Expansion of the Genetic Code Deciphered by Molecular Biological Analysis of Genus-Specific Transfer RNA

Kiyofumi Hamashima

A dissertation for the degree of *Doctor of Philosophy*

in the Graduate School of Media and Governance

Systems Biology Program

KEIO UNIVERSITY

2014

Abstract

The genetic code is a set of essential and fundamental rules for living cells, and is highly conserved among all organisms. During translation according to the genetic code, transfer RNA (tRNA) acts as an adaptor molecule by physically linking the nucleotide sequence of genetic information and the amino acid sequence of a protein. In some cases, tRNA changes have direct and specific effects on the decoding process, and hence a full investigation of the evolution and function of the tRNA molecule is necessary for a comprehensive understanding of the genetic code. In this study, we focused on the evolutionary divergence of eukaryotic tRNAs, and examined their structural and chemical properties based on bioinformatics and molecular biology approaches. We identified a novel type of tRNA (designated "nev-tRNA") with unusual structural characteristics that specifically diverged in the nematode lineage. Surprisingly, these tRNAs can translate nucleotides *in vitro* in a manner that transgresses the genetic code. We also confirmed that nev-tRNAs are expressed, matured, and exported from the nucleus *in vivo*. However, it is highly likely that nev-tRNAs are not used in protein biosynthesis, at least under normal growth conditions, and are therefore assumed to be involved in specific processes, such as responses to environmental changes. These findings provide the first example of unexpected tRNAs that can potentially alter the general translation rule for higher eukaryotes, and allow new insights into the genetic code and a new perspective on modern tRNA biology.

Keywords: Genetic code; Transfer RNA; Protein; Translation; Nematode; Molecular evolution

論文題目

種属特異的 tRNA の分子生物学的解析に基づく遺伝暗号多様化の理解

論文要旨

A, T, C, G の 4 文字で表現される DNA の暗号文をもとに, アミノ酸から成る生命活動に不可欠なタンパク質が翻訳される際には, 遺伝暗号と呼ばれる厳密な対応関係が存在する. 遺伝暗号の実体は 3 文字ずつの組み合わせが計 20 種類のアミノ酸を指定することであり, この生命システムの大原則は多くの生物で共通である. Transfer RNA (tRNA) は遺伝暗号を読み解く際にアダプター分子として働くため, 遺伝情報からタンパク質への架け橋というまさにセントラルドグマの核心を担う. それゆえ, 本分子の進化や機能を探求することは, 遺伝暗号の成り立ちや普遍性を議論する上で欠かすことができない. そこで本研究ではなかでも真核生物 tRNA に着目し, 生命情報学・実験生物学的手法を併用してその進化的多様性や構造的特徴, および化学的特性を詳細に調べた. その結果, 通常とは異なる分子構造をもつ奇妙な tRNA を, 線虫というある種属特異的に発見し, nev-tRNA と命名した. また試験管内の実験により, これらが普遍的と考えられている遺伝暗号を変則的な暗号へと変換する活性を有することを明らかにした. さらに線虫の生体内において, nev-tRNA が翻訳に使用可能な状態で存在していることも見出した. しかし, 少なくとも通常飼育環境下では翻訳に使用されている可能性は低く, 環境ストレス応答といった条件特異的な機構への関与が推察される. 以上の成果は, tRNA の分子構造の変化により遺伝暗号が拡張し得ることを示した高等真核生物における初めての例であり, 遺伝暗号の更なる理解や近代の tRNA 研究への貢献が期待される.

キーワード：遺伝暗号; Transfer RNA; タンパク質; 翻訳; 線虫; 分子進化

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1. Alternative genetic code for amino acids and transfer RNA

The genetic code is a set of essential and fundamental rules for living cells. In general, the genetic information in DNA is transcribed to messenger RNA (mRNA), and proteins are synthesized based on the genetic code using the information in the template mRNA. Because many organisms use the same genetic code and any change would produce widespread changes in the amino acid sequences of proteins, the code was thought to be invariable in all organisms ("frozen accident") (Crick 1968). However, in 1979, mammalian mitochondria were found to use a code that deviates from the universal genetic code (Barrell *et al*. 1979), and since then, not only further differences in the mitochondrial code but also in the nuclear code have been discovered (summarized in Table 1. 1). In many cases, stop codons are reassigned to various sense codons, which designate amino acids. For example, the UGA stop codon has been reassigned to either tryptophan (Trp) in a few Bacteria (such as *Mycoplasma*) (Yamao *et al*. 1985; Lovett *et al*. 1991; McCutcheon *et al*. 2009) and certain ciliated protozoans

(Lozupone *et al*. 2001), or cysteine (Cys) in the ciliate *Euplotes* (Lozupone *et al*. 2001). The UAA stop codon has been reassigned to glutamic acid (Glu) in three peritrich species, *Vorticella microstoma*, *Opisthonecta henneguyi*, and *O. matiensis* (Sánchez-Silva *et al*. 2003). Moreover, many ciliates (Lozupone *et al*. 2001), ulvophycean green algae (Schneider and de Groot 1991), diplomonads (Keeling and Doolittle 1997), and oxymonads (Keeling and Leander 2003) use UAR (UAA and UAG) for glutamine (Gln). Stop codons have also been used to expand the genetic code to include selenocysteine (Sec) (Söll 1988) and pyrrolysine (Pyl) (Hao *et al*. 2002; Srinivasan *et al*. 2002). In all kingdoms of life, the 21st proteinogenic amino acid, Sec, is usually an encoded amino acid in enzymes involved in oxidation–reduction reactions (Söll 1988). In some methanogenic Archaea and Bacteria, the 22nd proteinogenic amino acid, Pyl, is present in enzymes involved in methane-producing metabolism (Hao *et al*. 2002; Srinivasan *et al*. 2002). In contrast, in several species of the genera *Candida* and *Debaryomyces*, a sense codon, the leucine (Leu)-designating CUG codon, has been reassigned to serine (Ser) (Sugita and Nakase 1999). This widespread occurrence of deviant codes clearly indicates that the genetic code is not universal.

Transfer RNA (tRNA) is a small non-coding RNA of about 70–85 nucleotides that acts as an adapter molecule between the nucleotide sequences of mRNAs and the amino acid sequences of proteins during translation. In some cases, a single-nucleotide substitution on the tRNA molecule has direct and specific effects on the decoding process (Giegé *et al*. 1998), so tRNA is considered one of major factors involved in non-standard codon assignments. These tRNA changes can be categorized into two

types. The first involves tRNA/mRNA pairing. For instance, a specific mutation in the tRNA anticodon occurs when a stop codon is reassigned, as described above. In *Mycoplasma capricolum*, besides the common $tRNA^{Trp}$ with a CCA anticodon, a deviant $tRNA^{Trp}$ with a UCA anticodon is also encoded and decodes the UGA stop codon as Trp (Osawa *et al*. 1992). Similarly, RNA editing of the tRNA anticodon causes limited codon reassignment in several mitochondria. In the kinetoplastid protist *Leishmania tarentolae*, the host $tRNA^{Trp}$ with the UCA anticodon is imported into the mitochondrion and the anticodon is converted to UCA by RNA editing (Alfonzo *et al*. 1999). The resulting tRNA decodes both the UGG and UGA codons to Trp only in the mitochondria, and not in the nucleus (Alfonzo *et al*. 1999). Additionally, a number of alterations of post-transcriptional base modifications at tRNA anticodons modify the codon/anticodon base-pairing rules ("wobble rule"), resulting in codon reassignments (Knight *et al*. 2001). The second type of tRNA change that alters the genetic code affects tRNA/aminoacyl-tRNA synthetase (aaRS) recognition. The attachment of a specific amino acid to the 3′ end of each tRNA is catalysed by enzymes called aaRSs, through the appropriate tRNA recognition (McClain 1993). Therefore, mutations of the tRNA identity elements, which are specifically recognized by only one aaRS during aminoacylation, are also known to be one of the leading causes of deviations in the genetic code (Knight *et al*. 2001). These findings clearly suggest that the genetic code is still in a state of evolution, and that this evolution is closely related to tRNA diversification.

Table 1. 1. Summary of nuclear genetic code alternatives.

Codon	Standard to alternative code	Organism
UGA	Stop → Trp	Firmicutes *Mycoplasma* spp. *Spiroplasma citri* *Bacillus subtilis* Proteobacteria *Hodgkinia cicadicola* Ciliates *Colpoda inflata* *Blepharisma americanum*
UGA	Stop → Cys	Ciliates *Euplotes* spp.
UAA	Stop → Glu	Ciliates *Vorticella microstoma* *Opisthonecta henneguyi* *Opisthonecta matiensis*
UAR	Stop → Gln	Many Ciliates All diplomonads other than *Giardia* Oxymonads *Streblomatrix strix* Green algae *Acetabularia* spp. *Batophora oerstedii*
UGA	Stop → Sec	Many species in three domains
UGA	Stop → Sec/Cys	Ciliates *Euplotes crassus*
UAG	Stop → Pyl	Some methanogenic archaea and bacteria
CUG	Leu → Ser	Fungi Many *Candida* spp. Many Ascomycetes

1.2. Extensive modern tRNA biology

In the 1960s–1970s, the role of tRNA was considered to be restricted to the delivery of specific amino acids to the ribosome during protein synthesis. However, since the 1980s, a growing number of studies have demonstrated alternative tRNA functions (Figure 1. 1) (Banerjee *et al*. 2010). For example, uncharged tRNAs have been shown to function as *i*) primers for DNA synthesis (Saadatmand and Kleiman 2012); *ii*) sensors of amino acid concentrations (Hinnebusch 2005; Wendrich *et al*. 2002); and *iii*) regulators of gene transcription termination (Henkin 2008). During the replication of human immunodeficiency virus 1 (HIV-1), uncharged $tRNA^{Lys}$ is used to prime the reverse transcription of the viral RNA genome to double-stranded DNA, which is then integrated into the host genome (Saadatmand and Kleiman 2012). In certain eukaryotic cells under amino acid starvation, uncharged tRNAs interact with the protein Gcn2p, which induces eIF2 phosphorylation and reduces the overall level of translation (Hinnebusch 2005). This phosphorylation causes the activation of the transcriptional regulator Gcn4p and consequently increases amino acid production (Hinnebusch 2005). In Gram-negative bacteria, such as *Escherichia coli*, uncharged tRNAs block protein synthesis by penetrating the A site of the ribosome and induce the production of the global transcriptional regulator ppGpp in response to amino acid starvation, as in eukaryotes (Wendrich *et al*. 2002). In Gram-positive bacteria, uncharged tRNAs act as riboswitches in the T-box transcription termination system (Henkin 2008). When a tRNA is poorly charged, an uncharged tRNA binds to the 5′

untranslated region (UTR) on the mRNA and prevents the formation of a stable terminator helix, causing read-through of the termination site and the transcription of the downstream genes, such as those encoding aaRS and the proteins involved in amino acid biosynthesis and transport (Henkin 2008). It has also been shown that the ribosome is not the only destination to which tRNAs deliver amino acids. tRNAs are used in other amino acid addition pathways, ranging from lipid modification to antibiotic biosynthesis (Peschel *et al*. 2001).

Since 2005, pseudo-tRNAs, tDNAs, and tRNA-derived fragments have, like mature common tRNAs, also been attributed more extensive roles. Recent genome sequencing projects have identified a large number of tRNA isodecoders, which share the same anticodon but differ in their body sequences. However, many of these annotated tRNA genes were automatically predicted using a computational approach, such as tRNAscan-SE (Lowe and Eddy 1997), ARAGORN (Laslett and Canback 2004) and SPLITS (Sugahara *et al*. 2006, 2007), and parts of these tRNA molecules lack canonical features, such as the conserved secondary structure, and are classified as pseudo-tRNAs. Therefore, their functions were largely unknown. Rudinger-Thirion *et al*. have shown that one human $tRNA^{Asp}$ is poorly aminoacylated but is used in a regulatory role beyond translation (Rudinger-Thirion *et al*. 2011). This tRNA isodecoder binds directly to a partial Alu sequence in the 3′ UTR of the aspartyl-tRNA synthetase (AspRS) mRNA, and modulates the stability of the mRNA (Figure 1. 2A) (Rudinger-Thirion *et al*. 2011). Similarly, in *Bacillus cereus* and several other *Bacillus* species, even though a predicted pseudo-$tRNA^{Trp}$ is poorly aminoacylated and does not

associate with polysomes *in vitro*, it plays a role in the regulation of tryptophanyl-tRNA synthetase (TrpRS) gene expression in the stationary phase (Ataide *et al*. 2009). The deletion of this tRNA isodecoder led to significant changes in the cell's wall morphology and antibiotic resistance and was accompanied by changes in the expression of numerous genes involved in the cellular responses to oxidative stress (Rogers *et al*. 2012). Furthermore, expanded tRNA roles have been reported not only at the level of RNA but also at the DNA level (McFarlane and Whitehall 2009). In certain eukaryotic cells, tRNA genes (tDNAs) are repetitive sequences dispersed throughout the whole genome and function as chromatin insulators, helping to separate active chromatin domains from silenced ones (Ebersole *et al*. 2011; Raab *et al*. 2012). tDNAs can also block enhancers from activating promoters transcribed by RNA polymerase II (Ebersole *et al*. 2011; Raab *et al*. 2012). This finding indicates that tRNA (tDNA) can potentially mediate the spatial and functional organization of the genome and drive genome change and evolution (McFarlane and Whitehall 2009).

Recently, it has been demonstrated that tRNA fragments are often generated by endonucleolytic cleavage of tRNAs under specific conditions. However, there are significant differences between the mechanisms involved in this phenomenon in prokaryotes and eukaryotes (Thompson and Parker 2009a; Banerjee *et al*. 2010). In prokaryotes, tRNA cleavage mainly occurs to rapidly reduce the levels of tRNA, thereby reducing protein translation, as a cellular defence mechanism against competing organisms (Masaki and Ogawa 2002). For instance, in *E. coli*, the plasmid-encoded nuclease PrrC cleaves tRNAs in their anticodon loops and completely depletes

full-length tRNAs in response to bacteriophage infection (Levitz *et al*. 1990). In eukaryotes, although similar tRNA cleavage within the anticodon loop has been reported (Thompson and Parker 2009a; Banerjee *et al*. 2010), in most cases, full-length tRNA levels do not decline markedly and tRNA fragment levels are consistently lower than those of full-length tRNAs (Lee and Collins 2005; Thompson *et al*. 2008; Yamasaki *et al*. 2009). This suggests that eukaryotic tRNA cleavage has functions other than to reduce tRNA levels. Recent studies have reported regulatory roles for these tRNA fragments, as reviewed here.

In eukaryotic cells, tRNA cleavage is a conserved part of the responses to a variety of stresses (Thompson and Parker 2009a; Banerjee *et al*. 2010). For example, in *S. cerevisiae, Arabidopsis*, and humans, tRNA cleavage occurs during oxidative stress (Thompson *et al*. 2008). In yeast, Rny1, a member of the RNase T7 family, is activated and moves into the cytoplasm from the vacuole in response to oxidative stress, where it cleaves tRNAs into fragments (Thompson and Parker 2009b). tRNAs are also cleaved in mammalian cells and the potential impact of this cleavage there is more fully understood. Although mammalian cells express an orthologue of Rny1, called RNASET2, stress-induced tRNA cleavage depends on a member of the RNase A family, angiogenin, rather than on RNASET2 (Yamasaki *et al*. 2009; Fu *et al*. 2009; Saikia *et al*. 2012; Wang *et al*. 2012). Angiogenin is normally localized in the nucleus and is regulated by the ribonuclease inhibitor RNH1 (Shapiro and Vallee 1987; Tsuji *et al*. 2005). In response to oxidative stress, angiogenin is dissociated from RNH1 and enters the cytoplasm, where it cleaves the anticodon loops of specific tRNAs (Yamasaki *et al*.

2009). A subset of 5′ tRNA fragments, which contain 4–5 consecutive guanine residues at their 5′ ends, interacts with the translational silencer YB-1 (Ivanov *et al*. 2011). This complex inhibits the initiation of translation by recruiting the general translation initiation factors eIF4E/G/A from capped mRNAs or eIF4G/A from uncapped mRNAs (Figure 1. 2B) (Ivanov *et al*. 2011). In addition to oxidative stress, amino-acid-starvation-induced, age-associated or tissue-specific tRNA fragmentation has also been observed in many eukaryotic cells (Lee and Collins 2005; Kato *et al*. 2011; Peng *et al*. 2012), but its biological roles are still unclear.

Current RNA cloning and high-throughput sequencing methods are sufficiently sensitive to capture even RNA fragments that are present in the cell in very few copies. This has allowed novel secreted tRNA-derived sRNAs to be identified. In the ciliate *Tetrahymena thermophila*, RNA fragments derived from the 3′ ends of mature tRNAs were detected by deep sequencing 18–22-nt sRNAs co-purified with Twi12, which is a growth-essential *Tetrahymena* Piwi protein that forms a complex with the exonuclease Xrn2, which is involved in cellular ribosomal RNA processing (Couvillion *et al*. 2010, 2012). The binding of tRNA fragments to Twi12 is required for the stabilization, localization, and activation of this complex (Figure 1. 2C) (Couvillion *et al*. 2010, 2012). Against this background, the presence of prokaryotic tRNA fragments and their possible functions have been revised, and further experiments have suggested that specific tRNA fragmentation occurs more frequently than previously thought. For instance, Murakami *et al*. collected hot spring water containing uncultured organisms directly from the underground environment and analysed the sRNA sequences isolated

from it (Murakami *et al*. 2011). Their results demonstrated the presence of a large number of novel tRNA fragments and unique relationships between tRNA anticodons and their cleavage sites (Murakami *et al*. 2011). Taken together, these observations support the view that tRNA fragments have extensive biological functions.

These biological functions are accomplished by uncharged tRNAs (or pseudo-tRNA, tDNAs and tRNA fragments), but recent studies have demonstrated that aminoacylated tRNAs also have extended roles. As explained in the previous section, the most crucial role of tRNAs is to transfer the correct amino acids to the ribosome in accordance with the genetic code. Translational fidelity is essential for protein and cell integrity, which is achieved by accurate tRNA aminoacylation. The error rate of aminoacylation has been shown to be one per 10,000–1,000,000 couplings when purified aaRSs were used (Cochella and Green 2005). However, in certain eukaryotic cells, Met is misacylated to specific non-methionyl-tRNA families and these Met-misacylated tRNAs are used in translation in response to oxidative stress (Netzer *et al*. 2009; Jones *et al*. 2011; Wiltrout *et al*. 2012). The Met-misacylation function is thought to protect cells against oxidative stress by increasing Met incorporation into proteins because Met residues protect proteins from reactive oxygen species (ROS)-mediated damage (Figure 1. 2D) (Netzer *et al*. 2009). These unexpected functional repertoires of pseudo-tRNAs, tDNAs, aminoacylated tRNAs and tRNA fragments imply the existence of further regulatory roles for tRNA harboured in the genome (Hamashima and Kanai 2013).

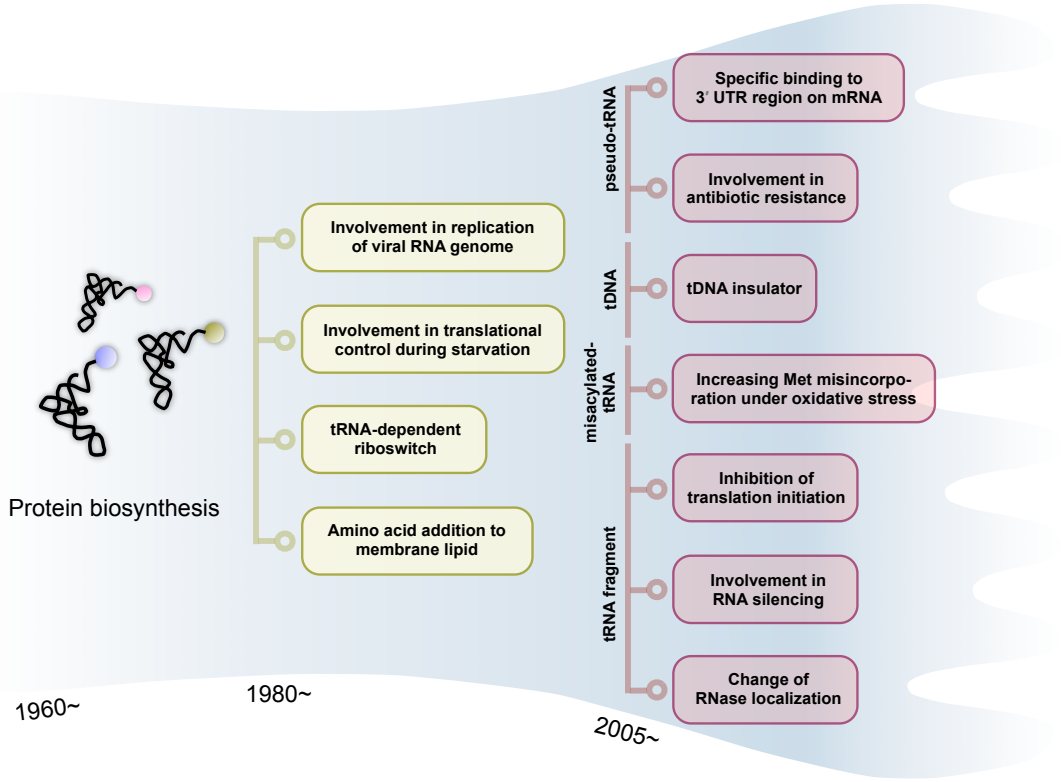

Figure 1. 1. Expansion of known tRNA functions during the past 50 years.

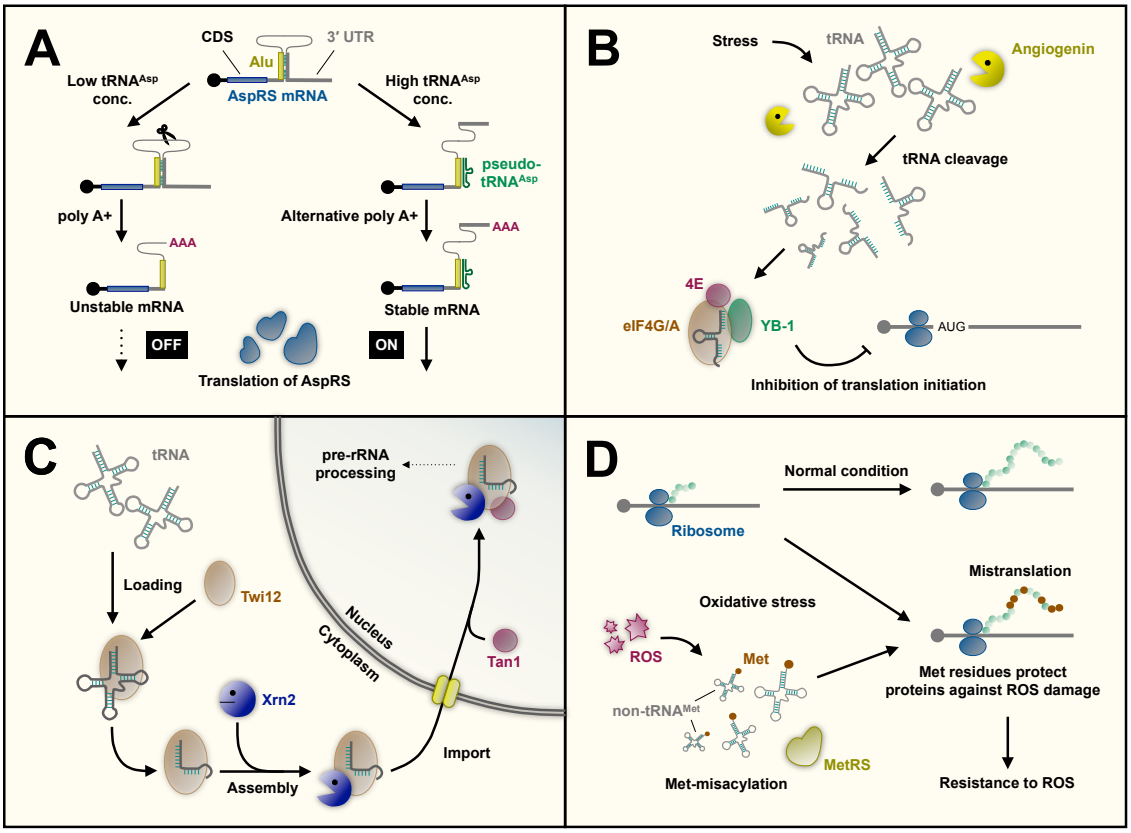

Figure 1. 2. Recently proposed novel aspects of tRNA as regulator RNAs.

(A) Regulation of target-specific gene expression by pseudo-tRNA. (B) Stress-induced tRNA-fragment-dependent inhibition of translation initiation. (C) tRNA-fragment-dependent activation of the exonuclease Xrn2 for RNA processing in *Tetrahymena*. (D) Modification of translational fidelity triggered by oxidative stress.

1.3. Objectives

The aim of this study is to clarify and characterize the alternative genetic code arising from tRNA diversification and to elucidate its impact on translation. As described above, the evolution of tRNA correlates strongly with the evolution of the genetic code and with biodiversity in general. Furthermore, the recently reported extensive functions of tRNAs provide important new perspectives on tRNA biology. However, the evolutionary divergence of the tRNAs of higher eukaryotes and its relevance to the genetic code have not yet been fully examined. A comprehensive understanding of the genetic code requires a detailed understanding of the tRNA molecule.

Therefore, we comprehensively reanalyzed the eukaryotic tRNA genes from 44 eukaryote species, and examined their structural and chemical properties. We identified a novel type of tRNA (nev-tRNA) that has specifically diverged in the nematode (worm) lineage and, surprisingly, translates nucleotides *in vitro* in a manner that transgresses the code (Chapter 2). However, a whole-cell proteomic analysis found no detectable level of nev-tRNA-induced mistranslation in *Caenorhabditis elegans*, suggesting that, contrary to our expectation, the nematode genetic code is not ambiguous and ensures high translational fidelity (Chapter 3). These findings provide the first example of unexpected tRNAs that can potentially alter the general translation rules, and further insight into the evolution of the genetic code and tRNA.

Chapter 2

Nematode-specific tRNAs that decode an alternative genetic code for leucine

2.1. Introduction

Transfer RNAs (tRNAs) can be classified into two groups based on structural differences in their variable regions: class I tRNAs have a short variable region of four to five nucleotides, whereas class II tRNAs (e.g., $tRNA^{Leu}$, $tRNA^{Ser}$, and bacterial $tRNA^{Tyr}$) have a long variable arm (V-arm) structure containing 10 or more nucleotides (Sprinzl *et al*. 1998). The attachment of the correct amino acid to the 3′ end of each tRNA is catalyzed by an aminoacyl-tRNA synthetase (aaRS), which collectively comprise a protein family, with very high accuracy (Freist 1989). The recognition elements for class I tRNAs are mainly found in their acceptor stems and anticodon domains (McClain 1993). In contrast, the recognition elements for class II tRNAs occur in the acceptor stem, D-stem, and long V-arm (Breitschopf *et al*. 1995; Soma *et al*. 1999; Biou *et al*. 1994; Yaremchuk *et al*. 2002), with the anticodons generally less important for their aminoacylation (Saks *et al*. 1994). The long V-arms of $tRNA^{Leu}$ and

$tRNA^{Ser}$ are conserved in all organisms and organelles, except animal mitochondria, whereas the long V-arms of archaeal and eukaryotic $tRNA^{Tyr}$ were probably lost soon after the separation of the domain Bacteria (Achsel and Gross 1993). Hence, the evolutionary divergence of class II tRNAs occurred after the differentiation of the three domains of life.

The comprehensive and phylogenetic analysis of the tRNA gene is becoming more important to understand the diversity of tRNA molecules in living organisms, as described in section 1.2. Recent studies found that various types of tRNAs are encoded in a unique manner. These include *i*) intron-containing tRNAs, which contain a maximum of three introns located at various nucleotide positions, and are found predominantly in crenarchaeal species (Sugahara *et al*. 2006, 2007, 2008, 2009); *ii*) tri-split tRNAs, comprising three transcripts of *Caldivirga maquilingensis* (Fujishima *et al*. 2009); and *iii*) permuted tRNAs, in which the 5′ and 3′ halves are encoded in a permuted orientation. These permuted tRNAs have been found in *Cyanidioschyzon merolae* and other unicellular red/green algae (Soma *et al*. 2007; Maruyama *et al*. 2009), and more recently, in the crenarchaeon *Thermofilum pendens* by Todd M. Lowe's group (Chan *et al*. 2011). The tRNA genes in higher eukaryotes were thought to contain a single intron at most, located in the canonical position (37/38), one nucleotide downstream from the anticodon (Sugahara *et al*. 2009; Marck and Grosjean 2002), but a comprehensive analysis of eukaryotic tRNA genes has revealed two putative $tRNA^{Gly}$ and $tRNA^{Ile}$ genes (with class I anticodons) in the *Caenorhabditis elegans* genome that harbor 15–16 nucleotide V-arm structures (Marck and Grosjean 2002). These tRNA-like

sequences have features found only in typical class II tRNAs, and their evolutionary origins and cellular functions remain unclear.

To assess the evolutionary origins, diversity, and functions of these putative tRNA genes, we comprehensively reanalyzed the eukaryotic tRNA genes and identified over 100 class II tRNA candidates with anticodons corresponding to neither leucine nor serine, but possessing the long V-arm. These sequences are specifically conserved among the nematodes and contain recognition elements for aaRSs similar to those of $tRNA^{Leu}$. Surprisingly, we found that these tRNAs in *C. elegans* can only be aminoacylated with leucine, but recognize codons corresponding to other amino acids. These findings identify a new type of tRNA that translates nucleotides in a manner that transgresses the genetic code, emphasizing the importance of considering the evolutionary processes underlying the general translation rules (Hamashima *et al.* 2012).

2.2. Materials and Methods

2.2.1. Phylogenetic analysis of the tRNA genes of three nematode species

All mature tRNA sequences, including nematode-specific V-arm-containing tRNAs (nev-tRNAs) in *C. elegans* and *C. brenneri*, were collected from the Genomic tRNA database (Chan and Lowe 2009), and the *C. japonica* and *Pristionchus pacificus* tRNA sequences were predicted using tRNAscan-SE. Any redundant tRNA sequences (multicopy tRNAs with identical sequences) were discarded. A neighbor-joining tree of nonredundant mature tRNA sequences was constructed using ClustalX (Larkin *et al*. 2007) and visualized with iTOL (Letunic and Bork 2007).

2.2.2. Preparation of total RNA

The N2 nematode strain and *Escherichia coli* strain OP50 used in this work were provided by the *Caenorhabditis* Genetics Center, which is funded by the NIH National Center for Research Resources. Total RNA was isolated from mixed-stage *C. elegans*, including eggs, larvae 1–4, and adults, using TRIzol Reagent (Invitrogen, Carlsbad, CA, USA), according to the manufacturer's instructions. For acid–urea polyacrylamide gel electrophoresis (PAGE) /northern blot analysis, the isolated RNA was incubated with 0.2 M Tris-HCl (pH 9.5) at 37°C for 2 h to deacylate the charged tRNAs.

2.2.3. Northern blot analysis

Total RNA (10 μg per lane) was separated on a denaturing 6% polyacrylamide gel containing 8 M urea and transferred onto Hybond-N+ membranes (GE Healthcare, Buckinghamshire, UK) by electroblotting. The membranes were hybridized with specific oligonucleotide probes (Table 2. 1), labeled with a Biotin 3′ End DNA Labeling Kit (Pierce Biotechnology, Rockford, IL, USA) in ULTRAhyb-Oligo Hybridization Buffer (Ambion, Austiin, TX, USA) at room temperature. The hybridized membranes were treated with 5× Wash Buffer (Ambion) at room temperature. The nonisotopic blots were detected with the BrightStar BioDetect Kit (Ambion) using the Enhanced ChemiFluorescence (ECF) substrate (GE Healthcare). The images were visualized with a Molecular Imager FX Pro (Bio-Rad Laboratories, Hercules, CA, USA).

2.2.4. RT–PCR analysis and nucleotide sequencing

Reverse transcription (RT)–PCR was performed with the enzymes, ReverTra Dash and KOD FX (Toyobo Biochemicals, Osaka, Japan). All amplification reactions, except those amplifying nev-tRNA^{Gly} (CCC), consisted of 30 cycles of denaturation at 98°C for 10 s, annealing at 55°C for 3 s, and extension at 74°C for 6 s. PCR of nev-tRNA^{Gly} was performed with stepdown amplification, consisting of 15 cycles of denaturation at 98°C for 10 s, annealing at 74°C (5 cycles), 72°C (5 cycles), 70°C (5 cycles), and 60°C (15 cycles) for 6 s, and extension at 74°C for 6 s. The PCR products were separated by electrophoresis on 3% NuSieve 3:1 agarose gels (Cambrex Bio Science, Rockland, ME, USA), which were stained with ethidium bromide. The bands

were visualized with a Molecular Imager FX Pro (Bio-Rad Laboratories). The primers used in this analysis are summarized in Table 2. 2. The RT–PCR products were further purified with an Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare), and subcloned into the pCR–Blunt II–TOPO vector (Invitrogen). The nucleotide sequences of the inserted DNAs were determined on an ABI PRISM 3100 DNA Sequencer (Applied Biosystems, Foster City, CA, USA).

Table 2. 1. List of oligonucleotides used in northern blot analysis.

Type	Anticodon	Length	Sequence - 5' to 3'
Control	Gly (GCC)	36	CCCGGGCCGCCCGCGTGGCAGGCGAGCATTCTACCA
	Ile (UAU)	38	TATAAGTACCACGCGCTAACCGACTGCGCCAATGGGGC
nev-tRNA	Gly (CCC)	35	CTCTGTTACCAGAATAGGATCGGGAGTCCTACGCC
	Gly (CCC)	35	TCGAACCCGCGCTCTGTTACCAGAATAGGATCGGG
	*Gly (CCC)	83	TGCGGTGGACGGGATTCGAACCCGCGCTCTGTTACCAGAATAG GATCGGGAGTCCTACGCCTTCACCGCTCGGCCACCACCGC
	Ile (UAU)	36	ACTGGTTTAACCCAATGAGATCATAAGTCTCACGCC
	Ile (UAU)	36	TCGAACCCGCGACTGGTTTAACCCAATGAGATCATA
	Lys (CUU)	36	TCGGTTAGAACCGTTTGAGATCAAGTGTCTCATGCC

* for acid-urea/northern blot analysis (see Figure 2. 8)

Table 2. 2. List of oligonucleotides used in RT–PCR analysis.

Type	Anticodon	Strand	Sequence - 5' to 3'
Control	Gly (GCC)	sense	GCATCGGTGGTTCAGTGGTAGA
	Gly (GCC)	antisense	TGCATCGACCGGGAATCGAACC
	Ile (UAU)	sense	GCCCCATTGGCGCAGTCGGTTAGC
	Ile (UAU)	antisense	TGCCCCATGCCAGGCTCGAACTG
nev-tRNA	Gly (CCC)	sense	GCGGTGGTGGCCGAGCGGTC
	Gly (CCC)	antisense	TGCGGTGGACGGGATTCGAACC
	Ile (UAU)	sense	GCCCCGGTGGCCGAGCGGTCGAAG
	Ile (UAU)	antisense	TGCCCCGGGCGGGATTCGAACCC
	Lys (CUU)	sense	GACACGGTGGCCGAGTGGTTT
	Lys (CUU)	antisense	TGACACGGGCAGGATTCGAACC

2.2.5. Preparation of recombinant aminoacyl-tRNA synthetases from *C. elegans*

C. elegans total RNA was used as the template for RT–PCR of four genes encoding aminoacyl-tRNA synthetases (aaRSs): glycyl-tRNA synthetase (GlyRS, GenBank accession no. NP_871640.1), isoleucyl-tRNA synthetase (IleRS, GenBank accession no. NP_501914.1), leucyl-tRNA synthetase (LeuRS, GenBank accession no. NP_497837.1), and seryl-tRNA synthetase (SerRS, GenBank accession no. NP_501804.1). RT–PCR was performed using the ReverTra Dash Kit and KOD FX (Toyobo Biochemicals), and the amplified products were cloned into the pET–23b expression vector (Novagen, Madison, WI, USA). Each resulting vector encoded a full-length aaRS with a six-histidine tag at its C-terminal end. The nucleotide sequences of the inserted DNA fragments were determined on an ABI PRISM 3100 DNA Sequencer (Applied Biosystems) and confirmed to be identical to those in the database. The primers and restriction enzymes used for cloning are summarized in Table 2. 3.

The resulting vectors were used to transform *E. coli* strain BL21 (DE3) pLysS (IleRS and LeuRS) or HMS174 (DE3) pLysS (GlyRS and SerRS). The transformants, growing logarithmically at 37°C in Luria–Bertani medium containing 50 μg/mL ampicillin and 35 μg/mL chloramphenicol, were supplemented with 0.1–0.4 mM isopropyl-β-D-thiogalactopyranoside. After 14–16 h of further growth at 16°C, the cells were pelleted by centrifugation (5000 × *g* for 5 min at 4°C) and sonicated (3 min) in 1 × phosphate buffered saline (PBS) with 10 mM imidazole. The extracts were centrifuged at 12,000 × *g* for 10 min at 4°C to remove any debris, and the recombinant proteins were purified with a Proteus IMAC Protein Purification Kit (Pro-Chem, Littleton, MA,

USA). The eluted protein solutions were pooled and gel filtered to remove any salt on a HiTrap column (GE Healthcare) with buffer A (50 mM Tris-HCl [pH 8.0], 1 mM EDTA, 0.02% Tween 20, 7 mM 2-mercaptoethanol, and 10% glycerol).

Table 2. 3. List of oligonucleotides for molecular cloning.

Gene Name	Providing Restriction Enzyme	Strand	Sequence - 5' to 3'
GlyRS	Nde I	sense	TAACATATGGCTACTCCGGAAATTGA
	Not I	antisense	TAAGCGGCCGCTTCAGTTGCGCTCGCTTC
IleRS	Nhe I	sense	TAAGCTAGCAGCGGCCTTTCGACCGT
	Xho I	antisense	TAACTCGAGTCGTACGAGTTGAAGAGTCT
LeuRS	Nde I	sense	TAACATATGTCGAAAATCAATAAGGA
	Not I	antisense	TAAGCGGCCGCGTTTTCTGGGACGTTAGC
SerRS	Nhe I	sense	TAAGCTAGCGTTCTCGACATTGACAT
	Xho I	antisense	TAACTCGAGTTTCTTTCCTGTCGCCTTTT
PF1549	Sgf I	sense	TAAGCGATCGCCATGATAATAATAGACGGAAG
	Pme I	antisense	TAAGTTTAAACTTAGTGGTGGTGGTGGTGGTGCCAGGCTTTCCTAACTACTC
luciferase	Sgf I	sense	TAAGCGATCGCCATGGAAGACGCCAAAAACAT
	Pme I	antisense	TAAGTTTAAACTTAGTGGTGGTGGTGGTGGTGAACCAATTTGGACTTTCCGC

2.2.6. *In vitro* aminoacylation assay

The tRNAs synthesized *in vitro* with T7 RNA polymerase were incubated at room temperature with purified recombinant aaRSs from *C. elegans* for 0, 2.5, 5, 10, 20, 40, and 60 min in a 25 μL reaction mixture containing 50 mM Hepes (pH 7.2), 50 mM KCl, 4 mM ATP, 15 mM $MgCl_2$, 3 mM dithiothreitol, 15 mM amino acid (glycine, isoleucine, leucine, or serine), 3 μg of tRNA, and 1 μg of aaRS (except IleRS, for which we used 5 μg to accommodate its lower aminoacylation efficiency). The reaction was stopped by the addition of an equal volume of acid–urea stop buffer (Köhrer and Rajbhandary 2008). The charged tRNAs were separated on acid–urea 6.5% polyacrylamide gels, which were stained with SYBR Green II (Lonza, Rockland, ME, USA). The bands were detected with a Molecular Imager FX Pro (Bio-Rad Laboratories).

2.2.7. Acid–urea PAGE/northern blot analysis

Alkali-treated total RNA (100 μg; see section 2.2.2) was used in the *in vitro* aminoacylation assay, with purified recombinant LeuRS and GlyRS titrated from 0.01 μg to 1 μg. The assay products were separated on acid–urea 6.5% polyacrylamide gels, followed by northern blot hybridization to detect the individual tRNAs.

2.2.8. Construction of expression vectors for cell-free protein synthesis

For *in vitro* transcription/translation, we prepared two expression vectors encoding PF1549 protein (RNA 3′-terminal-phosphate cyclase of *Pyrococcus furiosus*,

GenBank accession no. NP_579278) and firefly luciferase (GenBank accession no. ACF93193.1). *Pyrococcus furiosus* genomic DNA and luciferase ICE T7 Control DNA (Promega, Madison, WI, USA) were used as the templates for the PCR of PF1549 and luciferase, respectively. PCR was performed with KOD FX (Toyobo Biochemicals), and the amplified products were cloned into the pF25A ICE T7 Flexi Vector (Promega) containing the T7 promoter. The resulting expression vectors encoded full-length PF1549 or luciferase, with six-histidine tags at their C-terminal ends. The nucleotide sequences of the inserted DNA fragments were determined on an ABI PRISM 3100 DNA Sequencer (Applied Biosystems) and confirmed to be identical to those in the database. The primers and restriction enzymes used for cloning are summarized in Table 2. 3.

2.2.9. *In vitro* translation analysis

Each 46 μL cell-free protein expression reaction mixture, with coupled transcription (driven by the T7 promoter) and translation, contained 40 μL of cell extract from the *Spodoptera frugiperda* Sf21 cell line (Promega), 4 μg of the T7 expression vector encoding PF1549 protein or firefly luciferase, and 4 μg of *in vitro*-transcribed tRNA ($tRNA^{Leu}$ (AAG) or nev-$tRNA^{Gly}$ (CCC)). After incubation at 29°C for 4 h, the cell lysates were sonicated (1 min) with a Polytron sonicator (Kinematica, Bohemia, NY, USA). The samples containing the *in vitro*-translated PF1549 proteins were incubated at 80°C for 15 min to destroy any endogenous insect proteins. The homogenates were centrifuged at 13,000 × *g* for 10 min at 4°C to remove

any debris, and the His-tagged recombinant proteins were purified with a Talon Magnetic Beads Buffer Kit (Clontech Laboratories, Mountain View, CA, USA).

2.2.10. NanoLC–MS/ MS analysis

The purified proteins were extracted with 100 mM Tris-HCl (pH 9.0) containing 12 mM sodium deoxycholate (SDC) and 12 mM sodium lauroyl sarcosinate (SLS), then reduced with 10 mM dithiothreitol at room temperature for 30 min and alkylated with 55 mM iodoacetamide in the dark at room temperature for 30 min. The samples were diluted five-fold with 50 mM ammonium bicarbonate and digested with MS-grade lysyl endoprotease, followed by overnight trypsin digestion. The SDC and SLS were removed from the sample solutions (Masuda *et al*. 2008, 2009) by the addition of ethyl acetate, acidification with 0.5% trifluoroacetic acid, shaking for 1 min, and centrifugation at 15,700 × *g* for 2 min. The aqueous phase was desalted with StageTips on a C18 Empore Disk (Rappsilber *et al*. 2003).

An LTQ-Orbitrap (Thermo Fisher Scientific, Bremen, Germany) equipped with a nanoLC interface (Nikkyo Technos, Tokyo, Japan) and a Dionex Ultimate3000 pump with a FLM-3000 flow manager (Dionex, Germering, Germany) and an HTC-PAL autosampler (CTC Analytics, Zwingen, Switzerland) were used for nano liquid chromatography–tandem mass spectrometry (nanoLC–MS/MS) measurements. Reprosil C18 materials (3 μm; Dr Maisch, Ammerbuch, Germany) were packed into a self-pulled needle (150 mm length, 375 μm O.D. × 100 μm I.D.) with a nitrogen-pressurized column loader cell (Nikkyo Technos, Tokyo, Japan) to prepare an

analytical column needle with a "stone-arch" frit (Ishihama *et al*. 2002). The flow rate was 500 nL/min. The mobile phases consisted of (A) 0.5% acetic acid and (B) 0.5% acetic acid and 80% acetonitrile. A three-step linear gradient was used: 5% to 10% B in 5 min, 10% to 40% B in 60 min, 40% to 100% B in 5 min, and 100% B for 10 min. A spray voltage of 2400 V was applied. The MS scan range was *m/z* 300–1500 in the Orbitrap analyzer and the top 10 precursor ions were selected for subsequent MS/MS scans in the ion trap.

2.2.11. Data analysis for protein identification

The raw data files were analyzed with Mass Navigator v1.2 (Mitsui Knowledge Industry, Tokyo, Japan) to create peak lists based on the recorded fragmentation spectra. The peptides were identified with Mascot v2.2 (Matrix Science, London, UK) in a comparison with an in-house protein database containing PF1549, luciferase, and keratin proteins and trypsin sequences. A precursor mass tolerance of 3 ppm and a fragment ion mass tolerance of 0.8 Da were used with strict trypsin specificity, allowing up to two missed cleavages (Olsen *et al*. 2004). Carbamidomethylation of cysteine was set as a fixed modification, and methionine oxidation was allowed as a variable modification. Peptides were rejected if the Mascot score was below the 99% confidence limit based on the "identity" score of each peptide. False-positive rates were estimated by comparison with a randomized decoy database created with the Mascot Perl program and supplied by Matrix Science.

2.3. Results and Discussion

2.3.1. Nematode-specific class II tRNAs

Although a set of unique class II tRNA-like sequences with class I tRNA anticodons has been identified in the genome of *C. elegans*, the evolutionary and functional implications of these tRNA genes remain unclear. To assess whether these tRNAs are conserved in the eukaryotes and to determine their possible functions, we collected the sequences of 49,872 tRNAs from 44 eukaryote species registered in the Genomic tRNA Database (Chan and Lowe 2009). We also obtained the genomic sequences of *C. japonica* and *P. pacificus* from the University of California Santa Cruz (UCSC) Genome Bioinformatics database (Fujita *et al*. 2010). Application of tRNAscan-SE (Lowe and Eddy 1997) with default parameters predicted an additional 2,138 tRNA genes. We screened these 52,010 tRNA sequences for candidate class II tRNAs that contain the conserved basic tRNA features (Sprinzl *et al*. 1998; Marck and Grosjean 2002), including a 7-bp acceptor stem, a 3–4-bp D-stem, a 5-bp anticodon stem, a 5-bp T-stem, a 7-nt T-loop, a 7-nt anticodon loop, and a class I tRNA anticodon specifying an amino acid other than leucine or serine. We identified a total of 115 candidate tRNAs from six nematode species: *C. brenneri*, *C. briggsae*, *C. elegans*, *C. japonica*, *C. remanei*, and *P. pacificus* (Table 2. 4). All the candidate tRNAs were observed in nematodes, with none in any other eukaryotic genome. Therefore, we designated these unusual class II tRNAs “nematode-specific V-arm-containing tRNAs” (nev-tRNAs).

When we assessed the distribution of the predicted nev-tRNA genes among the six nematode species (Figure 2. 1), we found that three species, *C. remanei*, *C. briggsae*, and *C. brenneri*, encoded an average of 31 nev-tRNAs each, with *C. brenneri* encoding 56 nev-tRNAs, corresponding to seven different anticodons. This indicates the rapid development of many nev-tRNAs. In contrast, *C. elegans*, *C. japonica*, and *P. pacificus* each encoded only 6–8 nev-tRNA genes. These three species diverged earlier within the nematode lineage than the former three species, suggesting that the number of nev-tRNAs and their anticodon variations have increased during the evolution of the nematode taxon, and are especially prominent in *C. brenneri*. These findings also indicate that nev-tRNA^{Gly} (CCC) and nev-tRNA^{Ile} (UAU) are the two major types of nev-tRNAs and are widely conserved among the nematode genomes. Intriguingly, nev-tRNA^{Gly}s (CCC), found only in *C. briggsae*, *C. brenneri*, and *C. elegans*, are the only tRNAs that correspond to the codon GGG (shown as red boxes in Figure 2. 1). Although all nematode genomes encode a class I tRNA^{Ile} (UAU) containing a single intron, most tRNAs corresponding to the UAU anticodon are nev-tRNA^{Ile}s (UAU) (shown as yellow boxes in Figure 2. 1). Furthermore, three nev-tRNAs, Pro (GGG), Arg (CCU), and Arg (UCU), in *C. brenneri* constitute the entire population of tRNAs that possess these cognate anticodons. Therefore, the abundance and evolutionary conservation of nev-tRNAs suggest that they play important roles in cellular processes.

Table 2. 4. List of predicted nev-tRNAs in nematode genomes.

								Nearest Gene [b]	
ID	Chromosome	Strand	Start Position	End Position	Isotype	COVE [a]	Localization	Name	Annotation
CBEUV01	chrUn	-	49595801	49595718	Arg (CCU)	75.7	intergenic	sra-11	Serpentine Receptor, class A (alpha) family member
CBEUV02	chrUn	+	90481247	90481331	Arg (CCU)	51.8	intergenic	ZK418.11	hypothetical protein
CBEUV03	chrUn	+	135420606	135420689	Arg (CCU)	78.2	intergenic	pqn-71	Prion-like-(Q/N-rich)-domain-bearing protein family member
CBEUV04	chrUn	+	139901106	139901189	Arg (CCU)	68.5	intergenic	W02B8.1	hypothetical protein
CBEUV05	chrUn	-	173476580	173476497	Arg (CCU)	75.7	intergenic	sra-11	Serpentine Receptor, class A (alpha) family member
CBEUV06	chrUn	+	98280538	98280621	Arg (UCU)	56.5	intergenic	K08H10.3	hypothetical protein
CBEUV07	chrUn	-	157628871	157628788	Gly (GCC)	46.7	intron	K08H10.3	hypothetical protein
CBEUV08	chrUn	-	44899006	44898923	Gly (CCC)	66.0	intergenic	clec-234	C-type LECtin family member
CBEUV09	chrUn	-	44956004	44955921	Gly (CCC)	63.5	intron	Y69H2.14	hypothetical protein
CBEUV10	chrUn	+	52302398	52302481	Gly (CCC)	72.0	intergenic	F58E2.2	hypothetical protein
CBEUV11	chrUn	+	65756121	65756204	Gly (CCC)	69.6	intergenic	C31C9.7	hypothetical protein
CBEUV12	chrUn	+	65779668	65779751	Gly (CCC)	73.5	intergenic	Y48E1A.1	hypothetical protein
CBEUV13	chrUn	-	65784772	65784689	Gly (CCC)	69.6	intergenic	kqt-3	potassium channel, KvQLT family member
CBEUV14	chrUn	-	77278666	77278583	Gly (CCC)	63.5	intron-as	phg-1	PHarynx-associated GAS (growth arrest protein) related family member
CBEUV15	chrUn	-	77347572	77347489	Gly (CCC)	72.0	intron	cdc-25.4	Cell Division Cycle related family member
CBEUV16	chrUn	-	96214039	96213956	Gly (CCC)	72.0	intron	F54B3.1	hypothetical protein
CBEUV17	chrUn	-	96217761	96217678	Gly (CCC)	65.7	intron	F54B3.1	hypothetical protein
CBEUV18	chrUn	-	96223989	96223906	Gly (CCC)	70.5	intron	F54B3.1	hypothetical protein
CBEUV19	chrUn	+	96229404	96229487	Gly (CCC)	72.0	intron-as	F54B3.1	hypothetical protein
CBEUV20	chrUn	-	108459014	108458931	Gly (CCC)	72.0	intron-as	F54B3.1	hypothetical protein
CBEUV21	chrUn	+	108464259	108464342	Gly (CCC)	72.0	intron-as	C50F4.10	hypothetical protein
CBEUV22	chrUn	-	108466679	108466596	Gly (CCC)	72.0	intron-as	F54B3.1	hypothetical protein
CBEUV23	chrUn	+	108469757	108469840	Gly (CCC)	72.0	intron	F54B3.1	hypothetical protein
CBEUV24	chrUn	+	108472283	108472366	Gly (CCC)	70.0	intron	F54B3.1	hypothetical protein
CBEUV25	chrUn	+	123534372	123534455	Gly (CCC)	63.5	intergenic	clec-125	C-type LECtin family member
CBEUV26	chrUn	-	128630580	128630497	Gly (CCC)	59.0	intron-as	phg-1	PHarynx-associated GAS (growth arrest protein) related family member
CBEUV27	chrUn	+	151162459	151162542	Gly (CCC)	67.7	intergenic	srd-2	Serpentine Receptor, class D (delta) family member
CBEUV28	chrUn	+	174812053	174812136	Gly (CCC)	62.7	intergenic	C31C9.7	hypothetical protein
CBEUV29	chrUn	-	204709220	204709137	Gly (CCC)	63.5	intergenic	fipr-13	FIP (Fungus-Induced Protein) Related family member
CBEUV30	chrUn	+	81985208	81985292	Ile (UAU)	70.7	intron-as	ugt-27	UDP-GlucuronosylTransferase family member
CBEUV31	chrUn	+	81988507	81988591	Ile (UAU)	72.5	intron-as	ugt-28	UDP-GlucuronosylTransferase family member
CBEUV32	chrUn	+	82018557	82018641	Ile (UAU)	72.3	intergenic	W09C3.7	hypothetical protein
CBEUV33	chrUn	+	82019139	82019223	Ile (UAU)	71.2	intergenic	ugt-32	UDP-GlucuronosylTransferase family member
CBEUV34	chrUn	+	82092949	82093033	Ile (UAU)	73.6	intergenic	Y23H5B.4	hypothetical protein
CBEUV35	chrUn	+	82093543	82093627	Ile (UAU)	66.1	intergenic	Y23H5B.4	hypothetical protein
CBEUV36	chrUn	-	113365509	113365425	Ile (UAU)	71.2	intergenic	noah-1	NOmpA Homolog (Drosophila nompA: no mechanoreceptor potential A) family member
CBEUV37	chrUn	-	113367683	113367599	Ile (UAU)	71.2	intergenic	pqp-2	P-GlycoProtein related family member
CBEUV38	chrUn	-	113368278	113368194	Ile (UAU)	73.6	intergenic	pgp-2	P-GlycoProtein related family member
CBEUV39	chrUn	+	117389737	117389821	Ile (UAU)	73.6	intergenic	noah-1	NOmpA Homolog (Drosophila nompA: no mechanoreceptor potential A) family member
CBEUV40	chrUn	+	117390341	117390425	Ile (UAU)	71.3	intergenic	noah-1	NOmpA Homolog (Drosophila nompA: no mechanoreceptor potential A) family member
CBEUV41	chrUn	+	117392494	117392578	Ile (UAU)	71.2	intergenic	noah-1	NOmpA Homolog (Drosophila nompA: no mechanoreceptor potential A) family member
CBEUV42	chrUn	+	128873151	128873235	Ile (UAU)	74.6	intergenic	arx-5	ARp2/3 compleX component family member
CBEUV43	chrUn	+	139484203	139484287	Ile (UAU)	72.2	intron	gnrr-1	human GoNadotropin-Releasing hormone Receptor (GnRHR) related family member
CBEUV44	chrUn	-	162257307	162257223	Ile (UAU)	73.0	intron-as	ugt-29	UDP-GlucuronosylTransferase family member
CBEUV45	chrUn	-	162260979	162260895	Ile (UAU)	71.2	intron-as	ugt-28	UDP-GlucuronosylTransferase family member
CBEUV46	chrUn	+	179278696	179278780	Ile (UAU)	73.6	intergenic	Y23H5B.4	hypothetical protein
CBEUV47	chrUn	+	179279581	179279665	Ile (UAU)	71.3	intergenic	ent-3	Equilibrative Nucleoside Transporter family member
CBEUV48	chrUn	+	191197626	191197710	Ile (UAU)	71.2	intergenic	E01G4.6	hypothetical protein

Table 2. 4. (*Continued*)

ID	Chromosome	Strand	Start Position	End Position	Isotype	COVE [a]	Localization	Nearest Gene [b] Name	Nearest Gene [b] Annotation
CBEUV49	chrUn	-	115464800	115464715	Pro (GGG)	37.1	intergenic	F02E9.5	hypothetical protein
CBEUV50	chrUn	+	130684124	130684209	Pro (GGG)	40.6	intergenic	T10B11.2	hypothetical protein
CBEUV51	chrUn	+	163234209	163234292	Pro (GGG)	49.0	intergenic	fbp-1	Fructose-1,6-BiPhosphatase family member
CBEUV52	chrUn	+	182030997	182031080	Pro (GGG)	49.0	intergenic	fbp-1	Fructose-1,6-BiPhosphatase family member
CBEUV53	chrUn	+	129301206	129301290	Val (CAC)	52.5	intergenic	C03G6.6	hypothetical protein
CBEUV54	chrUn	-	129309145	129309061	Val (CAC)	45.5	intergenic	F44D12.8	hypothetical protein
CBEUV55	chrUn	+	148827680	148827764	Val (CAC)	50.4	intergenic	F44D12.8	hypothetical protein
CBEUV56	chrUn	-	148834876	148834792	Val (CAC)	52.5	intergenic	C03G6.6	hypothetical protein
CBIUV01	chrIII	+	6307513	6307596	Ala (CGC)	57.5	intergenic	ZK418.11	hypothetical protein
CBIUV02	chrII	+	3962928	3963011	Gly (CCC)	67.3	intron	bcs-1	BCS1 (mitochondrial chaperone) homolog family member
CBIUV03	chrII	-	3971981	3971898	Gly (CCC)	73.2	intron-as	F54C9.9	hypothetical protein
CBIUV04	chrII	+	3972134	3972217	Gly (CCC)	67.7	intron	F54C9.9	hypothetical protein
CBIUV05	chrII	-	3975781	3975698	Gly (CCC)	79.2	intron-as	arl-1	ARF-Like family member
CBIUV06	chrI_random	-	3498371	3498288	Ile (UAU)	80.6	intron-as	H37N21.1	hypothetical protein
CBIUV07	chrI_random	-	3498899	3498816	Ile (UAU)	76.1	intron-as	H37N21.1	hypothetical protein
CBIUV08	chrI	-	6501949	6501866	Ile (UAU)	73.9	intergenic	C36F7.2	hypothetical protein
CBIUV09	chrI	+	6502462	6502545	Ile (UAU)	73.9	intergenic	C36F7.2	hypothetical protein
CBIUV10	chrI	+	6506662	6506745	Ile (UAU)	73.9	intergenic	T22C1.9	hypothetical protein
CBIUV11	chrI	+	6509561	6509644	Ile (UAU)	68.0	intergenic	T22C1.9	hypothetical protein
CBIUV12	chrI	-	6510161	6510078	Ile (UAU)	73.9	intergenic	T22C1.9	hypothetical protein
CBIUV13	chrI	+	6511932	6512015	Ile (UAU)	51.2	intergenic	T22C1.9	hypothetical protein
CBIUV14	chrI	-	6520859	6520776	Ile (UAU)	73.9	intergenic	C28A5.2	hypothetical protein
CBIUV15	chrI	-	6642252	6642169	Ile (UAU)	73.9	intergenic	C54G4.4	hypothetical protein
CBIUV16	chrI	+	6642801	6642884	Ile (UAU)	73.9	intergenic	C54G4.4	hypothetical protein
CBIUV17	chrI	+	6643457	6643540	Ile (UAU)	69.8	intergenic	C54G4.4	hypothetical protein
CBIUV18	chrI	-	6665886	6665803	Ile (UAU)	73.9	intergenic	F58D5.7	hypothetical protein
CBIUV19	chrI	+	6744901	6744984	Ile (UAU)	74.9	intron-as	F55D12.2	hypothetical protein
CBIUV20	chrI	+	6746285	6746368	Ile (UAU)	73.9	intron-as	F55D12.2	hypothetical protein
CBIUV21	chrI	+	6748639	6748722	Ile (UAU)	57.9	intron-as	F55D12.2	hypothetical protein
CBIUV22	chrI	+	6785191	6785278	Ile (UAU)	66.7	intergenic	drh-3	Dicer Related Helicase family member
CBIUV23	chrIII	-	4211509	4211426	Pro (GGG)	32.8	intergenic	C54C6.6	hypothetical protein
CELUV01	chrI	-	10601202	10601120	Gly (CCC)	72.0	intron-as	pash-1	PArtner of DroSHa (DRSH-3 interactor) family member
CELUV02	chrI	+	10601838	10601920	Gly (CCC)	72.0	intron	pash-1	PArtner of DroSHa (DRSH-1 interactor) family member
CELUV03	chrI	-	10604203	10604121	Gly (CCC)	72.0	intron-as	pash-1	PArtner of DroSHa (DRSH-2 interactor) family member
CELUV04	chrI	+	7960830	7960914	Ile (UAU)	79.7	intron-as	T22C1.12	hypothetical protein
CELUV05	chrI	+	9562343	9562427	Ile (UAU)	79.7	intergenic	rig-5	neuRonal IGCAM family member
CELUV06	chrI	-	9563917	9563833	Ile (UAU)	79.7	intergenic	rig-5	neuRonal IGCAM family member
CELUV07	chrX	-	9829683	9829599	Ile (UAU)	79.7	intron	rrc-1	RhoGAP for Rac-1 and Cdc-42 family member
CELUV08	chrIII	-	6724158	6724073	Lys (CUU)	71.0	intron-as	F37A4.1	hypothetical protein
CJAUV01	chrUn	-	117342044	117341960	Ile (UAU)	75.7	intergenic	rskn-2	RSK-pNinety (RSK-p90 kinase) homolog family member
CJAUV02	chrUn	+	138538062	138538146	Ile (UAU)	75.7	intergenic	K04G2.9	hypothetical protein
CJAUV03	chrUn	-	138543058	138542974	Ile (UAU)	75.7	intergenic	K04G2.9	hypothetical protein
CJAUV04	chrUn	-	42974041	42973958	Thr (UGU)	51.3	intron	ugt-9	UDP-GlucuronosylTransferase family member
CJAUV05	chrUn	+	87595760	87595843	Trp (CCA)	55.8	intergenic	grl-26	GRound-Like (grd related) family member
CJAUV06	chrUn	+	126225166	126225249	Val (UAC)	53.1	intron-as	cyk-1	CYtoKinesis defect family member
CREUV01	chrUn	-	23853810	23853727	Arg (CCU)	47.8	intergenic	grl-29	GRound-Like (grd related) family member
CREUV02	chrUn	-	19286759	19286678	Gly (CCC)	75.2	intron-as	C46H11.6	hypothetical protein
CREUV03	chrUn	-	19323134	19323052	Gly (CCC)	74.1	intron	C05D11.8	hypothetical protein
CREUV04	chrUn	-	19325044	19324962	Gly (CCC)	74.1	intron	C05D11.8	hypothetical protein
CREUV05	chrUn	-	19671121	19671039	Gly (CCC)	74.1	intron-as	polq-1	POLQ (DNA polymerase theta) homolog family member
CREUV06	chrUn	+	21965339	21965423	Ile (UAU)	75.1	intron	smg-1	Suppressor with Morphological effect on Genitalia family member
CREUV07	chrUn	-	25627015	25626931	Ile (UAU)	73.2	intergenic	msh-5	MSH (MutS Homolog) family member
CREUV08	chrUn	-	25631915	25631831	Ile (UAU)	73.2	intron	msh-5	MSH (MutS Homolog) family member
CREUV09	chrUn	-	25636216	25636132	Ile (UAU)	73.2	intergenic	msh-5	MSH (MutS Homolog) family member
CREUV10	chrUn	-	25642385	25642301	Ile (UAU)	73.2	intergenic	C55C3.3	hypothetical protein
CREUV11	chrUn	-	25694155	25694071	Ile (UAU)	73.2	intron-as	Y48G1C.4	hypothetical protein
CREUV12	chrUn	+	109191950	109192033	Pro (GGG)	49.5	intergenic	K06B9.3	hypothetical protein

Table 2. 4. (*Continued*)

ID	Chromosome	Strand	Start Position	End Position	Isotype	COVE [a]	Localization	Nearest Gene [b]	
								Name	Annotation
CREUV13	chrUn	-	111749904	111749821	Pro (GGG)	49.5	intergenic	C36A4.4	hypothetical protein
CREUV14	chrUn	-	21921410	21921327	Val (CAC)	48.8	intergenic	F02E9.5	hypothetical protein
PPAUV1	chrUn	-	10269759	10269677	Ile (UAU)	72.51	intron	pxf-1	PDZ exchange facter
PPAUV2	chrUn	-	97319877	97319795	Ile (UAU)	74.48	intron-as	F25D7.2	hypothetical protein
PPAUV3	chrUn	-	133533498	133533416	Ile (UAU)	74.48	intron-as	nol-6	a nucleolar RNA-associated protein (NRAP)
PPAUV4	chrUn	+	160878669	160878748	Ile (UAU)	58.33	intron	F36H1.2a	hypothetical protein
PPAUV5	chrUn	+	160889950	160890032	Ile (UAU)	71.75	intron-as	cdc-42	Cell Division Cycle related protein
PPAUV6	chrUn	+	160892350	160892432	Ile (UAU)	66.04	intron-as	cdc-42	Cell Division Cycle related protein
PPAUV7	chrUn	-	160902693	160902611	Ile (UAU)	74.01	intron-as	unc-33	Collapsin response mediator protein-2
PPAUV8	chrUn	-	171980594	171980512	Ile (UAU)	70.73	intron-as	C32D5.3	hypothetical protein

[a] Covariance model (COVE) socres were calculated by tRNAscan-SE.
[b] The information of nearest genes from the candidates was searched using UCSC genome browser.

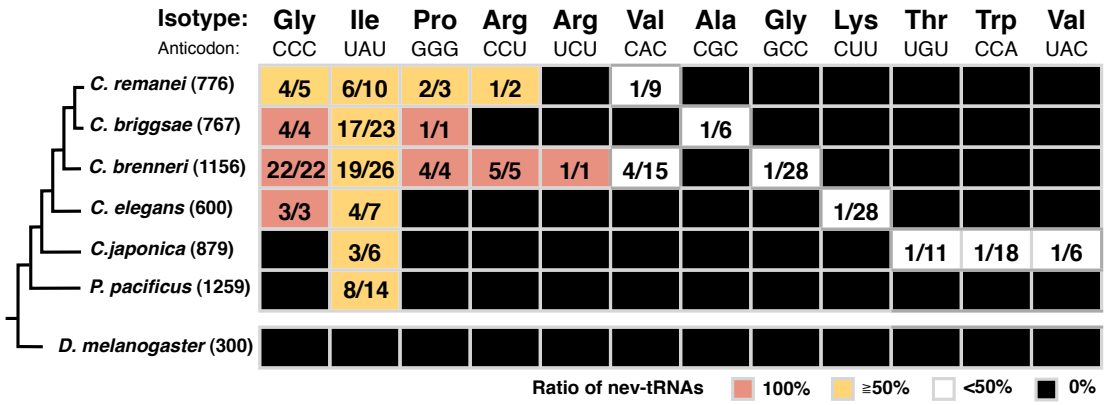

Figure 2. 1. Evolutionary conservation of the nev-tRNAs in Rhabditina.

For each nematode species, the ratios of nev-tRNA genes to tRNA genes corresponding to each specific anticodon are shown (red, 100%; yellow, > 50% but < 100%; white, < 50%). The total number of tRNA genes in each species is indicated in brackets.

2.3.2. Sequence characteristics of nev-tRNAs and their evolutionary background

We selected three nematode species to examine the origins and evolution of the nev-tRNA genes: *C. brenneri* (which has more nev-tRNAs than any other species), *C. elegans* (used for experimental verification, as described below), and *P. pacificus* (representing the earliest branch within the nematode lineage). Using 151 tRNA exons, including the exons of 13 nev-tRNAs from these three nematode species, we performed a phylogenetic analysis of the nonredundant tRNA sequences (Figure 2. 2). We found that four types of nev-tRNAs, Ile (UAU), Gly (CCC), Arg (CCU), and Lys (CUU), clustered into a class II tRNA clade, especially close to the $tRNA^{Leu}$ (UAA) family. Moreover, the class I tRNAs with these same anticodons clustered separately from these nev-tRNAs on the phylogenetic tree of nematode tRNAs. Because nev-$tRNA^{Ile}$ (UAU) and nev-$tRNA^{Gly}$ (CCC) are widely conserved in the nematode linage (Figure 2. 1), these results suggest that at least the major nev-tRNAs probably evolved from class II tRNAs, rather than by gene duplication of synonymous tRNAs. In contrast, six types of nev-tRNAs in *C. brenneri* clustered within the class I tRNA family, suggesting that nev-tRNAs emerged on two independent occasions during tRNA evolution. We also found that a single *C. brennneri* $tRNA^{Leu}$ (UAA) occurred within the nev-tRNA clade. Its sequence was almost identical (91.6%) to that of *C. brennneri* nev-$tRNA^{Arg}$ (CCU), but was only 55.9% similar to the synonymous $tRNA^{Leu}$ (UAA), indicating that it originated from the nev-tRNA gene (shown as an asterisk in Figure 2. 2).

To analyze the characteristics of the nev-tRNA sequences relative to their nucleotides at each position, we compared *C. elegans* nev-$tRNA^{Gly}$ (CCC) and

nev-tRNA^{Ile} (UAU) with their synonymous class I tRNA^{Gly} (GCC), tRNA^{Ile} (UAU), and tRNA^{Leu} (UAA) sequences (Figure 2. 3). Each of these nev-tRNAs forms a cloverleaf secondary structure, has endogenous A-box (5′-TGGCNNAGTGG-3′) and B-box (5′-GGTCGANNCC-3′) promoters (Galli *et al*. 1981), and contains a V-arm of 15–16 nucleotides that forms a stem–loop structure. We found that nev-tRNA^{Gly} (CCC) showed greater sequence similarity to tRNA^{Leu} (UAA) (64.3%) than to class I tRNA^{Gly} (GCC) (56.3%). The results for the nev-tRNA^{Ile} (UAU) sequence are similar. The nucleotides conserved among these three class II tRNAs were concentrated in the D-arm and V-arm regions, which are both recognition elements for LeuRS (Breitschopf *et al*. 1995; Soma *et al*. 1999) (shown as boxed nucleotides in Figure 2. 3). These results indicate that the nev-tRNAs may be recognized by LeuRS and charged with leucine.

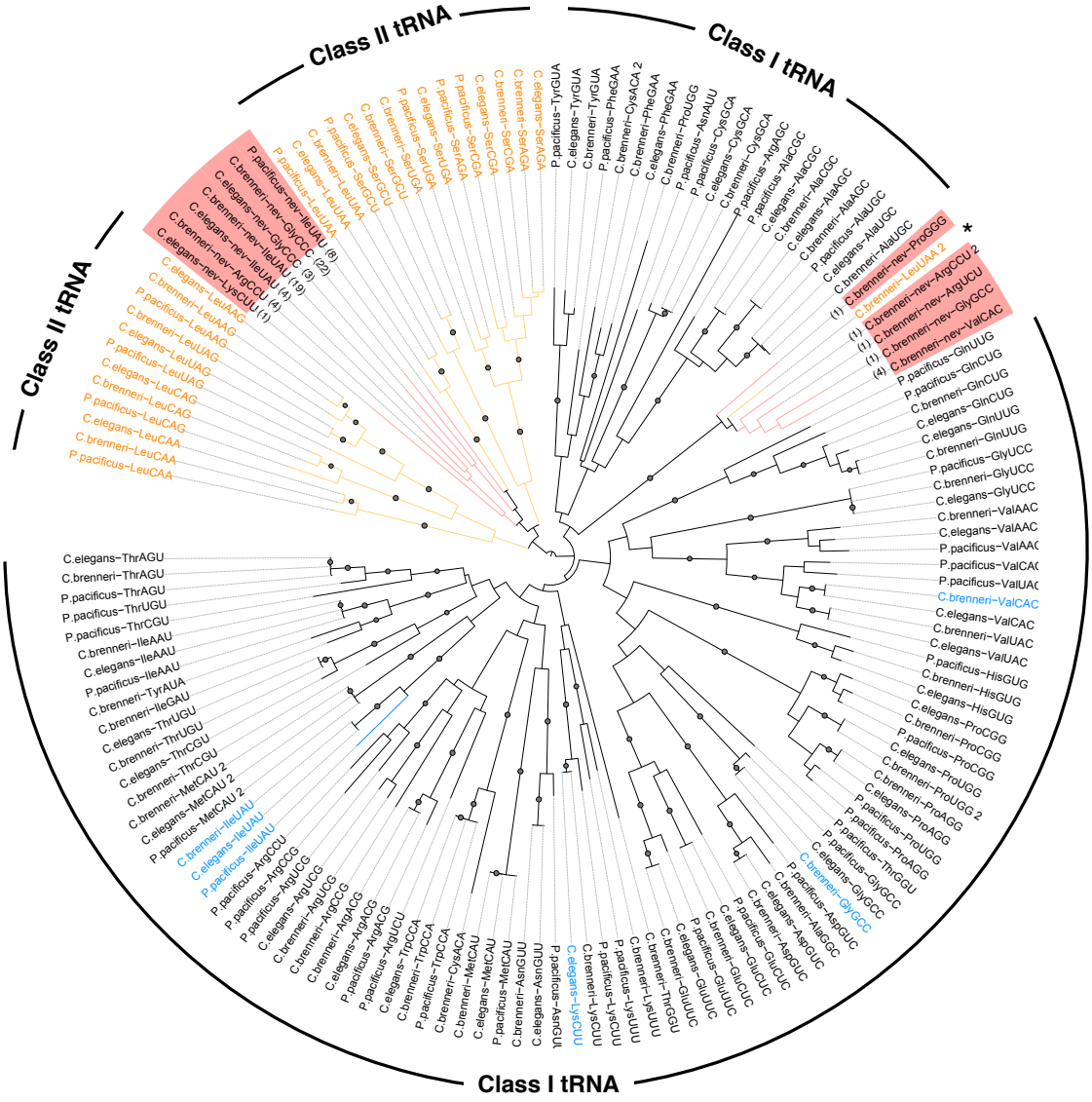

Figure 2. 2. Phylogeny of all tRNAs in the three nematode species.

A neibor-joining tree of all mature tRNA sequences in *C. elegans*, *C. brenneri*, and *P. pacificus* was constructed using ClustalX and visualized with iTOL. The phylogeny includes nev-tRNAs (red), $tRNA^{Leu}$ and $tRNA^{Ser}$ (yellow), tRNAs with anticodons synonymous to those of nev-tRNAs (blue) and other tRNAs (black). Copy numbers are indicated for each nev-tRNA. Black dots indicate bootstrap values above 80. *C. brenneri* encodes two types of $tRNA^{Leu}$ (UAA): the one with strong similarity to nev-$tRNA^{Arg}$ (CCU) is categorized as nev-tRNA (asterisk).

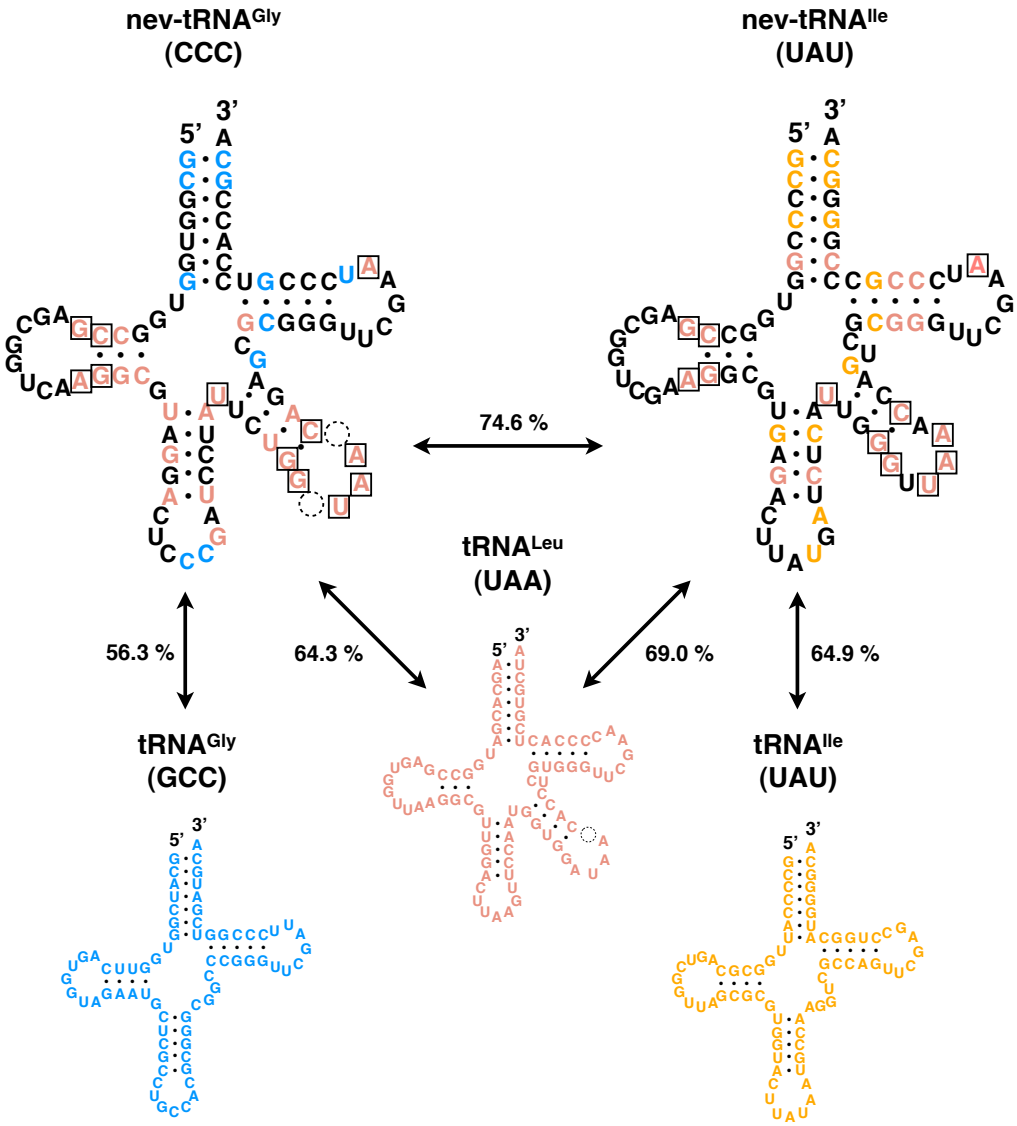

Figure 2. 3. Comparison of nucleotide sequences and secondary structures of two nev-tRNAs and their related tRNAs in *C. elegans*.

Nucleotides conserved between nev-tRNA^{Gly} (CCC) and the synonymous class I tRNA^{Gly} (GCC) are shown in blue; those between nev-tRNA^{Ile} (UAU) and the synonymous class I tRNA^{Ile} (UAU) are shown in yellow; and those between the two nev-tRNAs and the class II tRNA^{Leu} (UAA) are shown in red. Nucleotides conserved among the three V-arm-containing tRNAs are boxed. Sequence similarities are also shown for each pair of tRNAs.

2.3.3. nev-tRNAs are weakly expressed in *C. elegans*

Although these bioinformatics approach indicated that the nev-tRNAs probably originated from $tRNA^{Leu}$ and are expanding among the nematode genomes, it was unclear whether they are expressed in cells. Therefore, we assayed the expression of the nev-tRNA genes by northern blot analysis of the total RNA extracted from mixed stages of *C. elegans*, including eggs, L1–4 larvae, and adults, using specific oligonucleotide probes (Table 2. 1). We detected the expression of nev-$tRNA^{Gly}$ (CCC) and nev-$tRNA^{Ile}$ (UAU) at their predicted lengths, but no expression of nev-$tRNA^{Lys}$ (CUU) (Figure 2. 4). RT–PCR analysis confirmed the expression of all three nev-tRNAs at their expected lengths (Figure 2. 5), with the nucleotide sequences of the amplified products identical to those of the nev-tRNAs. Moreover, none of these nev-tRNAs contained a V-arm spliced tRNA band of approximately 70 bp, indicating that the V-arm domain is part of the mature tRNA, not an intronic sequence that is spliced out during tRNA processing. These results demonstrate that, although expressed, the levels of nev-tRNAs are lower than those of general tRNAs.

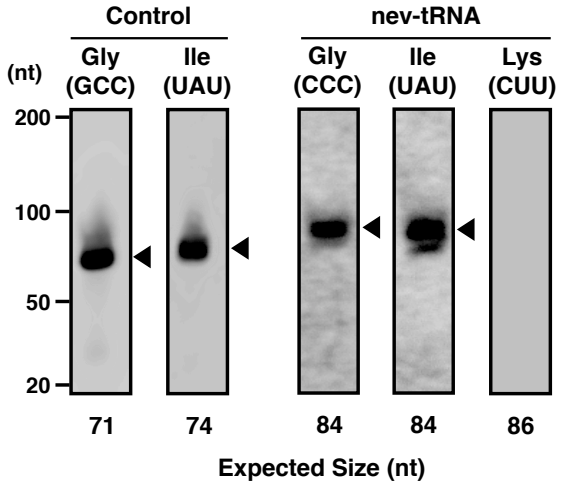

Figure 2. 4. Expression of nev-tRNAs in *C. elegans*.

The results of the northern blot analysis of the three nev-tRNA genes found in *C. elegans*, nev-tRNA^{Gly} (CCC), nev-tRNA^{Ile} (UAU), and nev-tRNA^{Lys} (CUU), and their synonymous class I tRNA genes, tRNA^{Gly} (GCC) and tRNA^{Ile} (UAU) (control), are shown. Black triangles indicate that the band sizes were approximately consistent with the predicted lengths of the nev-tRNAs based on the bioinformatics approach used in this study. nev-tRNA^{Lys} was not detected, even when the amount of total RNA was increased to 30 μg per lane. Contrast levels were adjusted for the nev-tRNA images.

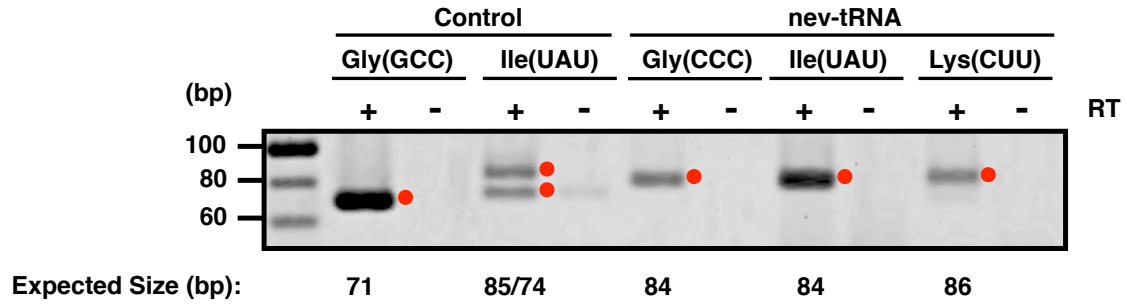

Figure 2. 5. RT–PCR analysis of the predicted nev-tRNAs of *C. elegans*.

Levels of expression of the three nev-tRNA genes and their synonymous tRNA genes (controls) were determined by RT–PCR. Samples without reverse transcriptase were used as the negative controls (RT–) to ensure that only RNA transcripts were amplified (RT+). The expected sizes are shown as red dots. Both the precursor and mature tRNAs were detected for the intron-containing $tRNA^{Ile}$ (UAU).

2.3.4. nev-tRNA^{Gly} and nev-tRNA^{Ile} are specifically aminoacylated with leucine

The V-arm domain and the other determinant sites for LeuRS that are conserved in the nev-tRNAs suggest that they are leucylated, whereas their anticodons correspond to other amino acids. To understand the functions of the cellular nev-tRNAs, we determined the amino acids with which they are charged. We attempted to aminoacylate *C. elegans* nev-tRNA^{Gly} (CCC) and nev-tRNA^{Ile} (UAU) *in vitro* with four purified recombinant *C. elegans* aaRSs: GlyRS, IleRS, SerRS, and LeuRS (Figure 2. 6). Each control tRNA was successfully aminoacylated with the aaRS and the amino acid corresponding to its anticodon (Figure 2. 7A, left). Under the same conditions, nev-tRNA^{Gly} (CCC) and nev-tRNA^{Ile} (UAU) were not aminoacylated with the amino acids corresponding to their respective anticodons, nor with serine, but with leucine in the presence of LeuRS (Figure 2. 7A, right). When we assessed the aminoacylation efficiency and specificity of each nev-tRNA at different time points, we found that both nev-tRNAs had a lower leucylation efficiency than tRNA^{Leu} (AAG) and that, conversely, almost no tRNA was aminoacylated by GlyRS or IleRS (Figure 2. 7B).

Because many examples have so far been reported of tRNA modifications that affect the recognition by aaRSs (Stern and Schulman 1978; Szweykowska-Kulinska *et al*. 1994), we conducted further aminoacylation assays using total RNA fractions, followed by the identification of the mature nev-tRNAs by northern blot analysis, to validate the leucylation of native nev-tRNAs *in vivo*. Because the expression levels of native nev-tRNAs are low, we designed an oligonucleotide complementary to full-length nev-tRNA^{Gly} as the northern hybridization probe (Table 2. 1) to improve the

signal strength. We used the most stringent hybridization conditions to detect nev-tRNA^{Gly} with the full-length probe and found that almost all the cross-hybridization signals could be reduced by increasing the hybridization and wash temperatures above 55°C (Figure 2. 8A). Accordingly, we performed the hybridization reactions for *in vivo* nev-tRNAs at 60°C and successfully demonstrated that native nev-tRNA^{Gly} is aminoacylated with leucine but not glycine (Figure 2. 8B). These results suggest that nev-tRNAs are only ever charged with leucine both *in vitro* and *in vivo*.

To determine the nucleotides in the nev-tRNAs recognized by LeuRS, we tested the ability of LeuRS to aminoacylate nev-tRNA^{Gly} (CCC) variants with mutations at several recognition sites. The V-arm domains are major determinants of the recognition of class II tRNAs (Breitschopf *et al*. 1995; Soma *et al*. 1999; Biou *et al*. 1994; Yaremchuk *et al*. 2002), and X-ray structural analysis of the *Pyrococcus horikoshii* LeuRS-tRNA^{Leu} complex showed that the protruding C-terminal domain of LeuRS specifically recognizes the bases at the tip of the V-arm (Fukunaga and Yokoyama 2005). *C. elegans* encodes five synonymous tRNA^{Leu} genes, and the V-arm sequences of its nev-tRNAs are most similar to that of tRNA^{Leu} (UAA), with the bases at the tip of the nev-tRNA V-arms (U47b and A47c) conserved in both tRNA^{Leu}s (UAA/CAA) (Figure 2. 9A). Complete deletion of the nev-tRNA^{Gly} (CCC) V-arm domain led to a severe loss of leucylation (Figure 2. 9B, M1), and replacement of the two nucleotides at the tip of the V-arm reduced the leucylation efficiency approximately two-fold (Figure 2. 9B, M2). This suggests that the V-arm domains of the nev-tRNAs, primarily the two nucleotides at their tips, are necessary for leucylation by *C. elegans*

LeuRS, as are the determinants in $tRNA^{Leu}$. In contrast, the discriminator base A73 of $tRNA^{Leu}$ is highly conserved throughout all living organisms and is important for leucylation in *E. coli* (Asahara *et al*. 1993), yeast (Soma *et al*. 1996), and humans (Breitschopf *et al*. 1995). A73 is also conserved in all $tRNA^{Leu}$ genes in the six nematode species examined, as well as in all 115 nev-tRNAs, whereas the class II $tRNA^{Ser}$ genes contain a conserved G73 (Figure 2. 9A). We found that the leucylation efficiency of nev-$tRNA^{Gly}$ (CCC) was strongly influenced by base substitutions at A73 (Figure 2. 9B, M3). The strong evolutionary conservation of A73 and the V-arm suggest that nev-tRNAs have been under evolutionary pressure to be specifically recognized by LeuRS.

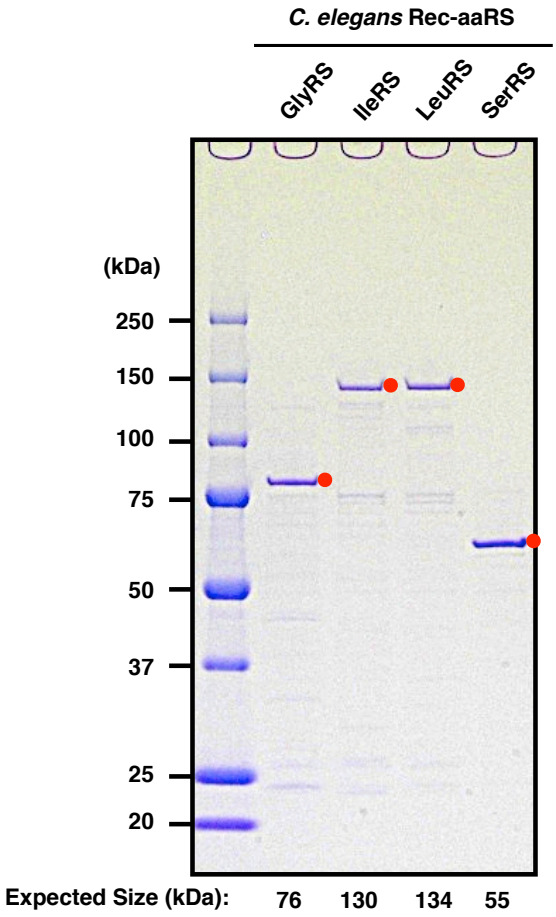

Figure 2. 6. Purified recombinant aminoacyl-tRNA synthetases from *C. elegans*.
His-tagged recombinant aminoacyl-tRNA synthetase (aaRS) enzymes were purified to near homogeneity using Ni^{2+} agarose resin spin columns followed by gel filtration (see section 2.2.5). The samples were analyzed by SDS/PAGE on a 10%–20% gradient gel and stained with Coomassie Brilliant Blue. Red dots indicate the positions of the purified recombinant aaRS proteins.

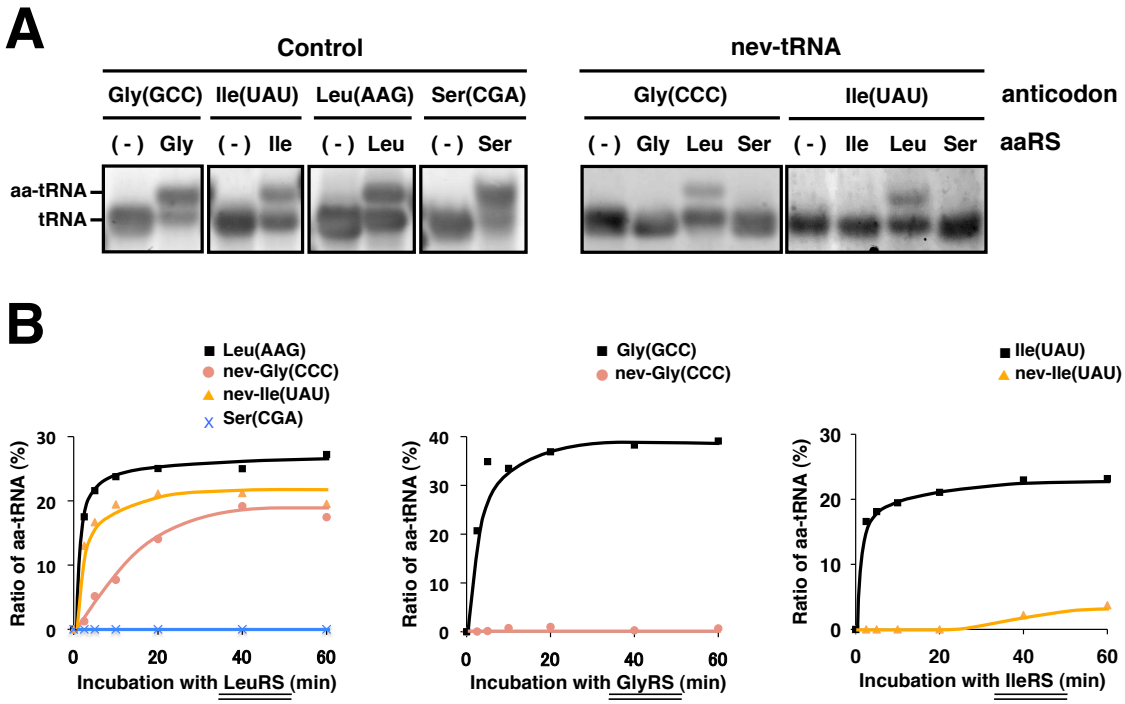

Figure 2. 7. Aminoacylation assay of nev-tRNAs.

(A) *In vitro*-transcribed tRNAs (four types of control tRNAs, nev-tRNA^{Gly} (CCC), and nev-tRNA^{Ile} (UAU)) were aminoacylated using four recombinant *C. elegans* aaRSs: GlyRS, IleRS, SerRS, and LeuRS (Figure 2. 6). The aminoacylated tRNAs (aa-tRNAs) were separated on an acid–urea polyacrylamide gel and stained with SYBR Green II. (B) Measurement of the aminoacylation efficiencies of the nev-tRNAs at 0, 2.5, 5, 10, 20, 40, and 60 min. (Left panel) Leucylation of tRNA^{Leu} (AAG), tRNA^{Ser} (CGA), and two nev-tRNAs by LeuRS. (Middle panel) Glycylation of tRNA^{Gly} (GCC) and nev-tRNA^{Gly} (CCC) by GlyRS. (Right panel) Isoleucylation of tRNA^{Ile} (UAU) and nev-tRNA^{Ile} (UAU) by IleRS.

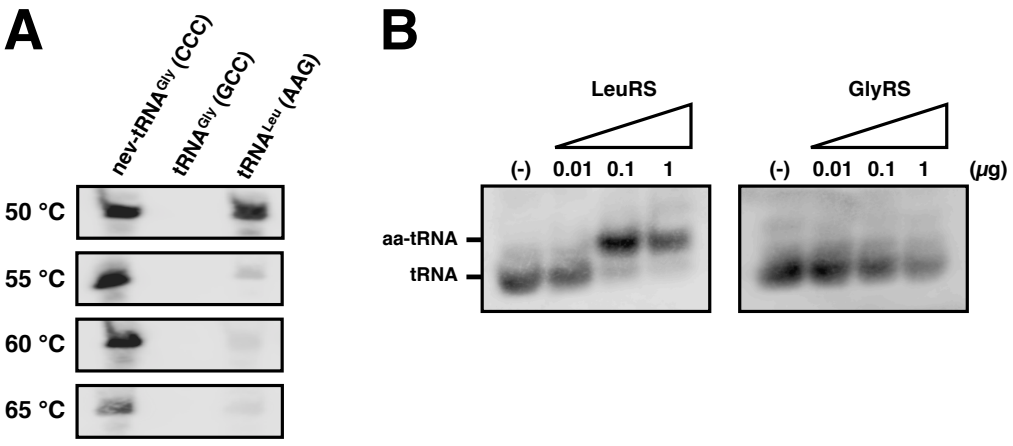

Figure 2. 8. *In vivo*-modified nev-tRNA^{Gly} (CCC) is specifically charged with leucine.

(A) Validation of the hybridization stringency conditions for detecting nev-tRNA^{Gly} (CCC) using the full-length probe. *In vitro*-transcribed nev-tRNA^{Gly} (CCC), tRNA^{Gly} (GCC), and tRNA^{Leu} (AAG) were detected using an oligonucleotide fully complementary to nev-tRNA^{Gly}. Northern hybridization and washing treatments were performed at 50, 55, 60, and 65°C. (B) The aminoacylation of native nev-tRNA^{Gly} (CCC) determined by acid–urea PAGE/northern hybridization. An aminoacylation assay was performed at room temperature for 60 min with purified recombinant LeuRS or GlyRS titrated from 0.01 μg to 1 μg.

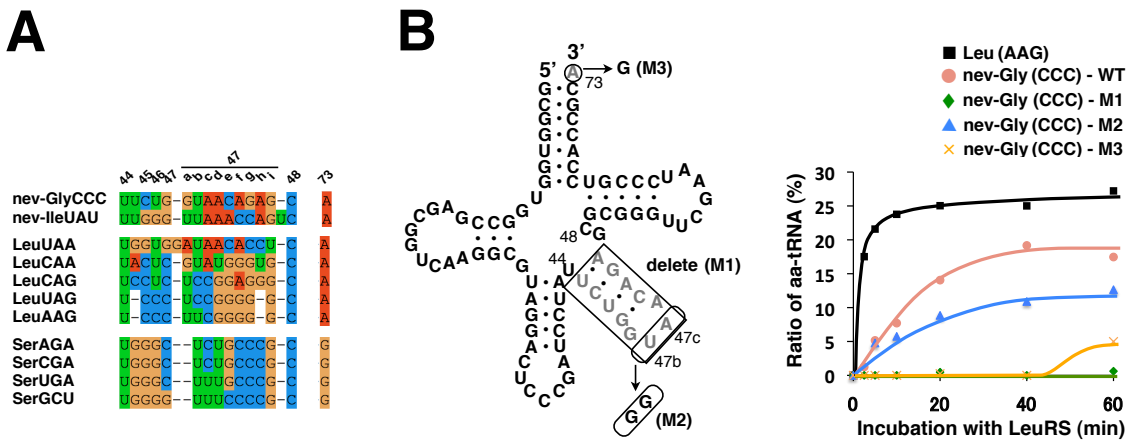

Figure 2. 9. Mutagenesis of nev-tRNA^{Gly} (CCC) sites recognized by LeuRS.

(A) Sequence alignments of the V-arm domains of class II tRNAs, including nev-tRNAs, in *C. elegans*. The nucleotide positions are numbered as described previously (Sprinzl *et al*. 1998). (B) (Left panel) The nev-tRNA^{Gly} (CCC) variants used in this study. (Right panel) Leucylation efficiencies of tRNA^{Leu} (AAG), wild-type nev-tRNA^{Gly} (WT), and the three nev-tRNA^{Gly} variants (M1–3). Each experiment was performed in duplicate to confirm the reproducibility of the enzymatic reactions.

2.3.5. nev-tRNA^{Gly} is incorporated into eukaryotic ribosomes during translation and decodes the GGG codon as leucine *in vitro*

Finally, to demonstrate whether nev-tRNAs are incorporated into ribosomes and function in translation, we used these nev-tRNAs in an *in vitro* translation assay and determined the sequences of the resultant peptides with MS. In these experiments, we attempted to synthesize two proteins, PF1549 and firefly luciferase, because *i*) their mRNAs contain 6–8 GGG codons, corresponding to glycine, and *ii*) PF1549, an RNA 3′-terminal-phosphate cyclase in the hyperthermophilic archaeon *Pyrococcus furiosus*, can be purified easily by heat treatment. The *in vitro* translation reactions were performed in an insect cell-free protein expression system using *in vitro*-transcribed tRNAs, either tRNA^{Leu} (AAG) (control) or nev-tRNA^{Gly} (CCC). After partial purification with a magnetic separator, SDS/PAGE analysis showed that the PF1549 protein and luciferase were successfully synthesized at their expected sizes (Figure 2. 10). To examine whether specific leucines, derived from the nev-tRNA decoding of GGG codons, were contained in the proteins synthesized in the presence of nev-tRNA^{Gly}, we analyzed the purified proteins with nanoLC–MS/MS. Surprisingly, proteins containing leucine residues translated from GGG codons were only observed in the presence of nev-tRNA^{Gly}, in both PF1549 and luciferase (Table 2. 5, Figure 2. 11). Although leucine is indistinguishable from isoleucine on MS, these amino acid residues were almost certainly leucines because *in vitro* aminoacylation assays showed that nev-tRNA^{Gly} is not charged with isoleucine. Peptides containing glycine translated from GGG codons were also observed in the presence of nev-tRNA^{Gly}, suggesting that the

GGG codon is also decoded by endogenous insect $tRNA^{Gly}$ (UCC) via G–U wobble base pairing (Varani and McClain 2000). These results strongly suggest that in insect cells, *i*) nev-$tRNA^{Gly}$ is aminoacylated with leucine, *ii*) leucylated nev-$tRNA^{Gly}$ is incorporated into ribosomes during translation, and *iii*) leucylated nev-$tRNA^{Gly}$ recognizes GGG codons, resulting in the incorporation of leucine, which does not correspond to their anticodons.

Table 2. 5. Identification of aberrant residues in *in vitro*-translated proteins.

Protein Name	Type	Sequence	Modification	Mascot Score (n = 4)	
				$tRNA^{Leu}$ (Control)	nev-$tRNA^{Gly}$
PF1549	GGG-to-Gly	GGGIVAGYVKPWIER		45.0	37.4
		GGGIVAGYVKPWIERK		57.6	44.5
		GKPAEEVGREAAQELLSQVK		69.1	63.3
		KGKPAEEVGREAAQELLSQVK		74.9	73.6
		KLANAKVEGAEVGSR		75.1	83.4
		LANAKVEGAEVGSR		49.2	61.8
	GGG-to-Leu	GGLIVAGYVKPWIER		-	46.1
		GGLIVAGYVKPWIERK		-	20.1
		KLKPAEEVGR		-	42.8
		KLKPAEEVGREAAQELLSQVK		-	20.3
		LKPAEEVGR		-	30.5
		LKPAEEVGREAAQELLSQVK		-	42.1
		VEGAEVLSR		-	63.0
luciferase	GGG-to-Gly	EVGEAVAK		25.7	29.1
		RFHLPGIR		39.3	43.5
		STLIDKYDLSNLHEIASGGAPLSK		110.2	113.4
		TIALIMNSSGSTGLPK	Oxidation@M:6	122.0	121.6
		VVDLDTGK		40.6	37.0
	GGG-to-Leu	EVLEAVAK		-	25.9
		RFHLPLIR		-	25.4
		STLIDKYDLSNLHEIASLGAPLSK		-	71.6
		TIALIMNSSGSTLLPK		-	91.6
		VVDLDTLK		-	47.3

PF1549 and luciferase was translated *in vitro* in the presence of $tRNA^{Leu}$ (AAG) or nev-$tRNA^{Gly}$ (CCC). The peptide sequences containing amino acid residues (red) arising from the decoding of the GGG codon are listed. The highest Mascot score was based on four independent experiments.

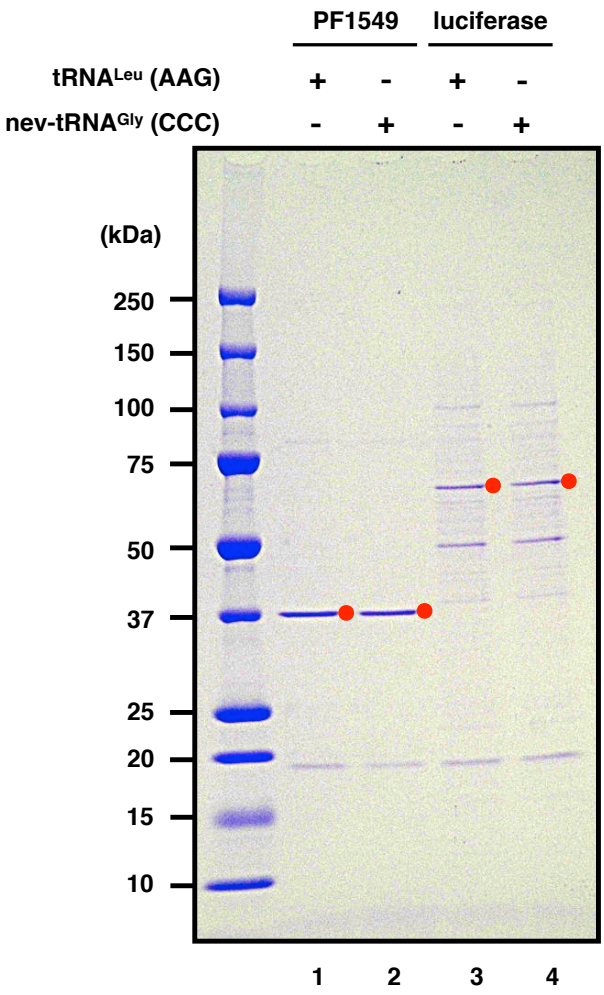

Figure 2. 10. SDS/PAGE analysis of purified *in vitro*-translated proteins.

His-tagged recombinant proteins synthesized in cell-free systems were purified with a magnetic separator (see section 2.2.9). The samples were analyzed by SDS/PAGE on a 10%-20% gradient gel and stained with Coomassie Brilliant Blue. Lanes 1–2, *in vitro*-translated PF1549 protein (37 kDa); lanes 3–4, *in vitro*-translated firefly luciferase (61 kDa). The tRNA types used in the translation reactions were: lanes 1 and 3, $tRNA^{Leu}$ (AAG); lanes 2 and 4, nev-$tRNA^{Gly}$ (CCC). Red dots indicate the positions of the purified recombinant proteins.

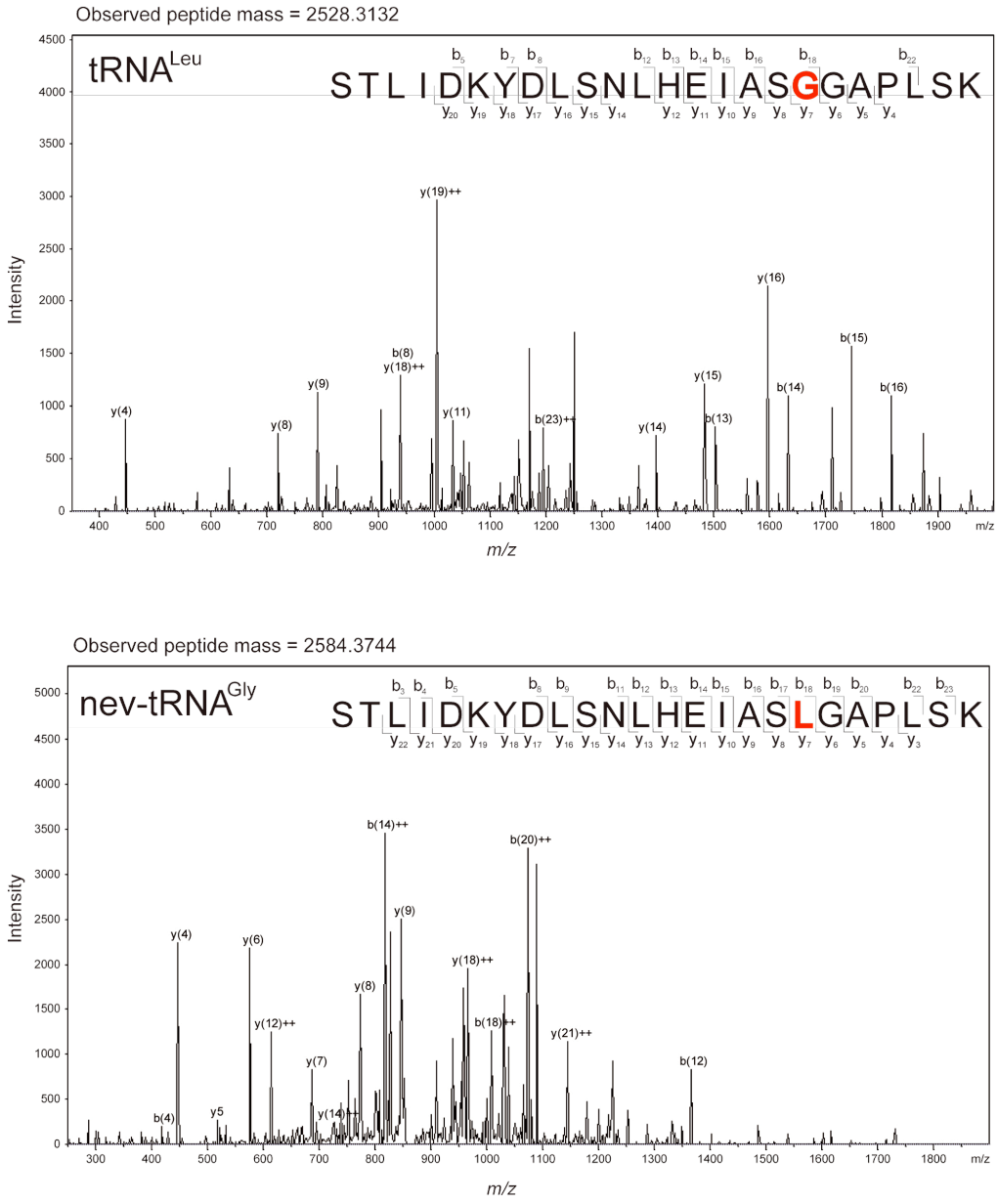

Figure 2. 11. Comparison of the MS/MS spectra of peptides with amino acid residues arising from the decoding of GGG codons.

Peptides with the sequence STLIDKYDLSNLHEIAS(G/L)GAPLSK from luciferase synthesized in cell-free systems containing $tRNA^{Leu}$ (AAG) or nev-$tRNA^{Gly}$ (CCC) were compared. Fragmented ions detected in this analysis are shown, with the top 15 ions annotated in the spectra. Amino acid residues arising from the decoding of the GGG codons are indicated in red.

2.3.6. nev-tRNAs tend to correspond to rare codons

In this work, we have identified a new group of class II tRNA genes, which probably originated and diversified from the $tRNA^{Leu}$ gene. These unique tRNAs are found only in several nematode species, and their numbers and the types of anticodons involved have tended to increase during nematode evolution. Based on the codon usage of *C. elegans,* we found that the major nev-tRNAs correspond to the three rarest codons in the organism, Gly (GGG), Pro (CCC), and Arg (AGG), which comprise 0.44%, 0.44% and 0.38%, respectively, of the codons in *C. elegans* (Figure 2.12). The usage of codon Ile (ATA) is lowest among the three synonymous isoleucine codons (0.94%), but is higher than that of the other rare codons. This may be attributable to the coexistence of both class I $tRNA^{Ile}$ (UAU) and nev-$tRNA^{Ile}$ (UAU) in the nematode genomes (Figure 2. 1). Because isoleucine and leucine are chiral amino acids with highly similar chemical properties, the penetration of nev-tRNA against the Ile (ATA) codon would have the least effect on the proteome. This hypothesis is also supported by the findings that other major nev-tRNAs tend to selectively penetrate the rare codons and are weakly expressed in the cell.

In conclusion, we have shown here that: *i*) nev-tRNAs are present only in the nematode lineage and tend to correspond to rare codons; *ii*) nev-tRNAs are weakly expressed in *C. elegans*; *iii*) nev-tRNAs are only charged with leucine *in vitro*; and *iv*) nev-tRNAs are incorporated into eukaryotic ribosomes during translation and decode rare codons to leucine *in vitro*, contrary to the genetic code. These results may provide further insight into the expansion of the genetic code in eukaryotes.

First base of codon		Second base of codon: T	Second base of codon: C	Second base of codon: A	Second base of codon: G		Third base of codon
T		Phe (TTT)	Ser (TCT)	Tyr (TAT)	Cys (TGT)	T	
		2.28	1.68	1.75	1.11		
		Phe (TTC)	Ser (TCC)	Tyr (TAC)	Cys (TGC)	C	
		2.36	1.06	1.37	0.90		
		Leu (TTA)	Ser (TCA)	End (TAA)	End (TGA)	A	
		0.97	2.07	0.11	0.08		
		Leu (TTG)	Ser (TCG)	End (TAG)	Trp (TGG)	G	
		2.00	1.23	0.04	1.09		
C		Leu (CTT)	Pro (CCT)	His (CAT)	Arg (CGT)	T	
		2.11	0.90	1.40	1.11		
		Leu (CTC)	Pro (CCC)	His (CAC)	Arg (CGC)	C	
		1.46	**0.44**	0.90	0.50		
		Leu (CTA)	Pro (CCA)	Gln (CAA)	Arg (CGA)	A	
		0.79	2.63	2.74	1.21		
		Leu (CTG)	Pro (CCG)	Gln (CAG)	Arg (CGG)	G	
		1.21	0.98	1.44	0.47		
A		Ile (ATT)	Thr (ACT)	Asn (AAT)	Ser (AGT)	T	
		3.24	1.92	3.03	1.22		
		Ile (ATC)	Thr (ACC)	Asn (AAC)	Ser (AGC)	C	
		1.88	1.03	1.82	0.83		
		Ile (ATA)	Thr (ACA)	Lys (AAA)	Arg (AGA)	A	
		0.94	2.03	3.74	1.53		
		i/eMet (ATG)	Thr (ACG)	Lys (AAG)	Arg (AGG)	G	
		2.61	0.89	2.58	**0.38**		
G		Val (GTT)	Ala (GCT)	Asp (GAT)	Gly (GGT)	T	
		2.43	2.26	3.64	1.09		
		Val (GTC)	Ala (GCC)	Asp (GAC)	Gly (GGC)	C	
		1.35	1.26	1.72	0.67		
		Val (GTA)	Ala (GCA)	Glu (GAA)	Gly (GGA)	A	
		0.99	2.02	4.14	3.16		
		Val (GTG)	Ala (GCG)	Glu (GAG)	Gly (GGG)	G	
		1.45	0.83	2.47	**0.44**		

Figure 2. 12. Codon usage of *C. elegans*.

The percentage usage of all sense codons is shown. Codons corresponding to the conserved nev-tRNAs in nematodes, Gly (GGG), Ile (ATA), Pro (CCC), and Arg (AGG), are shown in red. The five rarest codons, except the stop codons, are indicated in blue boxes.

Chapter 3

Analysis of genetic code ambiguity arising from nematode-specific misacylated tRNAs

3.1. Introduction

The translation of genes into proteins is based on the genetic code, a set of essential rules for living cells. The frequency of translational errors has been estimated to be approximately one misincorporated amino acid per 10,000 codons (Drummond and Wilke 2009; Reynolds *et al*. 2010; Harris and Kilby 2014; de Pouplana *et al*. 2014). This faithful translation of the genetic code is a central pillar of molecular biology, and is maintained by the accuracy of transfer RNA (tRNA) aminoacylation and codon–anticodon matching, as described in section 1.1. Therefore, any deviation from these sequential molecular recognition rules reflects the imperfection of the translation process, resulting in 'genetic code ambiguity' (Moura *et al*. 2009, 2010; Hamashima and Kanai 2013). A number of instances of natural codon reassignment suggest that genetic code ambiguity has evolutionary implications. Almost all genetic code ambiguities occur at stop codons, and ambiguity at sense codons has only been found in

several species of the genus *Candida*. The proteome of *Candida albicans* is unstable because deviant tRNAs carrying the CAG anticodon have a dual identity and decode the leucine (Leu) CUG codon not only as Leu but also as serine (Ser), with a frequency of ~5% (Sugita and Nakase 1999; Moura *et al*. 2009, 2010). Because these amino acid misincorporations alter the structural and biochemical characteristics of a large number of encoded proteins, the ambiguity of sense codons seems to have been strictly limited during the evolutionary diversification of most species.

Recent genome sequencing projects have identified a large number of eukaryotic tRNA isodecoders, tRNA molecules that share the same anticodons but differ in their molecular sequences, which sometimes contribute to the genome architecture and evolution (Iben and Maraia 2012; Hamashima and Kanai 2013). As described in the previous chapter, we identified novel nematode-specific tRNAs (nev-tRNAs) and demonstrated their weak expression in *Caenorhabditis elegans* (*C. elegans*) (Hamashima *et al*. 2012). *In vitro* aminoacylation assays have clearly shown that these tRNAs are solely charged with Leu, instead of glycine (Gly) or isoleucine (Ile). This is primarily attributable to the V-arm domains of the nev-tRNAs, which are very similar to that of $tRNA^{Leu}$ and are known to be a major determinant of recognition by leucyl-aminoacyl tRNA synthetase (Breitschopf *et al*. 1995; Soma *et al*. 1999; Fukunaga and Yokoyama 2005).

In contrast, each nematode species also express $tRNA^{Gly}$ (UCC) and $tRNA^{Ile}$ (UAU) with standard short variable loops, which are the cognate tRNAs of nev-$tRNA^{Gly}$ (CCC) and nev-$tRNA^{Ile}$ (UAU), respectively. These cognate tRNAs decode the GGG

codon and AUA codon, respectively, but $tRNA^{Gly}$ (UCC) is charged with Gly by glycyl-tRNA synthetase and $tRNA^{Ile}$ (UAU) with Ile by isoleucyl-tRNA synthetase (Hamashima *et al*. 2012). Therefore, if nev-$tRNA^{Gly}$ (CCC) participates in translation, GGG codons could be translated as either Gly or Leu in the competition between the nev-tRNA and its cognate tRNA. Similarly, if nev-$tRNA^{Ile}$ (UAU) participates in translation, the AUA codon could be translated as either Ile or Leu. Importantly, cell-free protein expression assays have demonstrated that nev-tRNAs can be incorporated into eukaryotic ribosomes and used in protein synthesis and therefore cause genetic code ambiguity, at least *in vitro* (see section 2.3.5). This raises several fundamental questions. Typically, many tRNA genes are encoded as precursor forms in the genome, and they must be processed to yield mature functional forms after transcription. The processing steps include trimming, CCA addition, intron splicing, base modification, and nuclear export (El Yacoubi *et al*. 2012). Although nev-tRNAs have unusual structural and aminoacyl properties that are inconsistent with the universal rules, it is unclear whether they are processed normally for translation like common tRNAs and whether they function in protein synthesis *in vivo*, which would confirm the ambiguity of the nematode genetic code.

To address these two questions, we next analyzed the functionality of nev-tRNAs in terms of their maturation by the addition of the 3′ CCA and their subcellular localization in *C. elegans*. These results show that nev-tRNAs are weakly expressed but mature normally, and are exported from the nucleus like their cognate tRNAs. We further performed a large-scale analysis of amino acid misincorporation

using high-resolution mass spectrometry (MS) to investigate whether nev-tRNAs participate in translation. However, no level of nev-tRNA-induced mistranslation was detected in the whole-cell proteome. These findings suggest that the nematode genetic code is not ambiguous, at least under normal growth conditions, and ensures high translational fidelity, contrary to our expectations (Hamashima *et al*. 2015).

3.2. Materials and Methods

3.2.1. Nematode culture and strain

The N2 strain of *C. elegans* and the OP50 strain of *Escherichia coli* used in this work were provided by the *Caenorhabditis* Genetics Center, which is funded by the NIH National Center for Research Resources. Mixed-stage worms, including eggs, larval stages 1–4, and adults, were grown with standard methods at 20°C (Brenner 1974).

A transgenic strain expressing *myo-3p::GFP-LacZ* was constructed by microinjecting wild-type worms with *myo-3p::GFP-LacZ* (pPD96.02) and coinjecting the marker *rol-6* (pRF4), generating an extrachromosomal array. We then randomly integrated the extrachromosomal array into the genome with UV irradiation (Mitani 1995). The transgenic worms containing the integrated array (msIs4) were back-crossed twice to the wild type, generating strain YK38.

3.2.2. Detection of the CCA sequence at the 3′ ends of tRNAs

Total RNA was isolated from mixed-stage *C. elegans* with TRIzol Reagent (Invitrogen, Carlsbad, CA, USA), according to the manufacturer's instructions. The RNAs were separated on a denaturing 6% polyacrylamide gel containing 8 M urea, and the tRNA fraction, ranging from 60 to 90 nt, was purified from the gel. These RNAs were treated with bacterial alkaline phosphatase and ligated with an adaptor sequence CATCGATCCTGCAGGCTAGAGAC at their 3′ ends using the Small RNA Cloning Kit

(TaKaRa, Shiga, Japan). The resulting RNAs were used as the templates for Reverse transcription (RT)–PCR.

To determine the sequences at the 3′ ends of the tRNAs, RT–PCR was performed with ReverTra Dash reverse transcriptase and KOD FX DNA polymerase (Toyobo Biochemicals, Osaka, Japan). The amplification reactions consisted of 30 cycles of denaturation at 98°C for 30 s, annealing at 60°C for 30 s, and extension at 74°C for 30 s, with specific primers (Table 3. 1; see also Figure 3. 1). The PCR products were separated on 10% polyacrylamide gel and then purified from the gel to remove the primers, primer dimers, and nonspecific PCR products. The purified PCR products were subcloned into the pCR-Blunt II-TOPO vector (Invitrogen). The nucleotide sequences of the inserted DNAs were determined with an ABI3100 DNA Sequencer (Applied Biosystems, Foster City, CA, USA).

Table 3. 1. Oligonucleotides used to detect the CCA sequence at the 3′ end of each tRNA.

Name	Type	Strand	Sequence
PA	anti-3' adaptor	Reverse	GTCTCTAGCCTGCAGGATCGATG
P1	$tRNA^{Gly}$ (UCC)	Forward	GCATGGATGCCTTCCAAGC
P2	$tRNA^{Ile}$ (UAU)	Forward	GCGCGTGGTACTTATAATGC
P3	nev-$tRNA^{Gly}$ (CCC)	Forward	CCTATTCTGGTAACAGAGCG
P4	nev-$tRNA^{Ile}$ (UAU)	Forward	CTCATTGGGTTAAACCAGTCG

3.2.3. Analysis of the subcellular localization of tRNAs

Whole *C. elegans* worms were subjected to subcellular fractionation according to the protocol of Zisoulis *et al*. (Zisoulis *et al*. 2012), with slight modifications. Briefly, mixed-stage worms were harvested, snap frozen, and lysed in ice-cold NP-40 lysis buffer (NLB) containing 10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM $MgCl_2$, 0.5% NP-40, 1 mM dithiothreitol (DTT), and 100 U/ml RNasin Plus (Promega, Madison, WI, USA) using a mortar and pestle precooled with liquid N_2. The lysates were centrifuged at 500 × *g* for 30 s at 4°C and one-fifth of the supernatant volume was collected (whole-cell fraction). The remainder was centrifuged at 2,000 × *g* for 5 min at 4°C. The supernatant (postnuclear fraction, i.e., cytoplasmic fraction) was centrifuged again (2,000 × *g* for 5 min at 4°C) and the pellet (nuclear fraction) was washed with NLB. The nuclear fraction was resuspended in a volume of NLB equivalent to that of the cytoplasmic fraction. The RNAs in each fraction were isolated with TRIzol Reagent and used as the templates for RT–PCR.

To examine the expression of tRNAs in each fraction, RT–PCR was performed with the enzymes ReverTra Dash and KOD FX Neo (Toyobo Biochemicals). The amplification reactions consisted of 20–35 cycles of denaturation at 98°C for 10 s, annealing at 55/60°C for 3 s, and extension at 74°C for 6 s, with specific primers (Table 3. 2). The PCR conditions were optimized by manipulating the number of cycles and the annealing temperature to determine the linear ranges. The PCR products were separated by electrophoresis on 3% NuSieve 3:1 Agarose gel (Cambrex Bio Science, Rockland, ME, USA). The bands were stained with ethidium bromide and visualized

with a Molecular Imager FX Pro (Bio-Rad Laboratories, Hercules, CA, USA), and then semiquantified with densitometry using the ImageJ v1.48 software (Schneider *et al*. 2012). The RT–PCR products were then purified with the Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare, Buckinghamshire, UK), and subcloned into the pCR-Blunt II-TOPO vector (Invitrogen). The nucleotide sequences of the inserted DNAs were determined with an ABI3100 DNA Sequencer (Applied Biosystems), and confirmed to be identical to those in the database.

Table 3. 2. Oligonucleotides used to analyze the subcellular localization of tRNAs.

				Optimized PCR condition [a]	
Name	**Type**	**Strand**	**Sequence**	**Annealing**	**Cycles**
P5	snU6	Forward	GTTCTTCCGAGAACATATAC	55°C	x 20
P6		Reverse	AACGCTTCACGAATTTGC		
P7	snoU3	Forward	ACTATACAGAATCATTTCTGCAG	60°C	x 30
P8		Reverse	ACTGCTCAGAAGAGCAGG		
P9	$tRNA^{iMet}$	Forward	AGCAGCGTGGCGCAGTGGAA	60°C	x 25
P10		Reverse	TAGCAGCGAGTGGTTTCGATCCA		
P11	$tRNA^{Gly}$ (UCC)	Forward	GCGTTCGTGGTGTAATGGTCAGC	60°C	x 35
P12		Reverse	TGCGTTCGGGGGGAATCGAA		
P13	$tRNA^{Ile}$ (UAU)	Forward	GCCCCATTGGCGCAGTCGGTTAGC	60°C	x 30
P14		Reverse	TGCCCCATGCCAGGCTCGAACTG		
P15	nev-$tRNA^{Gly}$ (CCC)	Forward	GCGGTGGTGGCCGAGCGGTCA	60°C	x 35
P16		Reverse	CCCGCGCTCTGTTACCAGAATAG		
P17	nev-$tRNA^{Ile}$ (UAU)	Forward	GCCCCGGTGGCCGAGCGGTCG	60°C	x 30
P18		Reverse	CCCGCGACTGGTTTAACCCAATGA		

[a] PCR reactions were performed using the indicated cycle numbers : denaturation for 10 s at 98°C, annealing for 3 s at the indicated temperatures and extension for 6 s at 74°C.

3.2.4. Immunoprecipitation of the GFP–LacZ protein

The GFP–LacZ protein was immunoprecipitated with GFP-Trap_A (ChromoTek, Martinsried, Planegg, Germany) from mixed-stage transgenic worms expressing *myo-3p::GFP-LacZ*, according to the manufacturer's instructions. The purified protein was detected with an immunoblotting analysis using an anti-GFP antibody (Clontech Laboratories, Mountain View, CA, USA, catalogue # 632380), and its molecular weight was identical to that in the database.

3.2.5. NanoLC–MS/MS analysis

The protein mixture was extracted from mixed-stage wild-type worms with 100 mM triethylammonium bicarbonate (TEAB; pH 8.5) containing 12 mM sodium deoxycholate and 12 mM sodium dodecanoyl sarcosinate. The resulting mixture or the purified GFP–LacZ protein was reduced with 10 mM DTT at room temperature for 30 min, and alkylated with 47 mM iodoacetamide at room temperature in the dark for 30 min. The samples were diluted five-fold with 100 mM TEAB (pH 8.5) and digested with sequence-grade lysyl endoprotease (Lys-C), trypsin, endprotease Glu-C (V8), or chymotrypsin, and then desalted with StageTips using a C18 Empore™ disk membrane (Rappsilber *et al*. 2003).

An LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a nanoLC interface (Nikkyo Technos, Tokyo, Japan), a Dionex UltiMate 3000 pump with an FLM-3000 Flow Manager (Germering, Germany), and an HTC-PAL® Autosampler (CTC Analytics, Zwingen, Switzerland) were used for the

nanoLC–MS/MS measurements. Reprosil C18 material (3 μm; Dr Maisch, Ammerbuch, Germany) was packed into a self-pulled needle (100 μm, I.D. × 130 mm; tip I.D. 5 μm) with a nitrogen-pressurized column load cell (Nikkyo Technos, Tokyo, Japan) to prepare an analytical column needle with a "stone-arch" frit (Ishihama *et al*. 2002). The flow rate was 500 nL/min. The mobile phases consisted of (A) 0.5% acetic acid and (B) 0.5% acetic acid and 80% acetonitrile. A three-step linear gradient was used: 5% to 10% B in 5 min, 10% to 40% B in 60 min, 40% to 100% B in 5 min, and maintained at 100% B for 10 min. A spray voltage of 2400 V was applied. The MS scan range was *m/z* 300–1500 in the Orbitrap mass spectrometer and the top 10 precursor ions (for whole-cell proteomics) or the targeted precursor ions (for targeted proteomics) were selected for subsequent MS/MS scans in the linear ion trap mass spectrometer.

3.2.6. Identification of peptides containing misincorporated amino acids

Peak lists were created using an in-house Perl script based on the recorded product ion mass spectra. Peptides and proteins were identified with Mascot v2.4 (Matrix Science, London) against the UniProt/Swiss-Prot database (downloaded 2013/06, subset *C. elegans*, 3428 protein entries) in error-tolerant mode (Creasy and Cottrell 2002) or in an in-house protein database (in which all Gly residues arising from the decoding of the GGG codons were altered to Leu or Ser). A precursor mass tolerance of 3 ppm and a fragment ion mass tolerance of 0.8 Da were used with strict enzyme specificity, allowing for up to two missed cleavages (Olsen *et al*. 2004). The carbamidomethylation of cysteine was set as a fixed modification, and methionine

oxidations were allowed as variable modifications. Peptides were rejected if the Mascot score was below the 95% confidence limit based on the "identity" score for each peptide. Some of the identified peptides were further validated with a targeted proteomics approach using synthesized peptides (AQUA™ Peptides; Sigma-Aldrich, St. Louis, MO, USA) that contained stable-isotope-labeled amino acids (see Table 3. 3 and Table 3. 4).

Table 3. 3. List of internal standards for the calibration of whole-cell proteomics.

Type	Name	Sequence [a]	Label	Calculated *m/z*
Gly-to-Leu	MADD_CAEEL_N	LNLLLEVK	L_C13N15	474.8140
	MADD_CAEEL_C	LNLLLEVK	K_C13N15	475.3125
	MED1_CAEEL	QPGEEPPAGRGKGLKK	K_C13N15	552.9809
	NDUS7_CAEEL	TALAVGTRR	R_C13N15	477.7898
	NGLY1_CAEEL	ERGELLESGPK	K_C13N15	611.8298
	PAFA_CAEEL	QVLCKDL	L_C13N15	413.2345
	IFA2_CAEEL	FEEAQRLR	R_C13N15	529.7847
	LE418_CAEEL	AFYLAVMR	R_C13N15	490.7670
	RIM_CAEEL	RTDTLK	K_C13N15	371.2211
	ACN1_CAEEL	ALEMISLK	K_C13N15	456.7696
	CADH3_CAEEL	RDLHINMAYL	L_C13N15	626.8329
	CLAP1_CAEEL	LAELNNTLIISIL	L_C13N15	717.4383
	CHITL_CAEEL	EDAATSVKVANLL	L_C13N15	669.3731
	SIR41_CAEEL_N	RSKDVLL	R_C13N15	419.2672
	SIR41_CAEEL_C	RSKDVLL	L_C13N15	420.7628
	NHR20_CAEEL	TPSMKVIL	L_C13N15	448.2736
	PS11A_CAEEL	SPASLDDDIK	K_C13N15	534.7688
	SRRT_CAEEL	LLIEK	K_C13N15	312.2148
	CLH_CAEEL	TLQIK	K_C13N15	305.7046
Gly-to-Ser	DHTK1_CAEEL	LSEEAILSF	L_C13N15	339.1859
	U520_CAEEL	SGIIQATELSR	R_C13N15	592.8294
	HUTU_CAEEL	AEKQVDSLR	R_C13N15	528.2899
	GALT9_CAEEL	HSLIR	R_C13N15	318.1970
	DGK3_CAEEL	MPSLFPMK	K_C13N15	479.7529
	SRB5_CAEEL	VTSQEGAR	V_C13N15	427.2242
	SRE37_CAEEL	FISSLPIIR	R_C13N15	528.3283

[a] Amino acid residues at the Gly (GGG) codon are shown in red. Stable isotopically labeled amino acids are underlined.

Table 3. 4. List of internal standards used in targeted proteomic analysis of purified GFP–LacZ.

Type	Name	Sequence [a]	Label	Calculated *m/z*
GGG-to-Gly	GFP2-Trp-G	DDGNYK	K_C13N15	711.2944
	GFP3-Trp-G	DHMVLLEFVTAAGITHGMDELYK	K_C13N15	2590.2680
	LacZ3-Trp-G	LSGQTIEVTSEYLFR	R_C13N15	1742.8959
	LacZ7-Trp-G	VNWLGLGPQENYPDR	R_C13N15	1757.8606
	GFP1-V8-G	VNGHKFSVSGE	V_C13N15	1160.5695
	GFP2-V8-G	GNYKTRAE	R_C13N15	938.4690
	GFP3-V8-G	FVTAAGITHGMD	V_C13N15	1219.5776
	LacZ2-V8-G	RNHPSVIIWSLGNE	R_C13N15	1621.8445
	LacZ7-V8-G	RVNWLGLGPQE	R_C13N15	1268.6746
	LacZ8-V8-G	NGLRCGTRE	R_C13N15	1062.5109
	LacZ9-V8-G	GFHMGIGGD	F_C13N15	890.3825
	LacZ1-TrpV8-G	GVNSAFHLWCNGR	R_C13N15	1517.7066
	LacZ4-TrpV8-G	SAGQLWLTVR	R_C13N15	1130.6317
	LacZ5-TrpV8-G	AGHISAWQQWR	R_C13N15	1339.6654
	LacZ6-TrpV8-G	AVLITTAHAWQHQGK	K_C13N15	1660.8918
GGG-to-Leu	GFP2-Trp-L	DDLNYK	K_C13N15	767.3570
	GFP3-Trp-L	DHMVLLEFVTAALITHGMDELYK	K_C13N15	2646.3306
	LacZ3-Trp-L	LSLQTIEVTSEYLFR	R_C13N15	1798.9585
	LacZ7-Trp-L	VNWLGLLPQENYPDR	R_C13N15	1813.9232
	GFP1-V8-L	VNLHKFSVSGE	V_C13N15	1216.6321
	GFP2-V8-L	LNYKTRAE	R_C13N15	994.5316
	GFP3-V8-L	FVTAALITHGMD	V_C13N15	1275.6402
	LacZ2-V8-L	RNHPSVIIWSLLNE	R_C13N15	1677.9071
	LacZ7-V8-L	RVNWLGLLPQE	R_C13N15	1324.7372
	LacZ8-V8-L	NGLRCLTRE	R_C13N15	1118.5735
	LacZ9-V8-L	GFHMLIGGD	F_C13N15	946.4451
	LacZ1-TrpV8-L	GVNSAFHLWCNLR	R_C13N15	1573.7692
	LacZ4-TrpV8-L	SALQLWLTVR	R_C13N15	1186.6943
	LacZ5-TrpV8-L	ALHISAWQQWR	R_C13N15	1395.7280
	LacZ6-TrpV8-L	AVLITTAHAWQHQLK	K_C13N15	1716.9544

[a] Amino acid residues at the Gly (GGG) codon are shown in red. Stable isotopically labeled amino acids are underlined.

3.2.7. Network analysis based on the sequence similarities of nev-tRNAs in the nematode taxon

To obtain large datasets of nev-tRNA sequences in a wide range of nematode species, 26 nematode genomic sequences were obtained from the WormBase database (ftp://ftp.wormbase.org/pub/wormbase/, last accessed January 12, 2015). These sequences were analyzed comprehensively using the criteria described in section 2.3.1, and a total of 195 nev-tRNA sequences were obtained. Sequence similarity scores were then calculated according to the strategy of Matsui *et al*. (Matsui *et al*. 2013), with slight modifications. Briefly, each similarity score was calculated for all pairs of nev-tRNAs based on a BLASTN (BLAST 2.2.17) analysis with a cutoff at *E* value < 1e–7. Using the resulting scores, a weighted undirected network graph was constructed using Cytoscape 2.8.3 (Smoot *et al*. 2011).

3.3. Results and Discussion

3.3.1. Expression, maturation, and subcellular localization of nev-tRNAs

As described in the previous chapter, we identified nematode-specific novel tRNA genes, designated as nev-tRNAs, e.g., nev-tRNA^{Gly} (CCC) and nev-tRNA^{Ile} (UAU), which contain 15–20-nt V-arm structures and are solely charged with Leu instead of Gly or Ile *in vitro* (Hamashima *et al*. 2012). To obtain further evidence of the functionality of nev-tRNAs in cells, the following two characteristics were analyzed: *i*) their maturation, with the addition of 3′ CCA; and *ii*) their subcellular localization. In these experiments, tRNA^{Gly} (UCC) and tRNA^{Ile} (UAU), which are the cognate tRNAs of nev-tRNA^{Gly} (CCC) and nev-tRNA^{Ile} (UAU), were used as the positive controls to test for GGG and AUA codon ambiguity in nematode cells.

The addition of CCA to the 3′ end of the tRNA molecule is one of its most important posttranscriptional modifications, and is essential for various tRNA functionalities, including other processing, aminoacylation, and tRNA–ribosome interactions (El Yacoubi *et al*. 2012). To determine the 3′ end sequences of the nev-tRNAs with RT–PCR, a set of template tRNAs was isolated from mixed stages of *C. elegans* (eggs, larval stages 1–4, and adults), and ligated with a 23-nt adaptor sequence at their 3′ ends. RT–PCR amplification was conducted with forward primers that annealed to a specific region on each tRNA (positions 22–40 and 23–42 for common tRNA^{Gly} and tRNA^{Ile}, respectively; positions 40–59 and 41–61 for nev-tRNA^{Gly} and nev-tRNA^{Ile}, respectively) and reverse primers that annealed to the 3′ adaptor region

(Figure 3. 1A). Targeted regions of the predicted lengths were successfully amplified, except for nev-tRNA^{Gly} (Figure 3. 1B). The amplification efficiency for nev-tRNA^{Gly} was considerably lower than that for the other templates but the amplified product was clearly detected with a second PCR analysis. The amplification efficiency for nev-tRNA^{Ile} was also slightly lower than that for the normal tRNAs. Taken together with our previous studies (see section 2.3.3), these data suggest that the abundance of the mature nev-tRNAs in the cells was low. The amplified products of the expected sizes were then subcloned and the nucleotide sequences at their 3′ ends were determined. Figure 3. 1C shows that not only the common tRNA^{Gly} (UCC) and tRNA^{Ile} (UAU) but also nev-tRNA^{Gly} (CCC) and nev-tRNA^{Ile} (UAU) matured normally, with the addition of CCA at their 3′ ends. These findings show that nev-tRNAs are processed to the functional form for translation, just like their cognate tRNAs, although the structural and biochemical properties of the nev-tRNAs differ from those of normal tRNAs.

We next analyzed the subcellular localization of the nev-tRNAs to determine whether they are exported from the nucleus after posttranscriptional modification. The whole *C. elegans* worm was subjected to subcellular fractionation with differential centrifugation (see section 3.2.3). Figure 3. 2 (upper panel) shows the subcellular localization of the control RNAs: U6 small nuclear RNA (snU6) and U3 small nucleolar RNA (snoU3) were enriched in the nucleus (~2.9-fold) relative to their levels in the cytoplasm, whereas tRNA^{iMet} was enriched in the cytoplasm (~2.8-fold) relative to its level in the nucleus, as previously reported (Lee *et al*. 2009; Zisoulis *et al*. 2012). Under the same conditions, nev-tRNA^{Gly} (CCC) and nev-tRNA^{Ile} (UAU) were detected at

higher levels (~2.0-fold) in the cytoplasm than in the nucleus (Figure 3. 2, lower panel), suggesting that the nev-tRNAs are exported from the nucleus and might therefore be used in translation. This experiment also confirmed that normal $tRNA^{Gly}$ (UCC) and $tRNA^{Ile}$ (UAU) are exported from the nucleus. Moreover, we determined the anticodon sequences of approximately 30 clones of each nev-tRNA, both in the nucleus and cytoplasm, and found that no anticodon was changed to a leucine codon by an RNA editing event. These results support the possibility that nev-tRNAs compete with their cognate tRNAs during translation. It must be noted that it is still unclear whether nev-tRNA anticodons are changed by specific chemical modifications so that they can read leucine codons.

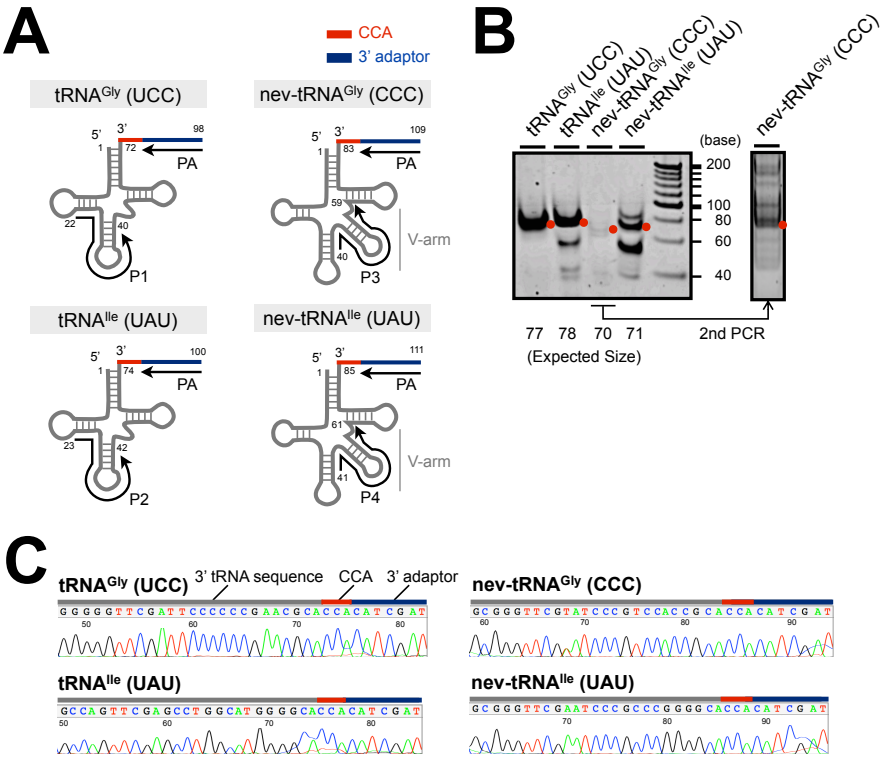

Figure 3. 1. Detection of the 3′ CCA end sequences of nev-tRNAs.

(A) PCR scheme for the detection of the 3′ ends of mature tRNAs: nev-tRNA^{Gly} (CCC) and nev-tRNA^{Ile} (UAU) and their cognates, tRNA^{Gly} (UCC) and tRNA^{Ile} (UAU), respectively. Numbers indicate the nucleotide positions relative to the 5′ end of each tRNA. (B) RT–PCR amplification of the 3′ end of each tRNA. PCR products of the expected sizes are shown as red dots. (C) Nucleotide sequence chromatograms of the 3′ end region of each tRNA.

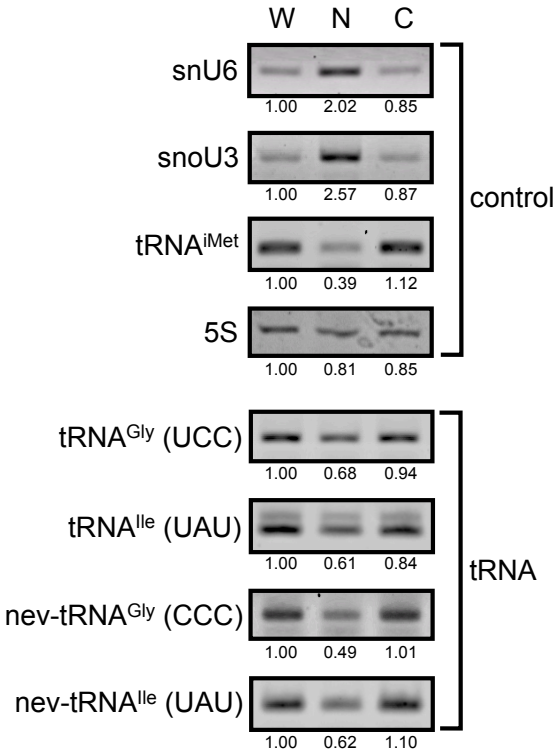

Figure 3. 2. Subcellular localization of nev-tRNAs in *C. elegans*.

RNA was isolated from each fraction of *C. elegans*: whole cell (W), nuclear (N), or cytoplasmic (C). RT–PCR analysis was used to detect snU6 and snoU3 RNAs (nuclear markers), $tRNA^{iMet}$ (cytoplasmic marker), and four tRNAs (nev-$tRNA^{Gly}$ and nev-$tRNA^{Ile}$, and their cognate tRNAs). 5S rRNA expression is shown as the loading control. Band densities were evaluated semiquantitatively with densitometry.

3.3.2. Analysis of amino acid misincorporation in the whole-cell proteome of *C. elegans*

Our previous studies have shown that nev-tRNA^{Gly} (CCC) can be incorporated into ribosomes and used for protein synthesis in an insect cell-free protein expression system (see section 2.3.5). This finding is evidence that nev-tRNAs cause genetic code ambiguity, at least *in vitro*. Because nev-tRNAs are exported from the nucleus and might compete with their cognate tRNAs in *C. elegans*, we assumed that nev-tRNAs are involved in protein synthesis *in vivo*, creating genetic code ambiguity. To address this hypothesis, we performed a shotgun proteomic analysis of *C. elegans* using liquid chromatography–tandem MS (LC–MS/MS), and examined the kinds of protein molecules within the whole-cell proteome that contained misincorporated amino acids. High-resolution MS can directly monitor very low levels of minor protein isoforms on a large scale (Yu *et al*. 2009; Zhang *et al*. 2013). In this experiment, we mainly focused on Gly-to-Leu (in which Gly at the GGG codon is replaced with Leu) and Gly-to-Ser (in which Gly at the GGG codon is replaced with Ser) misincorporations. Gly-to-Ser misincorporations were used as the negative control because nev-tRNA^{Gly} (CCC) cannot be completely charged with Ser *in vitro* (Hamashima *et al*. 2012), suggesting that it does not cause Gly-to-Ser misincorporation. We did not look for Ile-to-Leu (in which Ile at the AUA codon is replaced with Leu) misincorporation because the Leu residue is indistinguishable from the Ile residue on MS, as they are structural isomers with identical molecular weights.

For the whole-cell proteomic analysis, a protein mixture was extracted from

mixed-stage *C. elegans* and fragmented into small peptides by digestion with site-specific enzymes. After the LC–MS/MS analysis of the resulting peptides, the data were examined with Mascot v2.4 (Matrix Science, London) to identify the amino acid misincorporations, using two different approaches: (a) an error-tolerant search; and (b) an in-house database search (Figure 3. 3A). The error-tolerant search is one of the optional modes of the Mascot protein database search (Creasy and Cottrell 2002), in which the raw data are initially searched against a reference protein database, after which the MS/MS data that do not match the expected amino acid sequences of known proteins are checked against a database containing all possible amino acid misincorporations and posttranslational modifications. With the error-tolerant search, 295,216 nonredundant (unique) peptides were identified. The in-house database search was developed and optimized in this study to compare the raw data against modified protein databases containing only possible Gly-to-Leu or Gly-to-Ser misincorporations, with no initial search against a reference protein database. This search identified 12,719 and 12,502 unique peptides, respectively (Figure 3. 3A, Step 1).

After discarding the low-quality peptides, 75 (= 14 + 30 + 31) candidate Gly-to-Leu mutant peptides and 53 (= 6 + 33 + 14) candidate Gly-to-Ser mutant peptides were extracted (Figure 3. 3A, Step 2). The mean Mascot confidence score for the Gly-to-Leu candidates was 20.3 ± 6.3, which did not differ significantly from that of the Gly-to-Ser candidates ($p > 0.01$) (Figure 3. 3B). The candidate misincorporations were then further screened by the manual curation of their MS/MS spectra and isotope ratios, and 17 (= 1 + 10 + 6) and seven (= 0 + 3 + 4) mutant peptides were finally

obtained, respectively (Figure 3. 3A, Step 3, and summarized in Table 3. 5). To confirm that these peptides had identical amino acid sequences to those predicted with Mascot, a targeted proteome analysis was performed using an internal standard (IS) (Figure 3. 3A, Step 4). The IS was a synthesized peptide consisting of the same amino acid sequence as that identified with Mascot, in which one amino acid at the N- or C-terminus was labeled with a stable isotope (summarized in Table 3. 3). If the ions of both targeted peptides and the IS were detected at quite similar elution times with LC, indicating their almost equivalent chemical properties, peptide identification was deemed to be reliable. However, if their elution times differed by > 1.0 min, peptide identification was deemed to be unreliable. Validation with these criteria revealed that all the candidate misincorporations were false-positive Mascot identifications. One example is shown in Figure 3. 3C. This result means that no Gly-to-Leu mutant peptide was detectable, which was also true for the Gly-to-Ser negative control, suggesting that nev-tRNA^{Gly} (CCC) does not cause GGG codon ambiguity in the whole-cell proteome of *C. elegans*. This was also supported by the finding that no Gly-to-Leu candidate had a significantly higher Mascot score than the Gly-to-Ser candidates (Figure 3. 3B).

To gain more insight into the frequencies and variations of the amino acid misincorporations for each codon, we estimated the entire 64 × 19 possible codon-to-amino acid errors using data obtained with the error-tolerant search. Note that only a proportion of the identifications, with high Mascot confidence scores (> 30), was selected for this analysis because false-positive Gly-to-Leu misincorporations had low Mascot confidence scores (< 30), as described above. When the relationship between

the amino acids used in the whole proteome and the number of predicted misincorporations for each codon was investigated with Pearson's correlation coefficient, a strong significant correlation ($r = 0.917$) was observed (Figure 3. 4A). For example, the number of predicted misincorporations at frequent codons, such as the Glu (GAA) and Asp (GAU) codons, was up to 478, whereas fewer misincorporations were predicted at the Gly (GGG) and Ile (AUA) codons (approximately 4%). Furthermore, the Gly residues at the GGG codon showed little tendency to be substituted, not only with Leu (described as 'Xle' in the figure) but also with other amino acids (Figure 3. 4B). Similarly, there was no specific variation in the predicted misincorporations at the AUA codon. These observations show that nev-tRNAs do not seem to be involved in mistranslation at the corresponding codons in whole cells of *C. elegans*. However, in a single regression analysis, a dot corresponding to the Glu (GAG) codon was located outside the 95% confidence interval (Figure 3. 4A). As shown in Figure 3. 4B, Glu residues at the GAG codon tend to be substituted with Met residues at high levels (~7.3×10^{-4}). In bacterial, yeast, and mammalian cells, it has been reported that Met is misacylated to specific nonmethionyl tRNA families, such as $tRNA^{Glu}$ and $tRNA^{Lys}$, and that these Met-misacylated tRNAs are used for protein synthesis during some cellular responses (see section 1.2; Figure 1. 2D) (Netzer *et al*. 2009; Jones *et al*. 2011; Wiltrout *et al*. 2012; Pan 2013). Although nev-tRNAs cannot decode the GAG codon because at least one base pair is mismatched, the common $tRNA^{Glu}$ (CUC) encoded in the *C. elegans* genome can decode it. Therefore, the high Glu-to-Met error rate in *C. elegans* suggests the involvement of $tRNA^{Glu}$ (CUC) misacylation in this phenomenon, as in

bacterial, yeast, and mammalian cells.

Table 3. 5. Candidate peptides containing misincorporated Leu/Ser at Gly (GGG) codon.

						Retention time (min) [c]			Source protein for identified peptide	
Type	Sequence [a]	Modification	Observed *m/z*	Mascot Score	Screening [b]	Candidate	Internal Standard	Δ time	Swiss-Prot AC #	Description
Gly-to-Leu	LNLLLEVK		471.3051	23.67	common	70.1	59.5	10.6	MADD_CAEEL	MAP kinase-activating death domain protein
	QPGEEPPAGRGKGLKK		550.3103	14.13	common	60.8	25.2	35.6	MED1_CAEEL	Mediator of RNA polymerase II transcription subunit 1.1
	TALAVGTRR		472.7859	23.00	common	38.3	27.6	10.7	NDUS7_CAEEL	Probable NADH dehydrogenase [ubiquinone] iron-sulfur protein 7 mitochondrial
	ERGELLESGPK		607.8239	12.72	common	46.7	34.6	12.1	NGLY1_CAEEL	Peptide-N(4)-(N-acetyl-beta-glucosaminyl)asparagine amidase
	QVLCKDL		438.2366	17.49	common	ND	ND	-	PAFA_CAEEL	Platelet-activating factor acetylhydrolase homolog 2
	FEEAQRLR		524.7807	21.22	common	36.7	30.6	6.1	IFA2_CAEEL	Intermediate filament protein ifa-2
	AFYLAVMR		485.7629	19.27	common	57.1	58.3	1.2	LE418_CAEEL	Protein let-418
	RTDTLK		367.2139	12.70	common	ND	ND	-	RIM_CAEEL	Rab-3-interacting molecule unc-10
	ALEMISLK		452.7621	18.27	common	48.5	52.5	4.0	ACN1_CAEEL	Inactive angiotensin-converting enzyme-related protein
	RDLHINMAYL		623.3240	10.55	common	ND	54.9	-	CADH3_CAEEL	Cadherin-3
	LAELNNTLIISIL		713.9298	18.62	(a)	ND	ND	-	CLAP1_CAEEL	Protein CLASP-1
	EDAATSVKVANLL		665.8645	31.15	(b)	72.8	57.4	15.4	CHITL_CAEEL	Chitinase-like protein C25A8.4
	RSKDVLL		415.7591	23.69	(b)	40.8	31.2	9.6	SIR41_CAEEL	NAD-dependent protein deacetylase sir-2.2
	TPSMKVIL		444.7649	25.66	(b)	68.9	53.5	15.4	NHR20_CAEEL	Nuclear hormone receptor family member nhr-20
	SPASLDDDIK		530.7622	20.47	(b)	36.3	37.8	1.5	PS11A_CAEEL	Probable 26S proteasome regulatory subunit rpn-6.1
	LLIEK		308.2075	31.40	(b)	ND	ND	-	SRRT_CAEEL	Serrate RNA effector molecule homolog
	TLQIK		301.6973	25.82	(b)	ND	ND	-	CLH_CAEEL	Probable clathrin heavy chain 1
Gly-to-Ser	LSEEAILSF		336.8471	15.16	common	35.2	59.9	24.7	DHTK1_CAEEL	Probable 2-oxoglutarate dehydrogenase E1 component DHKTD1 homolog mitochondrial
	SGIIQATELSR		587.8251	27.20	common	ND	45.1	-	U520_CAEEL	Putative U5 small nuclear ribonucleoprotein 200 kDa helicase
	AEKQVDSLR		523.2858	22.76	common	34.1	27.0	7.1	HUTU_CAEEL	Probable urocanate hydratase
	HSLIR		313.1930	24.63	(b)	ND	ND	-	GALT9_CAEEL	Probable N-acetylgalactosaminyltransferase 9
	MPSLFPMK	Oxidation@M:7	483.7424	23.90	(b)	50.6	58.5	7.9	DGK3_CAEEL	Probable diacylglycerol kinase 3
	VTSQEGAR		424.2170	27.19	(b)	ND	22.6	-	SRB5_CAEEL	Serpentine receptor class beta-5
	FISSLPIIR		523.3242	26.38	(b)	73.9	58.8	15.1	SRE37_CAEEL	Serpentine receptor class epsilon-37

[a] Amino acid residues arising from the decording of Gly (GGG) codon are shown in red.
[b] Screening types of each mutant peptide candidate are shown (see Figure 3. 3A).
[c] Chromatographic retention times of the targeted peptides are shown.

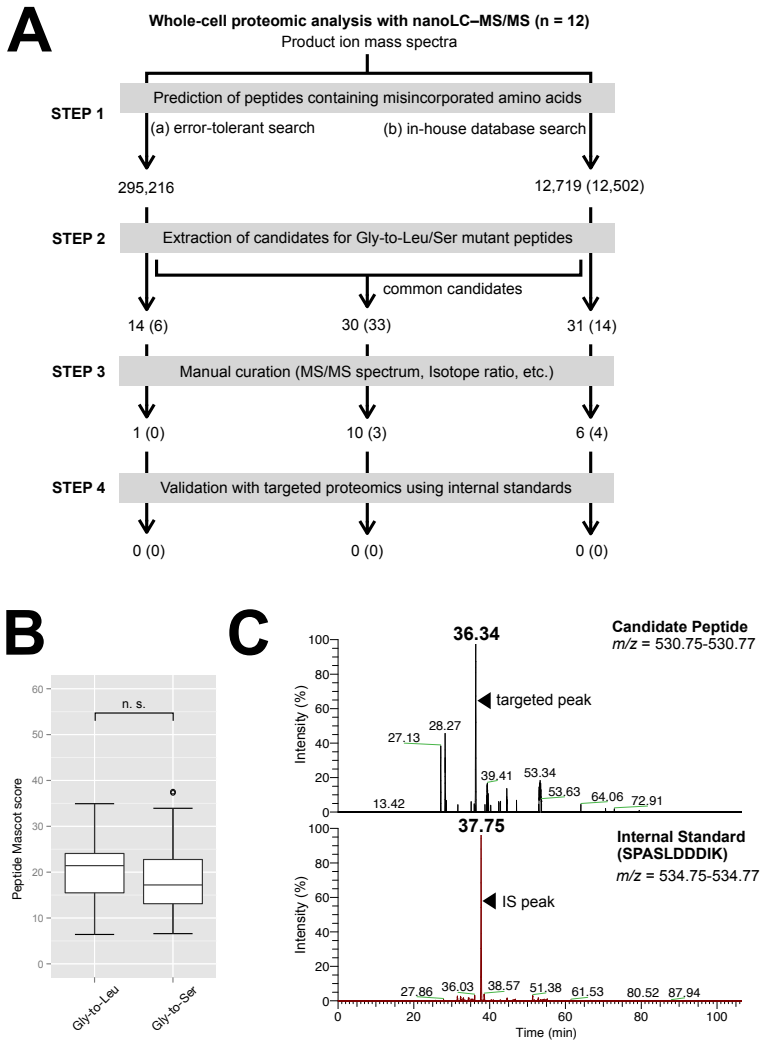

Figure 3. 3. Screening for mutant peptides resulting from nev-tRNA^{Gly}-dependent decoding.

(A) Summary of the whole-cell proteome analysis of mixed-stage *C. elegans*. Values are the unique peptide counts at each step. Values in parentheses are the count of candidate peptides containing misincorporated Ser at the Gly (GGG) codon (negative control). (B) Boxplot of the confidence scores for the candidate peptides in Step 2. Significant differences were determined with Student's two-sided *t* test. (C) Example of the validation of targeted proteomics. Extracted ion chromatograms of the candidate peptide and the synthetic peptide SPASLDDDIK (an internal standard) are shown. The candidate peptide ion was separated > 1.0 min earlier than the internal standard, indicating that the amino acid sequence of the candidate peptide was inconsistent with the sequence SPASLDDDIK.

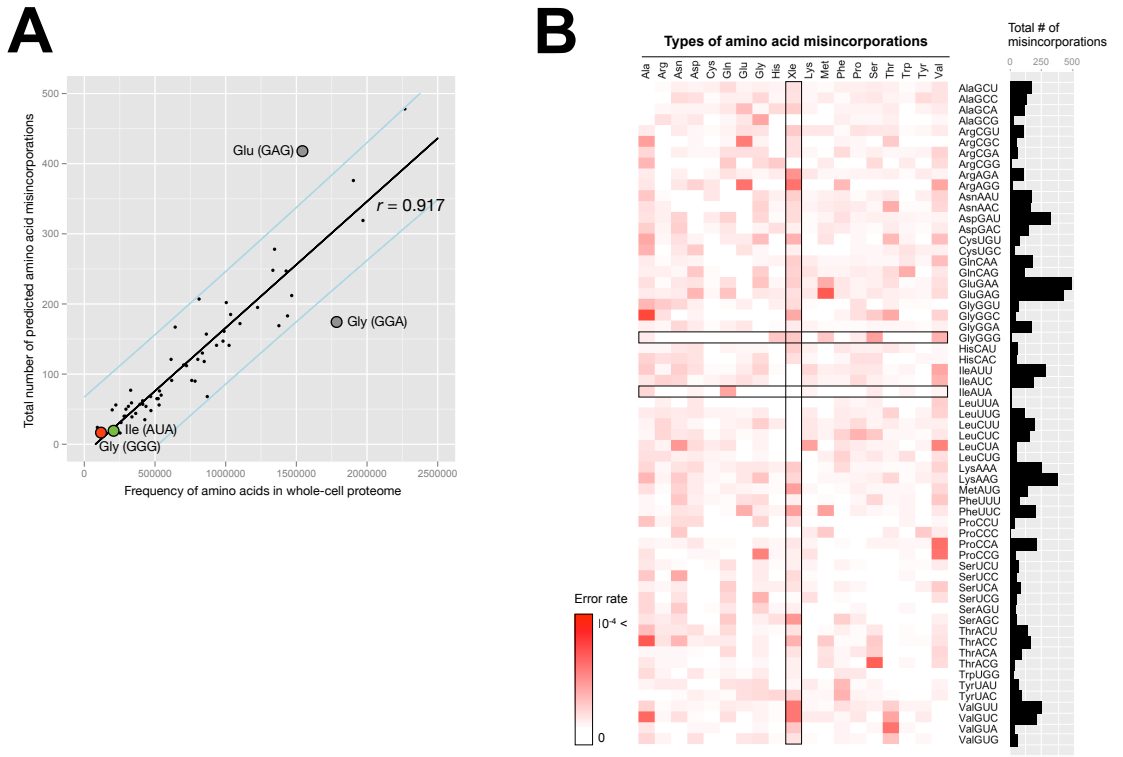

Figure 3. 4. Distribution of amino acid misincorporations predicted in the whole-cell proteome of *C. elegans*.

(A) Scatterplot of the frequencies of amino acids contained in all nonredundant peptides identified with a normal database search (x-axis) versus the total number of predicted amino acid misincorporations (y-axis) for each codon. The black line in the center denotes the linear regression line. The outer, light blue lines denote the 95% confidence interval for an individual predicted value. The red and green dots correspond to the GGG and AUA codons, respectively. The dots located outside the 95% confidence interval are shown in gray. (B) Heat map indicating the degree of predicted amino acid misincorporation (error rate) for each codon. The error rate was predicted by calculating the abundance of misincorporated amino acids relative to the total number of amino acids contained in the whole proteome. The matrix plots in the Gly (GGG) and Ile (AUA) row and in the 'Xle' (i.e., Ile or Leu) column are boxed. The total numbers of predicted misincorporations for each codon are indicated as a bar chart.

3.3.3. Possible explanations of the lack of genetic code ambiguity in *C. elegans*

We considered two possible reasons why no Gly-to-Leu mutant peptides were detected in this study, even though the nev-tRNAs matured normally and were exported from the nucleus. First, it is possible that nev-tRNAs are excluded from the protein synthesis process by a translation quality control mechanism. In Bacteria, one of the elongation factors, EF-Tu, selectively binds to the correct aminoacyl-tRNAs and delivers them into the A-site of the ribosome (LaRiviere *et al*. 2001; Fahlman *et al*. 2004; Reynolds *et al*. 2010). In human neural cells, if the translation process is stopped because a tRNA is mutated, one of the ribosome release factor, GTPBP2, interacts with the ribosome recycling protein Pelota, and releases the stalled ribosome (Ishimura *et al*. 2014). Although it is unclear whether homologues of EF-Tu and GTPBP2 act in *C. elegans*, as has been reported in other species, these findings allow the possibility that the translational errors induced by mischarged nev-tRNAs might be prohibited by such quality control systems.

Second, it is also possible that nev-tRNAs are used for protein synthesis in the cell, but that the frequency of amino acid misincorporations is below the level of MS detection. The MS-based method can directly measure a large number of amino acid misincorporations, down to a level of 0.01% (10^{-4}) (Yu *et al*. 2009; Zhang *et al*. 2013). However, because the abundance of mature nev-tRNAs in the cell is very low and they compete with highly expressed cognate tRNAs (Figure 3. 2), the incorporation of nev-tRNAs into ribosomes might be a rare and limited event compared with the incorporation of their cognate tRNAs. In addition to the low abundance of nev-tRNAs,

we noted the low usage of the codons with which nev-tRNAs are associated. For instance, the GGG codon to which nev-tRNA^{Gly} (CCC) corresponds is the second rarest codon (0.44%) in *C. elegans* (see section 2.3.6). Therefore, we assume that even if nev-tRNAs participate in translation, the identification of amino acid misincorporations at the GGG codon is statistically more difficult than at other more frequent codons. This hypothesis is supported by the observation of more abundant misincorporations at the more frequent codons (Figure 3. 4A). Collectively, these data demonstrate that there is no mutant protein containing misincorporated Leu at "high" frequency in the whole-cell proteome, whereas it is still unknown whether such Leu residues are misincorporated into low-abundance proteins and/or some specific sites in proteins at low frequency.

To determine whether nev-tRNA-induced mistranslations can occur at low frequencies, an overexpressed single recombinant protein was analyzed with targeted proteomics. In this experiment, we overexpressed a green fluorescent protein (GFP)–LacZ protein and purify it to improve the detectable level of Gly-to-Leu misincorporation, because *i*) the total 1284 codons of the *GFP–LacZ* mRNA contain 12 GGG codons (approximately 1% of the codons); and *ii*) the purified samples for MS include a small number of proteins, mainly GFP–LacZ, resulting in low background noise. For this analysis, we constructed a transgenic strain expressing *myo-3p::GFP-LacZ* and extracted the protein mixture. After immunoprecipitation with an anti-GFP antibody, the purified GFP–LacZ protein was fragmented into small peptides by digestion with site-specific enzymes. The LC–MS/MS analysis was performed using two types of ISs for calibration, a synthetic peptide consisting of the

same amino acid sequence as that in the database, and a synthetic peptide containing the Leu residue substituted for the Gly residue at the GGG codon (summarized in Table 3. 4). As shown in Table 3. 6, wild-type peptides containing the Gly residue at the GGG codon were detected at almost identical elution times as the ISs. In contrast, no aberrant peptide containing a misincorporated Leu residue at the GGG codon was detected. The fragmentation pattern in the mass spectrum of the identified peptide was consistent with that of the wild-type peptide rather than the aberrant peptide. One example is shown in Figure 3. 5. This result means that the Gly-to-Leu mutant peptides were not represented, even in the high-resolution targeted MS screen, suggesting that nev-tRNA^{Gly} (CCC) is not incorporated into ribosomes at a detectable level.

Table 3. 6. Summary of the identified peptides from transgenic worms expressing GFP–LacZ.

Type	Sequence	Observed *m/z*	Mascot Score	Retention time (min) [a]		
				Candidate	Internal Standard	Δ time
GGG-to-Gly	VNWLGLGPQENYPDR	879.4347	46.83	61.3	61.1	0.2
	SAGQLWLTVR	565.8201	45.41	54.9	54.8	0.1
	AGHISAWQQWR	447.2277	21.26	43.7	43.7	0.0
	RVNWLGLGPQE	634.8417	49.94	58.9	58.8	0.1
GGG-to-Leu	Not identified					

The peptide sequences containing amino acid residues (red) arising from the decoding of the GGG codon are listed.
[a] Chromatographic retention times of the targeted peptides are also shown.

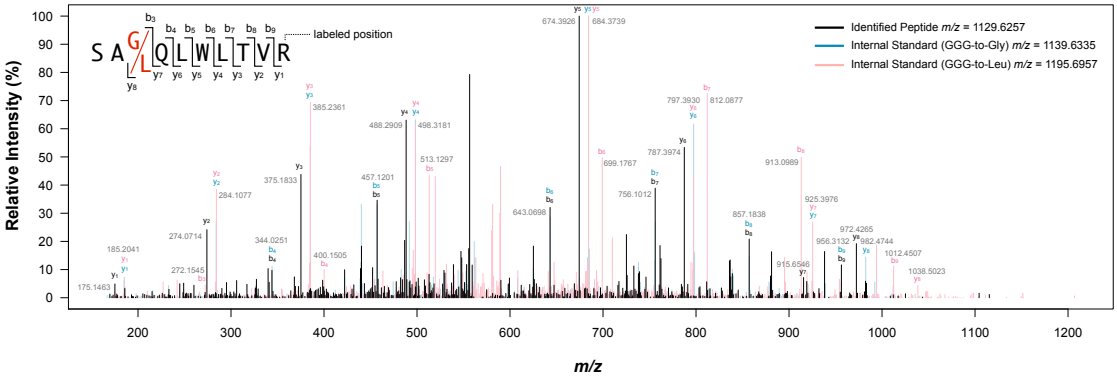

Figure 3. 5. Fragmentation pattern in the mass spectrum of the identified peptide from *C. elegans* is inconsistent with those of the peptides containing misincorporated Leu.

MS/MS spectrum of the identified peptide from the GFP-LacZ proteins expressed in *C. elegans* was compared with those of synthesized peptides (IS) with the sequence SA(G/L)QLWLTVR. The C-terminal arginine residue of each IS was labeled with a stable isotope ($\Delta m/z$ = 10.008269).

3.3.4. Evolutionary implications of nev-tRNAs for the nematode genetic code

In this chapter, we have demonstrated that nev-tRNAs are weakly expressed, mature normally, with the addition of the 3′ CCA, and are exported from the nucleus in *C. elegans*. However, no nev-tRNA-induced amino acid misincorporation was detected in the whole-cell proteome, and the nematode genetic code does not seem to be ambiguous. If this is the case, why does the *C. elegans* genome contain these deviant tRNAs, which could decode an alternative code? The possible explanation is that the nev-tRNAs appeared in the nematode genome as the result of 'neutral evolution'. Because their expression levels and the codon usage with which they are associated are very low, as described above, the misacylation of nev-tRNAs might have no apparent effect on cellular homeostasis. Nematode cells might also actively regulate errors in protein synthesis with specific translational quality control mechanisms.

To discuss the 'neutral evolution' hypothesis of the nev-tRNAs, we comprehensively reanalyzed the nematode genomes, and examined the evolutionary conservation and possible timing of the gains in nev-tRNAs. We obtained a total of 26 nematode genomic sequences, including four major lineages: Dorylaimia (clade I), Spirurina (clade III), Tylenchina (clade IV), and Rhabditina (clades Va, Vb, and Vc). We then identified 195 nev-tRNA genes from these nematode genomes, using the criteria described in section 2.3.1. When we assessed the distribution of the predicted nev-tRNA genes among the 26 nematode species (Figure 3. 6), we found that the copy numbers of nev-tRNAs and their anticodon variations increased rapidly during the evolution of the nematode taxon, and especially markedly in the clade Vc nematodes. For example,

nev-tRNA^{Ile} (UAU) occurred most broadly, across four nematode clades (clades I, IV, Va, and Vc). In addition to nev-tRNA^{Ile} (UAU), several isotypes of nev-tRNAs (Gly (CCC), Arg (CCU), Pro (GGG), Tyr (GUA), Val (UAC), and Val (CAC)) also tended to be enriched in the clade IV, Va, and Vc nematode genomes.

We next performed a network analysis based on the sequence similarities of the nev-tRNAs to assess how several isotypes of the nev-tRNAs emerged and evolved during nev-tRNA evolution in the nematode taxon (Figure 3. 7). A total of 195 nematode nev-tRNA sequences were classified into two clusters (groups a and b). Group a contained more than three quarters of the nev-tRNAs (149 nev-tRNAs). In particular, nev-tRNA^{Ile} (UAU) was located in the center of the network graph (shown in pink in Figure 3. 7A). Four types of nev-tRNAs (Gly (CCC), Arg (CCU), Arg (UCU), and Val (UAC)) also clustered into the same group, and were distributed throughout the surrounding areas of the Ile (UAU) network. They are assumed to have evolved from the nev-tRNA^{Ile} (UAU) family. On the other hand, group b contained other minor types of nev-tRNAs (e.g., Pro (GGG), Val (CAC), Gly (GCC), Pro (CGG), and Trp (CCA)). These results suggest that the nev-tRNAs emerged from distinct origins on at least two independent occasions during nematode tRNA evolution. This inference is supported by the results of the phylogenetic analysis of nematode tRNAs, described in section 2.3.2. We also noted that a single *T. suis* nev-tRNA^{Ile} (UAU) and *C. brenneri* nev-tRNA^{Ile} (UAU) did not cluster within group a, suggesting that they originated from other tRNA clades.

Finally, we focused on the isotype that emerged most broadly, nev-tRNA^{Ile}

(UAU), and have discussed the possible timing of its gain and the evolutionary implications for the nematode genetic code (Figure 3. 8). Within the nematode taxon, four nematode species (*T. suis*, *M. incognita*, *M. hapla*, and *C. angaria*) have a single copy of the nev-tRNA^{Ile} (UAU) gene, whereas the Rhabditina–Tylenchina clade species (*P. redivivus*, clade Va nematodes, and clade Vc nematodes, except *C. angaria*) have multiple copies of the nev-tRNA^{Ile} (UAU) gene (see also Figure 3. 6). Here, several important trends can be seen. First, nev-tRNA^{Ile} (UAU) seems to be duplicated specifically in free-living (nonparasitic) nematodes. In contrast, no duplication of the nev-tRNA^{Ile} (UAU) gene has occurred in the parasitic nematode genomes (*T. suis*, *M. incognita*, and *M. hapla*). Second, all such duplicated nev-tRNAs have the crucial determinants for leucylation. In eukaryotes, one nucleotide downstream from the anticodon G37 and the discriminator base A73 are necessary for recognition by LeuRS (see also section 2.3.4). These are strongly conserved in nev-tRNA^{Ile} (UAU) of the free-living nematodes, so they seem to have acquired their leucylation properties after the free-living nematodes diverged from the other nematodes. Based on these evolutionary features, we propose the following possible scenario for nev-tRNA^{Ile} (UAU) evolution in the nematode taxon: *i*) after the Rhabditina–Tylenchina clade diverged from the other nematodes, the major nev-tRNA^{Ile} (UAU) gene emerged from another tRNA family (shown with red arrowheads in Figure 3. 8); *ii*) a proportion of these genes was lost during evolutionary processes such as random genetic drift and natural selection (shown with blue arrowheads in Figure 3. 8); *iii*) in the free-living nematodes, they specifically evolved into deviant tRNAs with misacylation properties,

and duplicated rapidly in these genomes. These crucial tRNA mutations seem to have been acceptable in free-living nematodes, although they could potentially alter the general translation rule. It is noteworthy that the molecular divergence between the least closely related *Caenorhabditis* species (*C. briggsae* and *Caenorhabditis* sp. 5) is comparable to that between zebrafish and mouse (Kiontke and Fitch 2005). This heterogeneity supposes that the nematodes diverged from other animals in very ancient times, or that the molecular evolutionary rates are much higher in the nematodes than in the deuterostomes, or both. Therefore, the strong evolutionary conservation of nev-tRNAs among multiple nematode species indicates that, in the former scenario, they have been conserved in the long term, without loss during evolutionary processes. It is also noteworthy that in the latter scenario, they did not become pseudogenes under the influence of the high rate of molecular evolution. Based on these evolutionary implications, we assume that nev-tRNAs have been selected under positive evolutionary pressure during nematode evolution and that they confer a survival advantage on free-living nematodes.

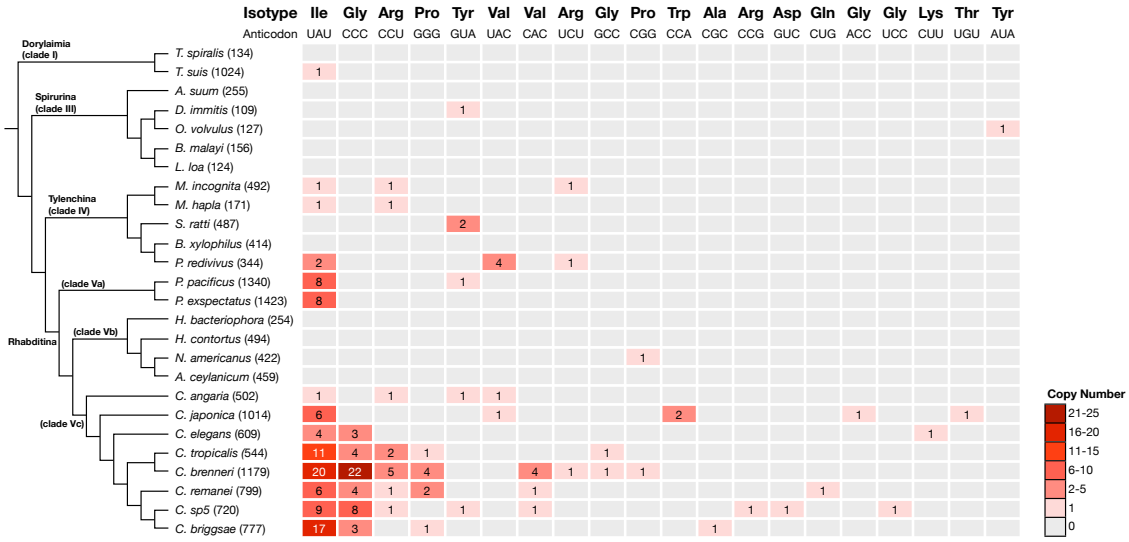

Figure 3. 6. Phylogenic distribution of nev-tRNAs among 26 nematode species.

The copy numbers of nev-tRNA genes corresponding to each specific anticodon are shown for each nematode species. The total number of tRNA genes encoded in each nematode genome is indicated in brackets. The phylogenetic dendrogram was generated in WormBase, as described.

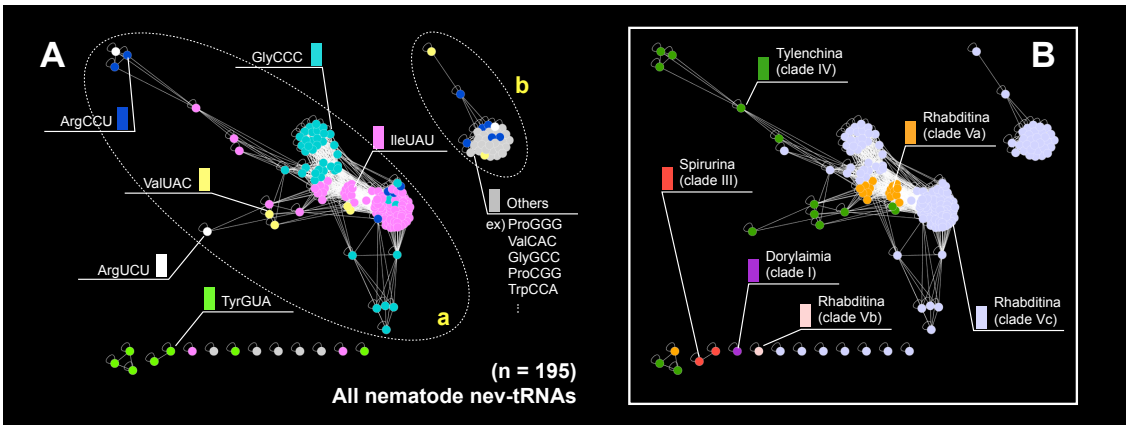

Figure 3. 7. Sequence similarity network of the nematode nev-tRNAs.

Two ways of coloring the same network of 195 nematode nev-tRNAs are shown. Symbols represent each nev-tRNA, and the edge lengths represent the sequence similarities. The sequences are classified into two clusters (a–b) with a threshold at *E* value < 1e–7. (A) Network colored by tRNA anticodons. (B) Network colored by the phylogenetic clades of the nematodes.

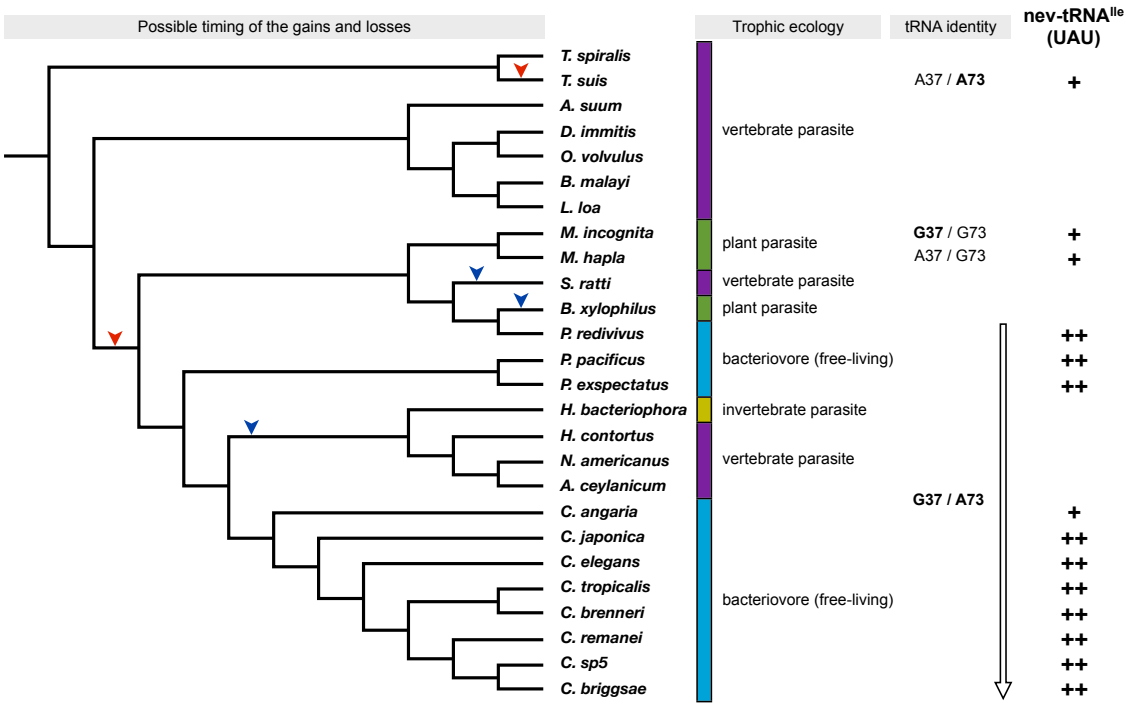

Figure 3. 8. Possible timing of the gains of nev-tRNA^{Ile} (UAU) and its evolutionary implications.

The conservation patterns of the nev-tRNA^{Ile} (UAU) genes among nematode species are illustrated with the symbols '+' (single copy) and '++' (multiple copies). Arrowheads show the possible evolutionary times of the gain (red) and loss (blue) of nev-tRNA^{Ile} (UAU). Major determinants of the recognition of class II tRNAs, base 37 (one nucleotide downstream from the anticodon) and base 73 (the fourth and unpaired nucleotide from the tRNA 3′ end), for each nev-tRNA are indicated. The trophic ecologies for all species are shown (purple, vertebrate parasite; green, plant parasite; light blue, bacteriovore; yellow, invertebrate parasite).

Chapter 4

Concluding remarks

The faithful translation of the genetic code requires the highly accurate aminoacylation of transfer RNAs (tRNAs). Changes in tRNAs sometimes cause nonstandard codon assignments. Therefore, we regard the evolutionary biology of the tRNA gene as an essential part of any comprehensive understanding of the genetic code. In this thesis, we analyzed and characterized the tRNA genes from 44 eukaryotic genomes, and discovered a novel class of tRNA (designated "nev-tRNA") in nematode species. Although the structural characteristics of nev-tRNAs are quite similar to those of common leucine tRNAs, the major nev-tRNAs have glycine and isoleucine anticodons. A series of *in vitro* aminoacylation assays confirmed that nev-tRNAs are only charged with leucine, which is inconsistent with their anticodons. An *in vitro* translation analysis showed that nev-tRNA^{Gly} decodes the GGG codon as leucine instead of glycine, indicating that nev-tRNAs can decode an alternative code *in vitro* (Chapter 2). Because the nematode genome also encodes common tRNAs that decode the universal code, we next assumed that nev-tRNAs cause 'genetic code ambiguity' in nematode cells. To test this hypothesis, we investigated the functionality of nev-tRNAs

and their impact on the proteome of *Caenorhabditis elegans*. Analysis of their expression, maturation, and subcellular localization demonstrated that nev-tRNAs are processed to their mature forms like common tRNAs and are available for translation *in vivo*. However, a whole-cell proteomic analysis found no detectable level of nev-tRNA-induced mistranslation, suggesting that the genetic code is not ambiguous, at least under normal growth conditions (Chapter 3). These findings indicate that the translational fidelity of the nematode genetic code is strictly maintained, contrary to our expectation, and we hypothesize that nev-tRNAs emerged in the nematode genomes as a result of neutral evolution rather than natural selection.

In contrast, pseudo-tRNA genes typically have several mismatched base pairings because of the high evolutionary rate (Marck and Grosjean 2002; Abe *et al*. 2011), but nev-tRNA genes do not contain such mutations and form a perfect cloverleaf secondary structure. Besides, the copy numbers of nev-tRNA genes and their anticodon variants have increased during the evolution of the nematode taxon, especially in the Rhabditina clade. From this feature of their evolutionary conservation, we also assume that they play important, if unexpected, roles, especially in certain biological processes. One such possible role is in the protective stress response. In bacterial, yeast, and mammalian cells, the level of Met-misacylation increases during the immune response (see also Figure 1. 2D). Because Met residues protect proteins from reactive oxygen species (ROS)-mediated damage (Levine *et al*. 1996), increased numbers of Met residues in proteins constitute a response mechanism, protecting cells against oxidative stress (Pan 2013). In addition to this pathway, recent studies have reported other

putative benefits of mistranslation under stress conditions. In *Saccharomyces cerevisiae* cells, tRNA-misacylation-dependent translation errors increase the ubiquitylation and aggregation of proteins, and enhance the expression of heat shock proteins and other stress proteins. Consequently, the cells can survive even lethal environmental conditions (Santos *et al*. 1999; Moura *et al*. 2009, 2010). Although nev-tRNAs are weakly expressed under normal growth conditions, their expression may be enhanced under some stress conditions, causing the synthesis of mistranslated proteins and the upregulation of the stress response to better cope with stress. For example, because some major nev-tRNA genes were duplicated and evolved specifically in free-living nematodes, they may be involved in the dauer stage, which is an alternative developmental stage of free-living worms that allows larvae to survive under harsh conditions for extended periods of time (Riddle *et al*. 1981).

Another possible role of nev-tRNAs is in the gain of novel protein functions through the production of mutant proteins. Although most mistranslated proteins will probably be deleterious or neutral in function, a minority of these proteins will acquire novel or altered functions arising from their chemical and/or structural changes, including new subcellular localization (Dunn *et al*. 2013), antibiotic resistance (Javid *et al*. 2014), or phenotypic diversification (Bezerra *et al*. 2013). Although these data suggest that whole nematode cells do not synthesize mutant proteins using nev-tRNAs, it is still possible that some cells or tissues do synthesize such novel functional mistranslated proteins. For instance, there are cell-specific physiological differences in the translational error rate in mice (Lee *et al*. 2006). Further studies are required to

clarify the extensive expression patterns of nev-tRNAs under various environmental conditions and in different cells and tissues, and to identify the cellular response during the induction of genetic code ambiguity by nev-tRNAs.

Recently, a growing number of studies have revealed a new series of tRNA functions, ranging from the regulation of translation to transcription and DNA replication. The existence of numerous tRNA-derived fragments has also offered a new perspective on tRNAs as regulatory RNAs. These findings highlight the need for further research to identify the complete population of these RNAs, which have been misinterpreted as mere pseudogenes or degradation products, and to characterize their regulatory mechanisms in detail (Chapter 1). We consider that the discovery of this novel class of tRNA genes, presented in this thesis, will contribute to the characterization of the role of tRNA as a key molecule in the processes of the central dogma, and will also provide new insight into the genetic code and extend our understanding of tRNA biology.

Acknowledgements

I would like to thank Professor Masaru Tomita, who has provided me with this great research environment at the Institute for Advanced Biosciences (IAB), Keio University. I would also like to thank Professor Akio Kanai, who is the best educator I have ever known, and has always helped me advancing my work. I would not be where I am today without him. My gratitude also goes to the members of my thesis committee, Professor Mitsuhiro Itaya and Associate Professor Kazuharu Arakawa, for their valuable comments and suggestions on this dissertation. I also thank Professor Tomoyoshi Soga, Professor Mitsuhiro Watanabe, Associate Professor Yasuhiro Naito, Associate Professor Hiroki Kuroda, Assistant Professor Hitomi Sano-Ito, and Dr. Rintaro Saito.

I extend my sincere thanks to my former and present research collaborators, Professor Yuji Kohara and Assistant Professor Yoshiki Andachi at the National Institute for Genetics, Dr. Takeshi Masuda at Harvard Medical School, and Dr. Masaru Mori at IAB. I also thank all the administrative staff at IAB and Taihei Building Service Inc., particularly Ayumi Mikami, Akiko Shiozawa, Maki Oike, Ayami Mizukami, Miwa Hiramoto, Hinako Hayashi, Satomi Yokoi, Miho Sato, Katsunori Komatsu, Nozomi Nakamura, and the staff of the Academic Affairs Office at Keio University Shonan Fujisawa Campus, for their kind help.

I would like to thank my research advisers, Dr. Kosuke Fujishima at the NASA Ames Research Center and Dr. Junichi Sugahara at Spiber Inc., and my research advisees Asaki Kobayashi and Yohei Chiwata. I am really lucky to have met such talented research partners. I am greatly indebted to all the members of the RNA project: Asako Soga-Sato, Kiriko Hiraoka, Kanako Takesue, Emiko Noro, Hiromi Shinoda-Kochiwa, Atsuko Fujishima-Kishi, Hyunchul Kim, Yuka Iwasaki-Watanabe, Motomu Matsui, Yoshiki Ikeda, Kahori Ikeda-Takane, Atsuko Ikeda-Shinhara, Kaoru Sugahara-Kikuta, Nobuto Saito, Hiromi Toyoshima, Hiromi Shinnabe, Shinnosuke Murakami, Keisuke Morita, Yuka Hirose, Junnosuke Imai, Yuki Usui, Gakuto Makino, Shohei Nagata, Atsuki Kawai, Momoko Hoshikawa, and Hikari Ikarashi.

I am grateful to my both past and present colleagues at IAB for their warm friendship and care, Dr. Shinichi Kikuchi, Dr. Yuri Matsuzaki, Noriyuki Kitagawa, Dr. Vincent Piras, Dr. Nobuaki Kono, Taro Ichinose, Dr. Haruna Kaneko-Imamura, Keita Ikegami, Tadasu Nozaki, Keiko Nozaki-Iino, Takayuki Ebi, Megumi Uetaki, Mana Nakagawa, Fujitaka Baba, Masaki Watanabe, Norikazu Saiki, Yuki Shindo, Nana Sugano, Yutaro Ito and Kyoko Ishino. They have made my life wonderful and passionate. I also give special thanks to the Power Surf fellows, in particular Chikara Niidate, Go Sato, Takashi Honma, and Yuji Saito, who have contributed to my personal development through priceless experiences.

Finally, I am enormously grateful to my parents Kazuo Hamashima and Ayako Hamashima, to my brothers and sister Yuki Hamashima, Mikio Hamashima, and Sayaka Hamashima, and to my dog Happy Hamashima, for supporting my PhD studies at Keio University. I will surely return the favor with my success in the future.

References

Abe T, Ikemura T, Sugahara J, Kanai A, Ohara Y, Uehara H, Kinouchi M, Kanaya S, Yamada Y, Muto A, Inokuchi H. 2011. tRNADB-CE 2011: tRNA gene database curated manually by experts. *Nucleic Acids Res* **39:** D210–213.

Achsel T, Gross HJ. 1993. Identity determinants of human tRNASer: sequence elements necessary for serylation and maturation of tRNA with a long extra arm. *EMBO J* **12:** 3333–3338.

Alfonzo JD, Blanc V, Estévez AM, Rubio MA, Simpson L. 1999. C to U editing of the anticodon of imported mitochondrial tRNA(Trp) allows decoding of the UGA stop codon in Leishmania tarentolae. *EMBO J* **18:** 7056–7062.

Asahara H, Himeno H, Tamura K, Hasegawa T, Watanabe K, Shimizu M. 1993. Recognition nucleotides of Escherichia coli tRNA(Leu) and its elements facilitating discrimination from tRNASer and tRNA(Tyr). *J Mol Biol* **231:** 219–229.

Ataide SF, Rogers TE, Ibba M. 2009. The CCA anticodon specifies separate functions inside and outside translation in Bacillus cereus. *RNA Biol* **6:** 479–487.

Banerjee R, Chen S, Dare K, Gilreath M, Praetorius-Ibba M, Raina M, Reynolds NM, Rogers T, Roy H, Yadavalli SS, Ibba M. 2010. tRNAs: cellular barcodes for amino acids. *FEBS Lett* **584:** 387–395.

Barrell BG, Bankier AT, Drouin J. 1979. A different genetic code in human mitochondria. *Nature* **282:** 189–194.

Bezerra AR, Simões J, Lee W, Rung J, Weil T, Gut IG, Gut M, Bayés M, Rizzetto L, Cavalieri D, Giovannini G, Bozza S, Romani L, Kapushesky M, Moura GR, Santos MAS. 2013. Reversion of a fungal genetic code alteration links proteome instability with genomic and phenotypic diversification. *Proc Natl Acad Sci U S A* **110:** 11079–11084.

Biou V, Yaremchuk A, Tukalo M, Cusack S. 1994. The 2.9 A crystal structure of T. thermophilus seryl-tRNA synthetase complexed with tRNA(Ser). *Science* **263:** 1404–1410.

Breitschopf K, Achsel T, Busch K, Gross HJ. 1995. Identity elements of human tRNA(Leu): structural requirements for converting human tRNA(Ser) into a leucine acceptor in vitro. *Nucleic Acids Res* **23:** 3633–3637.

Brenner S. 1974. The genetics of Caenorhabditis elegans. *Genetics* **77:** 71–94.

Chan PP, Cozen AE, Lowe TM. 2011. Discovery of permuted and recently split transfer RNAs in Archaea. *Genome Biol* **12:** R38.

Chan PP, Lowe TM. 2009. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res* **37:** D93–97.

Cochella L, Green R. 2005. Fidelity in protein synthesis. *Curr Biol* **15:** R536–540.

Couvillion MT, Bounova G, Purdom E, Speed TP, Collins K. 2012. A Tetrahymena Piwi Bound to Mature tRNA 3' Fragments Activates the Exonuclease Xrn2 for RNA Processing in the Nucleus. *Mol Cell* **48:** 509–520.

Couvillion MT, Sachidanandam R, Collins K. 2010. A growth-essential Tetrahymena Piwi protein carries tRNA fragment cargo. *Genes Dev* **24:** 2742–2747.

Creasy DM, Cottrell JS. 2002. Error tolerant searching of uninterpreted tandem mass spectrometry data. *Proteomics* **2:** 1426–1434.

Crick FH. 1968. The origin of the genetic code. *J Mol Biol* **38:** 367–379.

de Pouplana LR, Santos MAS, Zhu JH, Farabaugh PJ, Javid B. 2014. Protein mistranslation: friend or foe? *Trends Biochem Sci* **39:** 355–362.

Drummond DA, Wilke CO. 2009. The evolutionary consequences of erroneous protein synthesis. *Nat Rev Genet* **10:** 715–724.

Dunn JG, Foo CK, Belletier NG, Gavis ER, Weissman JS. 2013. Ribosome profiling reveals pervasive and regulated stop codon readthrough in Drosophila melanogaster. *Elife* **2:** e01179.

Ebersole T, Kim JH, Samoshkin A, Kouprina N, Pavlicek A, White RJ, Larionov V. 2011. tRNA genes protect a reporter gene from epigenetic silencing in mouse cells. *Cell Cycle* **10:** 2779–2791.

El Yacoubi B, Bailly M, de Crécy-Lagard V. 2012. Biosynthesis and Function of Posttranscriptional Modifications of Transfer RNAs. *Annu Rev Genet* **46:** 69–95.

Fahlman RP, Dale T, Uhlenbeck OC. 2004. Uniform binding of aminoacylated transfer RNAs to the ribosomal A and P sites. *Mol Cell* **16:** 799–805.

Freist W. 1989. Mechanisms of aminoacyl-tRNA synthetases: a critical consideration of recent results. *Biochemistry* **28:** 6787–6795.

Fu H, Feng J, Liu Q, Sun F, Tie Y, Zhu J, Xing R, Sun Z, Zheng X. 2009. Stress induces tRNA cleavage by angiogenin in mammalian cells. *FEBS Lett* **583:** 437–442.

Fujishima K, Sugahara J, Kikuta K, Hirano R, Sato A, Tomita M, Kanai A. 2009. Tri-split tRNA is a transfer RNA made from 3 transcripts that provides insight into the evolution of fragmented tRNAs in archaea. *Proc Natl Acad Sci U S A* **106:** 2683–2687.

Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M, Barber GP, Clawson H, Coelho A, Diekhans M, Dreszer TR, Giardine BM, Harte RA, Hillman-Jackson J, Hsu F, Kirkup V, Kuhn RM, Learned K, Li CH, Meyer LR, Pohl A, Raney BJ, Rosenbloom KR, Smith KE, Haussler D, Kent WJ. 2010. The UCSC Genome Browser database: update 2011. *Nucleic Acids Res* **39:** D876–882.

Fukunaga R, Yokoyama S. 2005. Aminoacylation complex structures of leucyl-tRNA synthetase and tRNALeu reveal two modes of discriminator-base recognition. *Nat Struct Mol Biol* **12:** 915–922.

Galli G, Hofstetter H, Birnstiel M. 1981. Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. *Nature* **294:** 626–631.

Giegé R, Sissler M, Florentz C. 1998. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res* **26:** 5017–5035.

Hamashima K, Fujishima K, Masuda T, Sugahara J, Tomita M, Kanai A. 2012. Nematode-specific tRNAs that decode an alternative genetic code for leucine. *Nucleic Acids Res* **40:** 3653–3662.

Hamashima K, Kanai A. 2013. Alternative genetic code for amino acids and transfer RNA revisited. *Biomol Concepts* **4:** 309–318.

Hamashima K, Mori M, Andachi Y, Tomita M, Kohara Y, Kanai A. 2015. Analysis of Genetic Code Ambiguity Arising from Nematode-Specific Misacylated tRNAs. *PLoS One* **10:** e0116981.

Hao B, Gong W, Ferguson TK, James CM, Krzycki JA, Chan MK. 2002. A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* **296:** 1462–1466.

Harris RP, Kilby PM. 2014. Amino acid misincorporation in recombinant biopharmaceutical products. *Curr Opin Biotechnol* **30:** 45–50.

Henkin TM. 2008. Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes Dev* **22:** 3383–3390.

Hinnebusch AG. 2005. Translational regulation of GCN4 and the general amino acid control of yeast. *Annu Rev Microbiol* **59:** 407–450.

Iben JR, Maraia RJ. 2012. tRNAomics: tRNA gene copy number variation and codon use provide bioinformatic evidence of a new anticodon:codon wobble pair in a eukaryote. *RNA* **18:** 1358–1372.

Ishihama Y, Rappsilber J, Andersen JS, Mann M. 2002. Microcolumns with self-assembled particle frits for proteomics. *J Chromatogr A* **979:** 233–239.

Ishimura R, Nagy G, Dotu I, Zhou H, Yang XL, Schimmel P, Senju S, Nishimura Y, Chuang JH, Ackerman SL. 2014. Ribosome stalling induced by mutation of a CNS-specific tRNA causes neurodegeneration. *Science* **345:** 455–459.

Ivanov P, Emara MM, Villen J, Gygi SP, Anderson P. 2011. Angiogenin-Induced tRNA Fragments Inhibit Translation Initiation. *Mol Cell* **43:** 613–623.

Javid B, Sorrentino F, Toosky M, Zheng W, Pinkham JT, Jain N, Pan M, Deighan P, Rubin EJ. 2014. Mycobacterial mistranslation is necessary and sufficient for rifampicin phenotypic resistance. *Proc Natl Acad Sci U S A* **111:** 1132–1137.

Jones TE, Alexander RW, Pan T. 2011. Misacylation of specific nonmethionyl tRNAs by a bacterial methionyl-tRNA synthetase. *Proc Natl Acad Sci U S A* **108:** 6933–6938.

Kato M, Chen X, Inukai S, Zhao H, Slack FJ. 2011. Age-associated changes in expression of small, noncoding RNAs, including microRNAs, in C. elegans. *RNA* **17:** 1804–1820.

Keeling PJ, Doolittle WF. 1997. Widespread and ancient distribution of a noncanonical genetic code in diplomonads. *Mol Biol Evol* **14:** 895–901.

Keeling PJ, Leander BS. 2003. Characterisation of a non-canonical genetic code in the oxymonad Streblomastix strix. *J Mol Biol* **326:** 1337–1349.

Kiontke K, Fitch DH. 2005. The phylogenetic relationships of Caenorhabditis and other rhabditids. *WormBook* **11:** 1–11.

Knight RD, Freeland SJ, Landweber LF. 2001. Rewiring the keyboard: evolvability of the genetic code. *Nat Rev Genet* **2:** 49–58.

Köhrer C, Rajbhandary UL. 2008. The many applications of acid urea polyacrylamide gel electrophoresis to studies of tRNAs and aminoacyl-tRNA synthetases. *Methods* **44:** 129–138.

LaRiviere FJ, Wolfson AD, Uhlenbeck OC. 2001. Uniform binding of aminoacyl-tRNAs to elongation factor Tu by thermodynamic compensation. *Science* **294:** 165–168.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23:** 2947–2948.

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

Lee JW, Beebe K, Nangle LA, Jang J, Longo-Guess CM, Cook SA, Davisson MT, Sundberg JP, Schimmel P, Ackerman SL. 2006. Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. *Nature* **443:** 50–55.

Lee SR, Collins K. 2005. Starvation-induced cleavage of the tRNA anticodon loop in Tetrahymena thermophila. *J Biol Chem* **280:** 42744–42749.

Lee YS, Shibata Y, Malhotra A, Dutta A. 2009. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev* **23:** 2639–2649.

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

Levine RL, Mosoni L, Berlett BS, Stadtman ER. 1996. Methionine residues as endogenous antioxidants in proteins. *Proc Natl Acad Sci U S A* **93:** 15036–15040.

Levitz R, Chapman D, Amitsur M, Green R, Snyder L, Kaufmann G. 1990. The optional E. coli prr locus encodes a latent form of phage T4-induced anticodon nuclease. *EMBO J* **9:** 1383–1389.

Lovett PS, Ambulos NP, Mulbry W, Noguchi N, Rogers EJ. 1991. UGA can be decoded as tryptophan at low efficiency in Bacillus subtilis. *J Bacteriol* **173:** 1810–1812.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25:** 955–964.

Lozupone CA, Knight RD, Landweber LF. 2001. The molecular basis of nuclear genetic code change in ciliates. *Curr Biol* **11:** 65–74.

Marck C, 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.

Maruyama S, Sugahara J, Kanai A, Nozaki H. 2009. Permuted tRNA genes in the nuclear and nucleomorph genomes of photosynthetic eukaryotes. *Mol Biol Evol* **27:** 1070–1076.

Masaki H, Ogawa T. 2002. The modes of action of colicins E5 and D, and related cytotoxic tRNases. *Biochimie* **84:** 433–438.

Masuda T, Saito N, Tomita M, Ishihama Y. 2009. Unbiased quantitation of Escherichia coli membrane proteome using phase transfer surfactants. *Mol Cell Proteomics* **8:** 2770–2777.

Masuda T, Tomita M, Ishihama Y. 2008. Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *J Proteome Res* **7:** 731–740.

Matsui M, Tomita M, Kanai A. 2013. Comprehensive computational analysis of bacterial CRP/FNR superfamily and its target motifs reveals stepwise evolution of transcriptional networks. *Genome Biol Evol* **5:** 267–282.

McClain WH. 1993. Rules that govern tRNA identity in protein synthesis. *J Mol Biol* **234:** 257–280.

McCutcheon JP, McDonald BR, Moran NA. 2009. Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genet* **5:** e1000565.

McFarlane RJ, Whitehall SK. 2009. tRNA genes in eukaryotic genome organization and reorganization. *Cell Cycle* **8:** 3102–3106.

Mitani S. 1995. Genetic regulation of mec-3 gene expression implicated in the specification of the mechanosensory neuron cell types in Caenorhabditis elegans. *Dev Growth Differ* **37:** 551–557.

Moura GR, Carreto LC, Santos MAS. 2009. Genetic code ambiguity: an unexpected source of proteome innovation and phenotypic diversity. *Curr Opin Microbiol* **12:** 631–637.

Moura GR, Paredes JA, Santos MAS. 2010. Development of the genetic code: insights from a fungal codon reassignment. *FEBS Lett* **584:** 334–341.

Murakami S, Fujishima K, Tomita M, Kanai A. 2011. Metatranscriptomic analysis of microbes in an ocean-front deep subsurface hot spring reveals novel small RNAs and type-specific tRNA degradation. *Appl Env Microbiol* **78:** 1015–1022.

Netzer N, Goodenbour JM, David A, Dittmar KA, Jones RB, Schneider JR, Boone D, Eves EM, Rosner MR, Gibbs JS, Embry A, Dolan B, Das S, Hickman HD, Berglund P, Bennink JR, Yewdell JW, Pan T. 2009. Innate immune and chemically triggered oxidative stress modifies translational fidelity. *Nature* **462:** 522–526.

Olsen JV, Ong SE, Mann M. 2004. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics* **3:** 608–614.

Osawa S, Jukes TH, Watanabe K, Muto A. 1992. Recent evidence for evolution of the genetic code. *Microbiol Rev* **56:** 229–264.

Pan T. 2013. Adaptive translation as a mechanism of stress response and adaptation. *Annu Rev Genet* **47:** 121–137.

Peng H, Shi J, Zhang Y, Zhang H, Liao S, Li W, Lei L, Han C, Ning L, Cao Y, Zhou Q, Chen Q, Duan E. 2012. A novel class of tRNA-derived small RNAs extremely enriched in mature mouse sperm. *Cell Res* **22:** 1609–1612.

Peschel A, Jack RW, Otto M, Collins LV, Staubitz P, Nicholson G, Kalbacher H, Nieuwenhuizen WF, Jung G, Tarkowski A, van Kessel KP, van Strijp JA. 2001. Staphylococcus aureus resistance to human defensins and evasion of neutrophil killing via the novel virulence factor MprF is based on modification of membrane lipids with l-lysine. *J Exp Med* **193:** 1067–1076.

Raab JR, Chiu J, Zhu J, Katzman S, Kurukuti S, Wade PA, Haussler D, Kamakaka RT. 2012. Human tRNA genes function as chromatin insulators. *EMBO J* **31:** 330–350.

Rappsilber J, Ishihama Y, Mann M. 2003. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **75:** 663–670.

Reynolds NM, Lazazzera BA, Ibba M. 2010. Cellular mechanisms that control mistranslation. *Nat Rev Microbiol* **8:** 849–856.

Riddle DL, Swanson MM, Albert PS. 1981. Interacting genes in nematode dauer larva formation. *Nature* **290:** 668–671.

Rogers TE, Ataide SF, Dare K, Katz A, Seveau S, Roy H, Ibba M. 2012. A pseudo-tRNA modulates antibiotic resistance in Bacillus cereus. *PLoS One* **7:** e41248.

Rudinger-Thirion J, Lescure A, Paulus C, Frugier M. 2011. Misfolded human tRNA isodecoder binds and neutralizes a 3' UTR-embedded Alu element. *Proc Natl Acad Sci U S A* **108:** E794–802.

Saadatmand J, Kleiman L. 2012. Aspects of HIV-1 assembly that promote primer tRNA(Lys3) annealing to viral RNA. *Virus Res* **169:** 340–348.

Saikia M, Krokowski D, Guan BJ, Ivanov P, Parisien M, Hu GF, Anderson P, Pan T, Hatzoglou M. 2012. Genome-wide identification and quantitative analysis of cleaved tRNA fragments induced by cellular stress. *J Biol Chem* **287:** 42708–42725.

Saks ME, Sampson JR, Abelson JN. 1994. The transfer RNA identity problem: a search for rules. *Science* **263:** 191–197.

Sánchez-Silva R, Villalobo E, Morin L, Torres A. 2003. A new noncanonical nuclear genetic code: translation of UAA into glutamate. *Curr Biol* **13:** 442–447.

Santos MA, Cheesman C, Costa V, Moradas-Ferreira P, Tuite MF. 1999. Selective advantages created by codon ambiguity allowed for the evolution of an alternative genetic code in Candida spp. *Mol Microbiol* **31:** 937–947.

Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9:** 671–675.

Schneider SU, de Groot EJ. 1991. Sequences of two rbcS cDNA clones of Batophora oerstedii: structural and evolutionary considerations. *Curr Genet* **20:** 173–175.

Shapiro R, Vallee BL. 1987. Human placental ribonuclease inhibitor abolishes both angiogenic and ribonucleolytic activities of angiogenin. *Proc Natl Acad Sci U S A* **84:** 2238–2241.

Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. 2011. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27:** 431–432.

Söll D. 1988. Genetic code: enter a new amino acid. *Nature* **331:** 662–663.

Soma A, Kumagai R, Nishikawa K, Himeno H. 1996. The anticodon loop is a major identity determinant of Saccharomyces cerevisiae tRNA(Leu). *J Mol Biol* **263:** 707–714.

Soma A, Onodera A, Sugahara J, Kanai A, Yachie N, Tomita M, Kawamura F, Sekine Y. 2007. Permuted tRNA genes expressed via a circular RNA intermediate in Cyanidioschyzon merolae. *Science* **318:** 450–453.

Soma A, Uchiyama K, Sakamoto T, Maeda M, Himeno H. 1999. Unique recognition style of tRNA(Leu) by Haloferax volcanii leucyl-tRNA synthetase. *J Mol Biol* **293:** 1029–1038.

Sprinzl M, Horn C, Brown M, Ioudovitch A, Steinberg S. 1998. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res* **26:** 148–153.

Srinivasan G, James CM, Krzycki JA. 2002. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* **296:** 1459–1462.

Stern L, Schulman LH. 1978. The role of the minor base N4-acetylcytidine in the function of the Escherichia coli noninitiator methionine transfer RNA. *J Biol Chem* **253:** 6132–6139.

Sugahara J, Fujishima K, Morita K, Tomita M, Kanai A. 2009. Disrupted tRNA gene diversity and possible evolutionary scenarios. *J Mol Evol* **69:** 497–504.

Sugahara J, Kikuta K, Fujishima K, Yachie N, Tomita M, Kanai A. 2008. Comprehensive analysis of archaeal tRNA genes reveals rapid increase of tRNA introns in the order thermoproteales. *Mol Biol Evol* **25:** 2709–2716.

Sugahara J, Yachie N, Arakawa K, Tomita M. 2007. In silico screening of archaeal tRNA-encoding genes having multiple introns with bulge-helix-bulge splicing motifs. *RNA* **13:** 671–681.

Sugahara J, Yachie N, Sekine Y, Soma A, Matsui M, Tomita M, Kanai A. 2006. SPLITS: a new program for predicting split and intron-containing tRNA genes at the genome level. *In Silico Biol* **6:** 411–418.

Sugita T, Nakase T. 1999. Non-universal usage of the leucine CUG codon and the molecular phylogeny of the genus Candida. *Syst Appl Microbiol* **22:** 79–86.

Szweykowska-Kulinska Z, Senger B, Keith G, Fasiolo F, Grosjean H. 1994. Intron-dependent formation of pseudouridines in the anticodon of Saccharomyces cerevisiae minor tRNA(Ile). *EMBO J* **13:** 4636–4644.

Thompson DM, Lu C, Green PJ, Parker R. 2008. tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *RNA* **14:** 2095–2103.

Thompson DM, Parker R. 2009a. Stressing out over tRNA cleavage. *Cell* **138:** 215–219.

Thompson DM, Parker R. 2009b. The RNase Rny1p cleaves tRNAs and promotes cell death during oxidative stress in Saccharomyces cerevisiae. *J Cell Biol* **185:** 43–50.

Tsuji T, Sun Y, Kishimoto K, Olson KA, Liu S, Hirukawa S, Hu GF. 2005. Angiogenin is translocated to the nucleus of HeLa cells and is involved in ribosomal RNA transcription and cell proliferation. *Cancer Res* **65:** 1352–1360.

Varani G, McClain WH. 2000. The G x U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep* **1:** 18–23.

Wang Q, Lee I, Ren J, Ajay SS, Lee YS, Bao X. 2012. Identification and Functional Characterization of tRNA-derived RNA Fragments (tRFs) in Respiratory Syncytial Virus Infection. *Mol Ther* **21:** 368–379.

Wendrich TM, Blaha G, Wilson DN, Marahiel MA, Nierhaus KH. 2002. Dissection of the mechanism for the stringent factor RelA. *Mol Cell* **10:** 779–788.

Wiltrout E, Goodenbour JM, Fréchin M, Pan T. 2012. Misacylation of tRNA with methionine in Saccharomyces cerevisiae. *Nucleic Acids Res* **40:** 10494–10506.

Yamao F, Muto A, Kawauchi Y, Iwami M, Iwagami S, Azumi Y, Osawa S. 1985. UGA is read as tryptophan in Mycoplasma capricolum. *Proc Natl Acad Sci U S A* **82:** 2306–2309.

Yamasaki S, Ivanov P, Hu GF, Anderson P. 2009. Angiogenin cleaves tRNA and promotes stress-induced translational repression. *J Cell Biol* **185:** 35–42.

Yaremchuk A, Kriklivyi I, Tukalo M, Cusack S. 2002. Class I tyrosyl-tRNA synthetase has a class II mode of cognate tRNA recognition. *EMBO J* **21:** 3829–3840.

Yu XC, Borisov OV, Alvarez M, Michels DA, Wang YJ, Ling V. 2009. Identification of codon-specific serine to asparagine mistranslation in recombinant monoclonal antibodies by high-resolution mass spectrometry. *Anal Chem* **81:** 9282–9290.

Zhang Z, Shah B, Bondarenko PV. 2013. G/U and certain wobble position mismatches as possible main causes of amino acid misincorporations. *Biochemistry* **52:** 8165–8176.

Zisoulis DG, Kai ZS, Chang RK, Pasquinelli AE. 2012. Autoregulation of microRNA biogenesis by let-7 and Argonaute. *Nature* **486:** 541–544.

Abbreviations

aa-tRNA	aminoacylated tRNA
aaRS	aminoacyl-tRNA synthetase
AspRS	aspartyl-tRNA synthetase
C. elegans	*Caenorhabditis elegans*
COVE	covariance model
Cys	cysteine
DTT	dithiothreitol
ECF	Enhanced ChemiFluorescence
GFP	green fluorescent protein
Gln	glutamine
Glu	glutamic acid
Gly	glycine
GlyRS	glycyl-tRNA synthetase
HIV-1	human immunodeficiency virus 1
Ile	isoleucine
IleRS	isoleucyl-tRNA synthetase
IS	internal standard
Leu	leucine
LeuRS	leucyl-tRNA synthetase
Lys-C	lysyl endoprotease
mRNA	messenger RNA

MS	mass spectrometry
nanoLC-MS/MS	nano liquid chromatography–tandem mass spectrometry
nev-tRNA	nematode-specific V-arm-containing tRNA
NLB	NP-40 lysis buffer
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
Pyl	pyrrolysine
ROS	reactive oxygen species
RT	reverse transcription
SDC	sodium deoxycholate
Sec	selenocysteine
Ser	serine
SerRS	seryl-tRNA synthetase
SLS	sodium lauroyl sarcosinate
snoU3	U3 small nucleolar RNA
snU6	U6 small nuclear RNA
TEAB	triethylammonium bicarbonate
tRNA	transfer RNA
Trp	tryptophan
TrpRS	tryptophanyl-tRNA synthetase
UTR	untranslated region
V-arm	variable arm
V8	endprotease Glu-C