

Transfer RNAs with novel cloverleaf structures

Takahito Mukai¹, Oscar Vargas-Rodriguez¹, Markus Englert¹, H. James Tripp², Natalia N. Ivanova², Edward M. Rubin², Nikos C. Kyrpides² and Dieter Söll^{1,3,*}

¹Department of Molecular Biophysics and Biochemistry, New Haven, CT 06520, USA, ²Department of Energy Joint Genome Institute (DOE JGI), Walnut Creek, CA 94598, USA and ³Department of Chemistry, Yale University, New Haven, CT 06520, USA

Received August 02, 2016; Revised September 22, 2016; Accepted September 30, 2016

ABSTRACT

We report the identification of novel tRNA species with 12-base pair amino-acid acceptor branches composed of longer acceptor stem and shorter T-stem. While canonical tRNAs have a 7/5 configuration of the branch, the novel tRNAs have either 8/4 or 9/3 structure. They were found during the search for selenocysteine tRNAs in terabytes of genome, metagenome and metatranscriptome sequences. Certain bacteria and their phages employ the 8/4 structure for serine and histidine tRNAs, while minor cysteine and selenocysteine tRNA species may have a modified 8/4 structure with one bulge nucleotide. In Acidobacteria, tRNAs with 8/4 and 9/3 structures may function as missense and nonsense suppressor tRNAs and/or regulatory non-coding RNAs. In δ -proteobacteria, an additional cysteine tRNA with an 8/4 structure mimics selenocysteine tRNA and may function as opal suppressor. We examined the potential translation function of suppressor tRNA species in *Escherichia coli*; tRNAs with 8/4 or 9/3 structures efficiently inserted serine, alanine and cysteine in response to stop and sense codons, depending on the identity element and anticodon sequence of the tRNA. These findings expand our view of how tRNA, and possibly the genetic code, is diversified in nature.

INTRODUCTION

Since the discovery of tRNA in the late-1950s (1), its role as an adaptor molecule during translation of the genetic information has been extensively investigated (2). While recent research has focused on the non-translational functions of tRNAs (3–6), the rapidly increasing expanse of genomic and metagenomic sequence information has revived the interests in the canonical function of tRNAs (7). Recent studies have identified a number of non-canonical tRNA species that were previously mis-annotated or undetected

due to their unusual recognition elements, anticodon sequences and irregular secondary structures (6–15). Interestingly, although the function of many of them remains unknown (5,6,8), some of these non-canonical tRNAs are known to be responsible for changes to the universal meaning of the genetic code (7,10–15). Therefore, proper identification of tRNA genes is important to identify genetic code variations in nature (10,13).

All tRNAs fold into an L-shaped tertiary structure which physically links the amino acid moiety attached to one end (amino-acid acceptor branch) to the genetic information of the anticodon sequence on the other end (the anticodon branch) (2,6). The amino-acid acceptor branch consists of a 7-bp acceptor stem and a 5-bp T-stem, and this 12-bp branch is recognized by the elongation factor (EF-Tu), whereas the anticodon branch consists of the D-arm, V-arm and anticodon arm. The size and structure of tRNAs are normally standardized in a particular genetic code system, as they share the same apparatus such as processing RNases, base modification enzymes, CCA-adding enzyme, EF-Tu and the ribosome. Although tRNA size reduction is common in the mitochondrial genomes, all prokaryotic and eukaryotic tRNA species are believed to have a 12-bp amino-acid acceptor branch (7/5) with a few exceptions. To date such exceptions are known for selenocysteine (Sec) tRNAs and histidine (His) tRNAs. Most tRNA^{His} species have an additional guanosine at the 5'-end (G-1) that produces a non-canonical 7/5 structure (16,17). In contrast, tRNA^{Sec} has a 13-bp amino-acid acceptor branch (8/5 or 9/4) (18,19) and are actually longer than the other tRNA species.

Recently, we and others identified bacterial tRNA^{Sec} species with 12-bp amino-acid acceptor branches (13,20). Among these, we found tRNA^{Sec} species with a 12-bp amino-acid acceptor branch composed of an 8-bp acceptor stem and a 4-bp T-stem and a bulge nucleotide at position 51a. The existence of such (8/4) tRNA^{Sec} structure in two different bacterial phyla (Actinobacteria and Chloroflexi) prompted us to hypothesize that tRNAs with an 8/4 structure are ubiquitous but remain largely uncharacterized. In this study, we identified, annotated and classified a large number of bacterial tRNA sequences with novel secondary

*To whom correspondence should be addressed. Tel: +1 203 432 6200; Fax: +1 203 432 6202; Email: dieter.soll@yale.edu

structures and evaluated their translational functions in *Escherichia coli*.

MATERIALS AND METHODS

Identification of tRNA sequences

The false positive sequences of our previous tRNA^{Sec} search study (13) were re-analyzed, and tRNA^{Sec}-like sequences with a non-canonical composition of the amino-acid acceptor branch were manually collected with the aid of the ARAGORN server (21) and the Clustal X program (22). Next, we performed a BLAST search of some soil and sediment metagenome data in the Integrated Microbial Genomes (IMG) system (23) and the National Center for Biotechnology Information (NCBI) for more allo-tRNA sequences. A number of metagenomic contigs were found to contain up to two allo-tRNA genes and frequently annotated as belonging to *Acidobacteria*. The resulting allo-tRNA sequences were classified into several groups, and representative sequences were used as query for BLAST searches of acidobacterial genomes. *Acidobacterium* strain C40 was re-named in this study as *Edaphobacter* strain C40, based on the 99.4% 16S rRNA sequence similarity with *Edaphobacter modestus* Jbg-1^T (24). Some of the *selC** tRNA sequences were found in the false positive sequences of the tRNA^{Sec} search. By using them as query for BLAST, most of the *selC** tRNA sequences were detected. The resulting *selC** tRNA sequences were classified into several groups with the aid of the ARAGORN server (21) and the Clustal X program (22). A few allo-tRNA sequences with a serine anticodon were found in tRNA gene clusters of unknown bacteriophages in the Macroalgal surface ecosystem from Botany Bay, Sydney, Australia. We manually predicted the secondary structures of all tRNAs in the clusters and found (8/4) tRNA^{Ser} and (8/4) tRNA^{His} species. Next, we did a BLAST search of all metagenomic assembled sequence data in IMG and NCBI for more (8/4) tRNA^{Ser} and (8/4) tRNA^{His} sequences. Some allo-tRNA sequences missing nucleotides 9–11 were found in some metatranscriptome reads of the Harvard Forest Long Term Ecological Research site (Petersham, MA, USA) and from the Peat soil microbial communities from Weissenstadt, Germany. We manually predicted the secondary structures of these tRNAs and performed BLAST search of all metatranscriptome data sets in IMG for more read sequences containing even a part of these tRNAs.

Identification and analysis of protein sequences

We performed BLASTp search of all genomic and metagenomic protein sequence data in the IMG and NCBI systems (search data: February 17, 2016) in order to identify *cysS*, *selA*, *selB* and *selD* genes and their homologs. To manually enlarge a metagenomic contig containing a *selC** gene from Wastewater microbial communities from Syncrude, Ft. McMurray, Alberta – Microbes from Suncor tailings pond 6 2012TP6.6, we performed BLAST searches of the raw data (SRR943333) using NCBI short read archive (SRA) BLAST. Likewise, we identified the amino-acid sequences for the *selB** and *cysS** genes of *Desulfonema limicola*

Jadebusen DSM 2076 by filling the gaps of the partially-sequenced genome using the raw data (SRR058919) in the NCBI SRA database. Sequence alignments and phylogenetic trees (Bootstrap N-J Tree) were made using Clustal X 2.1. The BoxShade Server was also used for the alignments.

Plasmids and *E. coli* strains

Plasmid pGF_iB (25) was used for cloning the *Desulfococcus biacutus selC** tRNA^{Cys} sequence and its variants using EcoRI and BamHI sites. The arabinose promoter cassette of pBAD-myc-HisA (Invitrogen) was previously transplanted into pRSFDuet1 (26) to make pBAD-RSF. The open reading frame (ORF) of the *Desulfomonile tiedjei cysS* gene was cloned from the genomic DNA obtained from DSMZ (German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany) and fused directly downstream of the *kan* marker gene of pBAD-RSF with a weak Shine–Dalgarno sequence of the *E. coli prfA* gene [TT TACAGGGTGCATTTACGCCT]. The cloned *cysS* gene was mutated using Infusion (Clontech) to make the *cysS* variant genes. The multiple cloning site (MCS) of pBAD-RSF was replaced by the MCS and the *rrnC* terminator sequence of pGF_iB with a modification of BamHI site to BglII site using Infusion to make pBAD-RSF5. Allo-tRNA sequences were cloned into the MCS of pBAD-RSF5 using EcoRI and BglII sites. Plasmid pBAD-sfGFP (27) was mutated using Infusion to make the sfGFP variant genes. Plasmid pACYC184 was mutated using Infusion to make the *cat* variant genes. The ORF of the *cat* marker gene in pACYC184 was replaced using Infusion with the ORF of *D. tiedjei selD* gene cloned from the genomic DNA to make pACYC-DtselD. The *E. coli* strain DH10B was used for allo-tRNA experiments. The *E. coli* WL400 (MC4100 *selD204::cat+*) (28) cells harboring pACYC-DtselD corresponded to *E. coli* $\Delta selD$ with *D. tiedjei selD*.

Mass analysis

sfGFP variants encoding a C-terminal His-tag were purified using nickel-nitrilotriacetic acid agarose (QIAGEN). Purified sfGFP solutions were concentrated by centrifugation using Amicon Ultra 10k (Merck Millipore) and subjected to peptide mass fingerprinting (PMF) analysis by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) performed by the Keck Foundation Biotechnology Resource Laboratory (Yale University).

selD reporter assay

WL400 [pACYC-DtselD] cells were transformed with pG-FiB or one of the tRNA-expressing plasmids together with pBAD-RSF or one of the *D. tiedjei cysS*-expressing plasmids. Their overnight cultures were spotted onto LB agar plates supplemented with 1 μ M Na₂MoO₄, 1 μ M Na₂SeO₃, 50 mM sodium formate, 100 μ g/ml carbenicillin, 30 μ g/ml kanamycin, 100 μ g/ml spectinomycin and 0.1% L-arabinose and grown anaerobically at 37°C overnight. These plates were overlaid under anaerobic condition with a 0.75% top agar containing 1 mg/ml benzyl viologen, 250 mM sodium formate and 25 mM KH₂PO₄ (pH 7.0). Within

a few seconds or less than a minute, spots on the plates became dark or light purple, depending on the formate dehydrogenase activity of cells.

***In vitro* cysteinylatation assay**

*D. tiedjei selC** tRNA^{Cys} was prepared by T7 RNA polymerase transcription, purified in preparative 10% urea PAGE and electroeluted. The tRNA was refolded by heating at 95°C for 3 min followed by cooling to 65°C at a rate of 0.5°C/s. During a 3 min hold at 65°C, MgCl₂ was added to a final concentration of 10 mM. The refolding was completed by cooling to 30°C at 0.2°C/s. tRNA was radiolabeled at the 3'-end using [α -³²P]ATP and *E. coli* CCA-adding enzyme as previously described (29). *D. tiedjei* cysteinyl-tRNA synthetase (CysRS) with a His-tag was recombinantly obtained using *E. coli*. Aminoacylation reactions were performed in [50 mM HEPES-NaOH, pH 7.2, 50 mM KCl, 10 mM MgCl₂, 10 mM ATP, 1 mM cysteine, 1 mM DTT]. At the indicated time points, 10 μ l aliquots were removed and digested by nuclease P1. One microliter of the quenched reaction was spotted on PEI cellulose plates. [³²P]AMP and [³²P]AMP-Cys were separated with 100 mM ammonium acetate – 5% acetic acid. The plates were analyzed using a PhosphorImager.

***In vitro* aminoacylation with *E. coli* aminoacyl-tRNA synthetases**

Escherichia coli cells harboring expression plasmids for *E. coli* threonyl-tRNA synthetase (ThrRS), glycyl-tRNA synthetase (GlyRS) (*glyQ* and *glyS* subunits) and histidyl-tRNA synthetase (HisRS) were obtained from the ASKA collections (30). Overnight-night cultures for each protein were used to inoculate 1 L of fresh LB media containing chloramphenicol. Cells were grown to an A₆₀₀ of 0.6 and protein overexpression was induced with 0.1 mM IPTG overnight at 25°C. Cells were harvested by centrifugation and the resulting pellet was lysed with buffer containing 50 mM Tris (pH 8), 300 mM NaCl and protease inhibitor cocktail tablets (cOmplete, Roche). Lysed cells were then centrifuged at 4°C for 45 min at 18 000 \times g. The lysate was loaded on a TALON metal affinity resin (Clontech), and the protein was eluted with varying concentrations of imidazole. The protein-containing fractions were pooled and stored in buffer containing 50 mM HEPES (pH 7.3) and 150 mM NaCl. tRNA genes were cloned into pUC18 using Gibson Assembly (New England Biolabs), and the tRNAs were prepared using *in vitro* transcription as previously described (31). Aminoacylation assays were carried out with 5 μ M tRNA and 0.5 μ M tRNA synthetase in buffer containing 50 mM Hepes (pH 7.3), 4 mM ATP, 10 mM MgCl₂, 0.1 mg/ml BSA, 1 mM dithiothreitol and 20 μ M [¹⁴C]His (590 cpm/pmol) (PerkinElmer), 100 μ M [¹⁴C]Gly (146 cpm/pmol) (PerkinElmer) or 25 μ M [³H]Thr (7779 cpm/mol) (American Radiolabeled Chemicals). Reactions were incubated at 37°C and after 15 min, 10 μ l of the reaction mixture was spotted on Whatman 3MM filters pre-soaked with 5% trichloroacetic acid (TCA). Filters were washed three times with 5% TCA, and the remaining radioactivity was quantified using a scintillation counter.

RESULTS AND DISCUSSION

tRNA^{Sec}-like tRNAs are widespread in bacteria

The tRNA^{Sec} search pipeline used in our previous work (13) produced tRNA sequences with high similarity to the tRNA^{Sec} covariance model (Figure 1A), but they were considered false positives after further curation. Upon reanalysis we found a series of tRNA sequences with non-canonical secondary structures. The basis of the cloverleaf structure prediction is discussed in the Supplementary Data. We classified these tRNA genes into two groups: ‘allo-tRNA’ (named after their irregular appearance) and ‘SelC* tRNA^{Cys}’ (Figure 1A, B and C). While allo-tRNA genes belong to bacteria from *Clostridia*, *Proteobacteria* and *Acidobacteria*, *selC** genes were found in anaerobic bacteria from the phyla *Firmicutes*, *Thermodesulfobacteria*, *Nitrospirae* and *Proteobacteria*. Both tRNA groups are structurally similar to tRNA^{Sec} as they have a long V-arm and longer anticodon and acceptor stems compared to canonical tRNAs (Figure 1B, C and Supplementary Figure S1). Moreover, the D-stem-loop of allo-tRNAs resembles that of tRNA^{Sec} with its long stem and tetraloop (Figure 1B). The most striking feature of allo-tRNAs is their 8/4 or 9/3 composition of the 12-bp amino-acid acceptor branch (Figure 1B), whereas SelC* tRNA^{Cys} species of certain δ -proteobacteria may have a modified 8/4 structure with a bulge base A51a like minor bacterial (8/4) tRNA^{Sec} species (Figure 1C, Supplementary Figure S1).

The presence of a long V-arm and the identity of the discriminator base (G73 or U73) in most allo-tRNAs indicate that these tRNAs may be serine tRNA isoacceptors, since these unique elements are essential for aminoacylation by seryl-tRNA synthetase (SerRS) (32–36). In addition, SerRS also recognizes not only (7/5) tRNA^{Ser} but also (8/5, 9/4 and 8/4) tRNA^{Sec} and even a variant of tRNA^{Sec} with a 9/3 structure (37). SelC* tRNAs were named after the *selC* gene, which encodes tRNA^{Sec} in *E. coli*. SelC* tRNA^{Cys} isoacceptors have an U73 discriminator base and cysteine GCA or opal UCA anticodons (Figure 1C). U73 and GCA are the most important identity elements for CysRS (38,39), and certain CysRS forms are known to cysteinylate tRNA^{Cys}_{UCA} (40).

We then analyzed metagenomic contigs containing allo-tRNA genes and discovered tRNA^{Ser} and tRNA^{His} species with 8/4 structure (Figure 1A and C) and polycistrons of irregular tRNAs with 7/5 and 8/4 structures. (8/4) tRNA^{Ser} was found in bacteriophages, proteobacteria and bovine rumen bacteria such as *Clostridia*, while (8/4) tRNA^{His} was found in bacteriophages and an α -proteobacterium. Interestingly, (8/4) tRNA^{Ser} and (8/4) tRNA^{His} genes coexist in the same tRNA gene clusters of α -proteobacterial phages (Supplementary Figure S2). The (8/4) tRNA^{Ser} species are included in the (8/4) allo-tRNA group. The (8/4) tRNA^{His} species lack G-1 and have A73 (or U73, see Supplementary Figure S3), which is characteristic of the (7/5) tRNA^{His} species of a group of α -proteobacteria (41). The polycistrons of irregular tRNAs are discussed below.

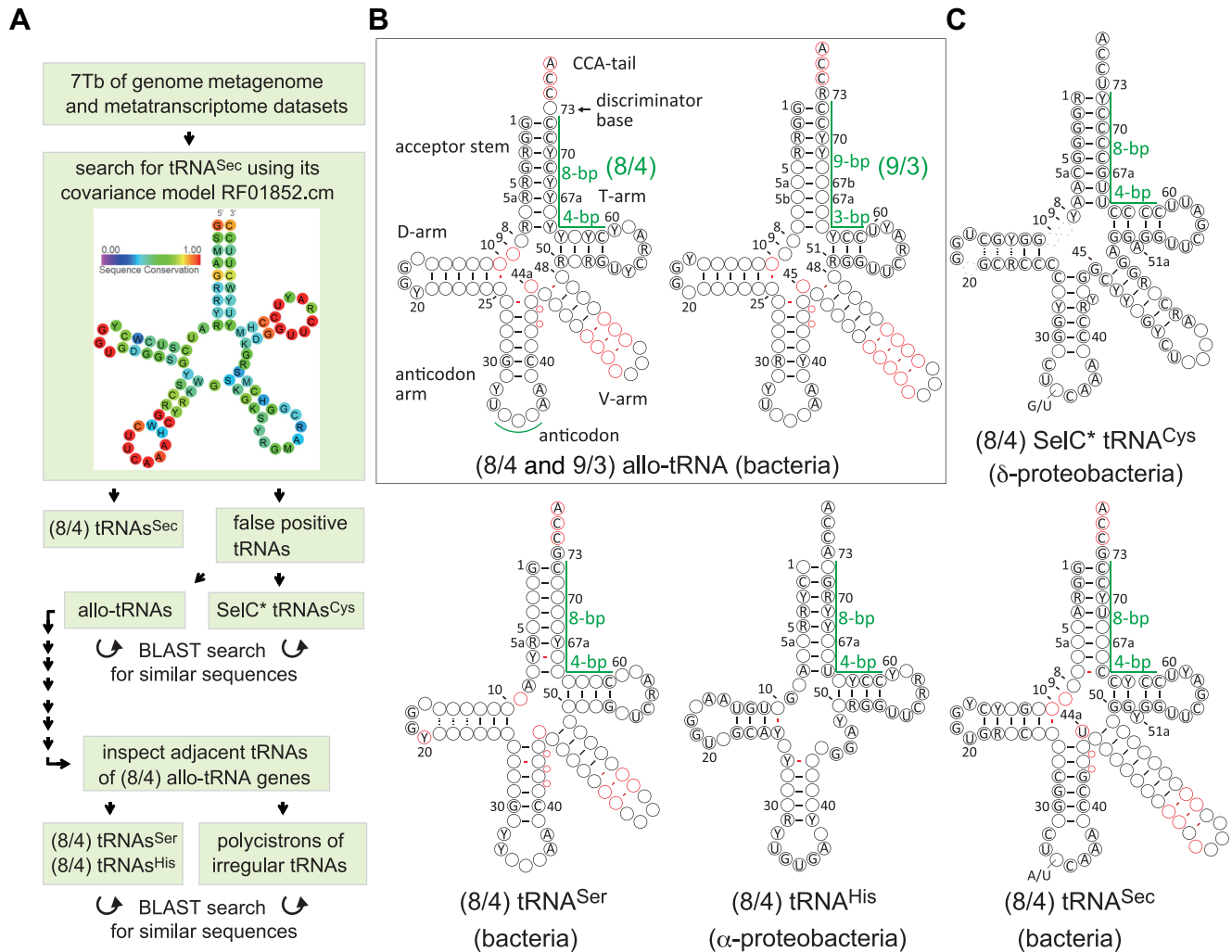


Figure 1. Identification of new tRNA classes. (A) Workflow used to identify new tRNA groups. The covariance model was derived from Rfam 12.0. (B) Inferred consensus cloverleaf structures of allo-tRNAs. Highly conserved nucleotides are shown. R and Y denote A/G and U/C, respectively. Red circles indicate nucleotides missing (larger) or inserted (smaller) in some of the tRNA genes. (C) Inferred consensus secondary structures of minor serine/histidine/cysteine/selenocysteine tRNA species. The (8/4) tRNA^{Ser} with Y20 are also classified as (8/4) allo-tRNA.

Diverse anticodon sequences of allo-tRNAs

Although the (8/4) tRNA^{Ser} species have anticodons corresponding to serine codons, most of allo-tRNA species have non-serine anticodons (Figure 2). In fact, their anticodon sequences are highly diverse and correspond to 35 distinct codons (Figure 2). Among them, the UAU, GCG and GUC anticodons corresponding to the AUA isoleucine (Ile) codon, the CGC arginine (Arg) codon and the GAC aspartic acid (Asp) codon, respectively, are predominant, whereas (8/4) allo-tRNAs with anticodons corresponding to phenylalanine (Phe), valine (Val), His and lysine (Lys) codons were only found once in the examined metagenomic data set (Figure 2). In contrast, the (9/3) allo-tRNA species have anticodons corresponding to the AGA/AGG Arg codons, the UUA/UUG/CUA/CUG leucine (Leu) codons and the UAA stop codon (Figure 2).

Allo-tRNAs are fully compatible with a bacterial translation system

To investigate whether allo-tRNAs are active translational adaptors, we used super-folder green fluorescence protein (sfGFP) as a reporter in *E. coli*. The Ser codon at position 2 of wild-type (WT) sfGFP was mutated to either CUC or UUA and the resulting mutants were expressed together with *Silvibacterium bohemicum* (8/4) allo-tRNA_{GAG} (42) or (9/3) allo-tRNA_{UAA} in *E. coli* (Figure 3A). Interestingly, both tRNAs efficiently inserted Ser in response to the CUC and UUA Leu codons, respectively, as confirmed by LC-MS/MS and PMF analyses of purified sfGFP (Figure 3B). Furthermore, induction of the allo-tRNA expression from the *araBAD* promoter led to severe cell growth arrest and ultimately to cell death, which is possibly caused by global mis-incorporation of Ser at Leu codons in the *E. coli* proteome.

We also identified (8/4) and (9/3) allo-tRNAs with a G3:U70 wobble pair, the most important structural element

Phe Leu	UUU UUC → GAA 1 UUA → UAA 15/1 UUG → CAA 5/2	Ser	UCU → AGA 1 UCC → GGA 8 UCA → UGA 6 UCG → CGA 5	Tyr stop	UAU → GUA 13 UAC → UUA 9/2 UAG → CUA 2	Cys stop Trp	UGU → GCA 17 UGC → UCA 8 UGG
Leu	CUU CUC → GAG 3 CUA → UAG 1 CUG → CAG 1	Pro	CCU CCC CCA CCG	His Gln	CAU → GUG 1 CAC CAA CAG	Arg	CGU → GCG 139 CGC CGA CGG
Ile Met	AUU AUC AUA → UAU 186 AUG	Thr	ACU ACC ACA ACG	Asn Lys	AAU → UUU 1 AAC AAA AAG	Ser Arg	AGU → GCU 12 AGC → UCU 1 AGG → CCU 1
Val	GUU → GAC 1 GUC GUA GUG	Ala	GCU GCC GCA GCG	Asp Glu	GAU → AUC 11 GAC → GUC 161 GAA → UUC 12 GAG → CUC 5	Gly	GGU → GCC 44 GGC GGA GGG

Figure 2. Anticodon diversity of allo-tRNAs. Possible codon–anticodon interactions are indicated with bars. The numbers of (8/4) and (9/3) allo-tRNA sequences are indicated with red and blue letters, respectively.

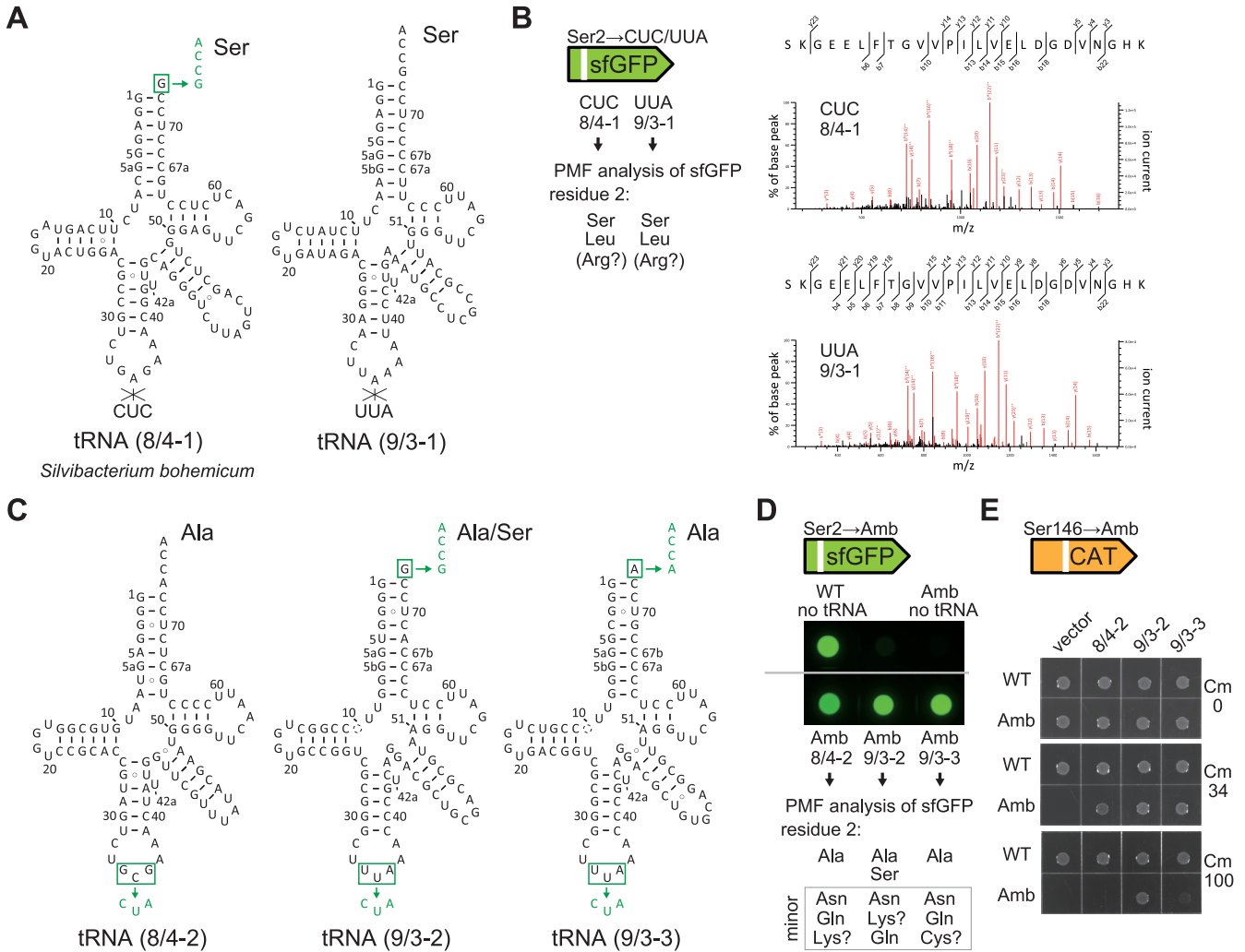


Figure 3. Compatibility of allo-tRNA species with the *E. coli* translation system. (A) Missense suppressor allo-tRNAs with Ser identity and Leu anticodons. (B) Missense suppression experiment in *E. coli* DH10B using sfGFP as reporter. Purified sfGFP variants were analyzed by LC-MS/MS and PMF. The amino-acid species incorporated into sfGFP in response to the CUC/UUA codons at position 2 are shown. The peaks corresponding to the serine-containing peptides are shown. (C) Amber suppressor variants of three allo-tRNA species. Their presumed amino-acid identities are indicated. (D) Amber suppression experiment in *E. coli* DH10B using sfGFP as reporter. The amino acids incorporated into sfGFP in response to the amber codon at position 2 by allo-tRNAs are shown. (E) Suppression of Ser146TAG CAT by allo-tRNAs in *E. coli*. Pre-cultured cells were spotted onto LB agar plates with various Cm concentrations (0, 34, 100 µg/ml). The plates were incubated overnight at 37°C.

for aminoacylation by alanyl-tRNA synthetase (AlaRS) (43–45). To test whether these allo-tRNAs can be acylated by AlaRS *in vivo*, we chose three examples, replaced their wild-type anticodons (GCG and UUA) with the amber anticodon CUA (Figure 3C) and mutated the Ser2 codon of sfGFP to an amber stop codon. The mutant sfGFP was then co-expressed with one of the three allo-tRNAs in *E. coli*. The three amber suppressor tRNAs efficiently translated the amber codon, leading to expression of the full-length sfGFP variant, and producing as much fluorescence as did cells expressing WT sfGFP (Figure 3D). The sfGFP variants were then purified, and the identity of the amino acid incorporated at position 2 was revealed by LC-MS/MS and PMF analyses (Figure 3D). The amber codon was mainly translated as Ala and Ser, as judged by the probability scores. Insertion of Asn, Gln, Lys and possibly Cys, Ile and Glu was also detected. Only the allo-tRNA 9/3-2 variant inserted Ser, probably due to its discriminator base G73.

To further confirm the allo-tRNA-mediated incorporation of Ala and Ser, we used the chloramphenicol (Cm) acetyltransferase (CAT) gene as a reporter since it contains a critical catalytic Ser residue at position 146 that only tolerates substitutions with Ala (46). We first validated that replacing Ser146 with Asn, Gln or Lys produced an inactive CAT, while the Ser146Ala CAT mutant retained activity (Supplementary Figure S4). Then, an amber codon was substituted for Ser146 and the resulting CAT variant was expressed with any of the three allo-tRNA variants. As expected, the allo-tRNAs suppressed the amber codon and conferred Cm resistance to the *E. coli* cells with the Ser146TAG *cat* gene (Figure 3E). However, only the allo-tRNA 9/3-2 variant conferred the resistance at a Cm concentration of 100 $\mu\text{g/ml}$, which corroborates that only this tRNA incorporates Ser. Together, these results clearly demonstrated that these allo-tRNA (both 8/4 and 9/3) sequences were properly folded, processed, aminoacylated and delivered to the ribosomes by EF-Tu in *E. coli*.

Association of active and inactive allo-tRNAs with toxin-antitoxin systems

The *Edaphobacter* strain C40 has an allo-tRNA_{UAU} pseudogene having several base-pair disruptions and overlapping with the ORF of a transposon-related protein (Figure 4A). In soil and sediment metagenomic sequences, allo-tRNA genes are often found in the vicinity of a variety of toxin-antitoxin systems (47) (Figure 4A). Among others, allo-tRNA_{UAU} species comprise the most abundant allo-tRNA group (Figure 2). Interestingly, they have cloverleaf structures slightly different from that of the standard allo-tRNA^{Ser}, stem-destabilizing mutations as in the *Edaphobacter* strain C40 allo-tRNA_{UAU} (Figure 4A), and a variety of possible five-stem-junction structures (Supplementary Figure S5). To assess their ability to serve in translation, we converted six allo-tRNA_{UAU} species with a potential Ser identity to allo-tRNA_{CUA} and examined their activity in *E. coli* using the CAT (Ser146TAG) reporter. However, none of the six variants conferred Cm resistance in response to the amber codon, and two of them caused cell death (Supplementary Figure S5). Thus, it is possible that most of

the allo-tRNA_{UAU} species are not used for translation and instead may be associated with transposable elements or toxin-antitoxin systems.

Moreover, we discovered polycistrons formed by allo-tRNA-like sequences and other irregular tRNA sequences (Figure 4A, B and Supplementary Figure S6) in metatranscriptome sequences of two forest/peat soil metatranscriptome projects. Surprisingly, both might be nonsense or missense suppressor tRNA (Figure 4A), and may also be associated with a toxin-antitoxin system (Figure 4A). Among the polycistronic tRNA groups (tRNA₀, tRNA₁ and tRNA₂) (Figure 4A, Supplementary Figure S6), 8/4 structures were predicted for tRNA_{GGU} species containing a G-1 and a group of tRNAs containing an extra-loop (E-loop) (Figure 4B). We examined three tRNA₁ species from the polycistrons; a (8/4) tRNA₁ with an E-loop and two tRNA₁ with G-1, one of which has an 8/4 structure (Figure 4B). The two G-1 tRNA₁ species with GGU anticodon were expected to be substrates for HisRS, whereas the E-loop tRNA₁ with U73 and GCU anticodon were predicted to be a poor substrate for *E. coli* ThrRS and/or GlyRS. The three tRNAs were transcribed *in vitro* and tested for aminoacylation with *E. coli* HisRS, GlyRS and ThrRS. As predicted, the two G-1 tRNA₁ species were aminoacylated by *E. coli* HisRS, even more efficiently than *E. coli* tRNA^{His}, despite the Thr GGU anticodon (Figure 4B). The GUG triplet at positions 35–37 (but not the anticodon positions 34–36 in tRNA^{His}) of the two G-1 tRNA₁ species might have been recognized by HisRS (35–37). Interestingly, the two G-1 tRNA₁ species did not insert His in response to the ACC Thr codon at position 2 in a sfGFP variant gene in *E. coli*. Thus, G-1 tRNA₁ species may be charged, but not used for translation. In contrast, the E-loop tRNA₁ was not aminoacylated either by *E. coli* GlyRS or ThrRS *in vitro* (Supplementary Figure S7). However, it is possible that the organisms encoding these irregular tRNAs encode a GlyRS and ThrRS capable of charging them. Alternatively, another aminoacyl-tRNA synthetase or homolog might charge them in a non-canonical manner, similar to the aminoacylation of a tRNA-like small RNA (tRNA^{Other}) by a complex of class I and II lysyl-tRNA synthetases in *Bacillus cereus* (48).

*selC** tRNA^{Cys} functions in translation

*selC** genes were found in isolated genomic sequences of Clostridia, Bacilli, Thermodesulfobacteria, Thermodesulfobacterium, δ -proteobacteria, a composite genome of Smithella and a few metagenomic contigs, probably derived from Nitrospirae or δ -proteobacteria (Supplementary Table S1 and Figures S1, S8–S13). Interestingly, *selC** tRNA^{Cys} species belonging to a particular phylum or class show a unique conserved secondary arrangement (Supplementary Figures S1 and S9). Furthermore, we identified the complete selenocysteine-inserting apparatus (consisting of the *selA*, *selB*, *selC* and *selD* genes) in all *selC**-containing genomes, whereas incomplete selenocysteine systems were also found in the metagenomic contigs (Supplementary Figure S13). Strikingly, in two δ -proteobacterial subgroups, *Syntrophobacterales* and *Desulfobacterales*, a second copy of *selB*, which we named *selB**, was found downstream

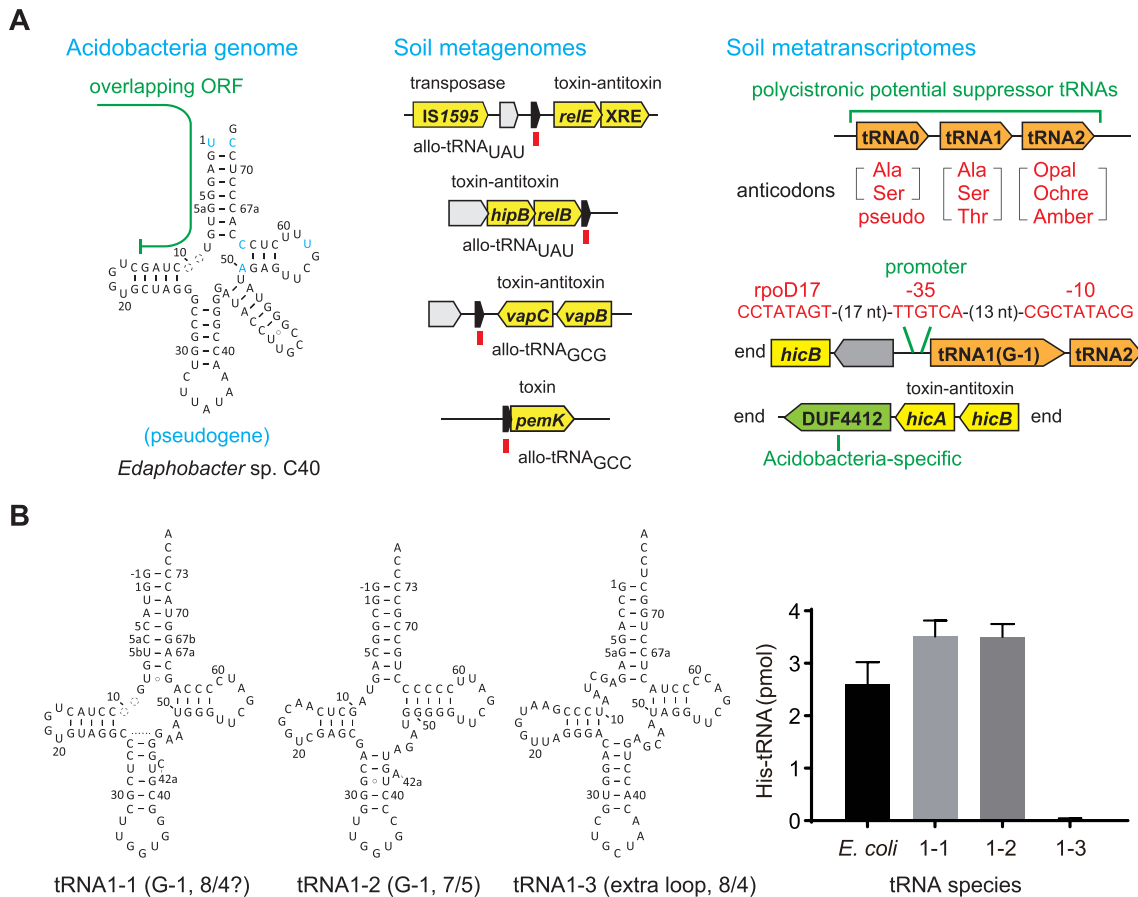


Figure 4. Active and inactive allo-tRNAs associated with toxin-antitoxin systems. (A) Genetic loci of allo-tRNA genes and polycistrons of irregular tRNA sequences. All these nucleotide sequences may be derived from acidobacteria. (B) *In vitro* aminoacylation assay of the three tRNA¹ species shown and *E. coli* tRNA^{His} by *E. coli* HisRS. Inferred cloverleaf structures of the three putative tRNA¹ species tested. Bars represent the average of three trials with the corresponding standard deviation shown.

of the (8/4) *selC** genes. Because *selB* encodes the Sec-tRNA^{Sec}-specific elongation factor, it is possible that (8/4) *selC** tRNA^{Cys} mimics (8/4) tRNA^{Sec}, both structurally and functionally and is recognized by SelB*. In addition to *selB**, in a subgroup of *Desulfobacterales*, we discovered an additional copy of the *cysS* gene, which encodes CysRS, downstream of the *selB** genes (Figure 5A). This second copy, named *cysS**, encodes CysRS* (Supplementary Figure S14) that lacks the anticodon binding domain (ABD) and has mutations in the CP1 domain responsible for the recognition of the first base pair of the tRNA^{Cys} acceptor stem (49). Intriguingly, genomes with *cysS** genes always encode *selC** tRNA^{Cys} species containing an A1:U72 base pair and an opal anticodon (UCA) (Figure 5A, B, Supplementary Figure S15).

To assess the ability of the (8/4) tRNA^{Cys}_{UCA} to translate the opal codon as cysteine, we developed an *E. coli* strain whose *selD* gene was replaced with the *Desulfomonile tiedjei* (Dt) *selD* gene intrinsically having an opal Sec codon at position 15 (Figure 5C). Translation of the opal codon as either Sec or Cys produces functional selenophosphate synthase (SelD), which allows the synthesis of Sec-containing formate dehydrogenase H (FDH_H) that converts benzyl viologen into a purple dye (Figure 5C) (50). Because CysRS*

didn't show any activity in *E. coli* (data not shown), and because Dt CysRS was functional *in vitro* (Supplementary Figure S16), we decided to develop a Dt CysRS variant that may recapitulate the activity of CysRS*. The variant named ΔABD/PTVS lacks the ABD and has the same CP1 motif as Db CysRS* (PTVS instead of MSGA) (Figure 5D). When Db tRNA^{Cys}_{UCA} was co-expressed with the ΔABD/PTVS variant and the Dt *selD* gene in the *E. coli* Δ*selD* strain, the cells were stained dark purple (Figure 5E), which indicated that Db tRNA^{Cys}_{UCA} was efficiently cysteinylated and used for translating the opal codon of the *selD* gene. Another Dt CysRS variant having PTVS alone was also functional, whereas ABD elimination alone produced a less active variant (Figure 5E). These results suggest that CysRS* may have evolved to specifically aminoacylate *selC** tRNA^{Cys}_{UCA} species with A1:U72. This hypothesis was explored by using a Db tRNA^{Cys}_{UCA} variant in which the A1:U72 base pair was mutated to G1:C72, the most common base pair in tRNA^{Cys} (49). As predicted, the CysRS ΔABD/PTVS was unable to efficiently aminoacylate the G1:C72 Db tRNA^{Cys}_{UCA} mutant (Figure 5F). Lastly, we experimentally confirmed the predicted 8/4 conformation of Db tRNA^{Cys}_{UCA}, by employing a previously developed method used to confirm the 9/4 structure of eu-

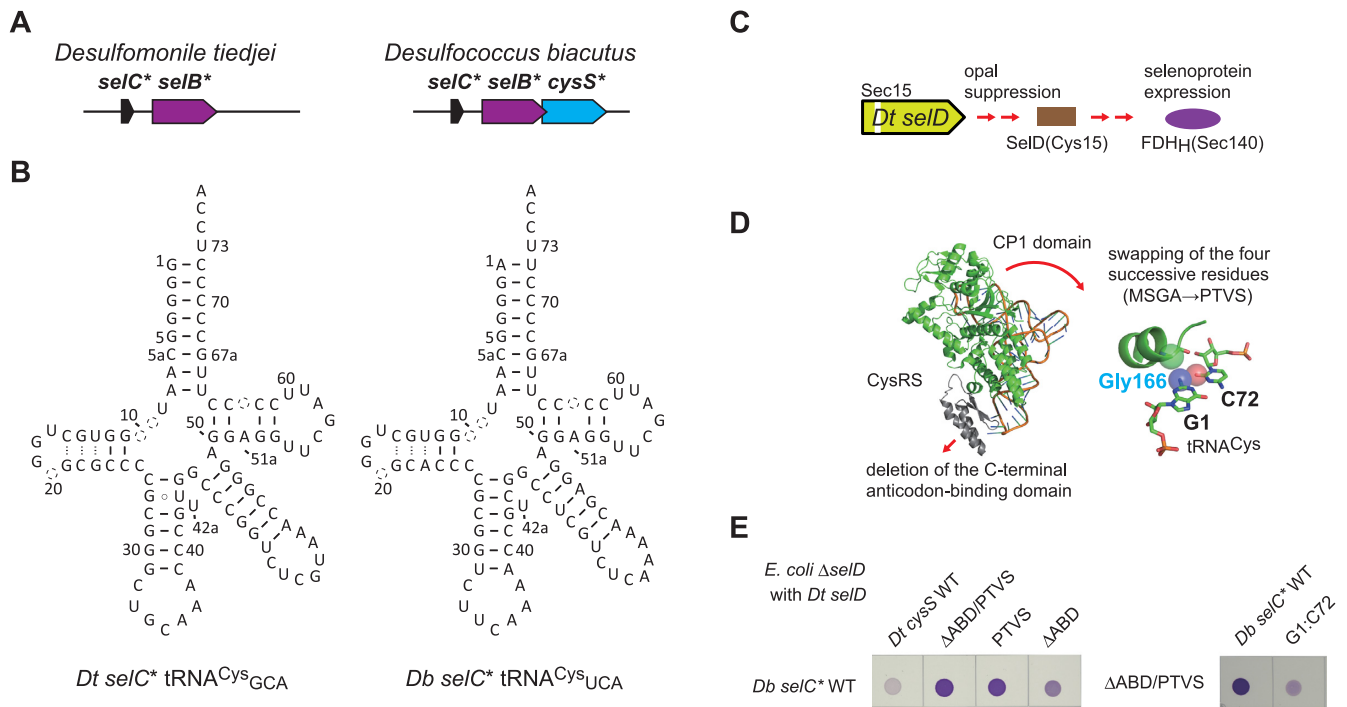


Figure 5. *D. biacutus selC** tRNA^{Cys} requires a dedicated CysRS. (A) Genetic locus representation of *selC** genes in two δ -proteobacteria. The *selC** gene is upstream of a putative *selB* gene (*selB**). In *D. biacutus*, the *selB** gene overlaps with a second copy *cysS* gene (*cysS**). (B) Inferred cloverleaf structures of the *selC** tRNAs. (C) Scheme for the opal suppression experiment using an *E. coli* $\Delta selD$ strain harboring a plasmid expressing *D. tiedjei selD*. Suppression of the opal codon at position 15 of *D. tiedjei selD* by *Db Cys*-tRNA^{Cys}_{UCA} leads to selenophosphate formation, required for Sec-tRNA^{Sec} synthesis followed by expression of active FDH_H. *E. coli* cells expressing FDH_H reduce benzyl viologen on agar plates resulting in a purple dye. (D) *Dt CysRS* mutant construction. In the *Dt CysRS* PTVS mutant, CP1 domain residues M164, S165, G166 and Ala167 ('MSGA') of *Dt CysRS* were mutated to Pro, Thr, Val and Ser ('PTVS'), respectively, in the context of *Db CysRS**. The *Dt CysRS* ΔABD /PTVS was designed by deleting the C-terminal anticodon binding domain (ABD) and replacing the 'MSGA' residues with 'PTVS' to resemble *Db CysRS**. The crystal structure of *E. coli CysRS* and tRNA^{Cys} (pdb: 1U0B) was used for modeling. (E) Opal suppression experiments using cells expressing *Dt CysRS* variants together with *Db selC** tRNA and using cells expressing the *Db selC** tRNA G1:C72 variant together with *Dt CysRS* ΔABD /PTVS.

karyotic tRNA^{Sec} (51). *Db tRNA*^{Cys}_{UCA} was engineered to have a fixed 7/5 or 8/4 composition by stabilizing the acceptor and T-stems with strong WC base pairing. It was revealed that the 8-bp acceptor stem was important for efficient opal suppression, whereas the bulge structure was dispensable (Supplementary Figure S17, see Supplementary text).

CONCLUSION

In the present study, we identified a large number of tRNAs with novel secondary structures. The function of these tRNAs is still uncharted: some of them may be involved in the translation of Ser and His codons, in opal suppression or recoding with Cys and in mis-translation of diverse codons with Ser or Ala; others may have non-translational roles. The (8/4) tRNA^{Cys} species may have co-evolved with its dedicated aminoacyl-tRNA synthetase and elongation factor. Future studies should elucidate the biological functions of these non-canonical tRNAs and protein components, and structural studies of these tRNAs will add to our knowledge of the structural plasticity of tRNA.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

T.M. is a Japan Society for the Promotion of Science Post-doctoral Fellow for Research Abroad. The authors thank Jeffrey Blanchard, Michael Pester, Bernhard Schink, Gerard Muyzer, Alexandra J. Roth Schulze, Barbara J Campbell, Jeanette Norton and many others for permission to use sequence data produced through the DOE-JGI's community sequencing program. The authors are grateful to Sergey Melnikov, Noah Reynolds, Corwin Miller, Anastasia Sevostyanova, Tateki Suzuki and Takuya Umehara for enlightened discussions

FUNDING

National Institute for General Medical Sciences [GM22854 to D.S.]; Division of Chemical Sciences, Geosciences and Biosciences, Office of Basic Energy Sciences of the Department of Energy [DE-FG02-98ER20311 to D.S.]; the work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility [Contract No. DE-AC02-05CH11231]. Funding for open access charge: NIGMS [GM22854 to D.S.].

Conflict of interest statement. None declared.

REFERENCES

- Hoagland, M.B., Stephenson, M.L., Scott, J.F., Hecht, L.I. and Zamecnik, P.C. (1958) A soluble ribonucleic acid intermediate in protein synthesis. *J. Biol. Chem.*, **231**, 241–257.
- RajBhandary, U.L. and Söll, D. (eds). (1995) *tRNA: Structure, Biosynthesis, and Function*. American Society for Microbiology, Washington, DC.
- Keam, S.P. and Hutvagner, G. (2015) tRNA-Derived Fragments (tRFs): Emerging new roles for an ancient RNA in the regulation of gene expression. *Life (Basel)*, **5**, 1638–1651.
- Raina, M. and Ibba, M. (2014) tRNAs as regulators of biological processes. *Front. Genet.*, **5**, 171.
- Hamashima, K. and Kanai, A. (2013) Alternative genetic code for amino acids and transfer RNA revisited. *Biomol. Concepts*, **4**, 309–318.
- Katz, A., Elgamal, S., Rajkovic, A. and Ibba, M. (2016) Non-canonical roles of tRNAs and tRNA mimics in bacterial cell biology. *Mol. Microbiol.*, **101**, 545–548.
- Ling, J., O'Donoghue, P. and Söll, D. (2015) Genetic code flexibility in microorganisms: Novel mechanisms and impact on physiology. *Nat. Rev. Microbiol.*, **13**, 707–721.
- Hamashima, K., Tomita, M. and Kanai, A. (2016) Expansion of noncanonical V-arm-containing tRNAs in eukaryotes. *Mol. Biol. Evol.*, **33**, 530–540.
- Marck, C. and Grosjean, H. (2002) tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA*, **8**, 1189–1232.
- Campbell, J.H., O'Donoghue, P., Campbell, A.G., Schwientek, P., Sczyrba, A., Woyke, T., Söll, D. and Podar, M. (2013) UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 5540–5545.
- Borrel, G., Gaci, N., Peyret, P., O'Toole, P.W., Gribaldo, S. and Brugere, J.F. (2014) Unique characteristics of the pyrrolysine system in the 7th order of methanogens: Implications for the evolution of a genetic code expansion cassette. *Archaea*, 374146.
- Ivanova, N.N., Schwientek, P., Tripp, H.J., Rinke, C., Pati, A., Huntemann, M., Visel, A., Woyke, T., Kyrpides, N.C. and Rubin, E.M. (2014) Stop codon reassignments in the wild. *Science*, **344**, 909–913.
- Mukai, T., Englert, M., Tripp, H.J., Miller, C., Ivanova, N.N., Rubin, E.M., Kyrpides, N.C. and Söll, D. (2016) Facile recoding of selenocysteine in nature. *Angew. Chem. Int. Ed. Engl.*, **55**, 5337–5341.
- Mühlhausen, S., Findeisen, P., Plessmann, U., Urlaub, H. and Kollmar, M. (2016) A novel nuclear genetic code alteration in yeasts and the evolution of codon reassignment in eukaryotes. *Genome Res.*, **26**, 945–955.
- Swart, E.C., Serra, V., Petroni, G. and Nowacki, M. (2016) Genetic codes with no dedicated stop codon: Context-dependent translation termination. *Cell*, **166**, 691–702.
- Cooley, L., Appel, B. and Söll, D. (1982) Post-transcriptional nucleotide addition is responsible for the formation of the 5' terminus of histidine tRNA. *Proc. Natl. Acad. Sci. U.S.A.*, **79**, 6475–6479.
- Orellana, O., Cooley, L. and Söll, D. (1986) The additional guanylate at the 5' terminus of *Escherichia coli* tRNA^{His} is the result of unusual processing by RNase P. *Mol. Cell. Biol.*, **6**, 525–529.
- Hubert, N., Sturchler, C., Westhof, E., Carbon, P. and Krol, A. (1998) The 9/4 secondary structure of eukaryotic selenocysteine tRNA: more pieces of evidence. *RNA*, **4**, 1029–1033.
- Schön, A., Böck, A., Ott, G., Sprinzl, M. and Söll, D. (1989) The selenocysteine-inserting opal suppressor serine tRNA from *E. coli* is highly unusual in structure and modification. *Nucleic Acids Res.*, **17**, 7159–7165.
- Cravedi, P., Mori, G., Fischer, F. and Percudani, R. (2015) Evolution of the selenoproteome in *Helicobacter pylori* and Epsilonproteobacteria. *Genome Biol. Evol.*, **7**, 2692–2704.
- Laslett, D. and Canback, B. (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.*, **32**, 11–16.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R. et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.
- Markowitz, V.M., Chen, I.M., Chu, K., Szeto, E., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., Pagani, I., Tringe, S. et al. (2014) IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res.*, **42**, D568–D573.
- Koch, I.H., Gich, F., Dunfield, P.F. and Overmann, J. (2008) *Edaphobactermodestus* gen. nov., sp. nov., and *Edaphobacteraggregans* sp. nov., acidobacteria isolated from alpine and forest soils. *Int. J. Syst. Evol. Microbiol.*, **58**, 1114–1122.
- Normanly, J., Masson, J.M., Kleina, L.G., Abelson, J. and Miller, J.H. (1986) Construction of two *Escherichia coli* amber suppressor genes: tRNA^{Phe}_{CUA} and tRNA^{Cys}_{CUA}. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 6548–6552.
- Haruna, K., Alkazemi, M.H., Liu, Y., Söll, D. and Englert, M. (2014) Engineering the elongation factor Tu for efficient selenoprotein synthesis. *Nucleic Acids Res.*, **42**, 9976–9983.
- Fan, C., Xiong, H., Reynolds, N.M. and Söll, D. (2015) Rationally evolving tRNA^{Pyl} for efficient incorporation of noncanonical amino acids. *Nucleic Acids Res.*, **43**, e156.
- Leinfelder, W., Forchhammer, K., Veprek, B., Zehelein, E. and Böck, A. (1990) In vitro synthesis of selenocysteinyl-tRNA_{UCA} from seryl-tRNA_{UCA}: Involvement and characterization of the selD gene product. *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 543–547.
- Ledoux, S. and Uhlenbeck, O.C. (2008) [3'-32P]-labeling tRNA with nucleotidyltransferase for assaying aminoacylation and peptide bond formation. *Methods*, **44**, 74–80.
- Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H. and Mori, H. (2006) Complete set of ORF clones of *Escherichia coli* ASKA library (A Complete Set of *E. coli* K-12 ORF Archive): Unique Resources for Biological Research. *DNA Res.*, **12**, 291–299.
- Ahel, I., Stathopoulos, C., Ambrogelly, A., Sauerwald, A., Toogood, H., Hartsch, T. and Söll, D. (2002) Cysteine activation is an inherent in vitro property of prolyl-tRNA synthetases. *J. Biol. Chem.*, **277**, 34743–34748.
- Wu, X.Q. and Gross, H.J. (1993) The long extra arms of human tRNA^{Ser/Sec} and tRNA^{Ser} function as major identify elements for serylation in an orientation-dependent, but not sequence-specific manner. *Nucleic Acids Res.*, **21**, 5589–5594.
- Suzuki, T., Ueda, T. and Watanabe, K. (1997) The 'polysemous' codon—a codon with multiple amino acid assignment caused by dual specificity of tRNA identity. *EMBO J.*, **16**, 1122–1134.
- Himeno, H., Hasegawa, T., Ueda, T., Watanabe, K. and Shimizu, M. (1990) Conversion of aminoacylation specificity from tRNA^{Tyr} to tRNA^{Ser} in vitro. *Nucleic Acids Res.*, **18**, 6815–6819.
- Tukalo, M., Yaremchuk, G., Kovalenko, O., Krikliviy, I. and Gudžera, O. (2013) Recognition of tRNAs with a long variable arm by aminoacyl-tRNA synthetases. *Biopolymers Cell*, **29**, 311–323.
- Biou, V., Yaremchuk, A., Tukalo, M. and Cusack, S. (1994) The 2.9 Å crystal structure of *T. thermophilus* seryl-tRNA synthetase complexed with tRNA^{Ser}. *Science*, **263**, 1404–1410.
- Mizutani, T., Kanaya, K., Ikeda, S., Fujiwara, T., Yamada, K. and Totsuka, T. (1998) The dual identities of mammalian tRNA^{Sec} for SerRS and selenocysteine synthase. *Mol. Biol. Rep.*, **25**, 211–216.
- Pallanck, L., Li, S. and Schulman, L.H. (1992) The anticodon and discriminator base are major determinants of cysteine tRNA identity in vivo. *J. Biol. Chem.*, **267**, 7221–7223.
- Komatsoulis, G.A. and Abelson, J. (1993) Recognition of tRNA^{Cys} by *Escherichia coli* cysteinyl-tRNA synthetase. *Biochemistry*, **32**, 7435–7444.
- Turanov, A.A., Lobanov, A.V., Fomenko, D.E., Morrison, H.G., Sogin, M.L., Klobutcher, L.A., Hatfield, D.L. and Gladyshev, V.N. (2009) Genetic code supports targeted insertion of two amino acids by one codon. *Science*, **323**, 259–261.
- Ardell, D.H. and Andersson, S.G. (2006) TFAM detects co-evolution of tRNA identity rules with lateral transfer of histidyl-tRNA synthetase. *Nucleic Acids Res.*, **34**, 893–904.
- Llado, S., Benada, O., Cajthaml, T., Baldrian, P. and Garcia-Fraile, P. (2016) *Silvibacteriumbohemicum* gen. nov. sp. nov., an acidobacterium isolated from coniferous soil in the Bohemian Forest National Park. *Syst. Appl. Microbiol.*, **39**, 14–19.
- Hou, Y.M. and Schimmel, P. (1988) A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature*, **333**, 140–145.

44. McClain, W.H. and Foss, K. (1988) Changing the identity of a tRNA by introducing a G-U wobble pair near the 3' acceptor end. *Science*, **240**, 793–796.
45. Naganuma, M., Sekine, S., Chong, Y.E., Guo, M., Yang, X.L., Gamper, H., Hou, Y.M., Schimmel, P. and Yokoyama, S. (2014) The selective tRNA aminoacylation mechanism based on a single G•U pair. *Nature*, **510**, 507–511.
46. Lewendon, A., Murray, I.A., Shaw, W.V., Gibbs, M.R. and Leslie, A.G. (1990) Evidence for transition-state stabilization by serine-148 in the catalytic mechanism of chloramphenicol acetyltransferase. *Biochemistry*, **29**, 2075–2080.
47. Wen, Y., Behiels, E. and Devreese, B. (2014) Toxin-Antitoxin systems: their role in persistence, biofilm formation, and pathogenicity. *Pathog. Dis.*, **70**, 240–249.
48. Ataide, S.F., Jester, B.C., Devine, K.M. and Ibba, M. (2005) Stationary-phase expression and aminoacylation of a transfer-RNA-like small RNA. *EMBO Rep.*, **6**, 742–747.
49. Liu, C., Sanders, J.M., Pascal, J.M. and Hou, Y.M. (2012) Adaptation to tRNA acceptor stem structure by flexible adjustment in the catalytic domain of class I tRNA synthetases. *RNA*, **18**, 213–221.
50. Lacourciere, G.M., Levine, R.L. and Stadtman, T.C. (2002) Direct detection of potential selenium delivery proteins by using an *Escherichia coli* strain unable to incorporate selenium from selenite into proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 9150–9153.
51. Mizutani, T. and Goto, C. (2000) Eukaryotic selenocysteine tRNA has the 9/4 secondary structure. *FEBS Lett.*, **466**, 359–362.