Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genomic organization of tRNA genes and distribution of tRNA charging and modification enzymes across Bacteria, Archaea and major eukaryotic lineages

Davis W. Cheng

dcheng@rockybiotech.org

Rocky Institute of Biotechnology

Research Article

Keywords: aminoacyl-tRNA synthetase, tRNA gene repertoire, wobble modification, indirect aminoacylation, tRNA introns, comparative genomics

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11631560/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Genomic organization of tRNA genes and distribution of tRNA charging and modification enzymes across Bacteria, Archaea and major eukaryotic lineages

Davis W. Cheng

Rocky Institute of Biotechnology, Clovis, CA 93619, USA

ORCID: 0009-0001-1306-8923

Corresponding author: Davis W. Cheng, dcheng@rockybiotech.org

Author contributions: D.W.C. conceived and designed the study, directed the analyses, interpreted the results and wrote and approved the manuscript.

Acknowledgments: None.

Keywords: aminoacyl-tRNA synthetase; tRNA gene repertoire; wobble modification; indirect aminoacylation; tRNA introns; comparative genomics

Abstract

Transfer RNA (tRNA) genes and the enzymes that charge and modify tRNAs are usually studied separately, and enzyme distributions are rarely reported with measured detection coverage. We analyzed, in one reporting framework, high-confidence tRNA gene predictions for 2,125 species and 31 aminoacyl-tRNA synthetase and modification-enzyme families in 18,612 UniProt reference proteomes, with per-phylum breakdowns. Three findings emerged. First, similar anticodon richness coexisted with large differences in gene copy number and anticodon composition: median anticodon richness was 40–48 in every group, whereas median tRNA gene number ranged from 47 in Archaea to 708 in flowering plants; eukaryotic genomes carried

adenosine-34 (A34) anticodons in eight families, prokaryotic genomes in at most one. Second, charging-enzyme families varied within domains, whereas C34-modification enzymes showed strong domain associations: glutaminyl-, class I lysyl- and bacterial-type glycyl-tRNA synthetases each ranged from 0% to nearly 100% among bacterial phyla, while the lysidine and agmatidine synthetases were detected in 98.3% of bacterial and 96.7% of archaeal proteomes, respectively, and in none of the other domain. At least one alternative charging-family configuration was detected in 98.1–99.4% of proteomes in each comparison. Third, detection coverage varied among markers and lineages: 99.4% and 98.4% overall for the alanyl- and methionyl-tRNA synthetase markers, but 64.6% for the histidyl-tRNA synthetase marker in fungi and 70.2% for the tyrosyl-tRNA synthetase marker in protists. These findings distinguish variation in tRNA copy number from variation in anticodon composition, resolve taxonomic differences in charging-enzyme distributions, and provide lineage-specific reference data for interpreting protein-family detection.

1. Introduction

Accurate translation requires both codon recognition and correct aminoacylation of transfer RNA (tRNA). Aminoacyl-tRNA synthetases (ARSs) attach amino acids to cognate tRNAs, whereas the ribosome primarily evaluates codon–anticodon pairing. Aminoacylation specificity therefore constitutes an essential level of genetic-code fidelity, often described as the second genetic code (Ibba and Söll 2000; Giegé et al. 1998).

Canonical tRNAs form a cloverleaf secondary structure comprising the acceptor stem, D arm, anticodon arm and T arm; tertiary interactions generate the characteristic L-shaped structure (Holley et al. 1965; Kim et al. 1974). Their genomic organization differs among lineages.

Bacterial tRNA genes are often transcribed from multigene operons and rRNA operons, whereas eukaryotic nuclear tRNA genes are transcribed individually by RNA polymerase III from internal promoter elements within the D- and T-arm sequences (Galli et al. 1981), and their genomic clustering and copy number scale with genome size in characteristic ways (Bermudez-Santana et al. 2010). Mitochondrial tRNAs can be structurally reduced, including loss of the D or T arm (Jühling et al. 2018). Comparisons of nuclear or prokaryotic tRNA repertoires must therefore be kept separate from analyses of organellar translation systems; a separate study of fungal mitochondrial decoding treats the organellar side (D.W. Cheng, unpublished data; manuscript submitted).

ARSs are divided into two structurally distinct classes (Eriani et al. 1990). Class I enzymes generally contain a Rossmann-fold catalytic domain with HIGH and KMSKS signatures, whereas Class II enzymes share a distinct catalytic architecture and conserved motifs 1–3. Class I enzymes usually aminoacylate the 2′ hydroxyl of A76 and Class II enzymes the 3′ hydroxyl, with recognized exceptions (Ibba and Söll 2000; Perona and Gruic-Sovulj 2014). LysRS is particularly informative because unrelated Class I and Class II enzymes catalyze the same reaction. Recognition determinants can also reside outside the anticodon: the G3:U70 pair is a major alanine identity element, and isolated alanine acceptor-stem minihelices can be aminoacylated (Hou and Schimmel 1988; Francklyn and Schimmel 1989; Schimmel et al. 1993).

Several lineage-associated alternatives are well established. Archaea generally synthesize Gln-tRNA^{Gln} through an indirect amidotransferase-dependent route rather than GlnRS. Many methanogenic archaea use SepRS and SepCysS for tRNA-dependent cysteine formation, and Class I LysRS is prevalent in Archaea (Ibba et al. 1997; Sauerwald et al. 2005; Sheppard et al. 2008). Bacterial TilS and archaeal TiaS modify C34 of tRNA^{Ile2} to lysidine and agmatidine,

66 respectively, allowing AUA to be read as isoleucine and distinguished from AUG (Soma et
67 al. 2003; Ikeuchi et al. 2010). These are predominant strategies rather than universal rules.
68 Genome annotations suggested that a nanoarchaeum and a korarchaeum carry a UAU-anticodon
69 $tRNA^{Ile}$ instead, as eukaryotes do (Mandal et al. 2010), and a genome-wide analysis of 167
70 archaeal genomes confirmed the UAU gene in *Nanoarchaeum equitans* and *Candidatus*
71 Korarchaeum cryptofilum and found tRNA-UAU and all 16 tRNAs with A-starting (ANN)
72 anticodons to be rare across Archaea, probably because they were absent from the archaeal
73 ancestor (Wu et al. 2018). In Bacteria, *Mycoplasma mobile* lacks TilS and reads AUA with an
74 unmodified $tRNA^{Ile}$(UAU) (Taniguchi et al. 2013).

75 Anticodon modification further diversifies decoding capacity. Wobble pairing allows some
76 tRNAs to recognize several synonymous codons, and chemical modifications at position 34 can
77 expand or restrict this recognition (Crick 1966; Ogle et al. 2001; Agris et al. 2007; Weixlbaumer
78 et al. 2007). Bacterial TadA deaminates A34 mainly in $tRNA^{Arg}$, whereas eukaryotic
79 ADAT2/ADAT3 acts on seven or eight tRNA families depending on the organism; in yeast, a
80 single activity (the Tad2/Tad3 heterodimer) deaminates all seven A34-containing cytoplasmic
81 tRNAs (Auxilien et al. 1996; Gerber and Keller 1999; Novoa et al. 2012). Additional bacterial
82 A34 tRNA genes have been reported, and A-to-I modification of tRNA-Leu(AAG) has been
83 shown in *Streptococcus* (Rafels-Ybern et al. 2018; Katanski et al. 2026). A genomic A34
84 anticodon identifies a potential modification substrate; it does not show inosine in the mature
85 RNA.

86 Comparative studies have described anticodon-sparing strategies, tRNA gene diversity, archaeal
87 tRNA introns and the evolution of tRNA copy number across cellular life (Marck and Grosjean
88 2002; Sugahara et al. 2008; Fujishima et al. 2010; Grosjean et al. 2010; Bermudez-Santana et

al. 2010; Fujishima and Kanai 2014; Wu et al. 2018; Santos and Del-Bem 2022), and the phylogenetic distribution of synthetase families has been reviewed (Woese et al. 2000; O'Donoghue and Luthey-Schulten 2003). These studies mostly examined tRNA genes or enzymes separately. GtRNAdb and UniProt now allow both to be summarized for the same taxonomic groups (Chan and Lowe 2016; Chan et al. 2021; UniProt Consortium 2025). Because the genome and proteome collections are sampled independently, such summaries describe groups; they do not establish species-level coupling between a tRNA and its modification or charging enzyme.

We examined two complementary questions: how tRNA gene repertoires and genomic organization vary across the three domains of life, and how the distributions of tRNA charging and modification protein families vary across the same taxonomic reporting groups. We quantified repertoire size, anticodon richness, gene copy number, introns and local gene proximity in 2,125 species and scored 31 protein-family markers in 18,612 reference proteomes. The common reporting framework allowed us to distinguish variation in copy number from variation in anticodon composition, compare the taxonomic scales of charging- and modification-enzyme distributions, and quantify lineage-associated marker detection coverage. We also evaluated candidate exceptions to predominant AUA-decoding configurations. The six principal groups—Bacteria, Archaea, fungi, non-flowering green plants, flowering plants and animals—are reporting categories rather than equivalent taxonomic ranks. The human genome and proteome are presented separately as references.

2. Materials and Methods

We analyzed public data downloaded on November 1–3, 2026, using Python 3.13 with pandas, NumPy and Matplotlib. The genome and proteome datasets were filtered and summarized separately within the same reporting framework (Figure 1). All scripts and intermediate tables are in the data archive (Data availability).

2.1 Data sources

Resource	Release used	What it supplied
GtRNAdb (Chan and Lowe 2016)	Release 22 (Sept 2024); 4,853 genomes	tRNAscan-SE 2.0 gene calls: isotype, anticodon, coordinates, strand, introns
UniProt Reference Proteomes (UniProt Consortium 2025)	2026_03	22,134 reference proteomes (641 Archaea, 17,972 Bacteria, 3,521 Eukarya), all with BUSCO scores and lineages
InterPro (Blum et al. 2025)	110.0 (Pfam 38.2)	Family assignments for 31 synthetase and modification-enzyme markers
NCBI Taxonomy	Accessed November 1–3, 2026	Lineages for GtRNAdb genomes
MODOMICS (Boccaletto et al. 2022)	Accessed November 1, 2026	Reference for modification chemistry and enzyme names

2.2 tRNA gene data

Per-genome GtRNAdb archives were parsed for isotype, anticodon, coordinates, strand and introns (4,852 of the 4,853 genomes could be parsed). We kept GtRNAdb's high-confidence predictions: in each genome's name map, high-confidence genes carry the prefix "tRNA-" and other predictions the prefix "tRX-". For the 167 genomes that also include a separate confidence-set file, we additionally required membership in that file; in 146 of these the two definitions were

identical. Records on contigs whose names identify mitochondrial, chloroplast, plastid or apicoplast sequences (chrM, MT, mito, chrC, Pt, plastid, chloroplast, apicoplast, as whole tokens) were removed at parsing; no retained gene lies on a contig matching these names. This filter cannot remove organellar sequence on unlabeled contigs or nuclear insertions of organellar DNA (NUMTs, NUPTs), which are most likely in fragmented assemblies (several flowering-plant assemblies in the set have 200–1,100 tRNA-bearing contigs; Supplementary Table S4), so plant copy numbers should be read as upper bounds. The analysis therefore concerns nuclear tRNA genes in eukaryotes and genomic tRNA genes in prokaryotes. Selenocysteine (Sec), suppressor and pyrrolysine tRNAs were excluded from repertoire metrics; Sec genes were scored separately for presence. Ile2 denotes the tRNAscan-SE class of CAU-anticodon tRNAs that read AUA after C34 modification (Chan et al. 2021); it is a predicted functional assignment, and its uncertainty was quantified by counting elongator Met-CAU genes per archaeal genome (Section 3.3).

To avoid counting strains of one species repeatedly, one assembly per species was selected, choosing the assembly whose tRNA set covered the most of the 20 standard amino acids (ties resolved by taking the first assembly in GtRNAdb's species list). No minimum gene number was imposed, because reduced genomes can have genuinely small repertoires; for example, *Candidatus* Hodgkinia cicadicola and *Candidatus* Tremblaya princeps carry 19 and 16 tRNA genes, consistent with their highly reduced genomes (McCutcheon et al. 2009; McCutcheon and von Dohlen 2011). The dataset contained 2,125 species: 164 Archaea, 1,539 Bacteria and 422 Eukarya (270 fungi, 22 non-flowering green plants, 18 flowering plants, 107 non-human animals, the human genome and 4 other protists). The bacterial genomes come from 42 phyla, of which Pseudomonadota (589), Bacillota (283), Actinomycetota (195) and Bacteroidota (97)

make up 76%; the archaeal genomes come from 15 classes (Supplementary Table S4). Group-level percentages are therefore weighted by this composition, and the main anticodon prevalences are also reported per phylum (Supplementary Table S5).

For each genome we recorded the number of tRNA genes, the number of distinct anticodons, gene copies per anticodon and the share of genes with a predicted intron. Intron positions were classified relative to the canonical site between positions 37 and 38 from tRNAscan-SE secondary-structure output. A34 anticodons were scored in nine amino-acid families, and Ile2-CAU and Ile-UAU genes separately. Adjacent genes on the same contig were scored as a proximal pair when they were on the same strand and 100 bp or less apart; this criterion does not establish co-transcription. Anticodons are given 5′ to 3′ in RNA notation.

Sensitivity of the assembly choice. Results were recomputed under three other rules: the first-listed assembly with at least 20 tRNA genes (2,120 species), the first-listed assembly without a minimum (2,125 species), and only assemblies with tRNAs for all 20 amino acids (2,044 species). A34 and Ile2 results were identical under every rule. Median anticodon numbers changed by at most one and median gene numbers by less than 2% in every group except fungi, where restricting to complete assemblies (221 of 270 species) raised the median from 148 to 165 genes and lowered the intron-containing share from 58% to 53% (Supplementary Table S1).

2.3 Synthetase and modification enzymes

Thirty-one InterPro families were used as markers: seven canonical synthetase families (AlaRS, MetRS, HisRS, TrpRS, TyrRS, the Class Ic family and CysRS), the alternative charging enzymes, the GatCAB and GatDE amidotransferase subunits, wobble-modification enzymes and the AIMP2 scaffold family (Supplementary Table S2; InterPro names are given there and in

Table 2). For each marker, all UniProtKB entries carrying the InterPro cross-reference and the Reference Proteome keyword (KW-1185) were assigned to their proteomes, giving a proteome × marker presence/absence matrix.

Every one of the 22,134 reference proteomes had a BUSCO score. One completeness threshold, BUSCO ≥90%, was applied to every group. To match the genome dataset, one proteome was kept per species (the one with the highest BUSCO score); entries without a species epithet (for example "sp.") were treated as separate species. This left **18,612 proteomes: 538 Archaea, 16,097 Bacteria and 1,977 Eukarya** (1,100 fungi, 33 non-flowering green plants, 206 flowering plants, 496 non-human animals, the human proteome and 141 other protists). The BUSCO value used is the "complete" percentage (C) reported by UniProt for each proteome, computed by UniProt with BUSCO v5 against the lineage-specific dataset it selects. The bacterial proteomes are even more concentrated than the genomes in four phyla (Pseudomonadota 6,157, Actinomycetota 3,597, Bacillota 2,921, Bacteroidota 1,974; 90% of 16,097), and the archaeal proteomes are dominated by Halobacteria (262 of 538), whereas the archaeal genome set is not (26 of 164); genome-based and proteome-based percentages are therefore not directly comparable (Supplementary Table S4). No gymnosperm, bryophyte or fern proteome met the threshold, so the non-flowering green-plant proteomes are green algae (32 chlorophytes and one charophyte), whereas the corresponding genome category also contains bryophytes and a fern.

A marker hit supports the presence of a protein family but does not establish pathway activity or completeness. We used co-detection of GatA and GatB (IPR004413) as a marker configuration consistent with the bacterial indirect glutamine-charging route; GatC was not assessed because no corresponding InterPro marker was available. The archaeal GatDE configuration required

both GatD (IPR011878) and GatE (IPR004414). For each pair of alternative families or configurations, we report the exact numbers of proteomes containing one, both or neither family.

Reference-marker detection coverage. We quantified detection of broadly expected synthetase markers as an empirical benchmark for interpreting protein-family distributions. Across 18,612 proteomes, detection coverage was 99.4% for AlaRS and 98.4% for MetRS; at least one of these markers was detected in 99.7% of proteomes. The corresponding archaeal values were 98.1%, 98.0% and 98.7%. Seven canonical marker categories were also scored together—AlaRS, CysRS, MetRS, HisRS, TrpRS, TyrRS and LysRS (Class I or Class II counted as one category); at least five of the seven were detected in 99.7% of proteomes (Supplementary Table S7). Coverage varied by lineage: the MetRS marker was detected in 75% of the 28 Deinococcota proteomes, and HisRS and TyrRS coverage also varied among eukaryotic groups (Section 3.4). These values describe the selected markers and sampled proteomes; establishing sensitivity for other families requires marker-specific validation. InterPro coverage can be affected by sequence divergence, protein fragmentation and annotation completeness. We therefore interpret detection patterns alongside established biochemistry and retain the distinction between marker detection and biological presence. The nematode AIMP2 results were additionally examined using profile-HMM and phmmer searches (Section 3.6).

2.4 Statistics

Continuous variables are summarized as medians and interquartile ranges (IQRs), and marker-presence data as counts and percentages. All comparisons are descriptive. No significance tests or phylogenetically adjusted models were fitted; shared ancestry precludes treating the sampled genomes or proteomes as independent evolutionary replicates. Medians are reported to one decimal place for groups with fewer than ten members and otherwise generally rounded to the

nearest integer; ratios and percentages retain additional precision where informative. The flowering-plant (n = 18) and non-flowering green-plant (n = 22) genome summaries are based on relatively small samples and are presented with their IQRs.

2.5 Eukaryotic lineages

GtRNAdb species were assigned to NCBI taxonomic lineages; nine genus names that are ambiguous or renamed were resolved by hand against UniProt lineages (for example, *Cryptococcus* is both a fungal and an insect genus). UniProt proteomes were assigned from their own lineage strings using the same rules. Eukaryotes were grouped into about 40 reporting lineages (protists, 11 fungal groups, green-plant groups and animal groups). These categories are descriptive and are not a reconstructed phylogeny. Lineage values are shown only where a lineage had at least three genomes (tRNA columns) or at least three proteomes (enzyme columns); the human reference is shown for comparison.

2.6 Use of AI tools

During this work the author used Claude (Anthropic) for data retrieval, analysis code, and figures, and ChatGPT (OpenAI) for language revision for readability. All code and intermediate tables are in the data archive. The author prepared, reviewed and edited all content and takes full responsibility for the data, analyses, references and text.

3. Results

3.1 Larger tRNA repertoires consist mainly of additional copies of existing anticodons

Eukaryotic genomes carried a median of 204 tRNA genes (IQR 120–377), compared with 52 (46–65) in Bacteria and 47 (46–49) in Archaea (Table 1; Figure 2A). Median anticodon richness

differed much less: 40 in Bacteria, 44 in Archaea and 45 in Eukarya (Figure 2B). Median gene copies per anticodon were 4.6, 1.3 and 1.1, respectively (Figure 2C). Larger eukaryotic repertoires therefore consist mainly of additional copies of existing anticodons. The eukaryotic median is dominated by the 270 fungi (64% of the 422 eukaryotic genomes; fungal median 147.5), so the per-group values in Table 1 are the informative ones. These values replicate those of Santos and Del-Bem (2022), who reported a median of about 200 tRNA loci in eukaryotes, averages of about 60 in bacteria and 48 in archaea, and median anticodon diversities of 39, 44 and 45 from all GtRNAdb genomes; the one-assembly-per-species filter used here changes none of these conclusions. Archaeal gene numbers were the most uniform (IQR of three genes), and bacterial anticodon richness the most variable (IQR 34–43), consistent with differences in decoding strategy among bacterial genomes of different sizes (Grosjean et al. 2010).

Among eukaryotes, flowering plants carried the most copies per anticodon (median 15.1; 708 genes for 47 anticodons), followed by animals (8.5), non-flowering green plants (5.9) and fungi (3.3), while median anticodon richness stayed between 45 and 48 in all four groups (Table 1).

3.2 Genomic proximity and introns differ among groups

The median share of adjacent tRNA gene pairs on the same strand within 100 bp was 26% in Bacteria (IQR 16–44%), 5% in Archaea and 1% in Eukarya (Figure 2D). Median spacers within such pairs were 27 bp in Bacteria and 32 bp in Archaea. The bacterial pattern is compatible with compact tRNA arrays but does not establish shared transcription units, and rRNA coordinates were not analyzed.

Every archaeal and eukaryotic genome had at least one gene with a predicted intron, against 5.5% of bacterial genomes. The archaeal median share of intron-containing genes was 8.5%, and

14.0% of archaeal introns lay outside the canonical 37/38 site, against 0.3% in Eukarya. Among eukaryotic groups, the median share was 57.8% in fungi, 6.4% in animals and about 5% in both plant groups (Table 1). These are tRNAscan-SE predictions; splicing of individual loci was not examined.

3.3 Anticodon composition shows domain-associated patterns and lineage-specific exceptions

A34 anticodons were distributed differently among the groups (Figure 3). Eukaryotic genomes carried A34 genes in a median of eight families (Ala, Arg, Ile, Leu, Pro, Ser, Thr and Val), although not uniformly: Leu-AAG occurred in 87% of fungal genomes. Bacterial genomes carried a median of one A34 family, mainly Arg-ACG (92% of genomes). That percentage is a mixture of phylum-level states rather than a property of Bacteria: Arg-ACG is present in 99.6–100% of Pseudomonadota, Bacillota, Actinomycetota, Bacteroidota and Cyanobacteriota genomes but absent from every Campylobacterota (n = 38), Thermotogota (n = 20) and Chloroflexota (n = 11) genome and from 90% of Spirochaetota (Supplementary Table S5). No Thermotogota or Chloroflexota genome, and only one of the 38 Campylobacterota genomes, carried an A34 anticodon in any of the nine families, implying that CGN arginine codons in these phyla are decoded without inosine 34. In Archaea, 161 of 164 genomes (98.2%) had no A34 gene in any of the nine families. The three exceptions are single genes in three different genomes (Leu-AAG in *Ferroplasma acidarmanus*, Arg-ACG in *Methanobrevibacter ruminantium*, Gly-ACC in *Natrinema pellirubrum*), each among the lowest-scoring tRNA predictions of its genome (tRNAscan-SE scores 37.6–54.0, compared with an archaeal median of 78.8), and no archaeal proteome contains a TadA/ADAT2-family deaminase; we interpret them as borderline predictions with misassigned anticodons rather than as A34 tRNAs (Supplementary Table S8). Gly-ACC occurred in 6% of eukaryotic genomes. These data concern encoded

anticodons and do not show the modification state of mature tRNAs (Novoa et al. 2012; Rafels-Ybern et al. 2019).

Leu-AAG and Thr-AGU genes occurred in about 7% and 3% of bacterial genomes, mostly in Lactobacillales and Mollicutes. Additional bacterial A34 tRNA genes have been reported (Rafels-Ybern et al. 2018), and A-to-I modification of tRNA-Leu(AAG) has been shown in *Streptococcus* (Katanski et al. 2026). The modification state of the other candidates, including the Thr-AGU genes, cannot be inferred from the anticodon sequence.

Ile2-CAU genes occurred in 96% of bacterial and 98% of archaeal genomes and in no eukaryotic nuclear genome; Ile-UAU occurred in 100% of flowering-plant and animal genomes, 94% of fungal genomes and 16 of 22 non-flowering green plants (Table 1; Figure 3). Of the 17 fungal genomes without Ile-UAU (6.3%; 14 filamentous ascomycetes, one rust fungus, one microsporidian and one budding yeast), 14 carry Ile-AAU, whose inosine-34 can pair with A at the third codon position (Crick 1966); the other three (*Cyphellophora europaea*, *Melampsora pinitorqua* and *Pseudocercospora pini-densiflorae*) have no annotated tRNA that can read AUA. These 17 gene sets are smaller than those of other fungi (median 111 vs 147.5 genes), so incomplete assembly or annotation is the more likely explanation (Supplementary Table S8e). All six green-plant genomes without Ile-UAU retain Ile-AAU, whose inosine-34 can read AUA, so none has an undecodable codon. Two of the six are fragmented assemblies with small tRNA sets (*Klebsormidium nitens*, 86 genes on 79 contigs; *Pleurozium schreberi*, 176 genes on 143 contigs), where incompleteness is the likely cause; the other four, including *Ceratodon purpureus* (1,369 genes) and *Closterium* sp. (1,100 genes), are large gene sets in which the absence of any UAU gene more likely reflects tRNAscan-SE anticodon assignment or the high-confidence filter (Supplementary Table S8).

Because Ile2-CAU and Met-CAU differ only by tRNAscan-SE classification, we counted elongator Met-CAU genes per archaeal genome: the median is 1 (mean 1.04), only 5 of 164 genomes (3.0%) have two or more, and Ile2-CAU is annotated in 160 of 164, so at most five genomes could conceal an Ile2 gene among their Met-CAU copies. This set does not show the ~3 Met-CAU copies per genome reported by Wu et al. (2018) from all GtRNAdb archaeal assemblies. Of the four genomes without an annotated Ile2 gene, *Caldivirga maquilingensis* is the only archaeal genome with two strong Met-CAU predictions (scores 95.5 and 93.6) and no Ile2 gene, and it has a TiaS-family protein, so one of its Met-CAU genes is probably the agmatidine-modified $tRNA^{Ile2}$. *Nanoarchaeum equitans*, *Candidatus* Korarchaeum cryptofilum and the Asgard archaeon *Candidatus* Prometheoarchaeum syntrophicum (Imachi et al. 2020) have one Met-CAU gene and carry an Ile-UAU gene instead. None of the three proteomes meets the BUSCO ≥90% threshold of the analysis set, but all three have UniProt reference proteomes (BUSCO 60.3% for *N. equitans* and 88.7% for the other two), and neither a TiaS- nor a TilS-family protein was detected in any of them (Supplementary Table S8f). Because these proteomes are less complete than those analyzed elsewhere, the absence of TiaS is weaker evidence here, most of all for *N. equitans*. *Caldivirga maquilingensis*, by contrast, has a TiaS-family protein in a proteome of 96.5% completeness. The UAU gene in the nanoarchaeum and the korarchaeum was noted from genome annotation by Mandal et al. (2010) and confirmed by Wu et al. (2018), who traced the *Korarchaeum* gene to an anticodon change in a GAU gene. The present analysis adds a third archaeon, from the Asgard lineage, with the same tRNA genotype, and adds the proteome-level finding that neither TiaS nor TilS is detected in any of the three.

In Bacteria, 36 of 1,539 genomes (2.3%) carry an Ile-UAU gene and 60 (3.9%) lack an Ile2-CAU gene, 13 having both features; 33 of the 36 have a single UAU gene, and these genes score

lower than the bacterial median (median 61.9 vs 76.7), so some are borderline calls. For 43 of these genomes a proteome of the same species was available (Supplementary Table S8). TiaS was absent from all 43, as expected. Four species combine the three features expected of the *Mycoplasma mobile* decoding system—no Ile2-CAU gene, no TilS and an Ile-UAU gene: *M. mobile* itself, the published case (Taniguchi et al. 2013), *Neorickettsia helminthoeca*, *Neorickettsia risticii* and *Bifidobacterium adolescentis*. To our knowledge, the latter three species have not previously been identified as candidates for TilS-independent AUA decoding. The 22 species with TilS but neither tRNA, by contrast, have an apparent decoding gap that is more plausibly a tRNA-annotation failure than biology. Across the bacterial proteomes as a whole, TilS was detected in 15,822 of 16,097 (98.3%), a coverage comparable to that of the reference synthetase markers (Section 2.3). The 275 proteomes without a TilS marker are distributed across phyla roughly in proportion to sampling (TilS 93.8–100% in every bacterial phylum with at least ten proteomes except Spirochaetota, 86.1%; Supplementary Table S5), so the bacterial genomes in which the *M. mobile*-type configuration is verified remain the four species above.

3.4 Charging-enzyme families vary within domains, whereas C34-modification enzymes show strong domain associations

Of the seven canonical synthetase families scored, AlaRS, MetRS, TrpRS and the Class Ic family were detected in 97–100% of proteomes in every group, and CysRS in 88–100% (Table 2). HisRS and TyrRS coverage varied more among lineages: the HisRS marker was detected in 64.6% of fungal proteomes (97–99% in Archaea, Bacteria and both plant groups) and the TyrRS marker in 70.2% of protist and 91.6% of archaeal proteomes. Cytoplasmic translation requires His-tRNA^{His} and Tyr-tRNA^{Tyr}, and no indirect route to either is known, so these lower values

reflect the coverage of the InterPro entries for divergent HisRS and TyrRS sequences in these lineages rather than loss of the enzymes. They show the range of marker coverage that can occur within a whole clade, and lineage-level non-detection of a synthetase family is therefore interpreted only with independent sequence-level evidence. The most informative variation lay in families associated with alternative routes for glutamine, cysteine and lysine charging and with C34 modification of $tRNA^{Ile2}$ (Figure 4A). The charging-enzyme families varied among phyla and classes as well as between domains, whereas the C34-modification enzymes showed strong domain associations (Supplementary Table S5).

Glutamine. Among 538 archaeal proteomes, GatD was detected in 529 (98.3%) and GatE in 530 (98.5%), with co-detection in 527, consistent with the predominant archaeal GatDE route. No GlnRS marker was detected in this set (0 of 538), consistent with the absence of GlnRS from Archaea. The GatB entry (IPR004413), detected in 85.9% of archaeal proteomes, represents the bacterial/mitochondrial subunit and was scored separately from GatE. Among 16,097 bacterial proteomes, 2,930 (18.2%) contained the GlnRS marker alone, 8,498 (52.8%) contained the GatA+GatB configuration alone, and 4,372 (27.2%) contained both (Figure 4B). Thus, at least one configuration was detected in 15,800 proteomes (98.2%). GatCAB also contributes to Asn-$tRNA^{Asn}$ formation, providing a functional context for co-occurrence. Bacterial GlnRS prevalence varied by phylum, reaching 89–96% in Bacteroidota, Myxococcota and Deinococcota and 0% in Cyanobacteriota, Thermotogota, Fusobacteriota, Aquificota and Chlorobiota. The remaining 297 proteomes require sequence-level examination to resolve their glutamine-charging configurations.

Cysteine. CysRS was detected in 99.1% of bacterial proteomes and in 97–100% of each eukaryotic group. Among 538 archaeal proteomes, 123 (22.9%) contained SepRS, and every one

of them also contained SepCysS (123 of 123; no proteome had either family alone). Of the 123 SepRS-containing proteomes, 59 lacked CysRS and 64 (52.0%) also contained CysRS; 409 contained CysRS alone and 6 neither (Figure 4B). At least one of the two routes was detected in 532 of 538 archaeal proteomes (98.9%); SepRS therefore occurs both instead of and alongside CysRS.

Lysine. Class I LysRS was detected in 85.3% of archaeal proteomes and Class II in 15.2%: 446 contained Class I alone, 69 Class II alone, and 13 (2.4%) both, yielding coverage by at least one class in 528 of 538 proteomes (98.1%). Bacteria showed the opposite class predominance: 2,884 contained Class I alone, 12,554 Class II alone, and 558 (3.5%) both, yielding combined coverage of 15,996 of 16,097 proteomes (99.4%). Bacterial Class I prevalence varied from 0% in nine phyla to 35.4% in Pseudomonadota and 44.9% in Spirochaetota (Supplementary Table S5). Eukaryotic groups contained Class II almost exclusively (98–100%). Archaeal Class I prevalence ranged from 34.4% in Thermoprotei and 66.3% in Methanomicrobia to 96.9–100% in the other six classes with at least ten proteomes. Within Thermoproteati (the kingdom-level name for the TACK group, including Thermoproteota, Nitrososphaerota and relatives), the 84 proteomes partitioned into 45 with Class I alone and 39 with Class II alone. This partition followed order boundaries—Class II in Sulfolobales and Class I in Desulfurococcales and Nitrososphaerales—extending the genus-level pattern reported by Shaul et al. (2006).

C34 of tRNA^{Ile2}. TilS was detected in 98.3% of bacterial proteomes and in no archaeal proteome; TiaS in 96.7% of archaeal proteomes and in no bacterial proteome. These distributions agree with the predominant lysidine and agmatidine strategies, and unlike the charging routes they are not phylum-structured: TilS is present in 86.1–100% of every bacterial phylum and absent from every archaeal class, and TiaS is present in 87.5–100% of every archaeal class and

absent from every bacterial phylum (Supplementary Table S5). Of the markers tested per phylum and class, TilS, TiaS and GatD (94.4–100% of archaeal classes, ≤0.8% of bacterial phyla) showed the strongest domain associations; GatE, tested at the domain level, has the same distribution as GatD (98.5% of archaeal and 0.1% of bacterial proteomes, co-occurring with GatD in 527 of 538 archaeal proteomes). TiaS was not detected in 18 archaeal proteomes (3.3%), comprising eight Thermoprotei (12.5% of that class), eight Halobacteria and two others; this coverage (96.7%) is slightly below that of the archaeal reference markers (98.0–98.1%). TilS also occurred in 84.5% of fungal, 97.0% of non-flowering green-plant and 98.5% of flowering-plant proteomes, but in only 1.6% of animal proteomes. Such hits are compatible with organellar functions; localization was not measured.

3.5 Modification enzymes and AIMP2 show lineage-associated distributions

Ctu1/Ncs6 and Elp3 were detected in 97.4% and 93.9% of archaeal proteomes and in 82–98% of each eukaryotic group, but in 0.4% of bacterial proteomes; the few bacterial Elp3 homologs are known (for example in *Dehalococcoides*) and are not annotation errors. Bacteria instead contained MnmA (98.3%) and MnmE (76.4%). The eukaryotic MnmA-type hits are the mitochondrial 2-thiouridylase Mtu1; its low fungal prevalence (32.8%, against 93–98% in plants and animals) is examined at the genome level in a separate study of fungal mitochondrial decoding (D.W. Cheng, unpublished data; manuscript submitted). Bacterial-type $\alpha_2\beta_2$ GlyRS ranges from 0–1% in Deinococcota, Chlorobiota, Verrucomicrobiota and Bacteroidota to 97–100% in Pseudomonadota, Cyanobacteriota and Campylobacterota, so its bacterial prevalence of 56.6% is a composition average (Supplementary Table S5). These enzymes act in different reactions and are not one-to-one equivalents; their presence alone does not define the full modification repertoire of a species.

The AIMP2 family was detected in 94.4% of the 496 animal proteomes and in the human proteome, and in no fungal or plant proteome. AIMP2 detection does not by itself establish assembly of a multisynthetase complex, and its absence does not exclude other synthetase complexes (Guo et al. 2010; Havrylenko and Mirande 2015).

3.6 Eukaryotic lineages differ in introns, gene copy number and enzyme families

Finer lineages showed large differences within the major eukaryotic groups (Figure 5; Table 3). Genome and proteome sample sizes differ among lineages, and the two datasets are not matched species by species.

Selenocysteine tRNA. A Sec tRNA gene was present in all 12 chlorophyte and charophyte algal genomes, but in none of the 9 bryophyte genomes, the single fern genome or the 18 flowering-plant genomes, and in none of the 270 fungal genomes. Among animals, it was present in 97–100% of mammal, bird, reptile and ray-finned fish genomes, in 13 of 14 fly and mosquito genomes and in 5 of 6 nematode genomes, but in none of 3 moth and butterfly genomes. The land-plant and insect distributions agree with reported losses of selenoprotein systems (Lobanov et al. 2007; Chapple and Guigó 2008). The fungal zero is a sampling result: complete Sec machinery, including $tRNA^{Sec}$, has been found in nine early-branching fungi of the Mucoromycota, Chytridiomycota and Zoopagomycota (Mariotti et al. 2019), phyla represented here by only 11 genomes, and no Dikarya genome is expected to carry it.

Introns. The median share of intron-containing tRNA genes was 5–8% in animal lineages and 5% in monocots and eudicots, against 63% in filamentous ascomycetes and 58–92% in the four basidiomycete lineages, reaching 92% in Tremellomycetes. The number of isotypes with at least one intron-containing gene was 2 in flowering plants, 3–6 in animal lineages and 18–20 in

filamentous ascomycetes and basidiomycetes. Chlorophyte algae had a median of 49% and 14.5 intron-bearing isotypes (n = 6; IQR 13.0–16.8), whereas charophyte algae had 6% and 6.5 (n = 6; IQR 4.2–9.5), closer to land plants. Ahammed and van Hoof (2024) reported that 92% of *Cryptococcus neoformans* tRNA genes contain introns and 3–4% in Mucorales; the Tremellomycetes value here confirms theirs, and the 270-species dataset places it at the upper end of the range of fungal lineages.

Gene copy number. Median copies per anticodon ranged from 1.1 in microsporidia and 2–7.5 in other fungal lineages to 6.6 in bryophytes, 12.0 in monocots and 15.1 in eudicots. Among animals, mammals carried about 9 copies per anticodon (human genome: 418 genes, 8.9 copies), birds 4.3, and nematodes and ray-finned fish 17.6–17.9.

A34 breadth. Most eukaryotic lineages carried A34 genes in eight families. Budding yeasts carried a median of seven, with Leu-AAG present in only 50% of their genomes; this agrees with the seven I34-containing tRNAs known in yeast (Auxilien et al. 1996). The TadA/ADAT2 deaminase family itself was detected in 94% of flowering-plant but only 0.1% of fungal and 33% of animal proteomes, although all carry A34 tRNAs and fungi and animals are known to deaminate them; this is another clade-specific limit of the InterPro entry (Supplementary Table S6).

Organelle-associated families. Bacterial-type $\alpha_2\beta_2$ GlyRS was detected in 100% of chlorophyte, 95% of monocot and 99% of eudicot proteomes, and in at most 25% of any fungal or animal lineage. TilS was detected in 95–100% of plant lineages, 75% of apicomplexan, 98% of filamentous ascomycete and 99% of mushroom-forming fungal proteomes, but in none of the Tremellomycetes (32), zygomycete-type fungi (26) or microsporidia (15). These annotations do not establish organellar localization or the decoding function of individual proteins.

AIMP2 in animals. The AIMP2 family was detected in 96–100% of insect and vertebrate lineages, in 75–76% of annelid and other arthropod proteomes, and in none of the 17 nematode proteomes. Searches of five nematode proteomes (*C. elegans*, *C. briggsae*, *Pristionchus pacificus*, *Brugia malayi*, *Onchocerca volvulus*) with human AIMP2 and the Pfam AIMP2 profiles PF16780 and PF18569 found no match ($E < 10^{-3}$), whereas the same searches detected AIMP2 in fly and flatworm controls (scripts in the data archive). Biochemical work has shown that *C. elegans* has a multisynthetase complex with a highly divergent AIMP2-like scaffold (Havrylenko et al. 2011; Havrylenko and Mirande 2015). The nematode result therefore reflects the limit of family-based detection, not loss of the complex, and the same limit applies to the AIMP2 zeros in fungi and plants, which we report as non-detection rather than absence.

4. Discussion

This study integrates tRNA gene organization and charging- and modification-enzyme distributions within a common taxonomic framework. Four findings summarize the comparative results: repertoire expansion is primarily associated with additional gene copies; charging-enzyme families and C34-modification enzymes exhibit different taxonomic patterns; protein-family detection coverage varies by lineage; and alternative charging-family configurations show broad complementary coverage with measurable co-occurrence. These findings extend established observations through species-filtered counts, phylum- and class-level summaries, and explicit characterization of marker coverage. Candidate alternative AUA-decoding configurations are considered in Section 4.3.

First, anticodon richness, gene copy number and anticodon composition show distinct patterns of variation. Across the principal reporting groups, median anticodon richness ranged from 40 to

48, whereas median gene counts ranged from 47 in Archaea to 708 in flowering plants and median copies per anticodon ranged from 1.1 to 15.1. Eukaryotic genomes typically encoded A34 anticodons in eight families, whereas bacterial genomes encoded a median of one A34 family and 161 of 164 archaeal genomes encoded none of the nine families examined. Ile-UAU and Ile2-CAU annotations also differed markedly among domains (Section 3.3). Conserved anticodon richness and domain-associated anticodon sets have precedents in Marck and Grosjean (2002), Wu et al. (2018) and Santos and Del-Bem (2022). The present comparison quantifies these dimensions together under a one-assembly-per-species rule: reporting groups with similar anticodon richness can differ substantially in gene copy number and anticodon composition. This descriptive contrast does not establish statistical independence among the three measures.

Second, charging-enzyme distributions vary substantially within domains, whereas TilS and TiaS show strong domain associations. Bacterial GlnRS and bacterial-type GlyRS prevalences span approximately 0–100% among phyla, and Class I LysRS prevalence also varies among bacterial phyla and archaeal classes (Supplementary Table S5). A pooled prevalence therefore reflects both lineage-specific distributions and the taxonomic composition of the sample. In contrast, TilS was detected in 98.3% of bacterial proteomes and no archaeal proteomes, whereas TiaS was detected in 96.7% of archaeal proteomes and no bacterial proteomes. GatD and GatE were also strongly associated with Archaea, although rare bacterial hits preclude describing their distributions as absolutely domain exclusive. The patchy distributions of GlnRS and LysRS classes (Woese et al. 2000; Shaul et al. 2006) and the predominant lysidine and agmatidine strategies (Soma et al. 2003; Ikeuchi et al. 2010) are established. Here, their contrasting taxonomic patterns are quantified within a common reporting framework. The order-associated

LysRS-class partition in the sampled TACK archaea provides an additional example of variation below the domain level (Section 3.4).

Third, protein-family detection coverage is strongly lineage dependent. Overall coverage of 99.4% for AlaRS and 98.4% for MetRS contrasts with 64.6% for HisRS in fungi and 70.2% for TyrRS in protists. AIMP2 further illustrates the importance of lineage-specific sequence validation: biochemical work established a divergent AIMP2-like scaffold in the *C. elegans* multisynthetase complex (Havrylenko et al. 2011), whereas the selected family markers yielded 0% coverage in the sampled nematode proteomes. These results identify lineages that would benefit from targeted marker refinement and annotation review. Marker coverage should therefore be evaluated within the relevant lineage, and biological loss should be supported by independent sequence-level and, where feasible, functional evidence.

Fourth, alternative charging-family configurations provide broad combined coverage, with measurable co-occurrence. At least one configuration was detected in 98.2% of bacterial proteomes for GlnRS versus GatA+GatB, 98.9% of archaeal proteomes for CysRS versus SepRS, and 98.1% and 99.4% of archaeal and bacterial proteomes, respectively, for the two LysRS classes. Both LysRS classes occurred in 2.4% of archaeal and 3.5% of bacterial proteomes, and GlnRS and GatA+GatB co-occurred in 27% of bacterial proteomes. Among SepRS-containing archaeal proteomes, 52% also contained CysRS. This combination of broad complementary coverage and variable overlap is consistent with established pathway biology (Woese et al. 2000; Sauerwald et al. 2005; Sheppard et al. 2008). Determining the functional consequences of overlap requires biochemical characterization.

The group-level patterns further contextualize these findings. Bacterial repertoires were small and frequently contained closely spaced, same-strand genes; archaeal gene counts were relatively

uniform; and eukaryotic repertoires contained more copies per anticodon, particularly in flowering plants. Fungi had the highest proportions of predicted intron-containing tRNA genes. Finer eukaryotic lineages differed in Sec tRNA detection, intron prevalence, gene copy number and associated protein-family markers. Most individual patterns have clear precedents (Table 4). The contribution of the present study is their quantitative comparison across 2,125 species and 18,612 proteomes, together with explicit taxonomic sampling and marker detection summaries.

4.1 Alternative enzyme families overlap

The exact co-detection counts show that alternative families are distributed in a complementary way but are not mutually exclusive. In Archaea, SepRS occurred without CysRS in 59 proteomes and with CysRS in 64; SepRS and SepCysS always occurred together. Coexistence of SepRS and CysRS may reflect different roles in cysteine formation and charging; the presence/absence data cannot resolve this (Sauerwald et al. 2005; Sheppard et al. 2008). Bacterial GlnRS and GatA+GatB coexisted in 27% of proteomes, as expected from the role of GatCAB in Asn-$tRNA^{Asn}$ formation. LysRS gave the clearest separation: only 2.4% of archaeal and 3.5% of bacterial proteomes contained both classes. TilS and TiaS were never detected in the same proteome. Whether a given family is active, and on which substrate, requires biochemical data.

The difference between bacterial and eukaryotic A34 repertoires agrees with the broader substrate range of eukaryotic ADAT2/ADAT3 (Novoa et al. 2012; Rafels-Ybern et al. 2019). Bacterial Leu-AAG genes and the seven-family repertoire of budding yeasts show that the one-versus-eight pattern is a predominant rule, not a categorical one (Auxilien et al. 1996; Katanski et al. 2026).

4.2 What the eukaryotic lineages add

The lineage summaries separate several dimensions of variation that a single eukaryotic average hides. Plant groups differ in copy number, fungal lineages in intron load and breadth, and animals share the AIMP2 family. The Sec tRNA pattern agrees with earlier analyses of selenoprotein evolution in land plants and insects (Lobanov et al. 2007; Chapple and Guigó 2008); the fungal result is a sampling statement, because Sec machinery has been found in early-branching fungal phyla that are sparsely represented here (Mariotti et al. 2019; Section 3.6). The human genome, with 418 tRNA genes, 47 anticodons and about nine copies per anticodon, lies within the mammalian range; it is a reference point, not an independent group.

4.3 Exceptions to the predominant AUA-decoding strategy

Four bacterial species—*Mycoplasma mobile*, *Neorickettsia helminthoeca*, *N. risticii* and *Bifidobacterium adolescentis*—and three archaea—*Nanoarchaeum equitans*, *Candidatus* Korarchaeum cryptofilum and *Candidatus* Prometheoarchaeum syntrophicum—combined an Ile-UAU annotation with non-detection of Ile2-CAU and the corresponding C34-modifying-enzyme markers (Section 3.3). In *M. mobile*, an unmodified $tRNA^{Ile}$(UAU) is experimentally established to read AUA (Taniguchi et al. 2013). The nanoarchaeal and korarchaeal tRNA configurations were previously reported (Mandal et al. 2010; Wu et al. 2018). The present study identifies the three additional bacterial species and the Asgard archaeon as candidates for an alternative decoding configuration. Their sequence annotations do not establish the *M. mobile* mechanism: the archaeal proteomes fall below the main completeness threshold, and the bacterial UAU predictions require validation. Confirming the tRNA sequences, enzyme complements, modification states and codon-recognition properties would determine whether these candidates represent functional alternatives to the predominant TilS- or TiaS-dependent systems.

5. Conclusions

This study compared tRNA gene repertoires and associated protein-family distributions across the three domains of life using a common taxonomic reporting framework. Three principal conclusions emerge from the descriptive analyses.

First, similar anticodon richness can coexist with substantial differences in tRNA gene copy number and anticodon composition. Repertoire expansion was primarily associated with additional copies of existing anticodons, whereas anticodon composition showed domain-associated patterns with identifiable exceptions. Second, charging-enzyme families varied markedly among bacterial phyla and archaeal classes, while TilS and TiaS showed strong bacterial and archaeal associations, respectively. Alternative charging-family configurations were usually detected in complementary combinations, but their co-occurrence was also quantified. Third, marker detection coverage varied substantially among lineages, establishing the value of lineage-specific reference measurements for comparative interpretation.

The phylum-, class- and lineage-level summaries provide reference data for interpreting comparative tRNA biology and prioritizing follow-up analyses (Supplementary Tables S4–S7). The candidate AUA-decoding configurations in an Asgard archaeon and three bacterial species beyond *M. mobile* offer specific targets for sequence validation and biochemical investigation.

The scope of inference reflects the sampling design and the annotation-based measurements. Genome and proteome datasets were summarized at the group level, with four bacterial phyla contributing most records; non-flowering green-plant proteomes represent green algae. The tRNA results describe predicted gene repertoires, and some plant copy counts may include residual organellar sequences. Protein-family detection provides evidence of family membership,

while activity, substrate specificity, cellular localization and divergent homologs require targeted validation. Because the sampled genomes and proteomes share ancestry, the comparisons are descriptive, and no significance tests assuming independent evolutionary observations were applied.

Priorities for follow-up are targeted sequence and profile-HMM searches to refine marker coverage, direct measurement of anticodon modifications and decoding properties in candidate species, and phylogenetically structured comparisons using species-matched genome–proteome datasets.

Statements and Declarations

Funding. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interests. The author has no relevant financial or non-financial interests to disclose.

Ethics approval. Not applicable; only public genome and proteome data were analyzed.

Consent to participate and consent to publish. Not applicable.

Data availability. All input data are public (Section 2.1). The data archive—analysis and figure scripts, the parsed table of 439,964 tRNA genes from 4,852 parsable GtRNAdb genomes, the 22,134-proteome × 31-marker matrix, the 18,612-proteome analysis set, per-genome metrics, co-detection tables, figure source data and Supplementary Tables S1–S8—is deposited in Zenodo (https://doi.org/10.5281/zenodo.23970272).

Author contributions. D.W.C. conceived and designed the study, directed the analyses, interpreted the results and wrote and approved the manuscript.

Tables

Table 1. tRNA gene repertoire and organization across six reporting groups and the human reference. Values are medians (IQR) unless stated; one assembly per species; GtRNAdb high-confidence nuclear genes.

Metric	Bacteria	Archaea	Fungi	Non-flowering green plants	Flowering plants	Animals	Human
Genomes analyzed	1,539	164	270	22	18	107	1
tRNA genes per genome	52 (46–65)	47 (46–49)	148 (101–217)	299 (179–706)	708 (534–765)	405 (296–581)	418
Distinct anticodons	40 (34–43)	44 (43–44)	45 (42–46)	48 (44–51)	47 (46–48)	47 (45–48)	47
Gene copies per anticodon	1.3 (1.1–1.7)	1.1 (1.0–1.2)	3.3 (2.3–4.9)	5.9 (3.8–13.4)	15.1 (11.5–16.6)	8.5 (6.7–12.0)	8.9
Genomes with ≥1 intron-containing gene (%)	6	100	100	100	100	100	Yes
Intron-containing genes per genome (%)	0.0 (0.0–0.0)	8.5 (5.9–27.8)	57.8 (28.9–72.0)	5.1 (3.4–34.3)	5.3 (4.8–5.8)	6.4 (5.3–7.4)	6.9
Introns outside the canonical 37/38 site (%)	32.3	14.0	0.3	2.4	0.0	0.0	0.0

Metric	Bacteria	Archaea	Fungi	Non-flowering green plants	Flowering plants	Animals	Human
Adjacent same-strand pairs ≤100 bp apart (%)	26 (16–44)	5 (0–13)	2 (0–5)	13 (4–19)	1 (1–2)	0 (0–1)	0
A34 families per genome	1 (1–1)	0 (0–0)	8 (8–8)	8 (8–8)	8 (8–8)	8 (8–8)	8
Genomes with Ile2-CAU tRNA (%)	96	98	0	0	0	0	No
Genomes with Ile-UAU tRNA (%)	2	2	94	73	100	100	Yes

Non-flowering green plants: chlorophyte and charophyte algae, bryophytes and one fern. Flowering plants: monocots and eudicots (GtRNAdb has no gymnosperm genome). Animals exclude humans. Human values use GRCh37 (GRCh38 gives 432 genes). Four other protist genomes are not shown. The noncanonical-intron row summarizes all predicted introns in the group.

Table 2. Prevalence of synthetase and modification-enzyme families (% of UniProt reference proteomes; BUSCO ≥90%; one proteome per species; human reference: Yes/No, shown for comparison only).

Marker (InterPro)	Role	Bacteria	Archaea	Fungi	Non-flowering green plants	Flowering plants	Animals	Human
Proteomes		16,097	538	1,100	33	206	496	1
AlaRS (IPR002318)	Reference synthetase	99.4	98.1	99.3	100.0	99.0	100.0	Yes
MetRS (IPR014758)	Reference synthetase	98.3	98.0	99.0	97.0	99.0	99.0	Yes
Class Ic family (TyrRS/TrpRS) (IPR002305)	Reference synthetase	99.8	98.9	99.7	100.0	99.0	100.0	Yes
TrpRS (IPR002306)	Reference synthetase	99.6	98.5	99.7	100.0	99.0	100.0	Yes
TyrRS (IPR002307)	Reference synthetase	99.5	91.6	99.7	100.0	97.1	100.0	Yes
HisRS (IPR015807)	Reference synthetase	99.0	98.3	64.6	97.0	98.5	95.6	Yes
LysRS Class I (IPR002904)	Lys charging, Class I fold	21.4	85.3	0.0	0.0	0.0	1.0	No
LysRS Class II (IPR002313)	Lys charging, Class II fold	81.5	15.2	98.9	100.0	99.0	98.4	Yes
CysRS (IPR015803)	Direct Cys charging	99.1	87.9	96.7	100.0	99.0	99.2	Yes

Marker (InterPro)	Role	Bacteria	Archaea	Fungi	Non-flowering green plants	Flowering plants	Animals	Human
SepRS (IPR005246)	Indirect Cys route	0.0	22.9	0.0	0.0	0.0	0.0	No
SepCysS (IPR013375)	Indirect Cys route	0.0	22.9	0.0	0.0	0.0	0.0	No
PylRS (IPR012739)	Pyrrolysine charging	0.0	4.5	0.0	0.0	0.0	0.0	No
GlnRS (IPR004514)	Direct Gln charging	45.4	0.0	94.1	97.0	94.7	95.4	Yes
AsnRS (IPR004522)	Direct Asn charging	45.0	15.1	98.5	100.0	99.5	99.0	Yes
GatA (IPR004412)	GatCAB subunit A	80.1	84.2	96.0	97.0	98.1	94.6	Yes
GatB (IPR004413)	GatCAB subunit B	81.8	85.9	96.5	93.9	97.6	95.2	Yes
GatE (IPR004414)	GatDE subunit E	0.1	98.5	0.0	0.0	0.0	0.0	No
GatD (IPR011878)	GatDE subunit D	0.1	98.3	0.0	0.0	0.0	0.0	No
TilS (IPR012094)	Lysidine synthetase (C34)	98.3	0.0	84.5	97.0	98.5	1.6	No
TiaS (IPR024913)	Agmatidine synthetase (C34)	0.0	96.7	0.0	0.0	0.0	0.0	No

Marker (InterPro)	Role	Bacteria	Archaea	Fungi	Non-flowering green plants	Flowering plants	Animals	Human
MnmA / Mtu1 (IPR004506)	s^2U34, bacterial/mitochondrial type	98.3	0.0	32.8	93.9	98.1	92.9	Yes
MnmE / Mss1 (IPR004520)	xm^5U34, bacterial/mitochondrial type	76.4	0.0	87.3	100.0	94.2	85.3	Yes
Ctu1/Ncs6 (IPR000541)	s^2U34, archaeal/eukaryotic type	0.4	97.4	95.1	81.8	81.6	84.5	Yes
Elp3 (IPR034687)	xm^5U34, archaeal/eukaryotic type	0.4	93.9	96.2	97.0	94.7	98.4	Yes
GlyRS $\alpha_2\beta_2$ (IPR006194)	Bacterial-type GlyRS	56.6	0.9	2.5	100.0	98.5	5.0	No
PheRS, class IIc, mitochondrial (IPR004530)	InterPro name; monomeric PheRS	0.0	0.0	98.1	81.8	66.5	83.3	Yes
AIMP2 (IPR042360)	Synthetase-complex scaffold family	0.0	0.0	0.0	0.0	0.0	94.4	Yes

Family detection does not establish a complete pathway, activity or cellular location. Reference-marker coverage was 99.4% for AlaRS and 98.4% for MetRS across all proteomes (Section 2.3); the HisRS and TyrRS rows show lower, lineage-dependent coverage. Eukaryotic TilS, MnmA

and MnmE hits may represent organellar enzymes. The non-flowering green-plant proteomes are green algae only. “Reference synthetase” rows are canonical, universally expected families. The full list of 31 markers with InterPro names, and prevalences by phylum and by eukaryotic group including protists, are in Supplementary Tables S2, S5 and S6.

Table 3. Additional tRNA gene features across the reporting groups. % of genomes or median (IQR); enzyme-family prevalences are in Table 2.

Feature	Bacteria	Archaea	Fungi	Non-flowering green plants	Flowering plants	Animals	Human
Genomes with Sec tRNA (%)	26	8	0	55	0	92	Yes
Genomes with initiator fMet tRNA (%)	99	0	0	0	0	0	No
Isotypes carrying introns	0 (0–0)	4 (3–10)	18 (12–19)	6 (3–12)	2 (2–3)	4 (4–5)	4
Main intron-bearing isotypes (% of intron-containing genes)	Ser 27, Phe 22, Leu 16, Pro 7	Trp 10, Met 10, Arg 8, Ser 7	Leu 10, Thr 7, Ala 7, Ser 7	Tyr 18, Met 15, Gly 9, Leu 8	Tyr 55, Met 42	Tyr 39, Leu 21, Arg 18, Ile 16	Tyr 48, Ile 17, Leu 17, Arg 17
Genomes with Leu-AAG (A34) (%)	7	1	87	100	100	99	Yes

Feature	Bacteria	Archaea	Fungi	Non-flowering green plants	Flowering plants	Animals	Human
Genomes with Gly-ACC (%)	0	1	6	0	0	9	No

The initiator row reports the fMet-type annotation, not measured formylation. Sec tRNA detection does not establish a complete Sec pathway; the fungal zero reflects the sampling of fungal phyla (Section 3.6; Mariotti et al. 2019). Bacterial "introns" are largely group I introns or tRNAscan-SE false positives rather than spliceosome-independent tRNA introns of the archaeal/eukaryotic type.

Table 4. Relationship of the findings to earlier work.

Result	Earlier work	This study
Larger eukaryotic tRNA repertoires with similar anticodon richness	Marck and Grosjean 2002; Santos and Del-Bem 2022; Bermudez-Santana et al. 2010	Replicated (medians 40–48 vs 39–45) with one assembly per species; domain-associated anticodon composition (Marck and Grosjean 2002; Wu et al. 2018) quantified across 2,125 species; copies per anticodon reach a median of 15.1 in flowering plants; groups with similar richness differ substantially in copy number and anticodon composition
Local tRNA gene proximity and noncanonical introns	Marck and Grosjean 2002; Sugahara et al. 2008; Fujishima et al. 2010; Fujishima and Kanai 2014	Quantified in 2,125 species; proximity does not establish co-transcription; archaeal noncanonical introns are concentrated in Thermoproteales as reported

Result	Earlier work	This study
Variable tRNA intron content in fungi	Ahammed and van Hoof 2024	The 92% value for *Cryptococcus* is confirmed; extended to 270 fungal species (median 58%) and 18–20 intron-bearing isotypes
Eight-family eukaryotic A34 repertoire and bacterial exceptions	Auxilien et al. 1996; Novoa et al. 2012; Rafels-Ybern et al. 2018, 2019; Katanski et al. 2026	Prevalence quantified; budding yeasts carry seven families; bacterial Leu-AAG and Thr-AGU mainly in Lactobacillales and Mollicutes; no A34 anticodon in Thermotogota or Chloroflexota genomes
Archaeal and bacterial Ile-UAU tRNAs	Mandal et al. 2010; Wu et al. 2018; Taniguchi et al. 2013	Replicated for *Nanoarchaeum* and *Korarchaeum*; the same configuration identified in the Asgard archaeon *Prometheoarchaeum*; TiaS not detected in any of the three proteomes (BUSCO 60.3–88.7%); three bacterial species besides *M. mobile* identified as candidates (Ile-UAU, with neither Ile2-CAU nor TilS detected)
Alternative charging-enzyme families	Woese et al. 2000; O'Donoghue and Luthey-Schulten 2003; Sauerwald et al. 2005; Shaul et al. 2006; Sheppard et al. 2008	Known distributions quantified with exact co-detection counts per group and per phylum; at least one configuration detected in 98.1–99.4% of proteomes per comparison; SepRS/CysRS co-occurrence (known) in 52% of SepRS-containing archaea; charging-enzyme families vary among phyla and classes, whereas TilS and TiaS show strong domain associations; TACK orders differ by LysRS class
Wobble-uridine modification enzymes	Grosjean et al. 2010; Suzuki 2021	Prevalence per group; activity and localization not measured
Sec tRNA distribution	Lobanov et al. 2007; Chapple and Guigó	Agreement at the level of tRNA genes; fungal absence is a sampling result

Result	Earlier work	This study
	2008; Mariotti et al. 2019	
AIMP2 and multisynthetase complexes	Guo et al. 2010; Havrylenko et al. 2011; Havrylenko and Mirande 2015	AIMP2 family in 94.4% of animal proteomes; not detected in nematodes, whose divergent scaffold (Havrylenko et al. 2011) lies outside the coverage of the family markers

641 **Supplementary tables** (in the data archive and the accompanying spreadsheet). **Table S1.**
642 Sensitivity of the tRNA results to four assembly-selection rules. **Table S2.** The 31 InterPro
643 marker families with InterPro names. **Table S3.** Exact co-detection counts for pairs of alternative
644 families. **Table S4.** Taxonomic composition of the genome and proteome sets by bacterial
645 phylum and archaeal class, with tRNA-gene and contig counts for the 18 flowering-plant
646 genomes. **Table S5.** Prevalence of Arg-ACG, Ile2-CAU and Ile-UAU genes and of GlyRS $\alpha_2\beta_2$,
647 LysRS Class I/II, GlnRS, GatD, TilS and TiaS by phylum and class ($n \geq 10$). **Table S6.**
648 Prevalence of all 31 markers by domain and by eukaryotic reporting group, including protists.
649 **Table S7.** Reference-marker detection coverage: proteomes without the AlaRS or MetRS
650 marker, by domain, phylum and class, and proteomes without three or more of the seven
651 canonical marker categories. **Table S8.** Gene-level detail for the exceptions discussed in Section
652 3.3: the three archaeal A34 genes, archaeal Met-CAU copy counts, bacterial genomes with Ile-
653 UAU or without Ile2-CAU cross-referenced to TilS/TiaS status, the six green-plant genomes
654 without Ile-UAU (S8a–S8d), the 17 fungal genomes without Ile-UAU (S8e), and the reference
655 proteomes of the three archaea with Ile-UAU and no Ile2-CAU and of *Caldivirga maquilingensis*
656 (S8f).

Figures

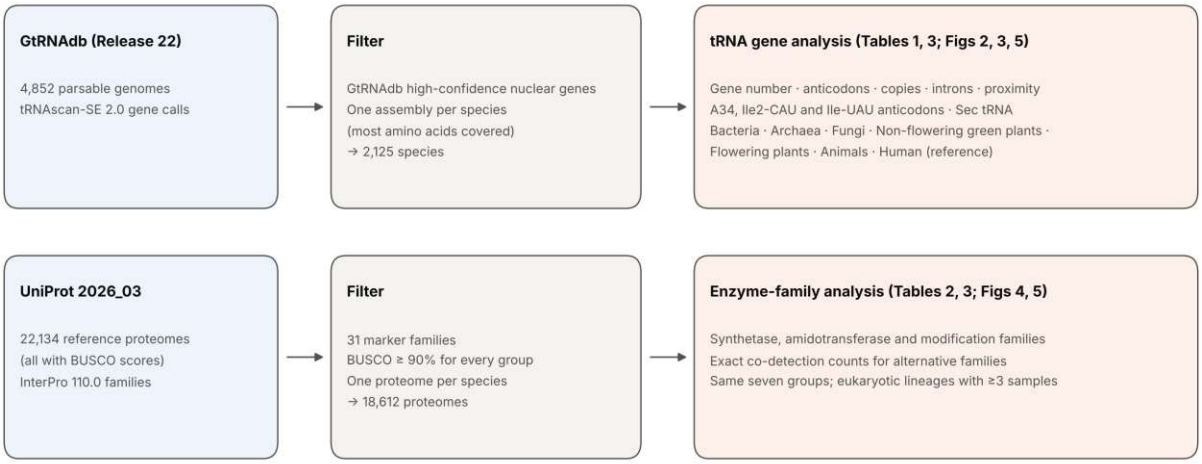


Figure 1. Study design. The tRNA and enzyme-family datasets were filtered separately and summarized in the same reporting groups. Numbers give the records retained at each step. The two datasets are not matched species by species.

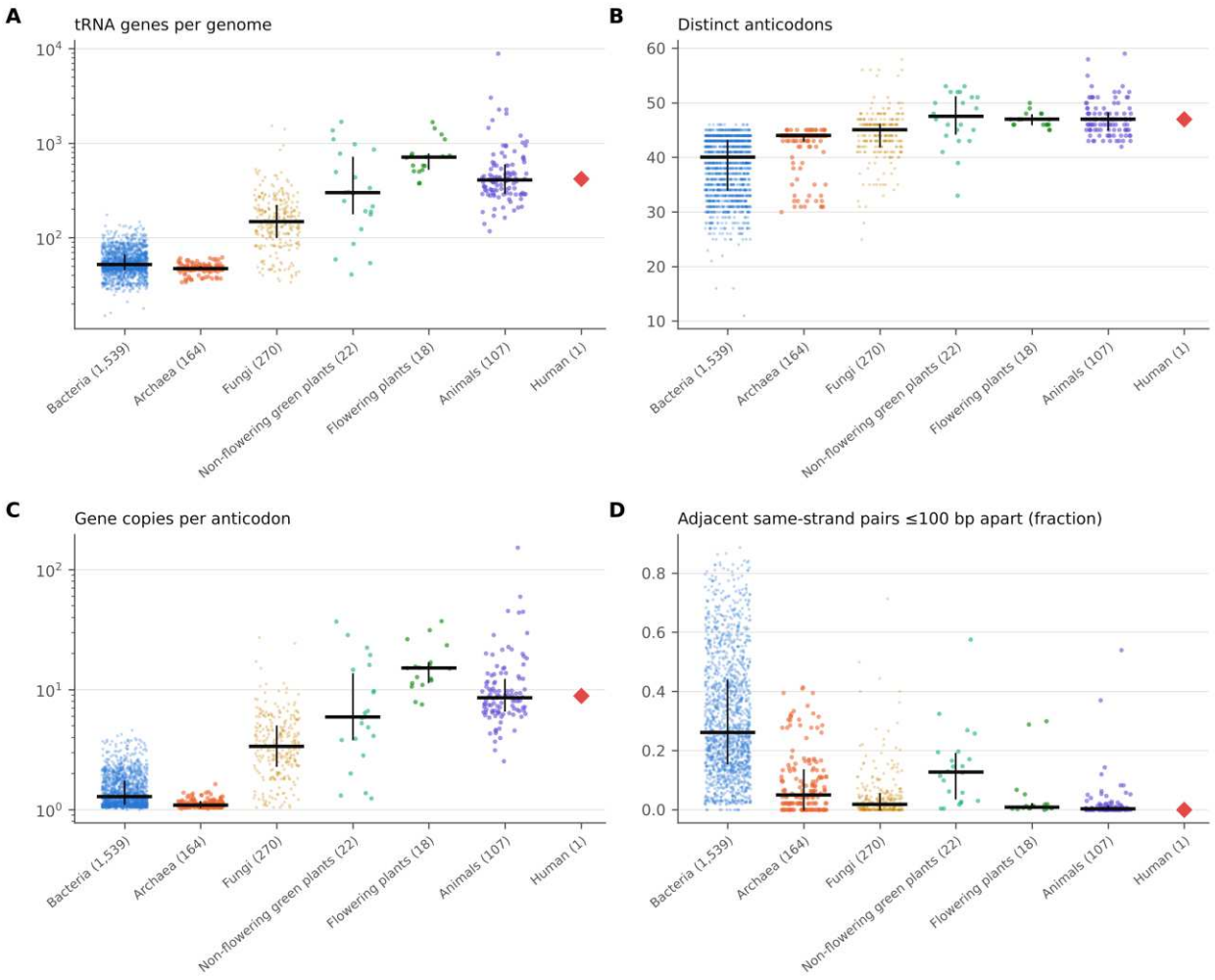
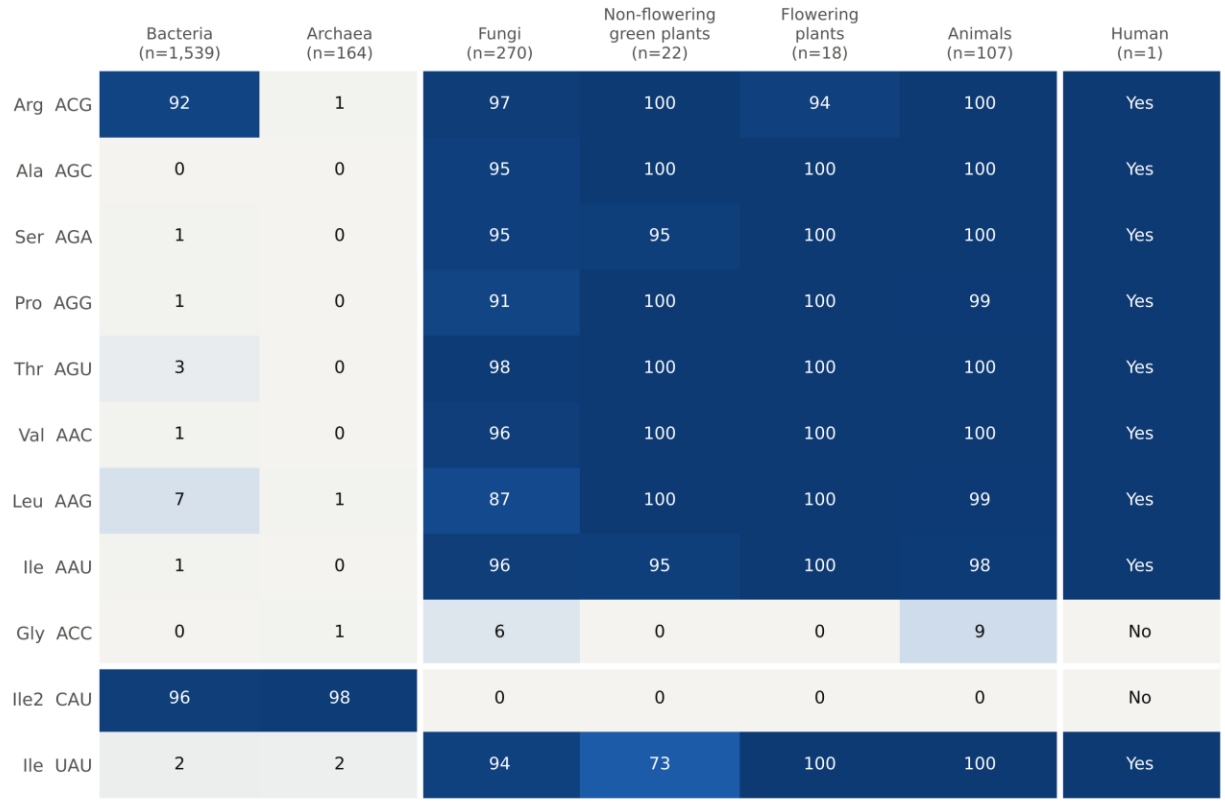
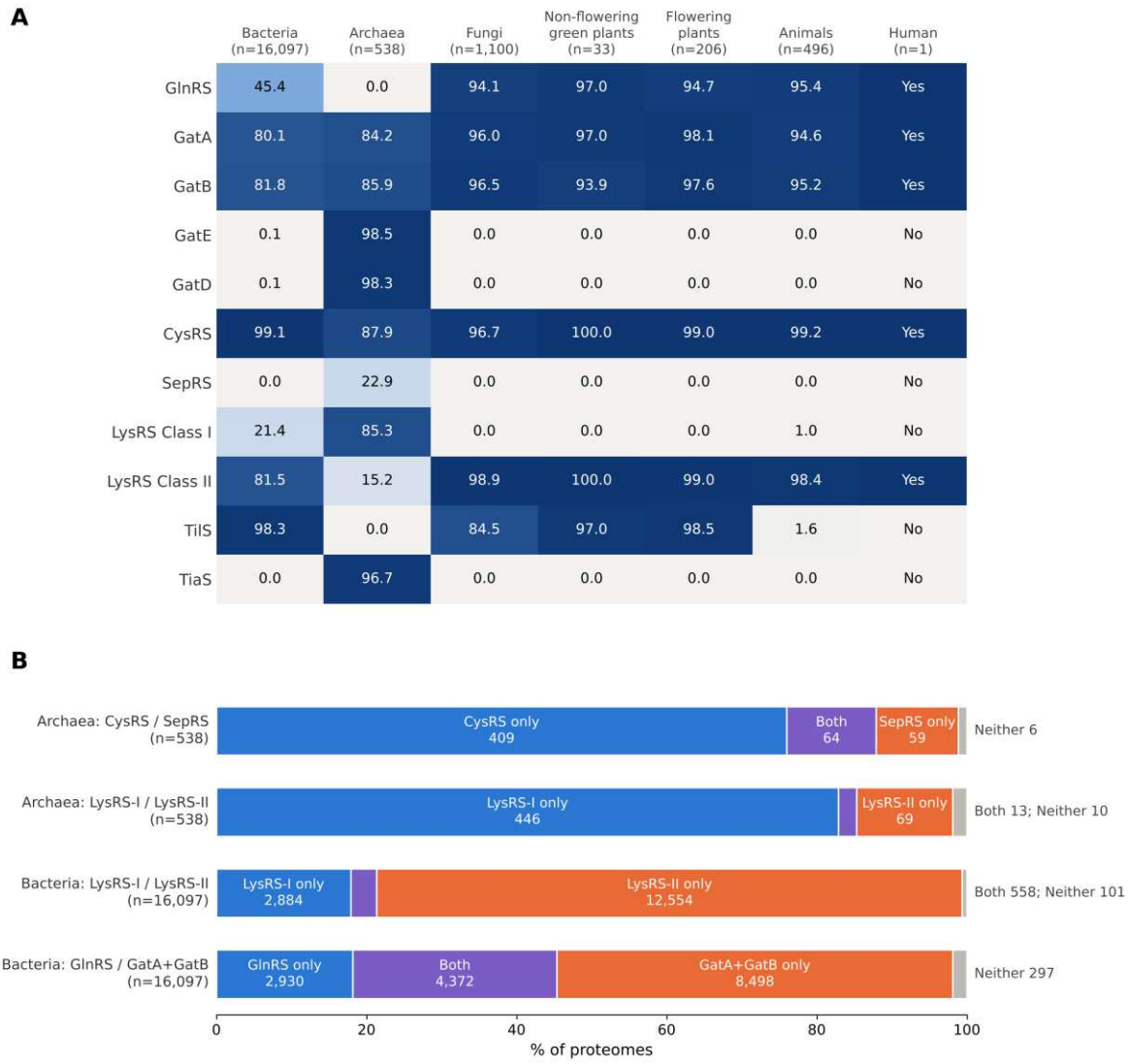


Figure 2. tRNA gene repertoire and genomic proximity across six reporting groups and the human reference. Each point is one species (one assembly per species; number of genomes in parentheses); bars show medians and IQRs; the diamond is the human reference (GRCh37). (A) tRNA genes per genome (log scale). (B) Distinct anticodons. (C) Gene copies per anticodon (log scale). (D) Fraction of adjacent gene pairs on the same strand within 100 bp; the denominator is all consecutive pairs of tRNA genes on the same contig.



669

670 **Figure 3. Encoded anticodons associated with wobble strategies.** Cells give the percentage of
671 genomes with at least one gene carrying the anticodon (RNA notation, 5′ to 3′); human:
672 Yes/No. Upper block: A34 anticodons, potential substrates for inosine formation, in nine
673 families. Lower block: the Ile2 (CAU) tRNA, which requires lysidine or agmatidine at C34 to
674 read AUA, and the UAU-anticodon Ile tRNA. Modification was not measured.



675

676 **Figure 4. Charging-enzyme families across six reporting groups and the human reference.**

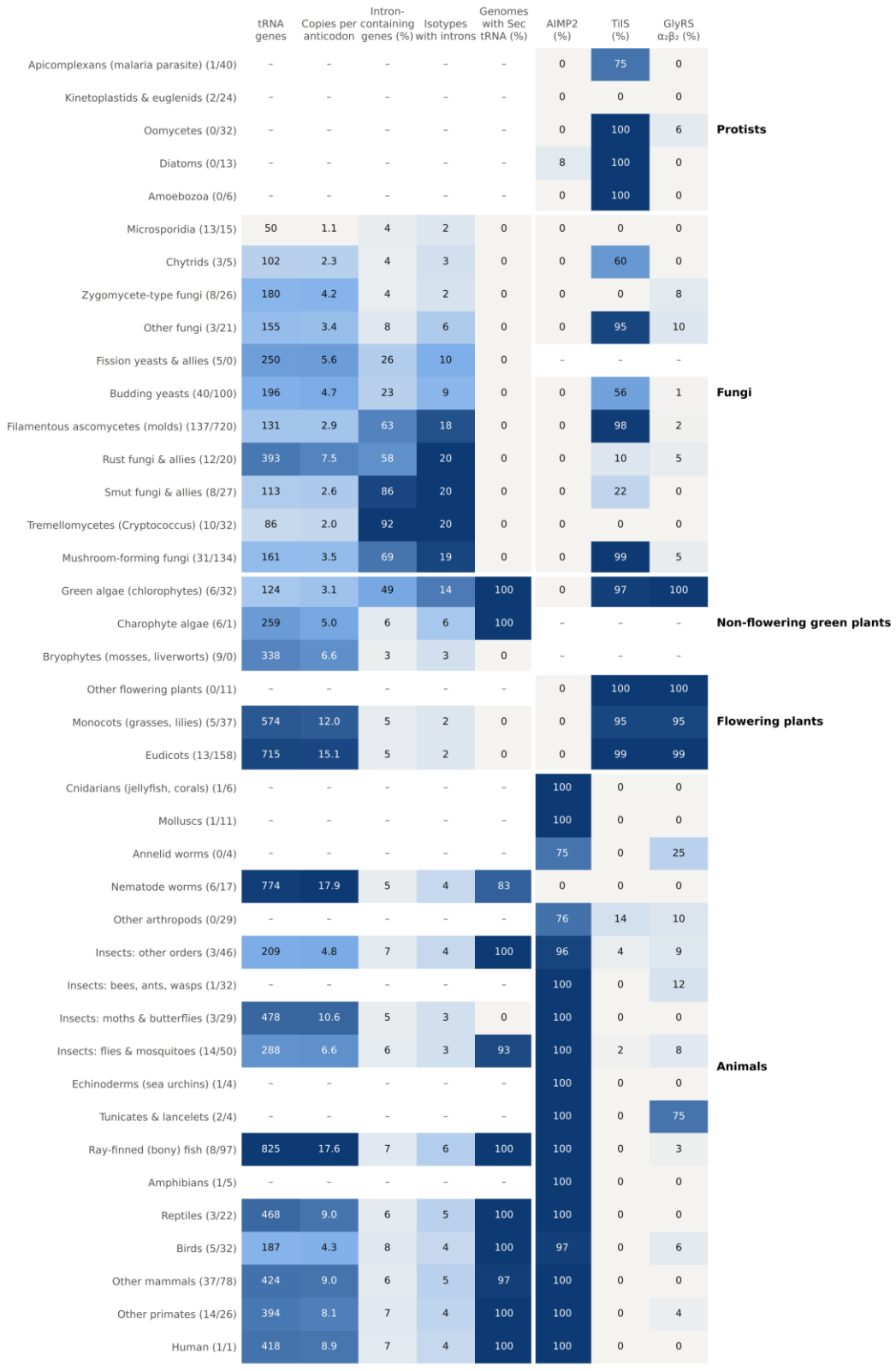
677 (A) Percentage of proteomes containing each family (BUSCO ≥90%; one proteome per species;

678 n given above each column). (B) Exact numbers of proteomes containing one, both or neither of

679 two alternative families. Bacterial indirect Gln charging is scored as GatA plus GatB

680 (IPR004413); the archaeal GatDE route is shown by GatD (IPR011878) and GatE (IPR004414).

681 Family detection does not establish pathway activity or localization.


tRNA genes
Copies per anticodon
Intron-containing genes (%)
Isotypes with introns
Genomes with Sec tRNA (%)
AIMP2 (%)
TilS (%)
GlyRS α2β2 (%)
Apicomplexans (malaria parasite) (1/40) – – – – – 0 75 0
Kinetoplastids & euglenids (2/24) – – – – – 0 0 0
Oomycetes (0/32) – – – – – 0 100 6
Diatoms (0/13) – – – – – 8 100 0
Amoebozoa (0/6) – – – – – 0 100 0
Protists
Microsporidia (13/15) 50 1.1 4 2 0 0 0 0
Chytrids (3/5) 102 2.3 4 3 0 0 60 0
Zygomycete-type fungi (8/26) 180 4.2 4 2 0 0 0 8
Other fungi (3/21) 155 3.4 8 6 0 0 95 10
Fission yeasts & allies (5/0) 250 5.6 26 10 0 – – –
Budding yeasts (40/100) 196 4.7 23 9 0 0 56 1
Filamentous ascomycetes (molds) (137/720) 131 2.9 63 18 0 0 98 2
Rust fungi & allies (12/20) 393 7.5 58 20 0 0 10 5
Smut fungi & allies (8/27) 113 2.6 86 20 0 0 22 0
Tremellomycetes (Cryptococcus) (10/32) 86 2.0 92 20 0 0 0 0
Mushroom-forming fungi (31/134) 161 3.5 69 19 0 0 99 5
Fungi
Green algae (chlorophytes) (6/32) 124 3.1 49 14 100 0 97 100
Charophyte algae (6/1) 259 5.0 6 6 100 – – –
Bryophytes (mosses, liverworts) (9/0) 338 6.6 3 3 0 – – –
Non-flowering green plants
Other flowering plants (0/11) – – – – – 0 100 100
Monocots (grasses, lilies) (5/37) 574 12.0 5 2 0 0 95 95
Eudicots (13/158) 715 15.1 5 2 0 0 99 99
Flowering plants
Cnidarians (jellyfish, corals) (1/6) – – – – – 100 0 0
Molluscs (1/11) – – – – – 100 0 0
Annelid worms (0/4) – – – – – 75 0 25
Nematode worms (6/17) 774 17.9 5 4 83 0 0 0
Other arthropods (0/29) – – – – – 76 14 10
Insects: other orders (3/46) 209 4.8 7 4 100 96 4 9
Insects: bees, ants, wasps (1/32) – – – – – 100 0 12
Insects: moths & butterflies (3/29) 478 10.6 5 3 0 100 0 0
Insects: flies & mosquitoes (14/50) 288 6.6 6 3 93 100 2 8
Echinoderms (sea urchins) (1/4) – – – – – 100 0 0
Tunicates & lancelets (2/4) – – – – – 100 0 75
Ray-finned (bony) fish (8/97) 825 17.6 7 6 100 100 0 3
Amphibians (1/5) – – – – – 100 0 0
Reptiles (3/22) 468 9.0 6 5 100 100 0 0
Birds (5/32) 187 4.3 8 4 100 97 0 6
Other mammals (37/78) 424 9.0 6 5 97 100 0 0
Other primates (14/26) 394 8.1 7 4 100 100 0 4
Human (1/1) 418 8.9 7 4 100 100 0 0
Animals

682

Figure 5. tRNA and enzyme-family features across eukaryotic lineages. n = genomes/proteomes. Columns 1–5 summarize tRNA genes (medians, or % of genomes); columns 6–8 give the percentage of proteomes containing the family. Values are shown only where a lineage had at least three genomes (columns 1–5) or three proteomes (columns 6–8); a dash marks a lineage with fewer than three samples, whereas "0" is a measured zero. Medians are printed as integers in the figure; the exact values for the two algal groups (14.5 and 6.5 intron-bearing isotypes, n = 6 each) are given in Section 3.6. The human reference is shown for comparison. Color is scaled within each column (log scale for the first two). Row order is descriptive and does not imply progression. "Zygomycete-type fungi" denotes Mucoromycota and Zoopagomycota; "Other fungi" the fungal genomes outside the named lineages.

References

Agris PF, Vendeix FA, Graham WD (2007) tRNA's wobble decoding of the genome: 40 years of modification. J Mol Biol 366:1–13. https://doi.org/10.1016/j.jmb.2006.11.046

Ahammed KS, van Hoof A (2024) Fungi of the order Mucorales express a "sealing-only" tRNA ligase. RNA 30:354–366. https://doi.org/10.1261/rna.079957.124

Auxilien S, Crain PF, Trewyn RW, Grosjean H (1996) Mechanism, specificity and general properties of the yeast enzyme catalysing the formation of inosine 34 in the anticodon of transfer RNA. J Mol Biol 262:437–458. https://doi.org/10.1006/jmbi.1996.0527

Bermudez-Santana C, Attolini CS, Kirsten T, Engelhardt J, Prohaska SJ, Steigele S, Stadler PF (2010) Genomic organization of eukaryotic tRNAs. BMC Genomics 11:270. https://doi.org/10.1186/1471-2164-11-270

Blum M, Andreeva A, Florentino LC, Chuguransky SR, Grego T, Hobbs E et al (2025) InterPro: the protein sequence classification resource in 2025. Nucleic Acids Res 53:D444–D456. https://doi.org/10.1093/nar/gkae1082

Boccaletto P, Stefaniak F, Ray A, Cappannini A, Mukherjee S, Purta E et al (2022) MODOMICS: a database of RNA modification pathways. 2021 update. Nucleic Acids Res 50:D231–D235. https://doi.org/10.1093/nar/gkab1083

Chan PP, Lowe TM (2016) GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. Nucleic Acids Res 44:D184–D189. https://doi.org/10.1093/nar/gkv1309

Chan PP, Lin BY, Mak AJ, Lowe TM (2021) tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Res 49:9077–9096. https://doi.org/10.1093/nar/gkab688

Chapple CE, Guigó R (2008) Relaxation of selective constraints causes independent selenoprotein extinction in insect genomes. PLoS One 3:e2968. https://doi.org/10.1371/journal.pone.0002968

Crick FH (1966) Codon–anticodon pairing: the wobble hypothesis. J Mol Biol 19:548–555. https://doi.org/10.1016/s0022-2836(66)80022-0

Eriani G, Delarue M, Poch O, Gangloff J, Moras D (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature 347:203–206. https://doi.org/10.1038/347203a0

724 Francklyn C, Schimmel P (1989) Aminoacylation of RNA minihelices with alanine. Nature
725 337:478–481. https://doi.org/10.1038/337478a0

726 Fujishima K, Kanai A (2014) tRNA gene diversity in the three domains of life. Front Genet
727 5:142. https://doi.org/10.3389/fgene.2014.00142

728 Fujishima K, Sugahara J, Tomita M, Kanai A (2010) Large-scale tRNA intron transposition in
729 the archaeal order Thermoproteales represents a novel mechanism of intron gain. Mol Biol Evol
730 27:2233–2243. https://doi.org/10.1093/molbev/msq111

731 Galli G, Hofstetter H, Birnstiel ML (1981) Two conserved sequence blocks within eukaryotic
732 tRNA genes are major promoter elements. Nature 294:626–631.
733 https://doi.org/10.1038/294626a0

734 Gerber AP, Keller W (1999) An adenosine deaminase that generates inosine at the wobble
735 position of tRNAs. Science 286:1146–1149. https://doi.org/10.1126/science.286.5442.1146

736 Giegé R, Sissler M, Florentz C (1998) Universal rules and idiosyncratic features in tRNA
737 identity. Nucleic Acids Res 26:5017–5035. https://doi.org/10.1093/nar/26.22.5017

738 Grosjean H, de Crécy-Lagard V, Marck C (2010) Deciphering synonymous codons in the three
739 domains of life: co-evolution with specific tRNA modification enzymes. FEBS Lett 584:252–
740 264. https://doi.org/10.1016/j.febslet.2009.11.052

741 Guo M, Yang XL, Schimmel P (2010) New functions of aminoacyl-tRNA synthetases beyond
742 translation. Nat Rev Mol Cell Biol 11:668–674. https://doi.org/10.1038/nrm2956

743 Havrylenko S, Mirande M (2015) Aminoacyl-tRNA synthetase complexes in evolution. Int J
744 Mol Sci 16:6571–6594. https://doi.org/10.3390/ijms16036571

745 Havrylenko S, Legouis R, Negrutskii B, Mirande M (2011) *Caenorhabditis elegans* evolves a
746 new architecture for the multi-aminoacyl-tRNA synthetase complex. J Biol Chem 286:28476–
747 28487. https://doi.org/10.1074/jbc.M111.254037

748 Holley RW, Apgar J, Everett GA, Madison JT, Marquisee M, Merrill SH et al (1965) Structure
749 of a ribonucleic acid. Science 147:1462–1465. https://doi.org/10.1126/science.147.3664.1462

750 Hou YM, Schimmel P (1988) A simple structural feature is a major determinant of the identity of
751 a transfer RNA. Nature 333:140–145. https://doi.org/10.1038/333140a0

752 Ibba M, Söll D (2000) Aminoacyl-tRNA synthesis. Annu Rev Biochem 69:617–650.
753 https://doi.org/10.1146/annurev.biochem.69.1.617

754 Ibba M, Morgan S, Curnow AW, Pridmore DR, Vothknecht UC, Gardner W et al (1997) A
755 euryarchaeal lysyl-tRNA synthetase: resemblance to class I synthetases. Science 278:1119–1122.
756 https://doi.org/10.1126/science.278.5340.1119

757 Ikeuchi Y, Kimura S, Numata T, Nakamura D, Yokogawa T, Ogata T et al (2010) Agmatine-
758 conjugated cytidine in a tRNA anticodon is essential for AUA decoding in archaea. Nat Chem
759 Biol 6:277–282. https://doi.org/10.1038/nchembio.323

760 Imachi H, Nobu MK, Nakahara N, Morono Y, Ogawara M, Takaki Y et al (2020) Isolation of an
761 archaeon at the prokaryote–eukaryote interface. Nature 577:519–525.
762 https://doi.org/10.1038/s41586-019-1916-6

763 Jühling T, Duchardt-Ferner E, Bonin S, Wöhnert J, Pütz J, Florentz C et al (2018) Small but
764 large enough: structural properties of armless mitochondrial tRNAs from the nematode

765 *Romanomermis culicivorax*. Nucleic Acids Res 46:9170–9180.
766 https://doi.org/10.1093/nar/gky593

767 Katanski CD, Pathirage C, Chang JC, Veseli I, Smith TJ, Foo M et al (2026) A transfer RNA
768 inosine modification drives genome-wide synonymous recoding across human commensal
769 bacterial families. PNAS Nexus 5:pgag187. https://doi.org/10.1093/pnasnexus/pgag187

770 Kim SH, Suddath FL, Quigley GJ, McPherson A, Sussman JL, Wang AH et al (1974) Three-
771 dimensional tertiary structure of yeast phenylalanine transfer RNA. Science 185:435–440.
772 https://doi.org/10.1126/science.185.4149.435

773 Lobanov AV, Fomenko DE, Zhang Y, Sengupta A, Hatfield DL, Gladyshev VN (2007)
774 Evolutionary dynamics of eukaryotic selenoproteomes: large selenoproteomes may associate
775 with aquatic life and small with terrestrial life. Genome Biol 8:R198. https://doi.org/10.1186/gb-
776 2007-8-9-r198

777 Mandal D, Köhrer C, Su D, Russell SP, Krivos K, Castleberry CM et al (2010) Agmatidine, a
778 modified cytidine in the anticodon of archaeal tRNA(Ile), base pairs with adenosine but not with
779 guanosine. Proc Natl Acad Sci U S A 107:2872–2877. https://doi.org/10.1073/pnas.0914869107

780 Marck C, Grosjean H (2002) tRNomics: analysis of tRNA genes from 50 genomes of Eukarya,
781 Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. RNA
782 8:1189–1232. https://doi.org/10.1017/s1355838202022021

783 Mariotti M, Salinas G, Gabaldón T, Gladyshev VN (2019) Utilization of selenocysteine in early-
784 branching fungal phyla. Nat Microbiol 4:759–765. https://doi.org/10.1038/s41564-018-0354-9

785 McCutcheon JP, von Dohlen CD (2011) An interdependent metabolic patchwork in the nested
786 symbiosis of mealybugs. Curr Biol 21:1366–1372. https://doi.org/10.1016/j.cub.2011.06.051

787 McCutcheon JP, McDonald BR, Moran NA (2009) Convergent evolution of metabolic roles in
788 bacterial co-symbionts of insects. Proc Natl Acad Sci U S A 106:15394–15399.
789 https://doi.org/10.1073/pnas.0906424106

790 Novoa EM, Pavon-Eternod M, Pan T, Ribas de Pouplana L (2012) A role for tRNA
791 modifications in genome structure and codon usage. Cell 149:202–213.
792 https://doi.org/10.1016/j.cell.2012.01.050

793 O'Donoghue P, Luthey-Schulten Z (2003) On the evolution of structure in aminoacyl-tRNA
794 synthetases. Microbiol Mol Biol Rev 67:550–573. https://doi.org/10.1128/MMBR.67.4.550-
795 573.2003

796 Ogle JM, Brodersen DE, Clemons WM, Tarry MJ, Carter AP, Ramakrishnan V (2001)
797 Recognition of cognate transfer RNA by the 30S ribosomal subunit. Science 292:897–902.
798 https://doi.org/10.1126/science.1060612

799 Perona JJ, Gruic-Sovulj I (2014) Synthetic and editing mechanisms of aminoacyl-tRNA
800 synthetases. Top Curr Chem 344:1–41. https://doi.org/10.1007/128_2013_456

801 Rafels-Ybern À, Torres AG, Grau-Bové X, Ruiz-Trillo I, Ribas de Pouplana L (2018) Codon
802 adaptation to tRNAs with inosine modification at position 34 is widespread among eukaryotes
803 and present in two bacterial phyla. RNA Biol 15:500–507.
804 https://doi.org/10.1080/15476286.2017.1358348

Rafels-Ybern À, Torres AG, Camacho N, Herencia-Ropero A, Roura Frigolé H, Wulff TF et al (2019) The expansion of inosine at the wobble position of tRNAs, and its role in the evolution of proteomes. Mol Biol Evol 36:650–662. https://doi.org/10.1093/molbev/msy245

Santos FB, Del-Bem LE (2022) The evolution of tRNA copy number and repertoire in cellular life. Genes (Basel) 14:27. https://doi.org/10.3390/genes14010027

Sauerwald A, Zhu W, Major TA, Roy H, Palioura S, Jahn D et al (2005) RNA-dependent cysteine biosynthesis in archaea. Science 307:1969–1972. https://doi.org/10.1126/science.1108329

Schimmel P, Giegé R, Moras D, Yokoyama S (1993) An operational RNA code for amino acids and possible relationship to genetic code. Proc Natl Acad Sci U S A 90:8763–8768. https://doi.org/10.1073/pnas.90.19.8763

Shaul S, Nussinov R, Pupko T (2006) Paths of lateral gene transfer of lysyl-aminoacyl-tRNA synthetases with a unique evolutionary transition stage of prokaryotes coding for class I and II varieties by the same organisms. BMC Evol Biol 6:22. https://doi.org/10.1186/1471-2148-6-22

Sheppard K, Yuan J, Hohn MJ, Jester B, Devine KM, Söll D (2008) From one amino acid to another: tRNA-dependent amino acid biosynthesis. Nucleic Acids Res 36:1813–1825. https://doi.org/10.1093/nar/gkn015

Soma A, Ikeuchi Y, Kanemasa S, Kobayashi K, Ogasawara N, Ote T et al (2003) An RNA-modifying enzyme that governs both the codon and amino acid specificities of isoleucine tRNA. Mol Cell 12:689–698. https://doi.org/10.1016/s1097-2765(03)00346-0

825 Sugahara J, Kikuta K, Fujishima K, Yachie N, Tomita M, Kanai A (2008) Comprehensive
826 analysis of archaeal tRNA genes reveals rapid increase of tRNA introns in the order
827 Thermoproteales. Mol Biol Evol 25:2709–2716. https://doi.org/10.1093/molbev/msn216

828 Suzuki T (2021) The expanding world of tRNA modifications and their disease relevance. Nat
829 Rev Mol Cell Biol 22:375–392. https://doi.org/10.1038/s41580-021-00342-0

830 Taniguchi T, Miyauchi K, Nakane D, Miyata M, Muto A, Nishimura S, Suzuki T (2013)
831 Decoding system for the AUA codon by $tRNA^{Ile}$ with the UAU anticodon in *Mycoplasma*
832 *mobile*. Nucleic Acids Res 41:2621–2631. https://doi.org/10.1093/nar/gks1344

833 UniProt Consortium (2025) UniProt: the Universal Protein Knowledgebase in 2025. Nucleic
834 Acids Res 53:D609–D617. https://doi.org/10.1093/nar/gkae1010

835 Weixlbaumer A, Murphy FV, Dziergowska A, Malkiewicz A, Vendeix FAP, Agris PF,
836 Ramakrishnan V (2007) Mechanism for expanding the decoding capacity of transfer RNAs by
837 modification of uridines. Nat Struct Mol Biol 14:498–502. https://doi.org/10.1038/nsmb1242

838 Woese CR, Olsen GJ, Ibba M, Söll D (2000) Aminoacyl-tRNA synthetases, the genetic code,
839 and the evolutionary process. Microbiol Mol Biol Rev 64:202–236.
840 https://doi.org/10.1128/MMBR.64.1.202-236.2000

841 Wu Y, Wu P, Wang B, Shao Z-Q (2018) Genome-wide analysis reveals ancestral lack of
842 seventeen different tRNAs and clade-specific loss of tRNA-CNNs in Archaea. Front Microbiol
843 9:1245. https://doi.org/10.3389/fmicb.2018.01245

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ChengPaper1SupplementaryTablesS1S8v12.xlsx