

# Frequent occurrence and predicted functions of tRNAs with 4-base-pair anticodon stems in bacteria: extended superwobble hypothesis

Fadel Fakh <sup>1,2,†</sup>, Satish Nandipati<sup>1,2,†</sup>, Ambar Kachale<sup>1,2,3,†</sup>, Jiří Heller<sup>1</sup>, Jakub Žváček<sup>4</sup>, Filip Brázdovič<sup>5</sup>, Nawal Al-Chamy<sup>5</sup>, Pragya Tripathi<sup>1,6</sup>, Zdeněk Paris <sup>1,2</sup>, Leoš Shivaya Valášek <sup>5</sup>, Michal H Kolář <sup>4</sup>, Vyacheslav Yurchenko <sup>7</sup>, Marek Eliáš<sup>7</sup>, Eugene V Koonin <sup>8</sup>, Julius Lukeš <sup>1,2,\*</sup>, Anzhelika Butenko <sup>1,2,7,\*</sup>

<sup>1</sup>Institute of Parasitology, Biology Centre, Czech Academy of Sciences, 37005 České Budějovice (Budweis), Czechia

<sup>2</sup>Faculty of Sciences, University of South Bohemia, 37005 České Budějovice (Budweis), Czechia

<sup>3</sup>Present address: Department of Biochemistry, University of Cambridge, CB21GA Cambridge, United Kingdom

<sup>4</sup>Department of Physical Chemistry, University of Chemistry and Technology, 16628 Prague, Czechia

<sup>5</sup>Institute of Microbiology, Czech Academy of Sciences, 14220 Prague, Czechia

<sup>6</sup>Present address: Department of Zoology, Banaras Hindu University, Varanasi 221005, India

<sup>7</sup>Department of Biology and Ecology, Faculty of Science, University of Ostrava, 71000 Ostrava, Czechia

<sup>8</sup>Computational Biology Branch, Division of Intramural Research, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, United States

\*To whom correspondence should be addressed. Email: [anzhelika.butenko@paru.cas.cz](mailto:anzhelika.butenko@paru.cas.cz)

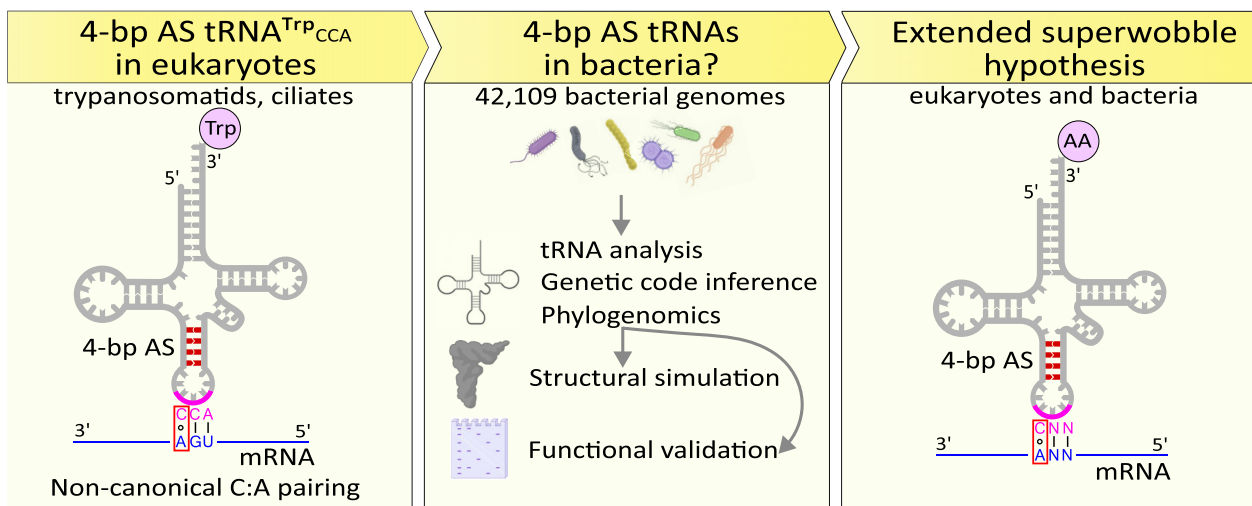
Correspondence may also be addressed to Julius Lukeš. Email: [jula@paru.cas.cz](mailto:jula@paru.cas.cz)

†The first three authors should be regarded as Joint First Authors.

## Abstract

Recently, a tRNA<sup>Trp</sup><sub>CCA</sub> with a 4-base-pair (bp) anticodon stem (AS) was shown to efficiently recognize a near-cognate UGA codon in unicellular eukaryotes, such as some trypanosomatids and ciliates, thereby representing a novel codon reassignment mechanism. To determine whether this mechanism also evolved in bacteria, we analysed a dataset of 42 109 genomes, including previously reported cases of stop-to-tryptophan UGA reassignment and a newly identified instance in the phylum Patescibacteriota. We show that the 4-bp AS tRNA<sup>Trp</sup> species are present across diverse bacteria and in some cases likely function in decoding in-frame UGA codons. Most notable is the endosymbiotic bacterium *Candidatus Zinderia insecticola*, which contains only the near-cognate 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub>, while lacking both canonical 5-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> and a tRNA<sup>Trp</sup><sub>UCA</sub>. The secondary structure of this 4-bp AS tRNA<sup>Trp</sup> resembles that of its eukaryotic counterpart, suggesting convergent evolution. We experimentally confirmed the UGA readthrough capacity of 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> in *Escherichia coli*, and applied molecular dynamics simulations to suggest the underlying mechanism. Furthermore, we tested several predictions based on accepting the previously excluded possibility of C:A base pairing at the 3rd codon position. These findings provide new insights into the structural diversity of transfer RNAs (tRNAs) and expand our understanding of genetic code evolution.

## Graphical abstract



Received: October 30, 2025. Revised: March 13, 2026. Accepted: March 23, 2026

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

## Introduction

Transfer RNAs (tRNAs) are key molecules in the translation process, responsible for delivering amino acids to the ribosomes, where they are incorporated into the growing polypeptide chain. Moreover, tRNAs are now also recognized as participants in translational control, transcription, splicing, retroelement regulation, nutrient sensing, and apoptosis, as well as carriers of intermediates in several biosynthetic pathways [1]. tRNAs are encoded by genes of variable structural organization, including split and circularly permuted ones [2].

Most mature tRNAs in bacteria, archaea, and eukaryotes adopt a highly conserved ‘cloverleaf’ secondary structure consisting of the D, anticodon, variable, T and acceptor arms, and an L-shaped tertiary fold [3–5]. Functional tRNAs typically range from 70 to 100 nucleotides (nt) in length, with variations attributed mainly to differences in the length of the variable and D arms, whereas the other parts of tRNA molecules are highly conserved, with only rare exceptions [6, 7]. Substantial deviations in the tRNA secondary structure have been documented in mitochondria, where the helical regions may vary in length or even be missing, with some of these alterations being reflected at the level of the tertiary structure [8]. Notably, mitochondrial tRNAs exhibit deviations in the number of paired bases within the anticodon stem (AS), demonstrating up to 10-bp AS [8]. The relationship between deviations in the length of structural elements within tRNAs and their functions remains largely unclear [6]. The tRNAs that depart from the structural consensus are generally perceived to be unable to function in translation [3]. The decoding capacity of tRNAs can be affected by several different mechanisms including post-transcriptional modifications, to which ~12% of nucleotides in a typical tRNA are subjected [9, 10].

A minimum set of 32 tRNAs is required to decode 61 sense codons of the genetic code according to the Crick’s ‘wobble rule’ [11, 12]. Under these premises, non-Watson–Crick base pairing, in particular, G:U, U:G, and I (inosine):A/C/U can occur at the wobble (3rd) position of the codon–anticodon mini-helix, allowing for flexibility in translation [12]. The wobble rule was later expanded after the discovery that tRNAs with an unmodified U in the wobble position of the anticodon can pair with any of the four bases at the third position of the codon via a phenomenon called ‘superwobbling’ [13]. More recently, the T-arm sequence in *Mycoplasma mycoides* tRNA<sup>Gly</sup> was identified as a novel pivotal element facilitating superwobbling [14]. Overall, the extension of possible codon–anticodon interactions through wobble and superwobble rules provides a mechanism, by which some bacterial and many organellar genomes, which lack a complete tRNA set, can achieve recognition of all four nucleotides at the third codon position [11].

The flexibility of tRNA decoding has been largely attributed to modifications in the anticodon loop, primarily at the wobble position 34. For example, in human mitochondria, 5-formylcytidine (f<sup>5</sup>C) at this position permits decoding of both AUG and AUA codons by tRNA<sup>Met</sup><sub>CAU</sub> [15]. In bacteria, as well as in plant plastids and mitochondria, tRNA<sup>Ile</sup><sub>CAU</sub> with lysidine (k<sup>2</sup>C) or its derivative 2-aminovaleramididine (ava<sup>2</sup>C) at position 34 favours AUA decoding as isoleucine, while restricting recognition of AUG [16–20]. A prominent group of wobble-base modifications are the xo<sup>5</sup>U derivatives, including the 5-carboxymethoxyuridine (cmo<sup>5</sup>U, also known as uridine-5-oxy acetic acid). These xo<sup>5</sup>U type modifications facilitate

non-Watson–Crick base pairing with guanosine and pyrimidines at the third positions of codons, thereby broadening the decoding range of the associated tRNAs [21–23]. Additional tRNA modifications that enable decoding of the near-cognate stop codons via non-canonical base pairing, while largely maintaining the ability of tRNAs to decode cognate or near-cognate sense codons, were recently reported [10]. Codon recognition by tRNAs can also be expanded by adenosine-to-inosine (A-to-I) edits that in both bacterial and eukaryotic tRNAs enable I<sup>34</sup> to read codons with A, C, and U bases [24]. Moreover, cytidine-to-uridine (C-to-U) editing in kinetoplast mitochondria generates a tRNA<sup>Trp</sup> variant capable of decoding UGA as tryptophan [25, 26].

Not surprisingly, modifications and/or specific point mutations can also alter tRNA’s ability to mediate translational readthrough of stop codons [27]. The so-called suppressor tRNAs with the anticodon mutated such as to enable recognition of stop codons as sense codons (thus, suppressing phenotypes resulting from the occurrence of in-frame termination codons in other genes) have been documented in viruses, bacteria, archaea, and eukaryotes, and are the most common molecular solution for decoding UGA, UAA, or UAG in organisms with an altered genetic code featuring stop-to-sense codon reassignments [28–32]. Point mutations occurring outside the anticodon, e.g. Hirsh mutation G24A and A9C, can also modulate the effectiveness of nonsense suppression [33, 34].

We have previously demonstrated that a relatively minor structural change in tRNA<sup>Trp</sup><sub>CCA</sub> of the trypanosomatid *Blastocrithidia nonstop* and the ciliate *Condylostoma magnum*, namely ‘shortening’ of its canonical 5-bp AS to 4-bp by unpinning the top base pair (26C:42U), dramatically increased UGA stop codon readthrough in the reporter systems in *Trypanosoma brucei* and *Saccharomyces cerevisiae* [35, 36]. Here, ‘shortening’ refers to the disruption of a Watson–Crick base pair. This shortening does not imply collapse of the AS, as one might infer from conventional 2D secondary structure diagrams, but rather a subtle and coordinated modulation of the tRNA structure and dynamics, often in a non-local manner as discussed below.

Importantly, this 4-bp-long AS tRNA<sup>Trp</sup><sub>CCA</sub> is a natural and the only form of tRNA<sup>Trp</sup><sub>CCA</sub> in *B. nonstop* and *C. magnum*, i.e. organisms with UGA reassigned as a tryptophan codon, in the latter species in a homonymous configuration preserving the ancestral function of UGA as stop when occurring at the coding sequence ends [32, 35, 37, 38]. The tRNA<sup>Trp</sup><sub>CCA</sub> with an identical structure, i.e. carrying the 4 bp-long AS, was subsequently identified in the genomes of other eukaryotes with reassigned UGA (fully or homonymously) to tryptophan in their nuclear genomes, namely, ciliates of the genera *Blepharisma* and *Loxodes* [39], other species of the genus *Blastocrithidia* [40], and the dinoflagellate *Amoebophrya* sp. ex *Karodinium veneficum* [41]. Moreover, the expected ‘mutant’ cytosolic tRNA<sup>Trp</sup><sub>UCA</sub> (fully cognate to UGA) could not be identified in *Amoebophrya* sp. ex *Karodinium veneficum*, *Blastocrithidia* spp., and *C. magnum* [35, 37, 40, 42]. Functional association of 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> with the UGA readthrough is further strengthened by the absence of this ‘unpinned’ tRNA in the genomes of ciliates *Stentor* and *Fabrea* (close relatives of *C. magnum*), both of which use UGA solely as a termination codon [39].

In light of these unexpected findings, the results of an early mutational screening of suppressor tRNAs in *Escherichia coli*, which linked 4 bp-long AS with readthrough of the UAA

stop codon, come to prominence. They demonstrated a normally disfavoured C:A base pairing at the tRNA wobble position [43], thereby further expanding the flexibility of decoding by tRNA. Since 1980s, bacteria that interpret in-frame UGA codon as an amino acid are known and new cases are being reported [32, 44]. Point mutations in the anticodon making tRNAs fully cognate to the respective stop codons appear to be the main mechanism of translational readthrough of in-frame stop codons in bacteria. However, a recent report identified a 4-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>* gene in a bacterial endosymbiont possibly hosted by dinoflagellates, providing an alternative putative mechanistic explanation for the stop-to-tryptophan UGA reassignment in this bacterium, where unpinning of the top base pair in the AS rather than anticodon mutation enables translational UGA readthrough [45].

Collectively, these data indicate that structural alterations of the AS can affect tRNA codon specificity in eukaryotes and bacteria, precluding prediction of codon specificity based solely on the Watson–Crick base pairing [46]. Although specific mutations at positions distant from the anticodon loop (AL) are known to change tRNA's specificity [27], the alteration of codon recognition properties via the above-described AS unpinning remains unprecedented.

Here, we analysed tRNAs encoded in the bacterial genomes where, in addition to eukaryotes, multiple instances of stop-to-tryptophan UGA reassignment have been identified ('Graphical abstract' section). Our results showed a surprisingly frequent occurrence of tRNAs with 4-bp-long AS that is not limited to *tRNA<sup>Trp</sup>*. We also demonstrate that the 4-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>* genes are naturally present in some bacterial genomes exhibiting stop-to-tryptophan UGA reassignment and could mediate decoding of UGA. By introducing an engineered 4-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>* into *E. coli*, we show that unpinning of the top base pair in the *tRNA<sup>Trp</sup><sub>CCA</sub>* AS indeed enables the UGA readthrough in this model bacterium.

## Materials and methods

### Identification of tRNAs in bacterial genome assemblies

Bacterial genome assemblies and the associated metadata (e.g. taxonomic affiliation, information on assembly size, GC content, contamination level, etc.) were downloaded from the Genome Taxonomy Database (GTDB) release 214.1 [47]. Genomes designated as representative for each species in the GTDB (typically those of a type strain with high-quality assembly) and demonstrating  $\leq 1\%$  contamination were retained for further analysis. Since a new version of the GTDB was released during the preparation of this manuscript, the taxonomic data have been updated to reflect the changes introduced in the GTDB release 226. Additionally, five genomes of endosymbiotic bacteria absent from the GTDB (*Ca. Carsonella ruddii*, *Ca. Hodgkinia cicadicola*, *Ca. Karelsulcia muelleri*, *Ca. Nasuia deltocephalinicola*, *Ca. Pinguicoccus supinus*), all exhibiting  $\leq 1\%$  of contamination, were downloaded from NCBI. For these genomes, contamination was assessed using CheckM v.1.2.2 [48], with the marker lineage set to Bacteria (UID2495). A total of 42 109 genome assemblies were retained for further analysis (Supplementary Table S1).

The tRNAs genes were predicted using the standalone version of tRNAscan-SE with the '-B' (bacteria) option, and other parameters left as default [49]. Identified genes labelled by

the software as putative pseudogenes and those of undetermined type (tRNA type column value 'undet' and/or anticodon column value 'NNN, NNA, TGR, GTN, or CCR') were excluded from subsequent analyses. Only tRNAs featuring anticodons within positions 33–37 were retained for the analyses and categorized based on their AS length, which was binned into 5-bp, 4-bp, and 'other' categories. When categorizing tRNAs with bulges (i.e. unpaired positions) in the AS, the total number of nucleotides between and including the outermost paired nucleotides, was considered. For example, a tRNA with four paired and one unpaired nucleotide in between in its AS was classified as a 5-bp AS molecule. The categorization of the tRNA structures in the output files of tRNAscan-SE was carried out using an in-house Python script available upon request. A separate category of 'Blastocritidia-like *tRNA<sup>Trp</sup><sub>CCA</sub>*' was defined as comprising tRNAs with 4-bp AS, 7-nt-long AL, and lacking A9C and G24A mutations. *tRNA<sup>Trp</sup><sub>CCA</sub>* were scanned for these features using a regular expression search in the output files of tRNAscan-SE containing secondary structure predictions. For some analyses, tRNAs were grouped based on their percentage out of all identified tRNAs, with those representing  $< 0.1\%$  of the total number of identified tRNAs designated as 'low abundance', and the remaining ones considered 'high abundance'.

For the analysis of *tRNA<sup>Arg</sup><sub>CCU</sub>* in the genomes with arginine-to-methionine AGG reassignment, six additional genomes (accessions GCA\_002404995.1, GCA\_002431755.1, GCA\_002439645.1, GCA\_002436825.1, GCA\_002451385.1, GCA\_002297105.1) were downloaded from NCBI (24 November 2024) and tRNAs were predicted using tRNAscan-SE and ARAGORN v.1.2.41 [50]. The plot demonstrating the correlation between the genome size and the fraction of 4-bp AS tRNAs along with the related statistics (Pearson correlation coefficient and *P*-values) were generated in MS Excel 365 (<https://www.microsoft.com/cs-cz/microsoft-365/excel>; last accessed: 3 September 2025). For the analysis of the presence of *tRNA<sup>Unn</sup>* along with the 4-bp AS *tRNA<sup>Cnn</sup>* or without (i.e. when only the 5-bp AS *tRNA<sup>Cnn</sup>* is present), only bacterial genomes possessing at least one *tRNA<sup>Cnn</sup>* gene were selected considering each of the 11 amino acids encoded by both NNA and NNG codons (Ala, Arg, Gln, Glu, Gly, Leu, Lys, Pro, Ser, Thr, and Val). Since high fraction of bacterial genomes lacks *tRNA<sup>Arg</sup><sub>UCG</sub>* (regardless of the presence of a 4-bp or 5-bp AS *tRNA<sup>Arg</sup><sub>CCG</sub>*), which is explained by the fact that bacteria generally decode the CGA codon with a *tRNA<sup>Arg</sup><sub>ICG</sub>* originating from *tRNA<sup>Arg</sup><sub>ACG</sub>* by post-transcriptional adenine-to-inosine conversion at the first anticodon position, we excluded the *tRNA<sup>Arg</sup><sub>CCG</sub>*/*tRNA<sup>Arg</sup><sub>UCG</sub>* pair from this analysis. In addition, the genomes where the meaning of the NNA and NNG codons differs, were excluded from the analysis. Statistical significance of the observed differences was evaluated using the two-sample proportion Z-test ([https://www.statskingdom.com/121proportion\\_normal2\\_0.html](https://www.statskingdom.com/121proportion_normal2_0.html); last accessed: 3 September 2025).

To determine whether the percentage of UGA stop codons (out of all identified stop codons) was lower in bacterial genomes encoding exclusively 4-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>* compared to those encoding only 5-bp AS tRNA, we analysed a dataset of 2677 bacterial genomes containing only the 4-bp AS variant (Supplementary Table S1). For statistical comparison, we selected bacterial orders containing  $\geq 10$  genomes that exclusively encoded 4-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>*

and compared them to genomes from the same orders that encoded only the 5-bp AS variant. To determine whether there are statistically significant differences in UGA percentages, the Shapiro–Wilk test was used to assess normality (with the significance level of 0.05). Depending on the distribution, either a Mann–Whitney U test (for non-normal distributions) or an independent *t*-test (for normally distributed data) was applied to assess statistical significance (both with the significance level of 0.05). All tests were performed using Statistics Kingdom web page ([https://www.statskingdom.com/121proportion\\_normal2\\_0.html](https://www.statskingdom.com/121proportion_normal2_0.html); last accessed: 3 September 2025).

### Mapping the distribution of the bacterial $tRNA^{Trp}_{CCA}$ variants on the phylogenomic tree

Bacterial phylogenomic reference tree based on the concatenated alignment of 120 proteins was retrieved from the GTDB release 214.1. For visualization purpose, branches corresponding to the species not included in the current dataset were removed from the tree. The tree was visualized using the Interactive Tree of Life (iTOL) web server, and the tRNA mapping was done via the annotation files containing the genome IDs and tRNA variant's presence/absence data [51].

### Comparison of 4- and 5-bp AS tRNA sequence conservation and synteny analysis

To compare sequence conservation levels between the cohorts of 4- and 5-bp AS  $tRNA^{Trp}_{CCA}$  genes, we selected bacterial genomes that contain 4- and 5-bp AS  $tRNA^{Trp}_{CCA}$  genes simultaneously. Using a custom Python script employing MAFFT v.7.505 with the L-INS-i algorithm [52] to perform pairwise sequence alignment (available upon request), we obtained pairwise sequence identity values for the sets of 4- and 5-bp AS  $tRNA^{Trp}_{CCA}$  genes. Synteny analysis was performed with Easyfig v.2.2.5 with the default settings [53].

### Gene expression analysis

For comparing the expression levels of 4- and 5-bp AS  $tRNA^{Trp}_{CCA}$  in the bacterial genomes where the respective genes co-occur, the available RNA-seq data were retrieved from NCBI SRA [54] and European Nucleotide Archive [55]. Data of sufficient coverage were available only for three species: *Amycolatopsis japonica* (RNA-seq read accession: SRX14259790; SRX14259791; SRX14259792), *Bacteroides oleiciplenus* (SRX9506087; SRX9506088), and *Variovorax paradoxus* (SRX7293081; SRX7293082; SRX7293083). RNA-seq reads were adapter- and quality-trimmed using Trimmomatic v.0.39 [56]. The trimming parameters applied were [TruSeq3-PE-2.fa]:2:20:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15. The 'MINLEN' parameter was set to 50 and 30 for paired-end and single reads, respectively. In the case of paired-end reads, only those that remained paired after trimming were used for further analyses. Read quality was assessed with FastQC v.0.11.8 before and after trimming [57]. Trimmed Illumina reads were mapped to the genome assemblies using HISAT2 v.2.2.1 [58] with default settings. The read mappings were sorted using SAMtools v.1.13 [59] and visualized in Integrative Genomics Viewer v.2.17.0 with the allele frequency threshold set to 0.3 [60]. TPM (transcripts per million) values for tRNAs were determined using TPM-Calculator [61] with the default parameters for single reads and the '-p' option for paired-end reads.

To determine whether *Blastocribidia*-like  $tRNA^{Trp}_{CCA}$  genes are expressed in bacteria, we downloaded all publicly available Illumina RNA-seq datasets for the respective genomes from NCBI SRA and European Nucleotide Archive. Only the genomes from bacteria with defined full species names were used for the analysis: *Campylobacter hepaticus* (accession: SRX4915420; SRX4915421), *Campylobacter jejuni* (SRX15890787; SRX15890788; SRX15890789), *Clostridium ljungdablii* (SRX3388079), *Peptacetobacter hiranonis* (SRX5713559), *Portiera aleyrodidarum* (SRX849002), and *Romboutsia ilealis* (ERX546147). The analysis was performed as described above.

### Genetic code inference

The inference of genetic codes for the genomes in the bacterial dataset was carried out using Codetta v.2.0 with default parameters [62]. The initial dataset of genome assemblies demonstrating stop-to-tryptophan UGA reassignment was filtered retaining only the genomes satisfying at least one of the following criteria: (i) suppressor  $tRNA_{UCA}$  is present, (ii) predicted Codetta stop-to-tryptophan UGA reassignment supported by  $\geq 100$  Pfam consensus columns (Supplementary Table S1). The absence of suppressor  $tRNA_{UCA}$  was confirmed with ARAGORN v.1.2.41.

### Validation of a putative stop-to-tryptophan UGA reassignment in the bacterial family JAKLIH01

In addition to the two genomes of bacteria from the family provisionally named JAKLIH01 in the GTDB release 214.1 used for most of our analyses and demonstrating stop-to-tryptophan reassignment according to the Codetta results (Supplementary Table S1), we incorporated two additional genomes assigned to the same family in the GTDB release 226 (accessions: GCA\_028702045.1 and GCA\_028691405.1). The GTDB annotations for these genomes appeared to be incorrect, likely due to gene prediction assuming the standard bacterial genetic code. To address this issue, we predicted protein-coding genes using Prodigal v.2.6.3 [63] with the '-g 4' option.

The phylogenetic position of these four JAKLIH01 genomes was initially assessed using the phylogenomic tree available in GTDB release 226. We then selected 27 additional genomes (Supplementary Table S2) representing a well-supported clade (96% bootstrap support) within the order UBA6257, identified as the closest known relatives of JAKLIH01, and used them for comparative analyses. To confirm the monophyly of the family JAKLIH01, we performed inference of orthologous groups of proteins with OrthoFinder v.2.5.5 (with the options -S blast, -T iqtree, and -M msa, and the remaining options left at their default values) [64] on the genome dataset of JAKLIH01 and their closest relatives. Phylogenomic analysis was performed on 71 proteins encoded by single-copy genes present in at least 80% of the genomes using IQ-TREE v.3.0.1 [65] with automatic model selection and 1000 ultrafast bootstrap replicates. Prior to this, protein sequences were aligned using MAFFT v.7.505 with the L-INS-i algorithm and trimmed with trimAl v.1.4.rev15 using the -gt 0.8 option [66]. The resulting tree was visualized with iTOL.

To assess whether UGA serves as a stop codon in addition to encoding tryptophan, we first identified orthogroups that included at least one protein from a genome with the stop-to-tryptophan UGA reassignment. These proteins were

temporarily excluded from the corresponding FASTA files, and the remaining protein sequences from the genomes with standard genetic code were aligned using MAFFT v.7.505 with the L-INS-i algorithm. Alignments showing no gaps within the last five alignment columns (a proxy for proteins with conserved C-terminal regions) were retained. The corresponding homologues from the reassigned genomes (predicted under the alternative genetic code, where UGA is not treated as a stop codon) were then added back to the respective alignments. After realignment, we examined proteins from the genomes with reassigned UGA for potential C-terminal extensions, which would indicate erroneous conceptual translation of a genuine UGA stop codon as tryptophan leading to an artificial extension of the coding sequence to a downstream UAA or UAG triplets.

The genes for suppressor tRNA<sub>UGA</sub> were identified using tRNAscan-SE and ARAGORN v.1.2.41. The homologues of bacterial release factor 2 (RF2) were searched using the respective sequence of *E. coli* (NCBI accession XQK66079.1) as a query in BLASTp and tBLASTn searches against the predicted proteins and the genome assemblies of JAKLIH01 and their closest relatives, respectively, with an *e*-value threshold set to 1. The putative RF2 candidates were checked by the backward BLASTp searches against the *E. coli* predicted proteins, as well as by inspecting the annotations of the top five hits obtained by BLASTp search against the non-redundant protein sequence NCBI database [67].

### Transformation of *E. coli*

An enhanced green fluorescent (eGFP) reporter with a premature stop (UGA) codon in pBAD plasmid (Invitrogen/Thermo Fisher Scientific) was synthesized by Gene Universal. The wild type (WT) 5-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> gene was amplified from *E. coli* DH10B genomic DNA using primers (forward primer: 5'-GATCGCATGCAAAAAGCCTGCTCGTTGAGCAGGCTTTTCGAATTTGAGGGGCGTAGTTCAATTGGT-3', reverse: 5'-GATCGCTTAGCTGGCAGGGGCGGAGAGA CTC-3') with SphI and BlnI restriction sites, respectively. The polymerase chain reaction product was cloned into pBK plasmid (kindly provided by Nanxi Wang, Nanjing University of Chinese Medicine). The mutant 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> was generated by Q5-driven site-directed mutagenesis according to the manufacturer's protocol (New England Biolabs) to introduce the C27A substitution (forward primer: 5'-TGGTAGAGCAACGGTCTCCAA-3', reverse: 5'-ATTGAACTACGCCCTTGG-3').

### Western blotting

*Escherichia coli* cells were induced with 0.2% L-arabinose and incubated for 4 h at 37°C. Whole cell lysates were separated on Mini-Protein TGX Stain-free gels (Bio-Rad) and transferred to the PVDF membrane. The membrane was probed with mouse anti-GFP antibody (Thermo Fisher Scientific) followed by an anti-mouse IgG peroxidase-labelled rabbit antibody (Sigma-Aldrich/Merck) at 1:1000 and 1:2000 dilutions, respectively. The membrane was re-probed with an anti-*E. coli* OmpA rabbit antibody (Antibody Res) and a secondary anti-rabbit IgG peroxidase-labelled goat antibody (Sigma-Aldrich/Merck) at 1:50 000 and 1:2 000 dilutions, respectively.

Three independent biological replicates were quantified using Image Lab (Bio-Rad) and analysed for each of

the following groups: eGFP-UGA (negative control), eGFP-UGA + C27A 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub>, and eGFP-UGA + WT 5-bp tRNA<sup>Trp</sup><sub>CCA</sub>. To assess the differences in full-length eGFP production among groups, we first performed a one-way ANOVA, which revealed a significant overall effect ( $F_{2,6} = 24.72$ ,  $P = 0.0013$ ). To identify differences among the conditions, we performed independent two-sample Welch's *t*-test for pairwise comparisons. All statistical analyses were performed in GraphPad Prism v.7.03 ([www.graphpad.com](http://www.graphpad.com); last accessed: 3 September 2025) and visualized as mean  $\pm$  standard deviation.

### Northern blotting

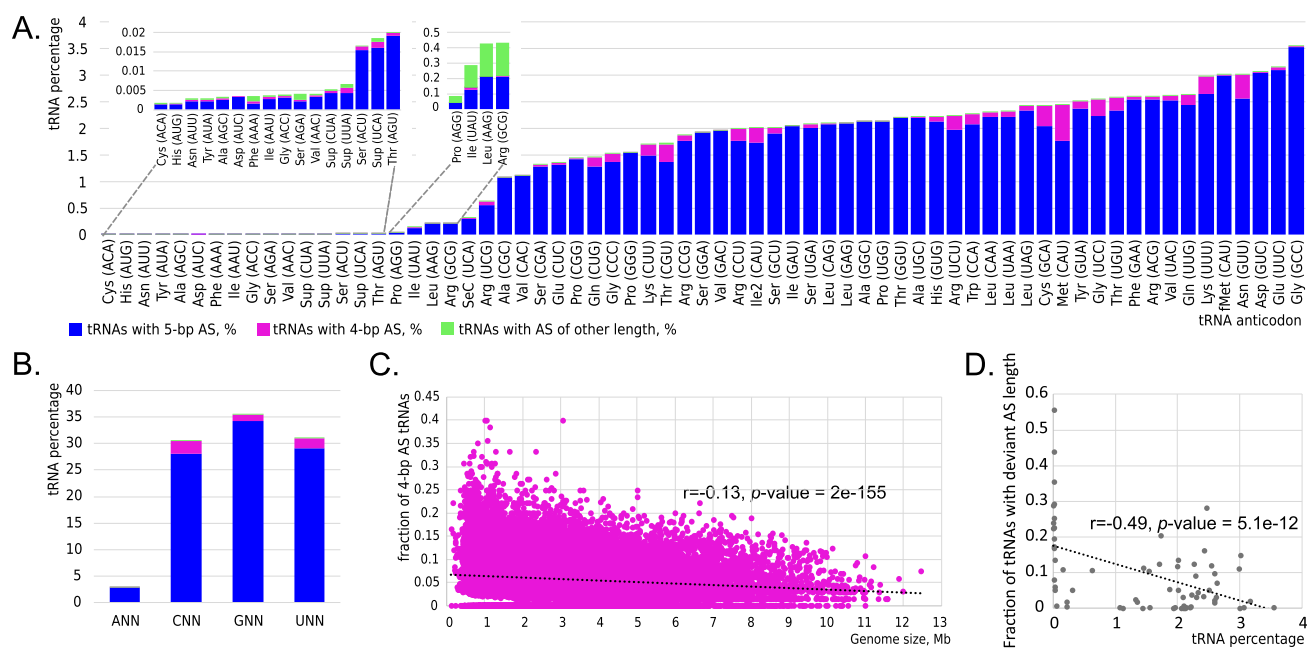
For northern blotting, 10  $\mu$ g of total RNA was denatured in urea loading dye and separated on 8% polyacrylamide gels containing 8 M urea. Electrophoresis was carried out in 1 $\times$  NNB buffer at 100 V for  $\sim$ 2 h. Gels were briefly stained with ethidium bromide to check RNA integrity, and RNA was transferred to Zeta-Probe nylon membranes in 0.5 $\times$  NNB buffer. After UV crosslinking (120 000  $\mu$ J/cm<sup>2</sup>), the membranes were hybridized with an *E. coli* tRNA<sup>Trp</sup> probe 5'-AGGGGCGGAGAGACTCGAAC-3' and a 5S rRNA probe 5'-TGAGTTCGGCATGGGGTCAGG-3' according to the standard protocol (Bio-Rad) and scanned on a Typhoon laser scanner (Cytiva).

### Molecular dynamics simulations of tRNA<sup>Trp</sup><sub>CCA</sub> variants

All-atom molecular dynamics (MD) simulations were performed for the WT tRNA<sup>Trp</sup><sub>CCA</sub> with 5-bp AS, mono-mutated variants unpinned at the top of AS denoted C27A, C27G, C27U, G43A, G43C, and G43U, and two 5-bp AS variants recognising UGA stop codon—Hirsh variant (G24A) and A9C. The initial tRNA tertiary structure was extracted in the A/T state from an *E. coli* X-ray crystallography ribosome-tRNA-EF-Tu model (PDB 4V5R [27]) including the non-canonical nucleotide modifications. The amino acid residue was removed and the nucleobase 9, 24, 27 or 43 was mutated by Open PyMOL v.2.5 using the Mutagenesis Wizard [68].

The isolated tRNA was immersed in a periodic rhombic dodecahedron box of water. The size of the box was chosen such that the distance between the tRNA and any box face was at least 1.2 nm. Sufficient K<sup>+</sup> ions were added to neutralize the system. Additionally, excess concentration of 150 mM KCl and 20 mM MgCl<sub>2</sub> was added to mimic the intracellular environment. Amber  $\chi_{OL3}$  force field [69, 70] and Aduri's force field [71] were used for canonical and non-canonical nucleotides, respectively. The simple point charge extended (SPC/E) water model [72] was used with Joung and Cheatham ion parameters [73].

All systems energy was minimized in 50 000 steps, gently heated to 300 K using 100 000 steps. The initial velocities were drawn randomly from the Maxwell-Boltzmann distribution at 10 K. The pressure was equilibrated to 1 bar using another 100 000 steps with 2 fs time step. The production simulations employed v-sites [74] that allowed for a time step of 4 fs. The temperature of 300 K was maintained by v-rescale thermostat [75], pressure of 1 bar by Parrinello-Rahman barostat [76]. All bonds were constrained to their equilibrium lengths by LINCS. The electrostatics was treated by Particle Mesh Ewald method [77] with a direct-space cut-off of 1.0 nm and van der



**Figure 1.** Distribution and properties of tRNA genes in bacteria. **(A)** Percentage of genes for tRNAs with the AS of various lengths in the total tRNA gene pool. **(B)** Percentage of genes for tRNAs with ANN, CNN, GNN, and UNN anticodons in the total tRNA gene pool. **(C)** Fraction of genes for 4-bp AS tRNAs (number of 4-bp AS tRNAs/total number of identified tRNAs) as a function of the genome size. The trend line demonstrates a negligibly weak negative correlation ( $r$ —Pearson correlation coefficient). **(D)** Fraction of tRNA genes with a deviant (i.e. different from 5 bp) AS (number of a certain tRNA species with a deviant AS/total number of identified tRNAs of the same species) as a function of a percentage of a given tRNA species out of all identified tRNAs.

Waals interactions were described by Lennard–Jones potential with a cut-off of 1.0 nm.

Twelve independent production trajectories, 250 ns-long each, were generated for each variant, differing in the initial set of velocities. All simulations were done in GROMACS 2021 [78]. The number of hydrogen bonds (H-bonds) was calculated by the *gmx hbond* tool with the default settings, namely with the donor–acceptor distance threshold of 0.35 nm and  $30^\circ$  as an angular threshold. The backbone root-mean-square deviations (RMSDs) were calculated using *gmx rms* tool after the least-square alignment of the backbone atoms to the initial tRNA conformation. The root-mean-square fluctuations (RMSFs) were calculated using *gmx rmsf* for all residues after the least-square alignment that involved non-hydrogen atoms excluding the stem residues 27–43. The RMSF profiles were averaged over the 12 independent trajectories. The relative  $\Delta$ RMSF in % was obtained as  $100 \cdot [\text{RMSF}(\text{mutant}) - \text{RMSF}(\text{WT})] / \text{RMSF}(\text{WT})$ . The analysis was done for 150–250 ns trajectory fragments, excluding the initial portions used for system equilibration. The  $P$ -values were calculated from the Mann–Whitney U test for sets of 12 values, representing each of the trajectories.

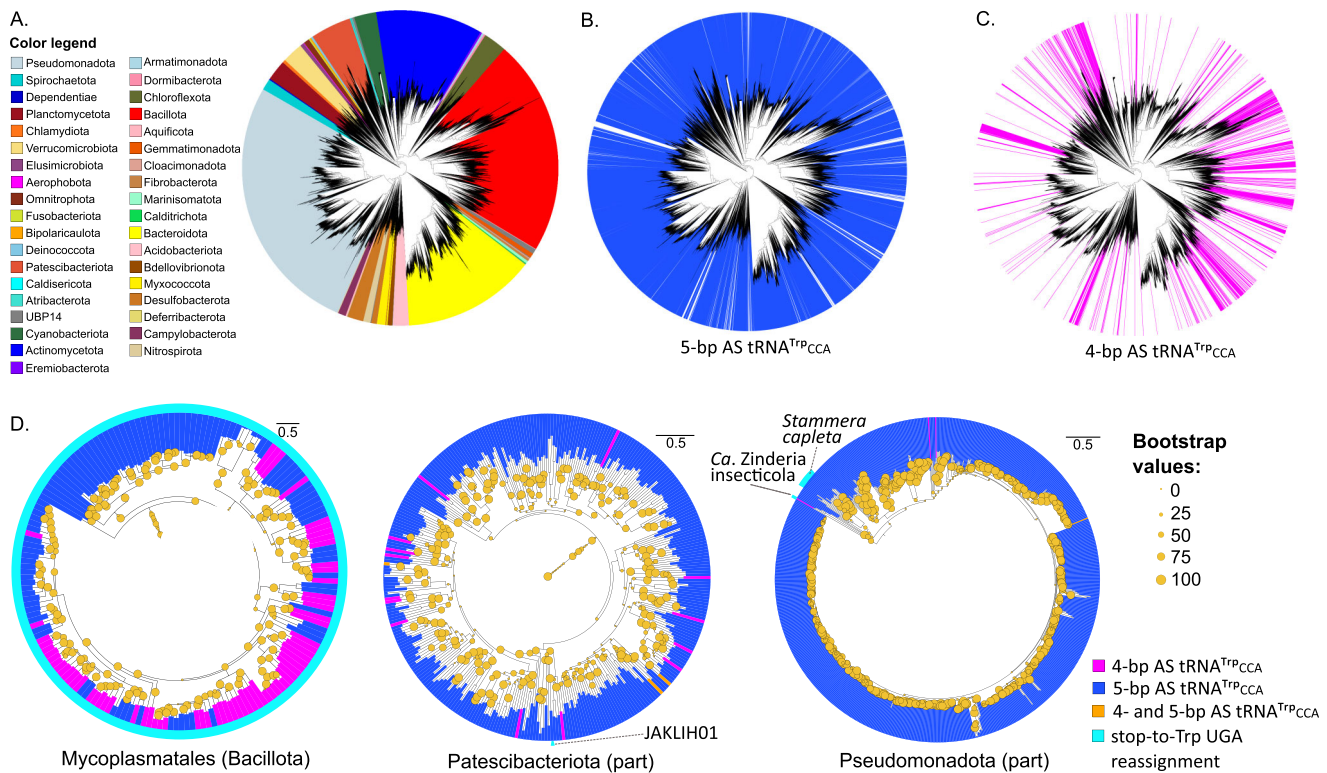
## Results

### tRNAs with 4-bp anticodon stem are widespread in bacteria

We conducted a systematic survey of tRNAs with a 4-bp AS beyond  $tRNA^{\text{TRP}}$  in bacteria, initially disregarding whether unpinning occurred at the very top or very bottom part of the AS. An analysis of 42 109 genomes, encompassing over 150 classes of bacteria as defined by the

GTDB [47], led to the identification of  $\sim 2.0$  million tRNAs (Supplementary Table S1). Following the removal of putatively truncated tRNA genes,  $\sim 1.7$  million tRNAs were analysed further. Bacterial genomes contain a median number of 39 tRNA genes (including copies of the same gene) with an average length of 77 nt (Supplementary Table S1). The most abundant tRNA genes in bacteria (here and below, abundance refers to the percentage of the respective genes out of all identified tRNAs) are those encoding  $tRNA^{\text{Gly}}_{\text{GCC}}$ , accounting for  $\sim 3.5\%$  of the whole tRNA gene dataset, whereas  $tRNA^{\text{His}}_{\text{AUG}}$  and  $tRNA^{\text{Cys}}_{\text{ACA}}$  are the least abundant, each at only 0.002%, and, thus, absent in most bacterial genomes (Fig. 1A). Overall, the  $tRNA_{\text{ANN}}$  species are among the least abundant, collectively making up just  $\sim 2.9\%$  of the tRNA pool, whereas  $tRNA_{\text{GNN}}$  comprise 35.4%, representing the most abundant tRNAs in the dataset (Fig. 1B).

We categorized bacterial tRNAs into three groups based on the length of the AS predicted using tRNAscan-SE: (i) 5-bp-long AS (1 596 864 tRNAs; 94.2%), (ii) 4-bp AS (95 748; 5.6%), and (iii) other AS lengths—2, 3, 6, and 7-bp (3136; 0.2%) (Supplementary Table S3). Only a negligible negative correlation ( $r = -0.13$ ,  $P$ -value =  $2e-155$ ) was observed between the fraction of genes encoding 4-bp AS tRNAs (number of 4-bp AS tRNAs/total number of identified tRNAs) and the bacterial genome size (Fig. 1C). When each tRNA species was considered separately, there was a moderate negative correlation ( $r = -0.49$ ,  $P$ -value =  $5.1e-12$ ) between the fraction of genes for tRNAs with a deviant AS length (number of a certain tRNA species with the AS different from five nucleotides/total number of identified tRNAs of the same species) and a percentage of a given tRNA gene in a total tRNA gene pool (Fig. 1D). Yet, only a weak negative correlation ( $r = -0.18$ ,  $P$ -value =  $1.4e-8$ ) was observed when exclusively 4-bp AS



**Figure 2.** Distribution of 4- and 5-bp  $tRNA^{Trp}_{CCA}$  variants across bacterial phylogeny. **(A)** Visualization of the bacterial phylogenomic tree from the GTDB depicting evolutionary relationships among bacteria. Only the 37 largest phyla, based on the number of species in the dataset, are shown. **(B)** The distribution of 5-bp  $tRNA^{Trp}_{CCA}$  genes across bacteria. **(C)** The distribution of 4-bp AS (unpinned at the top or bottom)  $tRNA^{Trp}_{CCA}$  genes across bacteria. **(D)** The distribution of 4-bp  $tRNA^{Trp}_{CCA}$  in some of the lineages where stop-to-tryptophan UGA reassignment has been documented, including the order Mycoplasmatales (phylum Bacillota) and particular lineages of the phyla Patescibacteriota and Pseudomonadota. For the latter two phyla, a broader part of the phylogenetic tree that contains the species with stop-to-tryptophan UGA reassignment and their relatives with the standard code is shown. We note that the Pseudomonadota tree does not display some of the cases with the same reassignment, including *Ca. Hodgkinia cicadicola* and *Ca. Nasuia deltocephalinicola*, as they are not included in the GTDB and hence are missing from the reference tree used for the display. For the same reason, we do not show data for the known Verrucomicrobiota lineage (e.g. *Ca. Pinguicoccus supinus*) with stop-to-tryptophan UGA reassignment. Species possessing 4- and 5-bp  $tRNA^{Trp}_{CCA}$  are highlighted in magenta and blue, respectively. Species containing both 4- and 5-bp  $tRNA^{Trp}_{CCA}$  are not visible due to their low number and compactness of the trees. Only the genomes with at least one identified  $tRNA^{Trp}_{CCA}$  gene are shown. Bootstrap supports are indicated by the size of the circles. Horizontal line represents 0.5 substitutions per site. Species demonstrating stop-to-tryptophan UGA reassignment are indicated with cyan.

tRNAs were considered. All tRNA types exhibit variability in the length of their AS, with the highest fraction of 4-bp AS tRNAs observed for the elongator  $tRNA^{Met}_{CAU}$  (27.3%),  $tRNA^{Sup}_{UUA}$  (20.5%), and  $tRNA^{Thr}_{CGU}$  (18.5%) (Fig. 1A and Supplementary Table S3).

The genome analysis indicates that unpinning of one of the terminal base pairs in the AS is a widespread, yet previously largely underappreciated deviation from the canonical secondary tRNA structure in bacteria. Although all tRNA species exhibit variations in the number of base pairs within their AS, the prevalence of these deviations is clearly non-random, being much higher in low-abundant tRNA species.

#### Distribution of 4-bp AS $tRNA^{Trp}_{CCA}$ across bacteria

The 4-bp AS  $tRNA^{Trp}_{CCA}$  (with AS unpinned at the top) was experimentally shown to decode the non-cognate UGA in eukaryotic translation systems [35, 36]. The 4-bp AS  $tRNA^{Trp}_{CCA}$  has also been recently proposed, although without experimental proof, to perform the same function in an endosymbiotic bacterium, dubbed XS4, possibly inhabiting a dinoflagellate [45]. Therefore, we first focused on the distribution of the respective tRNA form in bacteria.

In the analysed dataset of 42 109 bacterial genomes,  $tRNA^{Trp}_{CCA}$  comprise 38 615 (2.3%) of all identified tRNA genes, with 3438 (8.9%) of them possessing a 4-bp AS, out of which 3171 have AS unpinned at the top (Supplementary Tables S1 and S3). In contrast to the 5-bp AS variant, the distribution of 4-bp AS  $tRNA^{Trp}_{CCA}$  in bacteria is patchy (Fig. 2A–C), being more prominent in Hydrogenedentota (where exclusively 4-bp AS variant was identified), Campylobacterota, Babelota, Myxococcota, and Cloacimonadota (Supplementary Fig. S1). Overall, out of 142 bacterial phyla in the analysed dataset, 34 (23.9%) contained at least one representative with 4-bp AS  $tRNA^{Trp}_{CCA}$  (Supplementary Table S1). We zoomed onto several groups, where the stop-to-tryptophan UGA reassignment has been reported before or has been newly discovered in our analyses (Fig. 2D). The latter category corresponds to a family-level lineage of the phylum Patescibacteriota referred to as JAKLIH01 in the GTDB release 10-RS226 (see below). Among bacteria with stop-to-tryptophan UGA reassignment, there are species, which contain either 5-bp AS  $tRNA^{Trp}_{CCA}$  (some Mycoplasmatales, JAKLIH01, *Stammera capleta*) or its 4-bp variant (some Mycoplasmatales, *Ca. Zinderia insecticola*, and the uncultivated bacterium XS4 from the family

Fastidiosibacteraceae) (Fig. 2D and Supplementary Table S1) [45].

Among bacterial genomes with the standard genetic code (i.e. not showing signs of any type of codon reassignment), 2677 (~6.5%) encode exclusively *tRNA<sup>Trp</sup><sub>CCA</sub>* with 4-bp AS, indicating that this variant is sufficient for decoding standard UGG tryptophan codons, as has been demonstrated in the eukaryotes examined [35, 36]. In genomes that exclusively encode 4-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>*, the proportion of UGA codons is highly variable, ranging from as low as 1% to as high as 88% of all identified stop codons. At the taxonomic level of an order, analysis of bacterial groups with at least 10 genomes encoding only 4-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>* showed that in 13 of the 19 orders, the proportion of UGA stop codons was significantly lower compared to that in the genomes of bacteria from the same order possessing only the 5-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>* gene (Supplementary Table S4). Thus, the presence of 4-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>* is associated with a decreased UGA usage.

We performed intragenomic synteny analysis for 10 bacterial genera that included genomes encoding either 4- and 5-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>*. Among these, we found three cases in which the 4- and 5-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>* genes were located within a syntenic region, that is, appeared to be orthologous. Thus, at least in some cases, the conversion between 4- and 5-bp seems to occur via a point mutation in the AS (Supplementary Fig. S2).

The 4- and 5-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>* forms co-occur only in a small subset of bacteria with the standard genetic code (296 genomes representing ~0.7% of the analysed species with the standard code) (Supplementary Table S1). To minimize potential biases associated with species-specific evolutionary rates, we compared pairwise sequence identities between 4- and 5-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>* in bacteria encoding both variants (Supplementary Fig. S3). The median pairwise sequence identities were 63.9% and 64.4% for 4- and 5-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>*, respectively, with the difference being statistically significant (*t*-test *P*-value =  $7.2e-40$ ), but negligibly small (Cohen's *d* = 0.09). Thus, purifying selection acts to maintain sequence conservation in both tRNA variants. The analysis of RNA-seq data from several bacterial genomes confirmed that both 4-bp and 5-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>* genes were expressed, with the 4-bp AS variant showing lower expression level (Supplementary Fig. S4).

In addition to 4-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>* unpinned at the very top of the AS, we also identified 4-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>* unpinned at the very bottom, i.e. lengthening the AL by 2 nt (position 30A-38U in *B. nonstop* *tRNA<sup>Trp</sup><sub>CCA</sub>*), which were not subject of this study. Moreover, in any 4-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>* that harbour additional mutations (such as A9C and G24A), which were shown to facilitate UGA readthrough [33, 34], the contribution of 4-bp AS to the overall effect is unknown. Therefore, from here on, we refer to *tRNA<sup>Trp</sup><sub>CCA</sub>* with 4-bp AS and 7-nt AL (i.e. unpinned at the top) and lacking A9C and G24A mutations as the '*Blastocrithidia*-like *tRNA<sup>Trp</sup><sub>CCA</sub>*'. We categorized 3438 4-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>* based on the variation in sequence and structure, excluded 267 tRNAs with AS unpinned at the bottom and 2414 tRNAs carrying A9C and/or G24A, and focused on 757 genes encoding *Blastocrithidia*-like *tRNA<sup>Trp</sup><sub>CCA</sub>* representing 22.0% of all 4-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>* and 2.0% of all identified *tRNA<sup>Trp</sup><sub>CCA</sub>* (Supplementary Table S5). Only in 62 genomes, all with the standard genetic code, *Blastocrithidia*-like *tRNA<sup>Trp</sup><sub>CCA</sub>* co-occurred with the 5-bp AS variant, yet no RNA-seq data were available for these genomes. However, RNA-seq data

are available for several genomes that do not encode the 5-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>* gene along with its 4-bp AS variant (Fig. 3). In these genomes, *Blastocrithidia*-like *tRNA<sup>Trp</sup><sub>CCA</sub>* genes are covered by RNA-seq reads suggesting that they are expressed (Fig. 3). Moreover, analysis of RNA-seq read mapping did not reveal nucleotide substitutions resulting from the RNA editing events in the AS that would restore base pairing between the top unpinned nucleotides or change the anticodon to alter the codon specificity of the tRNA (Fig. 3).

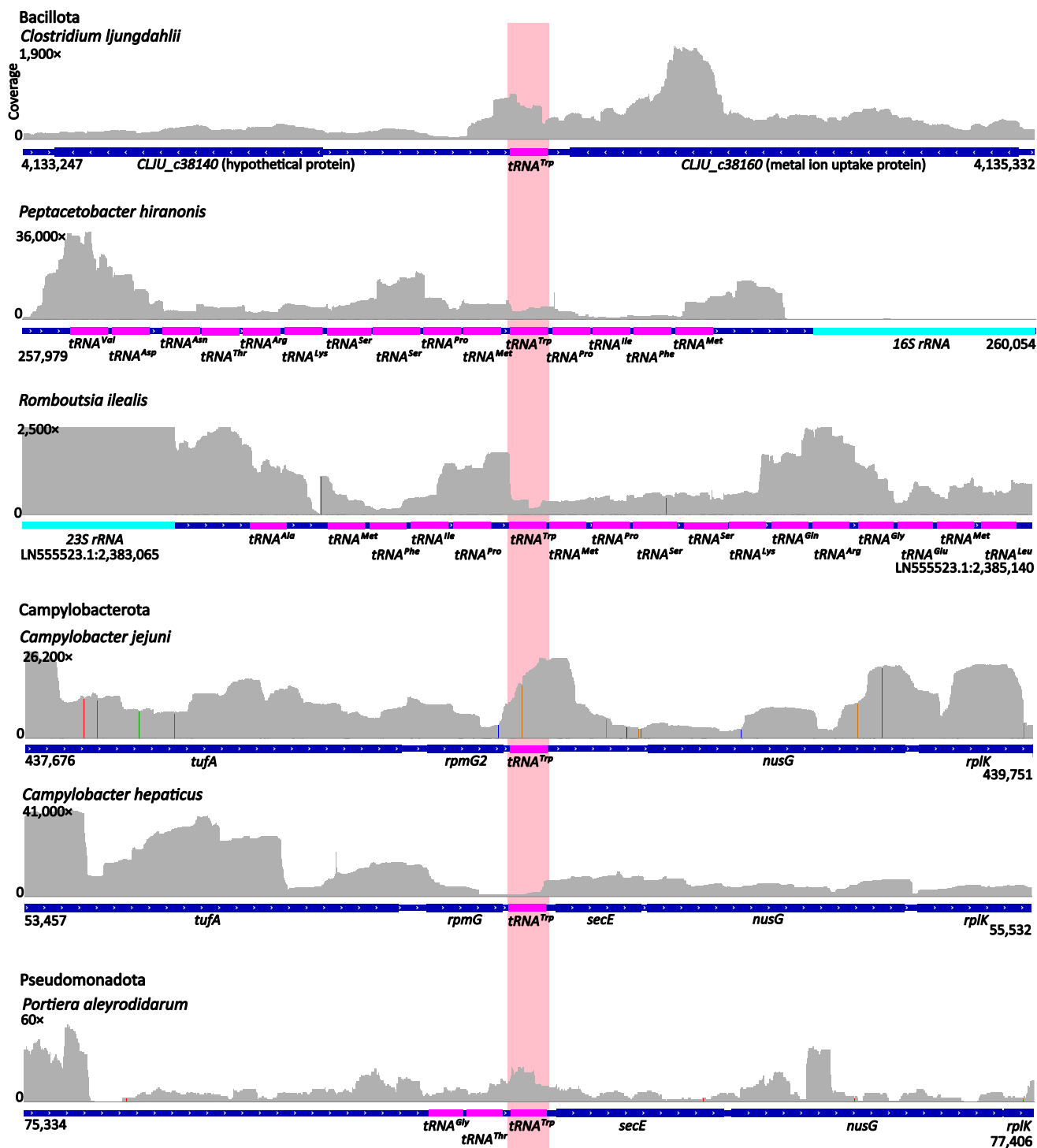
Overall, findings that some genomes encode exclusively 4-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>* and that its sequence conservation level is comparable to that of the 5-bp variant, combined with RNA-seq evidence of their expression, suggest that 4-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>* are functional, with the implication that they are capable of decoding canonical tryptophan UGG codon as tryptophan in bacteria.

### The distribution of 4-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>* across bacteria with alternative genetic codes

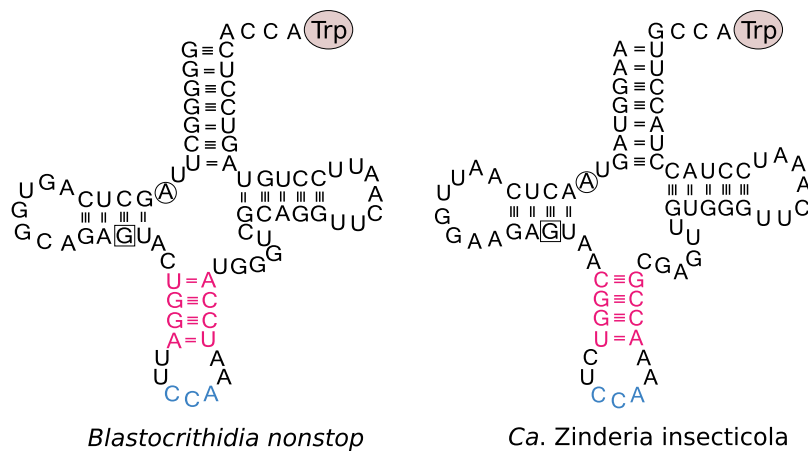
Given that the functionality of *Blastocrithidia*-like *tRNA<sup>Trp</sup><sub>CCA</sub>* was documented in eukaryotes with the stop-to-tryptophan UGA reassignment, we analysed the presence of such tRNAs in bacteria with alternative genetic codes.

The genetic code inference in the bacterial dataset led to the identification of a new independent case of the stop-to-tryptophan UGA reassignment in the phylum Patescibacteriota, specifically in a family-level lineage referred to as JAK-LIH01 and classified in the order UBA6257 and the class Minisyncoccia in the GTDB (release 10-RS226), which was not included in the previous large-scale analysis of codon reassignments in bacteria [44]. It exhibits a strong signal for UGA being decoded as tryptophan when analysed by Codetta (Supplementary Table S2). Analysis of four representatives of this family confirmed that they form a clade in a phylogenetic tree, suggesting that the reassignment in this group was a result of a single evolutionary event (Supplementary Fig. S5). The identification of a *tRNA<sup>UCA</sup>* gene in two out of the four genomes, the absence of RF2 genes, and the absence of UGA stop codons at the ends of coding sequences further support the presence of this reassignment (Supplementary Table S2).

Overall, among the bacterial genomes analysed, 201 (0.5%) were predicted to encompass stop-to-tryptophan UGA reassignment (Supplementary Table S1). These genomes belong to four phyla and jointly include around 30 bacterial genera (Supplementary Table S1). Of these genomes, 70 (34.8%) encode *tRNA<sup>Trp</sup><sub>CCA</sub>* with a 4-bp AS; in all cases, as the only *tRNA<sup>Trp</sup><sub>CCA</sub>* variant. Among these, 11 were *Blastocrithidia*-like *tRNA<sup>Trp</sup><sub>CCA</sub>*, and the remaining ones had deviations from the structure defined above. In 65 of the 70 genomes, 4-bp AS *tRNA<sup>Trp</sup><sub>CCA</sub>* co-occurs with a *tRNA<sup>UCA</sup>* fully cognate to the UGA stop codon (Supplementary Table S1). In the five remaining cases, namely *Candidatus* Zinderia insecticola and four representatives of the order Mycoplasmatales (Supplementary Table S1), we did not identify a *tRNA<sup>UCA</sup>* gene. However, only the genome of the endosymbiotic bacterium *Ca. Z. insecticola* was assembled into a single circular sequence and, thus, is likely complete. It suggests that, at least in this species, the *Blastocrithidia*-like *tRNA<sup>Trp</sup><sub>CCA</sub>* mediates decoding of the in-frame UGA codons as tryptophan in the absence of a fully cognate suppressor tRNA (Fig. 4). The representatives of Mycoplasmatales encoding *Blastocrithidia*-like *tRNA<sup>Trp</sup><sub>CCA</sub>* have fragmented genome assemblies



**Figure 3.** *Blastocrithidia*-like (4-bp AS) *tRNA<sup>Tp</sup><sub>CCA</sub>* are expressed in bacteria. A snapshot of RNA-seq data mapping onto the genome assembly at the *tRNA<sup>Tp</sup><sub>CCA</sub>* locus (highlighted in light pink) and 1000 nucleotides up- and downstream in various bacteria. Coordinates of the displayed region in the genome assembly are indicated (scaffold names are shown only when the assembly contains multiple scaffolds). Protein-coding, tRNA, and rRNA genes are depicted by blue, magenta, and turquoise bars, respectively. The upper track shows RNA-seq read coverage (with the maximum value indicated on Y axis). Nucleotide positions in the reads different from the reference are highlighted in the coverage track (supported by the read fraction  $\geq 0.3$ ): A, green; T, red; G, blue; C, orange. Transcriptomic data from public databases were used to generate coverage plots (see the ‘Materials and methods’ section for details). Products of protein-coding genes: *nusG*, N-utilization substance G; *rplK*, large ribosomal subunit protein uL11; *rpmG2*, large ribosomal subunit protein bL33B; *secE*, secretion protein E; *tufA*, elongation factor Tu 1.



**Figure 4.** Secondary structure of 4-bp AS  $tRNA^{\text{Trp}}_{\text{CCA}}$  from the trypanosomatid *Blastocrithidia nonstop* and the bacterium *Candidatus Zindieria insecticola*. In both cases, a 4-bp AS  $tRNA^{\text{Trp}}_{\text{CCA}}$  is present in the context of stop-to-tryptophan UGA reassignment and the absence of the ‘suppressor’  $tRNA_{\text{UCA}}$ .

(consisting of up to 73 contigs) and the absence of a  $tRNA^{\text{Trp}}_{\text{UCA}}$  gene might be an artefact of incomplete genome assembly. Nonetheless, these species with likely incomplete genome assemblies do not form a clade in the phylogenetic tree, which suggests that if the 4-bp AS  $tRNA^{\text{Trp}}_{\text{CCA}}$ -dependent mechanism for decoding UGA as tryptophan was functional, it must have evolved independently at least four times in Mycoplasmatales. Given that the vast majority of Mycoplasmatales encode a  $tRNA^{\text{Trp}}_{\text{UCA}}$ , it is parsimonious to assume that this particular tRNA form likely represents the shared adaptation underlying stop-to-tryptophan UGA reassignment in this order.

Collectively, these findings, along with the recently published data on the putative dinoflagellate endosymbiont XS4 from the family Fastidiosibacteraceae [45], imply that the stop-to-tryptophan UGA reassignment mechanism relying on a *Blastocrithidia*-like  $tRNA^{\text{Trp}}_{\text{CCA}}$  is not exclusive to eukaryotes but also functions in some bacteria. However, experimental validation is needed to support this observation, as well as to determine whether 4-bp AS  $tRNA^{\text{Trp}}_{\text{CCA}}$  decodes UGA as tryptophan in bacteria where it coexists with  $tRNA^{\text{Trp}}_{\text{UCA}}$ .

#### 4-bp AS $tRNA^{\text{Trp}}_{\text{CCA}}$ enables UGA readthrough in *E. coli*

Previously, in a bacterial system, it has been shown that the naturally unpinned top base pair of the AS of  $tRNA^{\text{Thr}}$  (by the unusual G27-A43 mismatch), when combined with the anticodon substitution from CGU to CUA, mediates the UAG readthrough more efficiently than its engineered 5-bp AS variant [79]. Similarly, mutations unpinning the top AS base pair of *E. coli*  $tRNA^{\text{Trp}}$  at the same time bearing the anticodon mutated from natural CCA to CUA could efficiently read through the UAA stop codons [43]. The ability of the *E. coli* *Blastocrithidia*-like  $tRNA^{\text{Trp}}_{\text{CCA}}$  (i.e. with the preserved CCA anticodon) to read through UGA has been tested in an *in vitro* system. Introducing C27A mutation alone (unpinning the top AS base pair) did not lead to UGA readthrough, however, C27A was essential for the function of the C27A, G24A, G59A triple mutant [80]. To test the impact of the single C27A substitution *in vivo*, we expressed an eGFP reporter with a premature UGA stop codon introduced at a randomly selected amino acid position 29 in *E. coli* in the presence of selected  $tRNA^{\text{Trp}}_{\text{CCA}}$  variants (Fig. 5A). Only

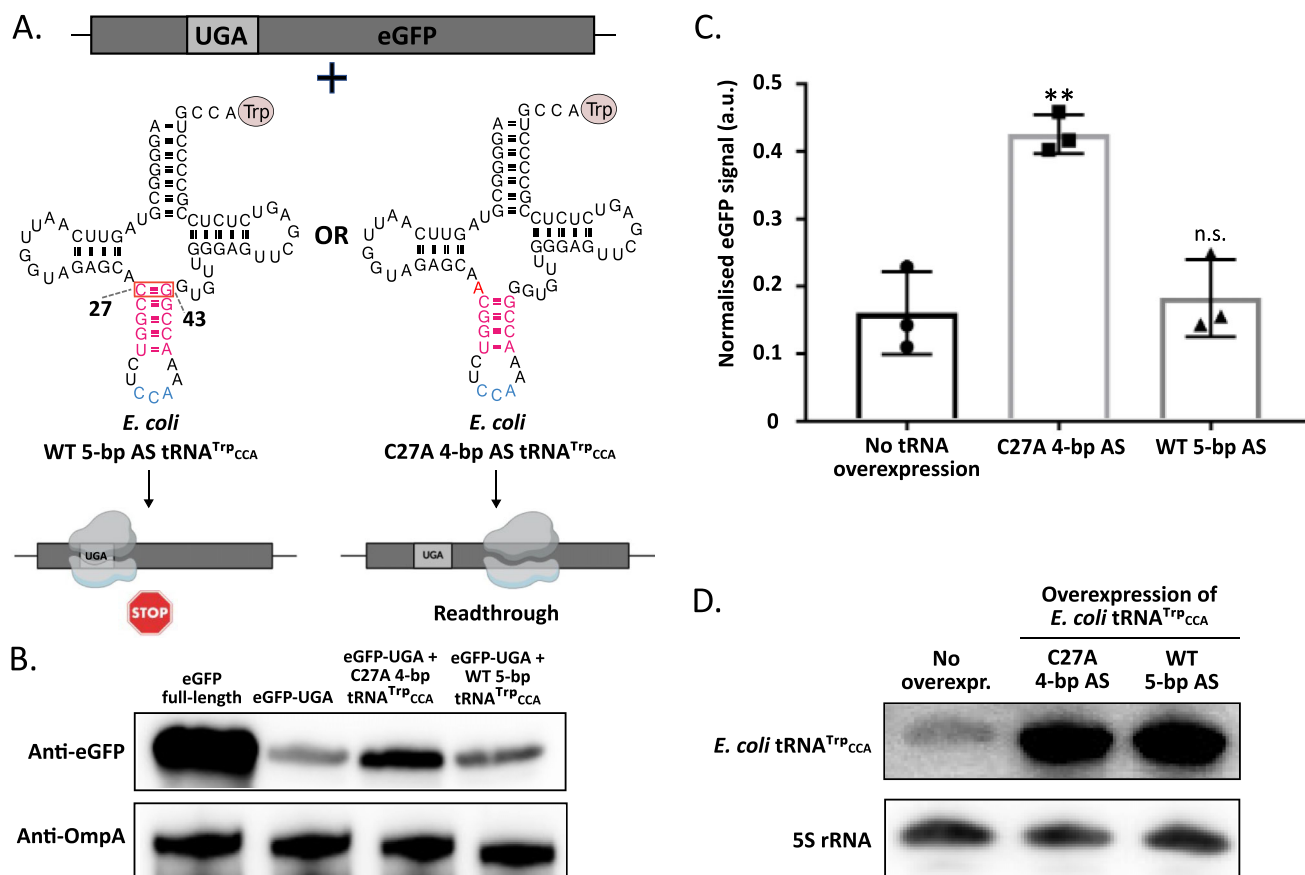
upon efficient UGA readthrough, the full-length eGFP can be made.

While overexpression of the WT 5-bp AS  $tRNA^{\text{Trp}}_{\text{CCA}}$  did not increase the full-length eGFP expression above the ‘no overexpression’ control (Fig. 5B and C), its unpinned 4-bp variant (the C27A substitution) substantially (~2.5-fold) boosted the mutant eGFP expression. The overexpression of both  $tRNA^{\text{Trp}}_{\text{CCA}}$  variants was confirmed by northern blot analysis (Fig. 5D). These findings clearly demonstrate that 4-bp AS  $tRNA^{\text{Trp}}_{\text{CCA}}$  enables higher UGA readthrough compared to the 5-bp AS variant in a prototypical bacterial system, strongly suggesting that the molecular mechanism underlying the action of this tRNA variant, naturally occurring in several species with the stop-to-tryptophan UGA reassignment [35, 39–41, 45], is conserved in at least two domains of life. However, because the overexpression of 4-bp AS  $tRNA^{\text{Trp}}_{\text{CCA}}$  in our reporter assay likely results in levels that exceed its endogenous abundance in bacteria that carry and express this form of tRNA from their genomes, the functional significance of its intrinsic capacity to promote more efficient stop codon readthrough remains to be assessed at native, physiologically relevant expression levels in these specific lineages.

#### Tertiary structure of *Blastocrithidia*-like 4-bp AS $tRNA^{\text{Trp}}_{\text{CCA}}$ variants

To gain insight into the mechanism of UGA readthrough in *E. coli*, we performed atomistic MD simulations of six variants unpinning the top bp of  $tRNA^{\text{Trp}}_{\text{CCA}}$ , including the 4-bp *Blastocrithidia*-like C27A:G43 variant, and compared them with the WT 5-bp variant that disfavours stop-to-tryptophan reassignment (Fig. 5A), and A9C and Hirsh G24A variants that enable UGA readthrough. All tRNAs were analysed in aqueous solution, i.e. outside the ribosome. Note that 2 out of 6 of our 4-bp variants (G43U and C27A) were shown experimentally to mediate UGA readthrough [43, 81].

Initial visual inspection of MD trajectories revealed that the overall tertiary structure of all tRNA variants was maintained. The number of H-bonds of the mutated pair is reduced relative to the WT, but remains non-zero in many cases (Supplementary Fig. S6). Consequently, unpinning a pair in the 2D secondary-structure diagram (Supplementary Fig. S6A) does not mean that H-bonds are absent in the 3D structure (Supplementary Fig. S6B). This might be attributed to the



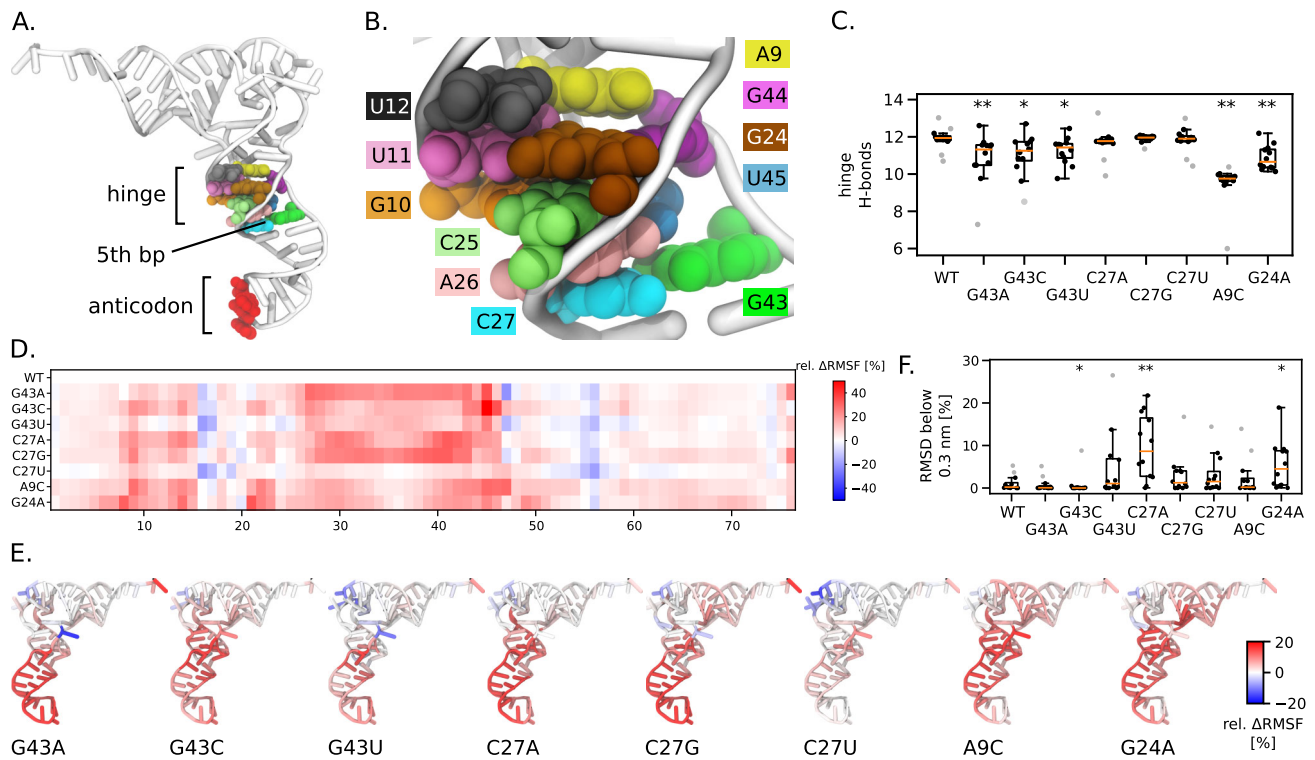
**Figure 5.** A confirmation of the ability of a 4-bp AS  $tRNA^{Trp}_{CCA}$  to enable translational UGA readthrough in *E. coli*. **(A)** Truncated eGFP with an in-frame UGA stop codon expressed in *E. coli* from pBAD plasmid in the presence of arabinose. The eGFP was co-expressed with either the *E. coli* WT 5-bp AS  $tRNA^{Trp}_{CCA}$  or the *E. coli* C27A 4-bp AS  $tRNA^{Trp}_{CCA}$  in pBK to assess the ability of the different versions of the tRNAs to mediate translational readthrough of the UGA codon. The outcome of the experiment predicted by the extended superwobble hypothesis is shown for the two tRNA variants. **(B)** Western blot analysis showing translational readthrough of the in-frame UGA codon. An anti-eGFP antibody is used along with an anti-OmpA antibody as a loading control. In the case of the anti-eGFP antibody, the second lane reflects the UGA readthrough at the background of the endogenous  $tRNA^{Trp}_{CCA}$  level, and the variant lanes show increased signal upon plasmid-driven overexpression of 4- and 5-bp AS  $tRNA^{Trp}_{CCA}$ . **(C)** Quantification of the UGA readthrough in the eGFP construct based on three replicates of the western blot analysis (mean  $\pm$  the standard deviation; the individual values are depicted as black circles, squares, and triangles). Differences were assessed with two-samples Welch's *t*-test for pairwise comparisons (eGFP-UGA + C27A 4-bp  $tRNA^{Trp}_{CCA}$  versus negative control:  $t = 6.76$ ,  $df = 2.86$ ,  $P = 0.0077$ ; eGFP-UGA + C27A 4-bp  $tRNA^{Trp}_{CCA}$  versus eGFP-UGA + WT 5-bp  $tRNA^{Trp}_{CCA}$ :  $t = 6.57$ ,  $df = 2.97$ ,  $P = 0.0074$ ). \*\* $P < .01$ . **(D)** Northern blot analysis showing the overexpression (compared to the endogenous level of 5-bp  $tRNA^{Trp}_{CCA}$  in the first lane) of the two versions of the *E. coli*  $tRNA^{Trp}_{CCA}$  upon transformation with pBK plasmid. The expression is detected using oligonucleotide probes specific to *E. coli*  $tRNA^{Trp}_{CCA}$  and the 5S rRNA. Abbreviations: AS, anticodon stem; n.s., not significant; WT, wild type.

stabilization of the tertiary tRNA structure by base stacking of purine residues and non-canonical H-bonding patterns of the unpinned pair.

Nevertheless, we identified perturbations caused by the unpinning. Notably, the top AS base pair is adjacent to a well-known hinge region at the interface of D arm, AS, and variable arm (Fig. 6A) that has previously been linked to the effects of the mutations A9C and G24A (Fig. 6B) [27]. To probe this connection, we analysed H-bonding within the tRNA hinge defined here by residues 9–12, 23–26, 44, and 45, using the average number of H-bonds as a proxy for thermodynamic stability. Mutations at position G43 increased the variance in the mean H-bond count similar to the Hirsh G24A (Fig. 6C). The loss of two H-bonds was observed for the A9C variant. This destabilization of the hinge region of G43 mutants was accompanied by an increase of the motion of the AS relative to the rest of the tRNA, as reflected by the changes in RMSF (Fig. 6D and E) that stands for a flexibility measure (see the

'Materials and methods' section). The increase in the flexibility of the A9C variant was previously linked to the adoption of the A/T tRNA state in the ribosomal ternary complex, thereby promoting GTP hydrolysis by EF-Tu [27]. We observe the same increase in AS flexibility of all G43 mutants and, therefore, conclude that G43 mutations may be prone to stop-codon readthrough via a mechanism similar to the A9C mutation.

Conversely, the C27 mutations did not alter the H-bonding count of the hinge region (Fig. 6C). Yet, at least the C27A variant supports stop-to-tryptophan reassignment. Inspired by the mechanism proposed for the Hirsh G24A mutation [27], we next focused on the tRNA structure. We calculated backbone RMSD values after superimposing the trajectories onto the initial A/T conformation. This metric reflects the global arrangement of the tRNA during the simulations, and consequently, low RMSD values correspond to stabilization in the A/T conformation required for successful stop codon



**Figure 6.** Atomistic MD simulations of several 4-bp *Blastocryptidia*-like tRNA<sup>Trp</sup><sub>CCA</sub> variants compared to the 5-bp AS WT, A9C and Hirsh G24A variants. (A) Tertiary structure of the tRNA. (B) Context of the unpinned residues C27:G43 at the top of the AS, and the tRNA hinge. (C) Number of hydrogen bonds (H-bonds) within the hinge inferred by MD simulations. The boxes indicate the interquartile range, the red line denotes the median, whiskers extend to the most extreme data points within 1.5× the interquartile range, and grey circles represent outliers. Statistics were calculated from 12 values, each obtained as the average over one of the 12 independent trajectories. \*\**P* < .01, \**P* < .05. (D) The residue-wise ΔRMSF relative to the WT. The colour scale goes from blue (more rigid than WT) through white to red (more flexible than WT). (E) A projection of the relative ΔRMSF onto the tertiary tRNA structure. The initial WT tRNA conformation was used for all projections. (F) A percentage of simulation time, where the tRNA backbone adopted the RMSD lower than 0.3 nm. High percentage means that the tRNA tends to stay in the A/T conformation, which is required for efficient stop codon readthrough.

reassignment. In the aqueous environment, the tRNA expectedly departed from the A/T conformation, but we noticed that the extent of this departure differed between variants. Strikingly, the three variants that sampled the A/T conformation most extensively were C27A, G24A, and G43U (Fig. 6F)—precisely those experimentally demonstrated to mediate UGA readthrough [43, 81]. By contrast, the G43C variant showed the largest departure from the A/T state (i.e. the lowest RMSD), likely precluding efficient readthrough despite its high AS flexibility.

To place our findings in a broader context, we analysed a dataset of 757 genes encoding *Blastocryptidia*-like tRNA<sup>Trp</sup><sub>CCA</sub> as well as the entire dataset of the tRNA<sup>Trp</sup><sub>CCA</sub> with the AS unpinned at the top (Supplementary Tables S5 and S6). The G43A and G43U variants (featuring C:A and C:U unpinned pairs), which our simulations identified as highly flexible, were present in 35.0 and 18.6% of sequences, respectively, and together with G:A unpinned pair represented three most populated variants. Moreover, the analysis revealed a marked ~4-fold increase in the frequency of the C:A unpinned pair in ‘*Blastocryptidia*-like’ tRNA<sup>Trp</sup><sub>CCA</sub> (35%) compared to the entire set of 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> (8.7%) with the AS unpinned at the top (Supplementary Table S6). In contrast, the G43C variant was found in only a single gene (0.1%) among the ‘*Blastocryptidia*-like’ sequences. In simulations of this variant, the hinge exhibited pronounced instability: within 250 ns, U45 detached completely from the hinge, affecting also

m<sup>7</sup>G46 and U47 making this region the major difference between G43C and the two readthrough-competent variants G43A and G43U (Fig. 6D and E).

Mutations at C27 were rare in both the ‘*Blastocryptidia*-like’ and the entire dataset of the 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> (with the frequency of such variants not exceeding 1.1%). C27U was absent in the datasets, which is in line with the MD results in which this variant was the only one that showed no increase in flexibility upon mutation (Fig. 6D and E). These results suggest that the G24A-like mechanism, which relies on pre-stabilization of the A/T tRNA conformation, is evolutionarily more difficult to achieve through mutations near the tRNA hinge.

## Discussion

Being essential for polypeptide chain formation, tRNA molecules retain highly conserved secondary and tertiary structures, with alterations occurring mostly in the variable arm or, less frequently, in the D- or T-stem/loop architecture [82]. The anticodon arm, known for its high conservation level, typically features a 5-bp AS and a 7-nt AL [6]. Recently, however, we reported the presence of a functional tRNA<sup>Trp</sup> species with a 4-bp AS in the trypanosomatid *B. nonstop* and the ciliate *C. magnum*, where this tRNA variant enables UGA stop-to-sense reassignment [35]. So far, the codon recognition

specificity of tRNAs was almost exclusively predicted on the basis of their anticodon [83]. However, this traditional rationale would not work for predicting the codon specificity of 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> [46], and potentially, the other 4-bp AS tRNAs with this (unpinned at the very top) or similar (unpinned at the very bottom) structural features. Here, we investigated the potential decoding capacity of 4-bp AS tRNAs in bacteria.

Although far less abundant than the canonical 5-bp AS tRNAs, 4-bp AS tRNAs (unpinned either at the top or at the bottom of the AS) are surprisingly common in bacteria. We identified this alteration in all tRNA species, with at least one 4-bp tRNA found in 85.6% of the analysed genomes. Thus, unpinning of one of the AS base pairs is a widespread, previously almost entirely overlooked deviation from the canonical tRNA structure in bacteria. The substantial prevalence of 4-bp AS tRNAs strongly suggests their functionality. Indeed, MD simulations revealed that (at least in the case of tRNA<sup>Trp</sup><sub>CCA</sub>) while unpinning the top base pair may reduce the number of H-bonds in the AS, the overall tRNA tertiary structure seems to be maintained by stacking interactions of surrounding bases and/or formation of the non-canonical H-bonding patterns in the unpinned base pair.

Considering that several unicellular eukaryotes carry a 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> to mediate translational UGA readthrough [35, 36, 41], here we specifically focused on the analysis of bacterial 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> to identify similar cases. This variant constitutes 8.9% of all identified tRNA<sup>Trp</sup><sub>CCA</sub> in bacteria, and its substantial subset (22%) is structurally closely similar to the *B. nonstop* tRNA<sup>Trp</sup><sub>CCA</sub>. In bacteria, 4- and 5-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> variants co-occur in the same genome extremely rarely, and, because the conservation levels of 4- and 5-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> in the genomes encoding both variants are comparable and both are transcribed, they appear to be subject to similar selection pressures.

Our analyses of 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> show that certain nucleotide combinations are more frequent in the unpinned position at the top of the AS, suggesting some sequence constraints at this site. Although the three-nucleotide anticodon of a tRNA is the primary determinant for base pairing with an mRNA codon, the ability of a tRNA to recognize and translate that codon is not dictated by the anticodon alone. In particular, it was previously proposed that residues on the 3' side of the anticodon (positions 37 and 38), as well as base pairs within the AS evolved in correlation with the identity of the cardinal anticodon nucleotide at position 36. Indeed, experiments aimed at testing this hypothesis demonstrated that stop codon suppression efficiency can be improved by modifying the AS-loop sequence to better match the identity of the cardinal nucleotide according to these rules [84–86]. In addition, *in vitro* biochemical analyses of bacterial tRNA binding to the ribosomal A site showed that the identity of 32 and 38 bases (pairing only in some tRNAs, and constituting an open AL in others) correlates with anticodon identity, similarly fine-tuning the interaction of each tRNA with the decoding centre in the ribosome [87]. Disruption of this correlation impaired ribosomal acceptance of cognate tRNAs and compromised translational fidelity, further highlighting the importance of coordinated alignment between the anticodon and the surrounding free bases and structural elements such as the AS. Structural studies of these tRNA variants bound to the ribosome reinforced this concept even further [88]. In line with these reports, our observations on the frequency of the nu-

cleotide combinations at the top of the AS combined with the MD simulations support a functional link between the identity of the unpinned base pair, the architecture of the AS-loop, and the propensity of tRNAs to promote stop or near-cognate codon readthrough. However, the situation seems to be highly complex, and additional structural studies will be required to fully uncover the molecular basis for the increased UGA codon readthrough by the 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> variants.

The distribution of 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> genes in bacteria implies that the ancestral 5-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> morphed into the 4-bp AS variant by means of point mutations, which is supported by the observation that the two variants are found within syntenic genomic regions. Moreover, in some bacterial lineages, the transition between 4- and 5-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> occurred notably more frequently than in others. Multiple cases of independent emergence of the 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> variant in lineages with the stop-to-tryptophan UGA reassignment (exemplified by Mycoplasmatales) suggest that it confers increased fitness, possibly, through augmenting the UGA to tryptophan recoding via the cognate tRNA<sup>Trp</sup><sub>UCA</sub>. An alternative explanation is that bacteria in which UGA encodes tryptophan tolerate the decoding of UGA by the 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> more readily than bacteria that use UGA as a stop codon, because in the latter the 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> may interfere with proper translation termination by competing with RF2.

In bacteria without the stop-to-tryptophan UGA reassignment, possessing exclusively 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> would be deleterious unless RF2 outcompetes this tRNA at the UGA codons. We identified a statistically significant reduction in the UGA stop codon usage in genomes encoding only 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> compared to their relatives possessing solely the 5-bp AS variant for 13 of the 19 analysed bacterial orders. Nevertheless, the fact that all 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub>-containing genomes still use UGA as a stop codon suggests that their RF2 is a sufficiently strong competitor of 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> for UGA recognition. Alternatively, some unknown modifications to this tRNA might prevent its interaction with UGA.

It has been previously demonstrated that unpinning the top base pair of the AS can alter the codon–anticodon interaction in bacteria [43, 81]. Here, we showed experimentally that 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> mediates higher UGA readthrough in *E. coli*, when compared to WT tRNA<sup>Trp</sup><sub>CCA</sub>, which further underscores the capacity of 4-bp AS tRNAs to superwobble in a bacterial system. Notably, a previous *in vitro* selection of *E. coli* tRNA<sup>Trp</sup><sub>CCA</sub> using UGA-programmed 70S ribosomes showed no effect of C27A alone (unpinning the top base pair) on UGA decoding [80]. These different outcomes likely reflect differences in an *in vitro* versus *in vivo* experimental setups. Consistent with the functional impact, our MD analysis indicates that C27A stabilizes the A/T conformation of the tRNA required for efficient GTP hydrolysis in the ribosome–EF-Tu–tRNA ternary complex.

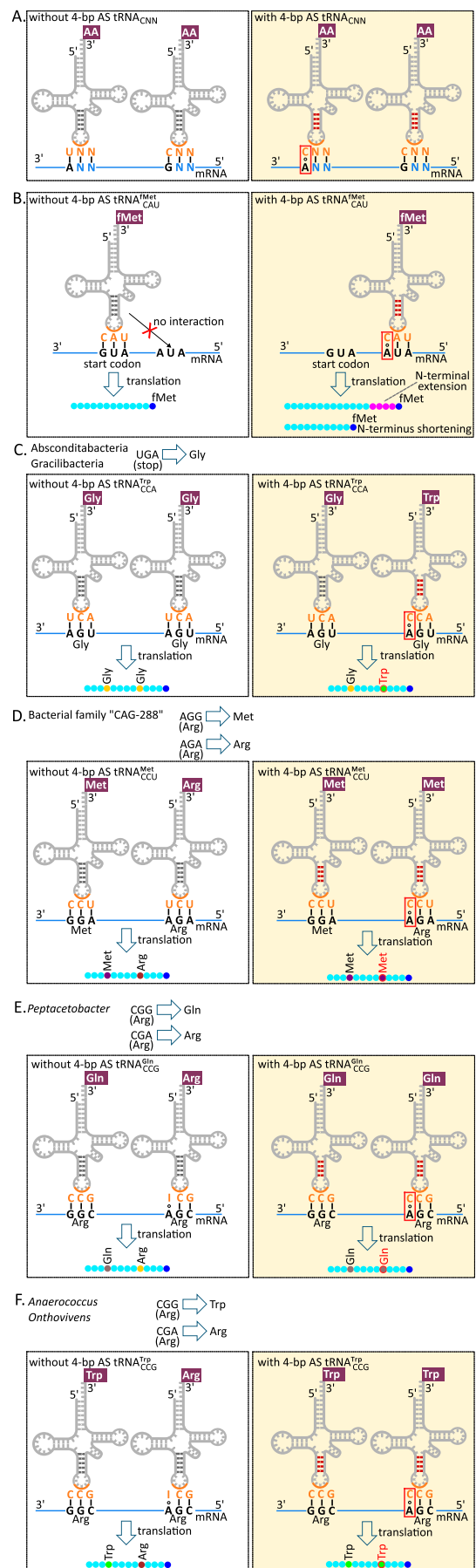
Our findings suggest that in bacteria, as in eukaryotes, 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> represents a dedicated, even if relatively rare mechanism for stop-to-tryptophan UGA reassignment. Most bacteria with a genome-wide UGA reassignment use an alternative strategy, in which the so-called suppressor tRNAs fully cognate to UGA decode this codon [44]. Why do eukaryotes and some bacteria favour the 4-bp AS tRNA-mediated mechanism of UGA decoding over the ‘suppressor’ tRNA-mediated one, remains an open question. ‘Suppressor’ tRNAs are also present in some bacteria with the standard genetic code (in

0.3% of the genomes analysed here), where they decode in-frame UGA (and other stop codons) that arise through non-sense mutations [89].

Our present results suggest an extension of the superwobble hypothesis. The minimal set of 32 tRNAs was initially proposed to be sufficient for decoding all sense codons of the genetic code [12]. Some codons are read by cognate tRNAs carrying anticodons that are exactly complementary to the codon, whereas others follow the wobble rules [12]. Under these rules, certain non-canonical base pairs, such as G:U, can occur in the wobble position, whereas other pairs, such as C:A, are less favourable. The superwobble hypothesis introduced later stated that tRNAs with U at the wobble position can decode all four codons in their respective four-codon families [11, 90]. Our results, reported here and previously [35], expand the superwobble hypothesis by including the C:A base pair, which, in particular, allows 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> to decode the UGA codon. The extended superwobble hypothesis thus postulates that, regardless of their anticodon, tRNAs with a 4-bp AS and a C at the wobble position have an expanded decoding range, recognising both G- and A-ending codons. This extension of the superwobble hypothesis assumes that the possibility of C:A pairing at the 3rd codon position is valid for other tRNA species, beyond 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub>. This hypothesis has several theoretical implications.

First, assuming the extended superwobble hypothesis is valid, bacteria that encode a particular 4-bp AS tRNA with a C at the first anticodon position do not need an isoacceptor tRNA with a U at the first anticodon position, whereas bacteria that only encode the 5-bp AS variant require such an isoacceptor tRNA to enable efficient decoding of NNA codons (Fig. 7A). We analysed tRNAs for the 11 amino acids that can be encoded by both NNG and NNA codons and decoded by tRNAs with the CNN and UNN anticodons, respectively (Supplementary Fig. S7). For 5 tRNA<sub>CNN</sub> species, the fraction of bacterial genomes encoding tRNA<sub>UNN</sub> in the presence of 4-bp AS tRNA<sub>CNN</sub> was significantly lower compared to the proportion of genomes encoding solely 5-bp AS, as predicted by the extended superwobble hypothesis. However, in four tRNA<sub>CNN</sub> cases, the opposite trend was observed, and in the three remaining cases, no statistically significant differences were detected (Supplementary Fig. S7).

Second, the 4-bp AS formylmethionine tRNA (tRNA<sup>fMet</sup><sub>CAU</sub>) is predicted to be eliminated by selection because it could cause production of proteins with potentially deleterious alterations of the N-terminal regions of proteins (Fig. 7B). The initiator tRNA<sup>fMet</sup><sub>CAU</sub> is the 5th most abundant tRNA in the bacterial dataset (Fig. 1). However, the proportion of 4-bp AS tRNA<sup>fMet</sup><sub>CAU</sub> is extremely low (82 out of the 50 849 identified tRNA<sup>fMet</sup><sub>CAU</sub> in bacteria, or 0.2%), with more than half of these (46 out of 82) having the AS unpinned at the top. For comparison, the percentage of 4-bp AS tRNAs for genes of similar abundance is 14.9% for tRNA<sup>Asn</sup><sub>GUU</sub> (4th most abundant tRNA in the bacterial dataset) and 11% for tRNA<sup>Lys</sup><sub>UUU</sub> (6th most abundant tRNA) (Fig. 1). Thus, 4-bp AS tRNA<sup>fMet</sup><sub>CAU</sub> indeed appears to be eliminated by purifying selection. Although this observation is compatible with our hypothesis, the selection pressure against 4-bp AS tRNA<sup>fMet</sup><sub>CAU</sub> might not be caused solely by its undesired putative propensity to decode AUA codons as methionine. It is highly likely that the major factor constraining the features of tRNA<sup>fMet</sup><sub>CAU</sub> is its binding to the P site and not the A site of the ribosome, in which the AS plays an important



**Figure 7.** Hypothetical implications of extending the superwobble hypothesis to include C:A base pairing. Panel (A) depicts functionally

role, contributing to accurate translation initiation [91, 92]. Consequently, any structural alteration of tRNA<sup>fMet</sup><sub>CAU</sub>, including changes in the number of base pairs in its AS, might be selected against to preserve efficient and error-free translation initiation. Furthermore, in reality tRNA<sup>fMet</sup><sub>CAU</sub> is expected to be capable of binding non-AUG codons, as translation initiation in bacteria quite frequently starts at codons other than AUG [93]; the NCBI translation table 11 used for annotation of bacterial genomes even allows AUA (i.e. the codon predicted to be recognized by a 4-bp AS tRNA<sup>fMet</sup><sub>CAU</sub> under the superwobble hypothesis) as an initiation codon.

In the case of the elongator tRNA<sup>Met</sup><sub>CAU</sub>, which binds to the A site of the ribosome, the 4-bp AS variant represents 27.3% of all identified tRNA<sup>Met</sup><sub>CAU</sub> (11 356 out of 41 644) genes in bacteria, of which 2272 (~20.0%) have AS unpinned at the top. It appears that bacteria are quite tolerant to the presence of 4-bp AS elongator tRNA<sup>Met</sup><sub>CAU</sub> given that misincorporation of methionine at the AUA codons (isoleucine) in the internal position of protein-coding regions is unlikely to cause substantial deleterious consequences [94]. Moreover, assuming that the levels of relative incorporation of isoleucine versus methionine at the AUA codons is regulated (e.g. via changing the relative abundance of 4-bp AS elongator tRNA<sup>Met</sup><sub>CAU</sub> and the natural AUA-cognate tRNA), decoding of the AUA codon as methionine might provide a protective mechanism

←

relevant deployment of 4-bp AS tRNA with a CNN anticodon for decoding NNA codons (note that the codons are depicted in the 3'-to-5' direction to preserve the pairing anticodon sequences in the conventional 5'-to-3' direction). Panels (B–F) present situations where the hypothetical ability of a 4-bp AS tRNA with a CNN anticodon to decode NNA codons would lead to detrimental effects and hence selection pressure against the occurrence of the respective 4-bp AS tRNA in the given bacterium. **(A)** Organisms that lack a 4-bp AS tRNA with C at the first anticodon position require the isoacceptor tRNA with U at the first anticodon position to enable efficient decoding of NNA codons (left). Organisms that encode a particular 4-bp AS tRNA with C at the first anticodon position do not need an isoacceptor tRNA with U at this position (right). **(B)** The 5-bp AS tRNA<sup>fMet</sup><sub>CAU</sub> recognizes the start codon (AUG) in mRNA sequence and incorporates formylmethionine to a polypeptide chain (left). The 4-bp AS tRNA<sup>fMet</sup><sub>CAU</sub> could hypothetically recognize AUA triplets up- or downstream of the start codon, which may lead to the production of proteins with an altered N-terminus (extended or shortened; right panel). **(C)** Organisms with stop-to-glycine UGA reassignment (those of the class JAEDAM01, uniting the groups also known as Absconditabacteria and Gracilibacteria) are expected to lack 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> (left) because such tRNAs could erroneously read UGA codons as tryptophan (right). **(D)** Organisms where arginine AGG codon was reassigned to encode methionine (e.g. in bacteria from the genus *Candidatus* Enterosoma of family CAG-288), while AGA preserves its original meaning and codes for arginine, should lack 4-bp AS tRNA<sup>Met</sup><sub>CCU</sub> decoding the AGG codon (left). Otherwise, mistranslation of the AGA as methionine could occur (right). **(E, F)** In organisms that have reassigned the CGG codon from arginine to another amino acid (either glutamine as in *Peptacetobacter*, or tryptophan as in *Anaerococcus* and *Onthovivens*), while decoding CGA as arginine, the tRNA decoding CGG is expected to have a 5-bp AS (left). The presence of 4-bp AS variant could lead to misreading of the CGA codon (right). The situations without and with C:A base pairing are in the boxes with transparent and yellow background, respectively. The tRNA anticodons are highlighted in orange; C:A base pairing, as assumed in the extended superwobble hypothesis, is shown with red rectangles. Watson–Crick base pairs and wobble interactions are shown as dashes and circles, respectively. The amino acids erroneously included in the polypeptide chain as a consequence of C:A base pairing are highlighted with red contour lines. Codon reassignments assumed in a particular case are shown at the top of each panel.

for coping with oxidative stress because many oxidants react with methionine residues to form methionine sulfoxide, which can be reduced back to methionine by methionine sulfoxide reductases [95].

Third, genomes with stop-to-glycine UGA reassignment are expected to lack 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> because it could misread the UGA codon as tryptophan (Fig. 7C). In our dataset, 53 bacterial genomes representing the class JAEDAM01 in the phylum Patescibacteriota demonstrate the stop-to-glycine UGA reassignment (Supplementary Table S1) as reported previously [44, 96, 97]. Among these, only four genomes (three forming a clade and one separate) were found to encode a 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> variant. All four genomes encode exclusively the 4-bp AS tRNA<sup>Trp</sup><sub>CCA</sub> variant. Of note, three of these differ from the *Blastocritidia*-like tRNA in the D-loop length, and one carries a G24A mutation; therefore, experimental validation is required to confirm that these variants decode UGA as tryptophan. In the 43 genomes, only the 5-bp AS variant is present, while in the six remaining genomes, we were unable to identify a tRNA<sup>Trp</sup><sub>CCA</sub> gene likely due to incomplete genome assembly.

Fourth, bacterial genomes, in which AGG and AGA code for different amino acids (methionine and arginine, respectively), should be depleted for 4-bp AS tRNA decoding the AGG codon to avoid misreading the AGA codon (Fig. 7D). In the analysed dataset, three genomes from the genus *Candidatus* Enterosoma of the family CAG-288 demonstrate the arginine-to-methionine AGG reassignment. Additionally, we included six genomes from the same bacterial group reported to have the same codon reassignment that are hypothesized to employ tRNA<sup>Arg</sup><sub>CCU</sub> with methionine tRNA identity elements to decode the AGG codon as methionine [44]. Indeed, these 9 genomes encode exclusively 5-bp AS tRNA<sub>CCU</sub> with methionine identity elements. Additionally, we examined tRNAs in 19 other genomes from the family CAG-288 that do not exhibit signs of the arginine-to-methionine AGG reassignment. Among these, we identified tRNA<sup>Arg</sup><sub>CCU</sub> genes in six genomes, all of these with the arginine identity elements and exclusively 5-bp AS. Because only 5-bp AS tRNA<sup>Arg</sup><sub>CCU</sub> are found in the subset of bacteria with the standard genetic code, it remains unclear whether the absence of 4-bp AS tRNA<sup>Arg</sup><sub>CCU</sub> in the group with the AGG reassignment reflects the results of negative selection or is an ancestral feature of this lineage.

Fifth, in the three separate lineages of phylum Bacillota that have reassigned the CGG codon from arginine to another amino acid (tryptophan or glutamine), while decoding CGA as arginine, the tRNA decoding CGG is expected to have a 5-bp AS to prevent misreading of the CGA codon (Fig. 7E and F). In contrast, this constraint should not apply to bacteria that interpret both CGG and CGA as the same amino acid. Indeed, in the two *Peptacetobacter* genomes, which evolved the arginine-to-glutamine CGG reassignment, tRNA<sub>CCG</sub> has a canonical 5-bp AS, but the situation is more complex in the other two lineages (the unrelated genera *Anaerococcus* and *Onthovivens*). Among the 17 genomes with predicted arginine-to-tryptophan CGG reassignment, 14 contained an identifiable tRNA<sub>CCG</sub> gene. Contrary to our prediction, eight of these genomes (1 and 7 in the genera *Onthovivens* and 7 *Anaerococcus*, respectively) encode a 4-bp AS tRNA<sub>CCG</sub> (Supplementary Table S3). Furthermore, in Absconditabacteriales that decode both CGG and CGA as tryptophan rather than the standard arginine and hence should tolerate 4-bp AS tRNA<sub>CCG</sub>, all tRNAs with the CCG anticodon have a 5-bp AS (Supplementary Table S3).

Thus, the occurrence of 4-bp AS tRNAs is subject to evolutionary constraints depending on the functional context. Although some of the results above further support, albeit indirectly, the extended superwobble hypothesis, experimental evidence is needed to confirm the substantial expansion of tRNA decoding capabilities beyond *Blastocystis*-like tRNA<sup>Trp</sup><sub>CCA</sub>.

## Acknowledgements

We are grateful to Ghadir Ibrahim (Institute of Parasitology) for assistance with the experimental work in *E. coli*.

*Author contributions:* Conceptualization: AB, EK, JL, LSV, ME, VY, ZP; data curation: AB, AK, FF, JL, JŽ, LSV, ME, MK, SN; formal analysis: AB, AK, FF, FB, JH, JŽ, ME, MK, NAC, PT, SN; funding acquisition: AB, JL, LSV, ME, MK, VY, ZP; methodology: AB, FF, ME, MK, JL, ZP, SN; supervision: AB, EK, JL, LSV, ME, MK, VY, ZP; visualization: AB, FF, JH, MK, SN; writing – original draft: AB, AK, EK, FF, JL, LSV, ME, MK, SN, VY, ZP; writing – review & editing – all authors.

## Supplementary data

Supplementary data is available at NAR online.

## Conflict of interest

None declared.

## Funding

This work was supported by the Czech Science Foundation grants 23-06479X (to J.L.), 23-07695S (to A.B.), and 23-05764S (to M.K.). L.S.V. and Z.P. were supported by the German Research Foundation & Czech Science Foundation Lead Agency Grant (23-08669L) and the European Regional Development Fund and the Czech Ministry of Education (CZ.02.01.01/00/22\_008/0004575 RNA for therapy). V.Y. and M.E. were supported by the European Union under the LERCO project CZ.10.03.01/00/22\_003/0000003 via the Operational Programme Just Transition. F.B. was supported by the European Union project MSCA Fellowships CZ No. 3 (Institute of Microbiology of the CAS, v. v. i.) “CZ.02.01.01/00/22\_010/0008767”. The computational resources were funded by the Ministry of Education, Youth and Sports of the Czech Republic through the e-INFRA CZ (project 90254).

## Data availability

The data underlying this article are available in the article and in its online supplementary material. The input data for reproducing the MD simulations, raw results files, and scripts needed for generating Fig. 6 are available at <https://github.com/kolarlab/fakih-4bp-trnas> and <https://doi.org/10.5281/zenodo.19220793>.

## References

- Su Z, Wilson B, Kumar P *et al.* Noncanonical roles of tRNAs: tRNA fragments and beyond. *Annu Rev Genet* 2020;54:47–69. <https://doi.org/10.1146/annurev-genet-022620-101840>
- Fujishima K, Kanai A. tRNA gene diversity in the three domains of life. *Front Genet* 2014;5:142. <https://doi.org/10.3389/fgene.2014.00142>
- Berg MD, Brandl CJ. Transfer RNAs: diversity in form and function. *RNA Biology* 2021;18:316–39. <https://doi.org/10.1080/15476286.2020.1809197>
- Zhang J. Recognition of the tRNA structure: everything everywhere but not all at once. *Cell Chemical Biology* 2024;31:36–52. <https://doi.org/10.1016/j.chembiol.2023.12.008>
- Kim SH, Quigley GJ, Suddath FL *et al.* Three-dimensional structure of yeast phenylalanine transfer RNA: folding of the polynucleotide chain. *Science* 1973;179:285–8. <https://doi.org/10.1126/science.179.4070.285>
- Krahn N, Fischer JT, Soll D. Naturally occurring tRNAs with non-canonical structures. *Front Microbiol* 2020;11:596914. <https://doi.org/10.3389/fmicb.2020.596914>
- Mukai T, Vargas-Rodriguez O, Englert M *et al.* Transfer RNAs with novel cloverleaf structures. *Nucleic Acids Res* 2017;45:2776–85. <https://doi.org/10.1093/nar/gkw898>
- Lang BF, Lavrov D, Beck N *et al.* Mitochondrial tRNA structure, identity, and evolution of the genetic code. In: Bullerwell CE (ed.), *Organelle Genetics: Evolution of Organelle Genomes and Gene Expression*. Berlin: Springer, 2012, 431–74.
- Phizicky EM, Alfonzo JD. Do all modifications benefit all tRNAs? *FEBS Lett* 2010;584:265–71. <https://doi.org/10.1016/j.febslet.2009.11.049>
- Blanchet S, Cornu D, Hatin I *et al.* Deciphering the reading of the genetic code by near-cognate tRNA. *Proc Natl Acad Sci USA* 2018;115:3018–23. <https://doi.org/10.1073/pnas.1715578115>
- Alkatib S, Scharff LB, Rogalski M *et al.* The contributions of wobbling and superwobbling to the reading of the genetic code. *PLoS Genet* 2012;8:e1003076. <https://doi.org/10.1371/journal.pgen.1003076>
- Crick FH. Codon-anticodon pairing: the wobble hypothesis. *J Mol Biol* 1966;19:548–55. [https://doi.org/10.1016/S0022-2836\(66\)80022-0](https://doi.org/10.1016/S0022-2836(66)80022-0)
- Rogalski M, Karcher D, Bock R. Superwobbling facilitates translation with reduced tRNA sets. *Nat Struct Mol Biol* 2008;15:192–8. <https://doi.org/10.1038/nsmb.1370>
- Kompatscher M, Bartosik K, Erharder K *et al.* Contribution of tRNA sequence and modifications to the decoding preferences of *E. coli* and *M. mycoides* tRNA<sup>Gly</sup><sub>UCC</sub> for synonymous glycine codons. *Nucleic Acids Res* 2024;52:1374–86. <https://doi.org/10.1093/nar/gkad1136>
- Lusic H, Gustilo EM, Vendeix FAP *et al.* Synthesis and investigation of the 5-formylcytidine modified, anticodon stem and loop of the human mitochondrial tRNA<sup>Met</sup>. *Nucleic Acids Res* 2008;36:6548–57. <https://doi.org/10.1093/nar/gkn703>
- Fabret C, Dervyn E, Dalmis B *et al.* Life without the essential bacterial tRNA<sup>Ile2</sup>-lysine synthetase Tils: a case of tRNA gene recruitment in *Bacillus subtilis*. *Mol Microbiol* 2011;80:1062–74. <https://doi.org/10.1111/j.1365-2958.2011.07630.x>
- Suzuki T, Numata T. Convergent evolution of AUA decoding in Bacteria and Archaea. *RNA Biology* 2014;11:1586–96. <https://doi.org/10.4161/15476286.2014.992281>
- Lavrov DV, Pett W. Animal mitochondrial DNA as we do not know it: mt-genome organization and evolution in nonbilaterian lineages. *Genome Biol Evol* 2016;8:2896–913. <https://doi.org/10.1093/gbe/evw195>
- Weber F, Dietrich A, Weil JH *et al.* A potato mitochondrial isoleucine tRNA is coded for by a mitochondrial gene possessing a methionine anticodon. *Nucl Acids Res* 1990;18:5027–30. <https://doi.org/10.1093/nar/18.17.5027>
- Miyauchi K, Kimura S, Akiyama N *et al.* A tRNA modification with aminovaleramide facilitates AUA decoding in protein synthesis. *Nat Chem Biol* 2025;21:522–31. <https://doi.org/10.1038/s41589-024-01726-x>
- Näsvall SJ, Chen P, Björk GR. The wobble hypothesis revisited: uridine-5-oxyacetic acid is critical for reading of G-ending codons. *RNA* 2007;13:2151–64. <https://doi.org/10.1261/rna.731007>

22. Weixlbaumer A, Murphy FV, Dziergowska A *et al.* Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines. *Nat Struct Mol Biol* 2007;14:498–502. <https://doi.org/10.1038/nsmb1242>
23. Sakai Y, Kimura S, Suzuki T. Dual pathways of tRNA hydroxylation ensure efficient translation by expanding decoding capability. *Nat Commun* 2019;10:2858. <https://doi.org/10.1038/s41467-019-10750-8>
24. Paris Z, Fleming IM, Alfonzo JD. Determinants of tRNA editing and modification: avoiding conundrums, affecting function. *Cell Dev Biol* 2012;23:269–74. <https://doi.org/10.1016/j.semcd.2011.10.009>
25. Paris Z, Svobodová M, Kachale A *et al.* A mitochondrial cytidine deaminase is responsible for C to U editing of tRNA to decode the UGA codon in *Trypanosoma brucei*. *RNA Biology* 2021;18:278–86. <https://doi.org/10.1080/15476286.2021.1940445>
26. Alfonzo JD, Blanc V, Estevez AM *et al.* C to U editing of the anticodon of imported mitochondrial tRNA(Trp) allows decoding of the UGA stop codon in *Leishmania tarentolae*. *EMBO J* 1999;18:7056–62. <https://doi.org/10.1093/emboj/18.24.7056>
27. Schmeing TM, Voorhees RM, Kelley AC *et al.* How mutations in tRNA distant from the anticodon affect the fidelity of decoding. *Nat Struct Mol Biol* 2011;18:432–6. <https://doi.org/10.1038/nsmb.2003>
28. Borges AL, Lou YC, Sachdeva R *et al.* Widespread stop-codon recoding in bacteriophages may regulate translation of lytic genes. *Nat Microbiol* 2022;7:918–27. <https://doi.org/10.1038/s41564-022-01128-6>
29. Engelhardt DL, Webster RE, Wilhelm RC *et al.* *In vitro* studies on the mechanism of suppression of a nonsense mutation. *Proc Natl Acad Sci USA* 1965;54:1791–7. <https://doi.org/10.1073/pnas.54.6.1791>
30. Capecchi MR, Gussin GN. Suppression *in vitro*: identification of a serine-sRNA as a “nonsense” suppressor. *Science* 1965;149:417–22. <https://doi.org/10.1126/science.149.3682.417>
31. Benzer S, Champe SP. A change from nonsense to sense in the genetic code. *Proc Natl Acad Sci USA* 1962;48:1114–21. <https://doi.org/10.1073/pnas.48.7.1114>
32. Lukeš J, Eliáš M, Kachale A *et al.* Natural and artificial variations of the standard genetic code. *Curr Biol* 2025;35:R1104–26. <https://doi.org/10.1016/j.cub.2025.09.071>
33. Smith D, Yarus M. Transfer-RNA structure and coding specificity. 1. Evidence that a D-arm mutation reduces transfer-RNA dissociation from the ribosome. *J Mol Biol* 1989;206:489–501. [https://doi.org/10.1016/0022-2836\(89\)90496-8](https://doi.org/10.1016/0022-2836(89)90496-8)
34. Hirsh D. Tryptophan tRNA of *Escherichia coli*. *Nature* 1970;228:57. <https://doi.org/10.1038/228057a0>
35. Kachale A, Pavlíková Z, Nenarokova A *et al.* Short tRNA anticodon stem and mutant eRF1 allow stop codon reassignment. *Nature* 2023;613:751–8. <https://doi.org/10.1038/s41586-022-05584-2>
36. Pavlíková ZC, Miletínová P, Roithová A *et al.* Ribosomal A-site interactions with near-cognate tRNAs drive stop codon readthrough. *Nat Struct Mol Biol* 2025;32:662–74. <https://doi.org/10.1038/s41594-024-01450-z>
37. Swart EC, Serra V, Petroni G *et al.* Genetic codes with no dedicated stop codon: context-dependent translation termination. *Cell* 2016;166:691–702. <https://doi.org/10.1016/j.cell.2016.06.020>
38. Lukeš J, Čapková Pavlíková Z, Yurchenko V *et al.* *Blastocrithidia*—a genetic alien from the planet Earth. *Cold Spring Harb Perspect* 2025;17:a041868. <https://doi.org/10.1101/cshperspect.a041868>
39. Swart EC, Emmerich C, Seah KBB *et al.* How did UGA codon translation as tryptophan evolve in certain ciliates? A critique of Kachale *et al.* 2023 *Nature*. *eLife* 2024;13:RP93502.
40. Záhonová K, Füßy Z, Albanaz ATS *et al.* Comparative genomic analysis of trypanosomatid protists illuminates an extensive change in the nuclear genetic code. *mBio* 2025;16:e0088525. <https://doi.org/10.1128/mbio.00885-25>
41. DeMontigny W, Bachvaroff T. The nuclear and mitochondrial genomes of *Amoebophrya* sp. ex *Karlodinium veneficum*. *G3* 2025;15:jkaf030. <https://doi.org/10.1093/g3journal/jkaf030>
42. Bachvaroff TR. A predated nuclear genetic code with all three termination codons reassigned as sense codons in the syndinean *Amoebophrya* sp. ex *Karlodinium veneficum*. *PLoS One* 2019;14:e0212912. <https://doi.org/10.1371/journal.pone.0212912>
43. Schultz DW, Yarus M. tRNA structure and ribosome function. II. Interaction between anticodon helix and other tRNA mutations. *J Mol Biol* 1994;235:1395–405. <https://doi.org/10.1006/jmbi.1994.1096>
44. Shulgina Y, Eddy SR. A computational screen for alternative genetic codes in over 250,000 genomes. *eLife* 2021;10:e71402. <https://doi.org/10.7554/eLife.71402>
45. Nakayama T, Harada R, Yabuki A *et al.* Marked genome reduction driven by a parasitic lifestyle: two complete genomes of endosymbiotic bacteria possibly hosted by a dinoflagellate. *Microb Environ* 2025;40:ME25005. <https://doi.org/10.1264/jsm.2025.005>
46. Baranov PV, Atkins JF. No stopping with a short-stem transfer RNA. *Nature* 2023;613:631–2. <https://doi.org/10.1038/d41586-022-04585-5>
47. Parks DH, Chuvochina M, Chaumeil PA *et al.* A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol* 2020;38:1079–86. <https://doi.org/10.1038/s41587-020-0501-8>
48. Parks DH, Imelfort M, Skennerton CT *et al.* CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015;25:1043–55. <https://doi.org/10.1101/gr.186072.114>
49. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997;25:955–64. <https://doi.org/10.1093/nar/25.5.955>
50. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 2004;32:11–6. <https://doi.org/10.1093/nar/gkh152>
51. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res* 2024;52:W78–82. <https://doi.org/10.1093/nar/gkae268>
52. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30:772–80. <https://doi.org/10.1093/molbev/mst010>
53. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. *Bioinformatics* 2011;27:1009–10. <https://doi.org/10.1093/bioinformatics/btr039>
54. Katz K, Shutov O, Lapoint R *et al.* The Sequence Read Archive: a decade more of explosive growth. *Nucleic Acids Res* 2022;50:D387–90. <https://doi.org/10.1093/nar/gkab1053>
55. O’Cathail C, Ahamed A, Burgin J *et al.* The European Nucleotide Archive in 2024. *Nucleic Acids Res* 2025;53:D49–55. <https://doi.org/10.1093/nar/gkae975>
56. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–20. <https://doi.org/10.1093/bioinformatics/btu170>
57. Andrews S. FastQC: a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (21 August 2025, date last accessed).
58. Kim D, Paggi JM, Park C *et al.* Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 2019;37:907–15. <https://doi.org/10.1038/s41587-019-0201-4>
59. Danecek P, Bonfield JK, Liddle J *et al.* Twelve years of SAMtools and BCftools. *GigaScience* 2021;10:giab008. <https://doi.org/10.1093/gigascience/giab008>
60. Robinson JT, Thorvaldsdottir H, Winckler W *et al.* Integrative genomics viewer. *Nat Biotechnol* 2011;29:24–6. <https://doi.org/10.1038/nbt.1754>

61. Vera Alvarez R, Pongor LS, Mariño-Ramírez L *et al.* TPMCalculator: one-step software to quantify mRNA abundance of genomic features. *Bioinformatics* 2019;35:1960–2. <https://doi.org/10.1093/bioinformatics/bty896>
62. Shulgina Y, Eddy SR, Codetta: predicting the genetic code from nucleotide sequence. *Bioinformatics* 2023;39:btac802. <https://doi.org/10.1093/bioinformatics/btac802>
63. Hyatt D, Chen GL, Locascio PF *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinf* 2010;11:119. <https://doi.org/10.1186/1471-2105-11-119>
64. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 2019;20:238. <https://doi.org/10.1186/s13059-019-1832-y>
65. Wong TKF, Ly-Trong N, Ren H *et al.* IQ-TREE 3: phylogenomic inference software using complex evolutionary models. *ecoevoRxiv*, <https://doi.org/10.32942/X2P62N>, 7 April 2025, preprint: not peer reviewed.
66. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;25:1972–3. <https://doi.org/10.1093/bioinformatics/btp348>
67. Sayers EW, Beck J, Bolton EE *et al.* Database resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Res* 2025;53:D20–9. <https://doi.org/10.1093/nar/gkae979>
68. Schrödinger L, DeLano W. The PyMOL Molecular Graphics System. <http://www.pymol.org/pymol> (12 February 2026, date last accessed).
69. Cornell WD, Cieplak P, Bayly CI *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 1995;118:2309. <https://doi.org/10.1021/ja955032e>
70. Zgarbová M, Otyepka M, Sponer J *et al.* Refinement of the Cornell *et al.* nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J Chem Theory Comput* 2011;7:2886–902. <https://doi.org/10.1021/ct200162x>
71. Aduri R, Psciuk BT, Saro P *et al.* AMBER force field parameters for the naturally occurring modified nucleosides in RNA. *J Chem Theory Comput* 2007;3:1464–75. <https://doi.org/10.1021/ct600329w>
72. Berendsen HJC, Grigera JR, Straatsma TP. The missing term in effective pair potentials. *J Phys Chem* 1987;91:6269–71. <https://doi.org/10.1021/j100308a038>
73. Joung IS, Cheatham TE. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J Phys Chem B* 2008;112:9020–41. <https://doi.org/10.1021/jp8001614>
74. Feenstra KA, Hess B, Berendsen HJC. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J Comput Chem* 1999;20:786–98. [https://doi.org/10.1002/\(SICI\)1096-987X\(199906\)20:8<786::AID-JCCS>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1096-987X(199906)20:8<786::AID-JCCS>3.0.CO;2-B)
75. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys* 2007;126:014101. <https://doi.org/10.1063/1.2408420>
76. Parrinello M, Rahman A. Polymorphic transitions in single-crystals—a new molecular-dynamics method. *J Appl Phys* 1981;52:7182–90. <https://doi.org/10.1063/1.328693>
77. Darden T, York D, Pedersen L. Particle mesh Ewald—an N.Log(N) method for Ewald sums in large Systems. *J Chem Phys* 1993;98:10089–92. <https://doi.org/10.1063/1.464397>
78. Abraham MJ, Murtola T, Schulz R *et al.* GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 2015;1-2:19–25. <https://doi.org/10.1016/j.softx.2015.06.001>
79. Komine Y, Inokuchi H. Importance of the G27-A43 mismatch at the anticodon stem of *Escherichia coli* tRNA<sup>Thr2</sup>. *FEBS Lett* 1990;272:55–7. [https://doi.org/10.1016/0014-5793\(90\)80447-Q](https://doi.org/10.1016/0014-5793(90)80447-Q)
80. Ortiz-Meoz RF, Green R. Functional elucidation of a key contact between tRNA and the large ribosomal subunit rRNA during decoding. *RNA* 2010;16:2002–13. <https://doi.org/10.1261/rna.2232710>
81. Schultz DW, Yarus M. tRNA structure and ribosome function. I. tRNA nucleotide 27–43 mutations enhance first position wobble. *J Mol Biol* 1994;235:1381–94. <https://doi.org/10.1006/jmbi.1994.1095>
82. Biela A, Hammermeister A, Kaczmarczyk I *et al.* The diverse structural modes of tRNA binding and recognition. *J Biol Chem* 2023;299:104966. <https://doi.org/10.1016/j.jbc.2023.104966>
83. Chan PP, Lowe TM. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res* 2016;44:D184–9. <https://doi.org/10.1093/nar/gkv1309>
84. Kleina LG, Masson JM, Normanly J *et al.* Construction of *Escherichia coli* amber suppressor tRNA genes. II. Synthesis of additional tRNA genes and improvement of suppressor efficiency. *J Mol Biol* 1990;213:705–17. [https://doi.org/10.1016/S0022-2836\(05\)80257-8](https://doi.org/10.1016/S0022-2836(05)80257-8)
85. Raftery LA, Yarus M. Systematic alterations in the anticodon arm make tRNA<sup>Glu-Suoc</sup> a more efficient suppressor. *EMBO J* 1987;6:1499–506. <https://doi.org/10.1002/j.1460-2075.1987.tb02392.x>
86. Yarus M, Cline S, Raftery L *et al.* The translational efficiency of tRNA is a property of the anticodon arm. *J Biol Chem* 1986;261:10496–505. [https://doi.org/10.1016/S0021-9258\(18\)67412-5](https://doi.org/10.1016/S0021-9258(18)67412-5)
87. Olejniczak M, Dale T, Fahlman RP *et al.* Idiosyncratic tuning of tRNAs to achieve uniform ribosome binding. *Nat Struct Mol Biol* 2005;12:788–93. <https://doi.org/10.1038/nsmb978>
88. Nguyen HA, Sunita S, Dunham CM. Disruption of evolutionarily correlated tRNA elements impairs accurate decoding. *Proc Natl Acad Sci USA* 2020;117:16333–8. <https://doi.org/10.1073/pnas.2004170117>
89. Eggertsson G, Soll D. Transfer ribonucleic acid-mediated suppression of termination codons in *Escherichia coli*. *Microbiol Rev* 1988;52:354–74. <https://doi.org/10.1128/mr.52.3.354-374.1988>
90. Osawa S, Jukes TH, Watanabe K *et al.* Recent evidence for evolution of the genetic code. *Microbiol Rev* 1992;56:229–64. <https://doi.org/10.1128/mr.56.1.229-264.1992>
91. Laursen BS, Sørensen HP, Mortensen KK *et al.* Initiation of protein synthesis in bacteria. *Microbiol Mol Biol Rev* 2005;69:101–23. <https://doi.org/10.1128/MMBR.69.1.101-123.2005>
92. Shetty S, Shah RA, Chembazhi UV *et al.* Two highly conserved features of bacterial initiator tRNAs license them to pass through distinct checkpoints in translation initiation. *Nucleic Acids Res* 2017;45:2040–50. <https://doi.org/10.1093/nar/gkx854>
93. Hecht A, Glasgow J, Jaschke PR *et al.* Measurements of translation initiation from all 64 codons in *E. coli*. *Nucleic Acids Res* 2017;45:3615–26. <https://doi.org/10.1093/nar/gkx070>
94. Ohmura T, Ueda T, Hashimoto Y *et al.* Tolerance of point substitution of methionine for isoleucine in hen egg white lysozyme. *Protein Eng* 2001;14:421–5. <https://doi.org/10.1093/protein/14.6.421>
95. Levine RL, Mosoni L, Berlett BS *et al.* Methionine residues as endogenous antioxidants in proteins. *Proc Natl Acad Sci USA* 1996;93:15036–40. <https://doi.org/10.1073/pnas.93.26.15036>
96. Campbell JH, O'Donoghue P, Campbell AG *et al.* UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. *Proc Natl Acad Sci USA* 2013;110:5540–5. <https://doi.org/10.1073/pnas.1303090110>
97. Rinke C, Schwientek P, Sczyrba A *et al.* Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 2013;499:431–7. <https://doi.org/10.1038/nature12352>