4438 74.7%	3525 / 4806 73.3%	3311 / 4563 72.6%	3675 / 5081 72.3%	6614 / 9364 70.6%	5585 / 7918 70.5%	1890 / 2713 69.7%	2335 / 3424 68.2%	5858 / 8776 66.8%	1707 / 2574 66.3%	6331 / 9568 66.2%	4060 / 6147 66.0%	2192 / 3359 65.3%	3084 / 4781 64.5%	1661 / 2575 64.5%	3859 / 6019 64.1%	1767 / 2762 64.0%	4779 / 7506 63.7%	6968 / 11186 62.3%	3834 / 6171 62.1%	1685 / 2717 62.0%	7943 / 12894 61.6%	505 / 566 89.2%	0 / 0 0.0%
69	▾ 4.5	Cyanobacteria	153 / 3264 4.7%	41 / 64 64.1%	1 / 84 1.2%	11 / 85 12.9%	4 / 144 2.8%	1 / 211 0.5%	3 / 179 1.7%	0 / 73 0.0%	19 / 74 25.7%	26 / 267 9.7%	2 / 77 2.6%	3 / 304 1.0%	2 / 111 1.8%	3 / 80 3.8%	6 / 87 6.9%	3 / 82 3.7%	3 / 213 1.4%	1 / 56 1.8%	2 / 230 0.9%	2 / 287 0.7%	3 / 82 3.7%	4 / 93 4.3%	12 / 367 3.3%	0 / 0 0.0%	0 / 0 0.0%
11	▸ 7.2	Nostocales	49 / 600 8.2%	8 / 9 88.9%	0 / 15 0.0%	7 / 19 36.8%	0 / 28 0.0%	0 / 35 0.0%	1 / 25 4.0%	0 / 12 0.0%	5 / 12 41.7%	4 / 62 6.5%	0 / 13 0.0%	0 / 50 0.0%	2 / 29 6.9%	0 / 17 0.0%	5 / 19 26.3%	2 / 16 12.5%	2 / 39 5.1%	1 / 10 10.0%	2 / 39 5.1%	0 / 48 0.0%	0 / 15 0.0%	3 / 17 17.6%	6 / 65 9.2%	0 / 0 0.0%	0 / 0 0.0%
12	▸ 1.9	Prochlorales	16 / 467 3.4%	11 / 12 91.7%	0 / 12 0.0%	0 / 12 0.0%	0 / 14 0.0%	0 / 30 0.0%	0 / 26 0.0%	0 / 12 0.0%	0 / 12 0.0%	0 / 28 0.0%	0 / 12 0.0%	1 / 48 2.1%	0 / 12 0.0%	3 / 12 25.0%	0 / 12 0.0%	0 / 12 0.0%	1 / 26 3.8%	0 / 12 0.0%	0 / 39 0.0%	0 / 48 0.0%	0 / 12 0.0%	0 / 12 0.0%	0 / 52 0.0%	0 / 0 0.0%	0 / 0 0.0%
3	▸ 5.7	Subsection II	5 / 127 3.9%	1 / 3 33.3%	0 / 3 0.0%	0 / 3 0.0%	0 / 7 0.0%	0 / 8 0.0%	0 / 7 0.0%	0 / 3 0.0%	2 / 3 66.7%	2 / 10 20.0%	0 / 3 0.0%	0 / 12 0.0%	0 / 5 0.0%	0 / 3 0.0%	0 / 3 0.0%	0 / 3 0.0%	0 / 9 0.0%	0 / 2 0.0%	0 / 9 0.0%	0 / 10 0.0%	0 / 3 0.0%	0 / 4 0.0%	0 / 14 0.0%	0 / 0 0.0%	0 / 0 0.0%
1	▸ 4.7	Gloeobacteria	2 / 44 4.5%	0 / 0 0.0%	0 / 1 0.0%	0 / 1 0.0%	0 / 1 0.0%	0 / 3 0.0%	0 / 3 0.0%	0 / 1 0.0%	0 / 1 0.0%	1 / 3 33.3%	0 / 1 0.0%	1 / 4 25.0%	0 / 2 0.0%	0 / 1 0.0%	0 / 2 0.0%	0 / 1 0.0%	0 / 3 0.0%	0 / 0 0.0%	0 / 3 0.0%	0 / 4 0.0%	0 / 2 0.0%	0 / 2 0.0%	0 / 5 0.0%	0 / 0 0.0%	0 / 0 0.0%
42	▾ 4.5	Oscillatoriophycideae	81 / 2026 4.0%	21 / 40 52.5%	1 / 53 1.9%	4 / 50 8.0%	4 / 94 4.3%	1 / 135 0.7%	2 / 118 1.7%	0 / 45 0.0%	12 / 46 26.1%	19 / 164 11.6%	2 / 48 4.2%	1 / 190 0.5%	0 / 63 0.0%	0 / 47 0.0%	1 / 51 2.0%	1 / 50 2.0%	0 / 136 0.0%	0 / 32 0.0%	0 / 140 0.0%	2 / 177 1.1%	3 / 50 6.0%	1 / 58 1.7%	6 / 231 2.6%	0 / 0 0.0%	0 / 0 0.0%
34	▸ 4.0	Chroococcales	52 / 1614 3.2%	14 / 32 43.8%	1 / 41 2.4%	2 / 39 5.1%	3 / 79 3.8%	0 / 107 0.0%	0 / 95 0.0%	0 / 37 0.0%	10 / 37 27.0%	14 / 132 10.6%	0 / 38 0.0%	0 / 151 0.0%	0 / 47 0.0%	0 / 37 0.0%	0 / 39 0.0%	1 / 40 2.5%	0 / 110 0.0%	0 / 29 0.0%	0 / 112 0.0%	2 / 145 1.4%	1 / 39 2.6%	0 / 45 0.0%	4 / 183 2.2%	0 / 0 0.0%	0 / 0 0.0%
8	▸ 6.5	Oscillatoriales	29 / 412 7.0%	7 / 8 87.5%	0 / 12 0.0%	2 / 11 18.2%	1 / 15 6.7%	1 / 28 3.6%	2 / 23 8.7%	0 / 8 0.0%	2 / 9 22.2%	5 / 32 15.6%	2 / 10 20.0%	1 / 39 2.6%	0 / 16 0.0%	0 / 10 0.0%	1 / 12 8.3%	0 / 10 0.0%	0 / 26 0.0%	0 / 3 0.0%	0 / 28 0.0%	0 / 32 0.0%	2 / 11 18.2%	1 / 13 7.7%	2 / 48 4.2%	0 / 0 0.0%	0 / 0 0.0%
1033	▸ 3.9	Proteobacteria	59095 / 61487 96.1%	2361 / 2409 98.0%	2261 / 2309 97.9%	2084 / 2139 97.4%	2756 / 2803 98.3%	4164 / 4299 96.9%	3801 / 3925 96.8%	1119 / 1157 96.7%	1483 / 1580 93.9%	4395 / 4529 97.0%	1097 / 1145 95.8%	4050 / 4234 95.7%	2691 / 2814 95.6%	1346 / 1414 95.2%	2132 / 2180 97.8%	1089 / 1158 94.0%	2680 / 2789 96.1%	1257 / 1313 95.7%	3127 / 3305 94.6%	5159 / 5461 94.5%	2741 / 2815 97.4%	1192 / 1281 93.1%	5763 / 6036 95.5%	338 / 378 89.4%	0 / 0 0.0%
497	▾ 3.2	Bacillus/Clostridium	16226 / 34003 47.7%	573 / 1147 50.0%	951 / 1512 62.9%	943 / 1525 61.8%	588 / 1207 48.7%	1544 / 2677 57.7%	1090 / 1930 56.5%	403 / 775 52.0%	564 / 1025 55.0%	921 / 2019 45.6%	349 / 586 59.6%	1246 / 2422 51.4%	868 / 1822 47.6%	501 / 1125 44.5%	619 / 1336 46.3%	285 / 571 49.9%	538 / 1264 42.6%	282 / 690 40.9%	951 / 1949 48.8%	934 / 2601 35.9%	645 / 1948 33.1%	283 / 653 43.3%	1067 / 3124 34.2%	67 / 72 93.1%	0 / 0 0.0%
360	▾ 3.1	Bacilli	8443 / 25395 33.2%	330 / 893 37.0%	652 / 1189 54.8%	630 / 1186 53.1%	383 / 976 39.2%	878 / 1950 45.0%	620 / 1411 43.9%	234 / 585 40.0%	334 / 772 43.3%	511 / 1566 32.6%	208 / 432 48.1%	699 / 1828 38.2%	393 / 1301 30.2%	246 / 852 28.9%	331 / 1020 32.5%	112 / 384 29.2%	184 / 870 21.1%	131 / 498 26.3%	485 / 1431 33.9%	335 / 1941 17.3%	253 / 1506 16.8%	95 / 452 21.0%	382 / 2325 16.4%	3 / 4 75.0%	0 / 0 0.0%
169	▾ 4.0	Bacillales	8275 / 13415 61.7%	330 / 494 66.8%	640 / 712 89.9%	628 / 679 92.5%	379 / 487 77.8%	858 / 1152 74.5%	618 / 783 78.9%	223 / 353 63.2%	324 / 387 83.7%	511 / 774 66.0%	172 / 214 80.4%	697 / 981 71.0%	393 / 679 57.9%	231 / 455 50.8%	327 / 519 63.0%	112 / 185 60.5%	184 / 463 39.7%	131 / 298 44.0%	474 / 727 65.2%	333 / 910 36.6%	253 / 825 30.7%	95 / 216 44.0%	345 / 1104 31.2%	3 / 3 100.0%	0 / 0 0.0%
46	▸ 2.8	Staphylococcaceae	1598 / 2742 58.3%	0 / 97 0.0%	130 / 175 74.3%	137 / 137 100.0%	79 / 79 100.0%	271 / 274 98.9%	134 / 134 100.0%	44 / 94 46.8%	54 / 92 58.7%	134 / 134 100.0%	47 / 47 100.0%	97 / 229 42.4%	46 / 135 34.1%	1 / 92 1.1%	45 / 91 49.5%	46 / 47 97.9%	0 / 90 0.0%	0 / 47 0.0%	136 / 136 100.0%	41 / 181 22.7%	47 / 139 33.8%	0 / 47 0.0%	95 / 230 41.3%	0 / 0 0.0%	0 / 0 0.0%
84	▸ 4.7	Bacillaceae	5084 / 7546 67.4%	276 / 290 95.2%	391 / 397 98.5%	369 / 379 97.4%	250 / 292 85.6%	460 / 629 73.1%	370 / 459 80.6%	143 / 170 84.1%	182 / 201 90.5%	240 / 468 51.3%	87 / 123 70.7%	487 / 506 96.2%	288 / 372 77.4%	194 / 267 72.7%	231 / 326 70.9%	24 / 90 26.7%	134 / 260 51.5%	109 / 176 61.9%	223 / 411 54.3%	213 / 507 42.0%	165 / 507 32.5%	84 / 129 65.1%	162 / 585 27.7%	2 / 2 100.0%	0 / 0 0.0%
1	▸ 4.0	Caryophanaceae	83 / 94 88.3%	3 / 3 100.0%	5 / 5 100.0%	7 / 7 100.0%	3 / 3 100.0%	8 / 8 100.0%	6 / 6 100.0%	2 / 2 100.0%	2 / 2 100.0%	3 / 3 100.0%	1 / 1 100.0%	7 / 7 100.0%	5 / 6 83.3%	4 / 4 100.0%	3 / 3 100.0%	1 / 1 100.0%	3 / 3 100.0%	1 / 3 33.3%	4 / 4 100.0%	5 / 7 71.4%	3 / 8 37.5%	1 / 1 100.0%	6 / 7 85.7%	0 / 0 0.0%	0 / 0 0.0%
20	▸ 2.9	Listeriaceae	329 / 1375 23.9%	0 / 42 0.0%	42 / 63 66.7%	42 / 82 51.2%	0 / 62 0.0%	18 / 101 17.8%	21 / 81 25.9%	0 / 42 0.0%	42 / 42 100.0%	60 / 81 74.1%	21 / 21 100.0%	21 / 105 20.0%	0 / 81 0.0%	0 / 42 0.0%	0 / 42 0.0%	21 / 21 100.0%	0 / 39 0.0%	0 / 21 0.0%	39 / 81 48.1%	0 / 102 0.0%	2 / 83 2.4%	0 / 21 0.0%	0 / 120 0.0%	0 / 0 0.0%	0 / 0 0.0%
1	▸ 3.6	Sporolactobacillaceae	68 / 68 100.0%	2 / 2 100.0%	4 / 4 100.0%	4 / 4 100.0%	2 / 2 100.0%	5 / 5 100.0%	4 / 4 100.0%	2 / 2 100.0%	2 / 2 100.0%	4 / 4 100.0%	1 / 1 100.0%	4 / 4 100.0%	3 / 3 100.0%	2 / 2 100.0%	3 / 3 100.0%	1 / 1 100.0%	2 / 2 100.0%	2 / 2 100.0%	4 / 4 100.0%	5 / 5 100.0%	5 / 5 100.0%	1 / 1 100.0%	5 / 5 100.0%	1 / 1 100.0%	0 / 0 0.0%
11	▸ 3.6	Paenibacillaceae	816 / 1205 67.7%	37 / 45 82.2%	53 / 53 100.0%	52 / 53 98.1%	30 / 34 88.2%	69 / 102 67.6%	68 / 77 88.3%	25 / 34 73.5%	38 / 38 100.0%	53 / 58 91.4%	12 / 15 80.0%	56 / 103 54.4%	39 / 66 59.1%	18 / 36 50.0%	38 / 42 90.5%	12 / 18 66.7%	30 / 54 55.6%	13 / 40 32.5%	48 / 62 77.4%	47 / 78 60.3%	15 / 61 24.6%	6 / 11 54.5%	57 / 125 45.6%	0 / 0 0.0%	0 / 0 0.0%
3	▸ 3.2	Alicyclobacillaceae	169 / 184 91.8%	6 / 6 100.0%	6 / 6 100.0%	8 / 8 100.0%	6 / 6 100.0%	17 / 17 100.0%	8 / 11 72.7%	3 / 3 100.0%	4 / 4 100.0%	12 / 12 100.0%	3 / 3 100.0%	13 / 13 100.0%	5 / 7 71.4%	6 / 6 100.0%	6 / 6 100.0%	4 / 4 100.0%	9 / 9 100.0%	3 / 3 100.0%	14 / 17 82.4%	15 / 15 100.0%	7 / 10 70.0%	3 / 3 100.0%	11 / 15 73.3%	0 / 0 0.0%	0 / 0 0.0%
3	▸ 3.0	Bacillales incertae sedis	128 / 201 63.7%	6 / 9 66.7%	9 / 9 100.0%	9 / 9 100.0%	9 / 9 100.0%	10 / 16 62.5%	7 / 11 63.6%	4 / 6 66.7%	0 / 6 0.0%	5 / 14 35.7%	0 / 3 0.0%	12 / 14 85.7%	7 / 9 77.8%	6 / 6 100.0%	1 / 6 16.7%	3 / 3 100.0%	6 / 6 100.0%	3 / 6 50.0%	6 / 12 50.0%	7 / 15 46.7%	9 / 12 75.0%	0 / 3 0.0%	9 / 17 52.9%	0 / 0 0.0%	0 / 0 0.0%
191	▸ 2.2	Lactobacillales	168 / 11980 1.4%	0 / 399 0.0%	12 / 477 2.5%	2 / 507 0.4%	4 / 489 0.8%	20 / 798 2.5%	2 / 628 0.3%	11 / 232 4.7%	10 / 385 2.6%	0 / 792 0.0%	36 / 218 16.5%	2 / 847 0.2%	0 / 622 0.0%	15 / 397 3.8%	4 / 501 0.8%	0 / 199 0.0%	0 / 407 0.0%	0 / 200 0.0%	11 / 704 1.6%	2 / 1031 0.2%	0 / 681 0.0%	0 / 236 0.0%	37 / 1221 3.0%	0 / 1 0.0%	0 / 0 0.0%
131	▸ 3.4	Clostridia	7446 / 8271 90.0%	235 / 246 95.5%	284 / 308 92.2%	299 / 325 92.0%	195 / 221 88.2%	634 / 695 91.2%	449 / 498 90.2%	162 / 183 88.5%	222 / 245 90.6%	391 / 434 90.1%	134 / 147 91.2%	521 / 568 91.7%	458 / 504 90.9%	246 / 264 93.2%	276 / 304 90.8%	165 / 179 92.2%	339 / 379 89.4%	143 / 184 77.7%	447 / 499 89.6%	572 / 633 90.4%	374 / 424 88.2%	182 / 195 93.3%	654 / 768 85.2%	64 / 68 94.1%	0 / 0 0.0%

Legend

0% >0% >5% >15% >25% >35% >45% >55% >65% >75% >85% >95%



Num. Genomes	Avg. Genome Size (Mb)	Taxon	All	Ini	Asp	Asn	Ile	Gly	Val	His	Tyr	Ala	Trp	Ser	Lys	Phe	Gln	Cys	Pro	Met	Thr	Arg	Glu	xIle	Leu	SeC	Pyl
47	2.3	Archaea	650 / 2140 30.4%	16 / 56 28.6%	14 / 53 26.4%	17 / 51 33.3%	16 / 51 31.4%	35 / 141 24.8%	43 / 136 31.6%	19 / 47 40.4%	14 / 47 29.8%	41 / 146 28.1%	15 / 48 31.2%	58 / 180 32.2%	30 / 90 33.3%	20 / 51 39.2%	29 / 85 34.1%	13 / 59 22.0%	45 / 127 35.4%	11 / 47 23.4%	35 / 137 25.5%	63 / 213 29.6%	32 / 95 33.7%	10 / 47 21.3%	70 / 222 31.5%	0 / 7 0.0%	4 / 4 100.0%
13	2.0	Crenarchaeota	302 / 598 50.5%	6 / 13 46.2%	7 / 13 53.8%	6 / 13 46.2%	5 / 13 38.5%	18 / 39 46.2%	17 / 39 43.6%	5 / 13 38.5%	5 / 13 38.5%	22 / 39 56.4%	9 / 13 69.2%	28 / 52 53.8%	16 / 26 61.5%	6 / 13 46.2%	15 / 26 57.7%	7 / 13 53.8%	16 / 39 41.0%	5 / 13 38.5%	17 / 39 43.6%	38 / 65 58.5%	13 / 26 50.0%	5 / 13 38.5%	36 / 65 55.4%	0 / 0 0.0%	0 / 0 0.0%
13	2.0	Crenarchaeota	302 / 598 50.5%	6 / 13 46.2%	7 / 13 53.8%	6 / 13 46.2%	5 / 13 38.5%	18 / 39 46.2%	17 / 39 43.6%	5 / 13 38.5%	5 / 13 38.5%	22 / 39 56.4%	9 / 13 69.2%	28 / 52 53.8%	16 / 26 61.5%	6 / 13 46.2%	15 / 26 57.7%	7 / 13 53.8%	16 / 39 41.0%	5 / 13 38.5%	17 / 39 43.6%	38 / 65 58.5%	13 / 26 50.0%	5 / 13 38.5%	36 / 65 55.4%	0 / 0 0.0%	0 / 0 0.0%
5	2.0	Thermoproteales	139 / 230 60.4%	1 / 5 20.0%	4 / 5 80.0%	2 / 5 40.0%	3 / 5 60.0%	6 / 15 40.0%	7 / 15 46.7%	2 / 5 40.0%	1 / 5 20.0%	10 / 15 66.7%	5 / 5 100.0%	11 / 20 55.0%	8 / 10 80.0%	4 / 5 80.0%	7 / 10 70.0%	3 / 5 60.0%	8 / 15 53.3%	1 / 5 20.0%	8 / 15 53.3%	23 / 25 92.0%	7 / 10 70.0%	2 / 5 40.0%	16 / 25 64.0%	0 / 0 0.0%	0 / 0 0.0%
4	2.0	Thermoproteaceae	95 / 184 51.6%	1 / 4 25.0%	3 / 4 75.0%	1 / 4 25.0%	2 / 4 50.0%	3 / 12 25.0%	4 / 12 33.3%	1 / 4 25.0%	1 / 4 25.0%	7 / 12 58.3%	4 / 4 100.0%	7 / 16 43.8%	6 / 8 75.0%	3 / 4 75.0%	5 / 8 62.5%	2 / 4 50.0%	5 / 12 41.7%	0 / 4 0.0%	5 / 12 41.7%	18 / 20 90.0%	5 / 8 62.5%	1 / 4 25.0%	11 / 20 55.0%	0 / 0 0.0%	0 / 0 0.0%
1	1.8	Thermofilaceae	44 / 46 95.7%	0 / 1 0.0%	1 / 1 100.0%	1 / 1 100.0%	1 / 1 100.0%	3 / 3 100.0%	3 / 3 100.0%	1 / 1 100.0%	0 / 1 0.0%	3 / 3 100.0%	1 / 1 100.0%	4 / 4 100.0%	2 / 2 100.0%	1 / 1 100.0%	2 / 2 100.0%	1 / 1 100.0%	3 / 3 100.0%	1 / 1 100.0%	3 / 3 100.0%	5 / 5 100.0%	2 / 2 100.0%	1 / 1 100.0%	5 / 5 100.0%	0 / 0 0.0%	0 / 0 0.0%
4	2.5	Sulfolobales	7 / 184 3.8%	1 / 4 25.0%	0 / 4 0.0%	1 / 4 25.0%	0 / 4 0.0%	1 / 12 8.3%	0 / 12 0.0%	0 / 4 0.0%	0 / 4 0.0%	0 / 12 0.0%	0 / 4 0.0%	2 / 16 12.5%	1 / 8 12.5%	0 / 4 0.0%	0 / 8 0.0%	0 / 4 0.0%	0 / 12 0.0%	1 / 4 25.0%	0 / 12 0.0%	0 / 20 0.0%	0 / 8 0.0%	0 / 4 0.0%	0 / 20 0.0%	0 / 0 0.0%	0 / 0 0.0%
4	2.5	Sulfolobaceae	7 / 184 3.8%	1 / 4 25.0%	0 / 4 0.0%	1 / 4 25.0%	0 / 4 0.0%	1 / 12 8.3%	0 / 12 0.0%	0 / 4 0.0%	0 / 4 0.0%	0 / 12 0.0%	0 / 4 0.0%	2 / 16 12.5%	1 / 8 12.5%	0 / 4 0.0%	0 / 8 0.0%	0 / 4 0.0%	0 / 12 0.0%	1 / 4 25.0%	0 / 12 0.0%	0 / 20 0.0%	0 / 8 0.0%	0 / 4 0.0%	0 / 20 0.0%	0 / 0 0.0%	0 / 0 0.0%
3	2.6	Sulfolobus	4 / 138 2.9%	1 / 3 33.3%	0 / 3 0.0%	1 / 3 33.3%	0 / 3 0.0%	0 / 9 0.0%	0 / 9 0.0%	0 / 3 0.0%	0 / 3 0.0%	0 / 9 0.0%	0 / 3 0.0%	2 / 12 16.7%	0 / 6 0.0%	0 / 3 0.0%	0 / 6 0.0%	0 / 3 0.0%	0 / 9 0.0%	0 / 3 0.0%	0 / 9 0.0%	0 / 15 0.0%	0 / 6 0.0%	0 / 3 0.0%	0 / 15 0.0%	0 / 0 0.0%	0 / 0 0.0%
1	2.2	Metallosphaera	3 / 46 6.5%	0 / 1 0.0%	0 / 1 0.0%	0 / 1 0.0%	0 / 1 0.0%	1 / 3 33.3%	0 / 3 0.0%	0 / 1 0.0%	0 / 1 0.0%	0 / 3 0.0%	0 / 1 0.0%	0 / 4 0.0%	1 / 2 50.0%	0 / 1 0.0%	0 / 2 0.0%	0 / 1 0.0%	0 / 3 0.0%	1 / 1 100.0%	0 / 3 0.0%	0 / 5 0.0%	0 / 2 0.0%	0 / 1 0.0%	0 / 5 0.0%	0 / 0 0.0%	0 / 0 0.0%
4	1.6	Desulfurococcales	156 / 184 84.8%	4 / 4 100.0%	3 / 4 75.0%	3 / 4 75.0%	2 / 4 50.0%	11 / 12 91.7%	10 / 12 83.3%	3 / 4 75.0%	4 / 4 100.0%	12 / 12 100.0%	4 / 4 100.0%	15 / 16 93.8%	7 / 8 87.5%	2 / 4 50.0%	8 / 8 100.0%	4 / 4 100.0%	8 / 12 66.7%	3 / 4 75.0%	9 / 12 75.0%	15 / 20 75.0%	6 / 8 75.0%	3 / 4 75.0%	20 / 20 100.0%	0 / 0 0.0%	0 / 0 0.0%
3	1.5	Desulfurococcaceae	110 / 138 79.7%	3 / 3 100.0%	2 / 3 66.7%	2 / 3 66.7%	1 / 3 33.3%	8 / 9 88.9%	7 / 9 77.8%	2 / 3 66.7%	3 / 3 100.0%	9 / 9 100.0%	3 / 3 100.0%	11 / 12 91.7%	5 / 6 83.3%	1 / 3 33.3%	6 / 6 100.0%	3 / 3 100.0%	5 / 9 55.6%	2 / 3 66.7%	6 / 9 66.7%	10 / 15 66.7%	4 / 6 66.7%	2 / 3 66.7%	15 / 15 100.0%	0 / 0 0.0%	0 / 0 0.0%
1	1.6	Staphylothermus	20 / 45 44.4%	1 / 1 100.0%	0 / 1 0.0%	0 / 1 0.0%	0 / 1 0.0%	2 / 3 66.7%	1 / 3 33.3%	0 / 1 0.0%	1 / 1 100.0%	3 / 3 100.0%	1 / 1 100.0%	3 / 4 75.0%	0 / 1 0.0%	0 / 1 0.0%	2 / 2 100.0%	1 / 1 100.0%	0 / 3 0.0%	0 / 1 0.0%	0 / 3 0.0%	0 / 5 0.0%	0 / 2 0.0%	0 / 1 0.0%	5 / 5 100.0%	0 / 0 0.0%	0 / 0 0.0%
1	1.3	Igneococcus	44 / 47 93.6%	1 / 1 100.0%	1 / 1 100.0%	1 / 1 100.0%	0 / 1 0.0%	3 / 3 100.0%	3 / 3 100.0%	1 / 1 100.0%	1 / 1 100.0%	3 / 3 100.0%	1 / 1 100.0%	4 / 4 100.0%	3 / 3 100.0%	0 / 1 0.0%	2 / 2 100.0%	1 / 1 100.0%	2 / 3 66.7%	1 / 1 100.0%	3 / 3 100.0%	5 / 5 100.0%	2 / 2 100.0%	1 / 1 100.0%	5 / 5 100.0%	0 / 0 0.0%	0 / 0 0.0%
1	1.7	Aeropyrum	46 / 46 100.0%	1 / 1 100.0%	1 / 1 100.0%	1 / 1 100.0%	1 / 1 100.0%	3 / 3 100.0%	3 / 3 100.0%	1 / 1 100.0%	1 / 1 100.0%	3 / 3 100.0%	1 / 1 100.0%	4 / 4 100.0%	2 / 2 100.0%	1 / 1 100.0%	2 / 2 100.0%	1 / 1 100.0%	3 / 3 100.0%	1 / 1 100.0%	3 / 3 100.0%	5 / 5 100.0%	2 / 2 100.0%	1 / 1 100.0%	5 / 5 100.0%	0 / 0 0.0%	0 / 0 0.0%
1	1.7	Pyrodictiaceae	46 / 46 100.0%	1 / 1 100.0%	1 / 1 100.0%	1 / 1 100.0%	1 / 1 100.0%	3 / 3 100.0%	3 / 3 100.0%	1 / 1 100.0%	1 / 1 100.0%	3 / 3 100.0%	1 / 1 100.0%	4 / 4 100.0%	2 / 2 100.0%	1 / 1 100.0%	2 / 2 100.0%	1 / 1 100.0%	3 / 3 100.0%	1 / 1 100.0%	3 / 3 100.0%	5 / 5 100.0%	2 / 2 100.0%	1 / 1 100.0%	5 / 5 100.0%	0 / 0 0.0%	0 / 0 0.0%
32	2.4	Euryarchaeota	331 / 1454 22.8%	10 / 41 24.4%	6 / 38 15.8%	10 / 36 27.8%	11 / 36 30.6%	17 / 96 17.7%	25 / 91 27.5%	14 / 32 43.8%	8 / 32 25.0%	16 / 101 15.8%	6 / 32 18.8%	27 / 120 22.5%	14 / 61 23.0%	13 / 36 36.1%	14 / 56 25.0%	6 / 44 13.6%	27 / 82 32.9%	6 / 32 18.8%	17 / 92 18.5%	25 / 140 17.9%	18 / 65 27.7%	5 / 32 15.6%	32 / 148 21.6%	0 / 7 0.0%	4 / 4 100.0%
3	1.8	Archaeobacteria	3 / 116 2.6%	0 / 3 0.0%	0 / 3 0.0%	0 / 4 0.0%	0 / 3 0.0%	0 / 7 0.0%	0 / 7 0.0%	3 / 3 100.0%	0 / 3 0.0%	0 / 9 0.0%	0 / 3 0.0%	0 / 12 0.0%	0 / 3 0.0%	0 / 4 0.0%	0 / 4 0.0%	0 / 3 0.0%	0 / 4 0.0%	0 / 3 0.0%	0 / 9 0.0%	0 / 12 0.0%	0 / 4 0.0%	0 / 3 0.0%	0 / 10 0.0%	0 / 0 0.0%	0 / 0 0.0%
6	1.7	Methanococci	64 / 228 28.1%	5 / 11 45.5%	1 / 11 9.1%	5 / 6 83.3%	1 / 6 16.7%	2 / 13 15.4%	10 / 13 76.9%	5 / 6 83.3%	2 / 6 33.3%	1 / 13 7.7%	1 / 6 16.7%	6 / 18 33.3%	0 / 11 0.0%	1 / 6 16.7%	4 / 6 66.7%	1 / 6 16.7%	8 / 12 66.7%	1 / 6 16.7%	2 / 12 16.7%	4 / 18 22.2%	0 / 12 0.0%	1 / 6 16.7%	3 / 18 16.7%	0 / 6 0.0%	0 / 0 0.0%
4	3.2	Halomebacteria	0 / 184 0.0%	0 / 4 0.0%	0 / 4 0.0%	0 / 4 0.0%	0 / 4 0.0%	0 / 12 0.0%	0 / 12 0.0%	0 / 4 0.0%	0 / 4 0.0%	0 / 14 0.0%	0 / 4 0.0%	0 / 16 0.0%	0 / 9 0.0%	0 / 4 0.0%	0 / 8 0.0%	0 / 4 0.0%	0 / 11 0.0%	0 / 4 0.0%	0 / 12 0.0%	0 / 18 0.0%	0 / 8 0.0%	0 / 4 0.0%	0 / 20 0.0%	0 / 0 0.0%	0 / 0 0.0%
3	1.6	Thermoplasmata	2 / 138 1.4%	0 / 3 0.0%	0 / 3 0.0%	0 / 3 0.0%	0 / 3 0.0%	0 / 9 0.0%	0 / 9 0.0%	0 / 3 0.0%	0 / 3 0.0%	0 / 9 0.0%	0 / 3 0.0%	1 / 12 8.3%	0 / 6 0.0%	0 / 3 0.0%	0 / 6 0.0%	0 / 3 0.0%	0 / 9 0.0%	0 / 3 0.0%	0 / 9 0.0%	1 / 15 6.7%	0 / 6 0.0%	0 / 3 0.0%	0 / 15 0.0%	0 / 0 0.0%	0 / 0 0.0%
4	1.9	Protoarchaea	181 / 184 98.4%	4 / 4 100.0%	4 / 4 100.0%	4 / 4 100.0%	4 / 4 100.0%	12 / 12 100.0%	12 / 12 100.0%	4 / 4 100.0%	4 / 4 100.0%	12 / 12 100.0%	4 / 4 100.0%	16 / 16 100.0%	8 / 8 100.0%	4 / 4 100.0%	8 / 8 100.0%	4 / 4 100.0%	12 / 12 100.0%	4 / 4 100.0%	12 / 12 100.0%	17 / 20 85.0%	8 / 8 100.0%	4 / 4 100.0%	20 / 20 100.0%	0 / 0 0.0%	0 / 0 0.0%
1	2.2	Archaeoglobea	0 / 46 0.0%	0 / 1 0.0%	0 / 1 0.0%	0 / 1 0.0%	0 / 1 0.0%	0 / 3 0.0%	0 / 3 0.0%	0 / 1 0.0%	0 / 1 0.0%	0 / 3 0.0%	0 / 1 0.0%	0 / 4 0.0%	0 / 2 0.0%	0 / 1 0.0%	0 / 2 0.0%	0 / 1 0.0%	0 / 3 0.0%	0 / 1 0.0%	0 / 3 0.0%	0 / 5 0.0%	0 / 2 0.0%	0 / 1 0.0%	0 / 5 0.0%	0 / 0 0.0%	0 / 0 0.0%
1	1.7	Methanopyri	29 / 35 82.9%	1 / 1 100.0%	1 / 1 100.0%	1 / 1 100.0%	1 / 1 100.0%	2 / 2 100.0%	1 / 2 50.0%	1 / 1 100.0%	1 / 1 100.0%	2 / 2 100.0%	1 / 1 100.0%	3 / 3 100.0%	1 / 1 100.0%	1 / 1 100.0%	1 / 1 100.0%	1 / 1 100.0%	1 / 2 50.0%	1 / 1 100.0%	2 / 2 100.0%	1 / 3 33.3%	2 / 2 100.0%	0 / 1 0.0%	3 / 3 100.0%	0 / 1 0.0%	0 / 0 0.0%
10	3.3	Methanomicrobia	52 / 523 9.9%	0 / 14 0.0%	0 / 11 0.0%	0 / 13 0.0%	5 / 14 35.7%	1 / 38 2.6%	2 / 33 6.1%	1 / 10 10.0%	1 / 10 10.0%	1 / 39 2.6%	0 / 10 0.0%	1 / 39 2.6%	5 / 21 23.8%	7 / 13 53.8%	1 / 21 4.8%	0 / 22 0.0%	6 / 29 20.7%	0 / 10 0.0%	1 / 33 3.0%	2 / 49 4.1%	8 / 23 34.8%	0 / 10 0.0%	6 / 57 10.5%	0 / 0 0.0%	4 / 4 100.0%
1	0.5	Nanoarchaeota	0 / 43 0.0%	0 / 1 0.0%	0 / 1 0.0%	0 / 1 0.0%	0 / 1 0.0%	0 / 3 0.0%	0 / 3 0.0%	0 / 1 0.0%	0 / 1 0.0%	0 / 3 0.0%	0 / 2 0.0%	0 / 4 0.0%	0 / 1 0.0%	0 / 1 0.0%	0 / 1 0.0%	0 / 1 0.0%	0 / 3 0.0%	0 / 1 0.0%	0 / 3 0.0%	0 / 4 0.0%	0 / 2 0.0%	0 / 1 0.0%	0 / 4 0.0%	0 / 0 0.0%	0 / 0 0.0%
1	2.0	Aigarchaeota	17 / 45 37.8%	0 / 1 0.0%	1 / 1 100.0%	1 / 1 100.0%	0 / 1 0.0%	0 / 3 0.0%	1 / 3 33.3%	0 / 1 0.0%	1 / 1 100.0%	3 / 3 100.0%	0 / 1 0.0%	3 / 4 75.0%	0 / 2 0.0%	1 / 1 100.0%	0 / 2 0.0%	0 / 1 0.0%	2 / 3 66.7%	0 / 1 0.0%	1 / 3 33.3%	0 / 4 0.0%	1 / 2 50.0%	0 / 1 0.0%	2 / 5 40.0%	0 / 0 0.0%	0 / 0 0.0%

Legend

0% >0% >5% >15% >25% >35% >45% >55% >65% >75% >85% >95%