# Supporting Online Material

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

Kiyofumi Hamashima[1,2], Kosuke Fujishima[1], Takeshi Masuda[1], Junichi Sugahara[1], Masaru Tomita[1,2,3] and Akio Kanai[1,2,3,]

[1]Institute for Advanced Biosciences, Keio University, Tsuruoka 997-0017, Japan
[2]Systems Biology Program, Graduate School of Media and Governance, Keio University, Fujisawa 252-8520, Japan
[3]Faculty of Environment and Information Studies, Keio University, Fujisawa 252-8520, Japan

## Table S1. List of oligonucleotides used in this study

A. For northern blot analyses:

| 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 | TGCGGTGGACGGGATTCGAACCCGCGCTCTGTTACCAGAATAGGATCGGGAGTC CTACGCCTTCACCGCTCGGCCACCACCGC |
| | Ile(UAU) | 36 | ACTGGTTTAACCCAATGAGATCATAAGTCTCACGCC |
| | Ile(UAU) | 36 | TCGAACCCGCGACTGGTTTAACCCAATGAGATCATA |
| | Lys(CUU) | 36 | TCGGTTAGAACCGTTTGAGATCAAGTGTCTCATGCC |

* for acid-urea/northern blot analysis (Figure 5)

B. For RT–PCR analyses:

| 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 |

C. 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 |

**Table S2. Identification of aberrant residues in *in vitro*-translated luciferase.** 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. Regions of nev-tRNA-dependent peptide sequences and their Mascot scores are shaded. The highest Mascot score was based on four independent experiments.

| | | Mascot Score (n = 4) | |
|---|---|---|---|
| **Sequence (Luciferase)** | **Modification** | **$tRNA^{Leu}$ (Control)** | **nev-$tRNA^{Gly}$** |
| 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 |
| EVLEAVAK | | - | 25.9 |
| RFHLPLIR | | - | 25.4 |
| STLIDKYDLSNLHEIASLGAPLSK | | - | 71.6 |
| TIALIMNSSGSTLLPK | | - | 91.6 |
| VVDLDTLK | | - | 47.3 |

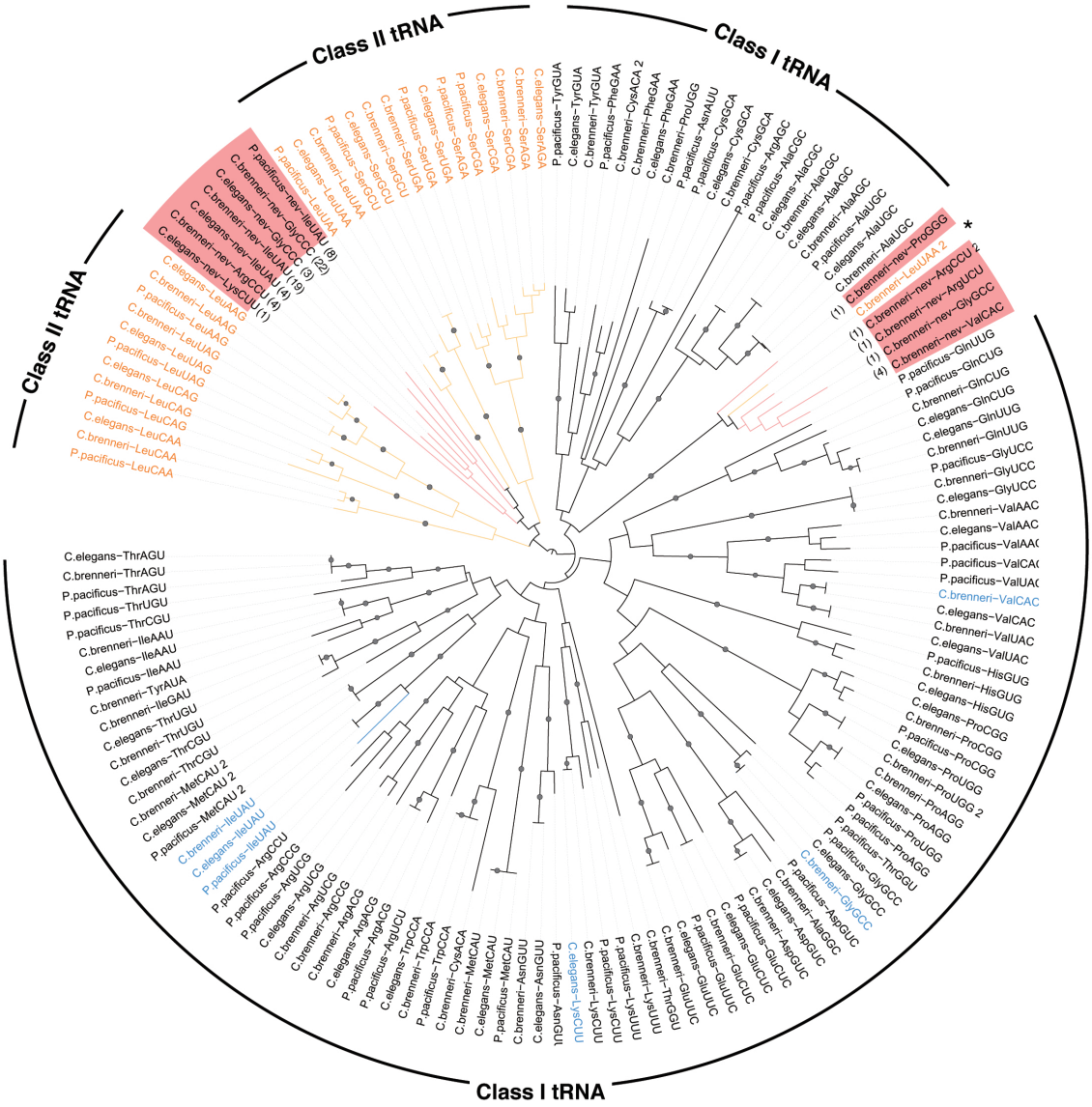


**Figure S1. Phylogeny of all tRNAs in the three nematode species.** A neibor-joining tree of all mature tRNA sequences in *C. elegans*, *C. brenneri*, and *P. pacificus* was constructed using ClustalX (20) and visualized with iTOL (21). The phylogeny includes nev-tRNAs (red), $tRNA^{Leu}$ and $tRNA^{Ser}$ (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).

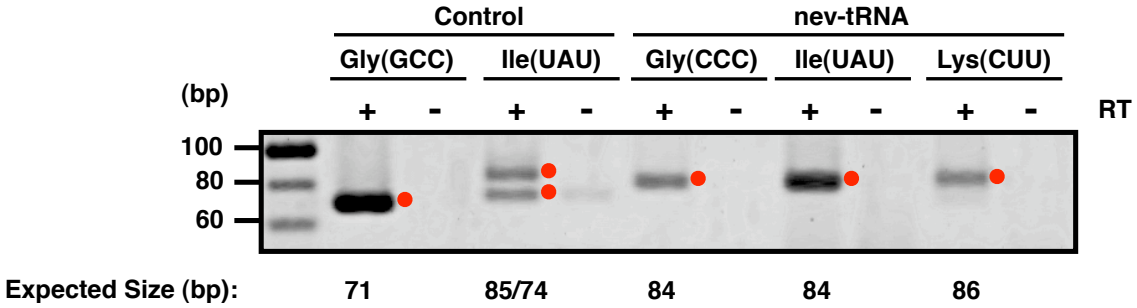


**Figure S2. RT–PCR analysis of the predicted nev-tRNAs of *C. elegans*.** Levels of expression of the three nev-tRNA genes and their synonymous tRNA genes (controls) were determined by RT–PCR. Samples without reverse transcriptase were used as the negative controls (RT–) to ensure that only RNA transcripts 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).

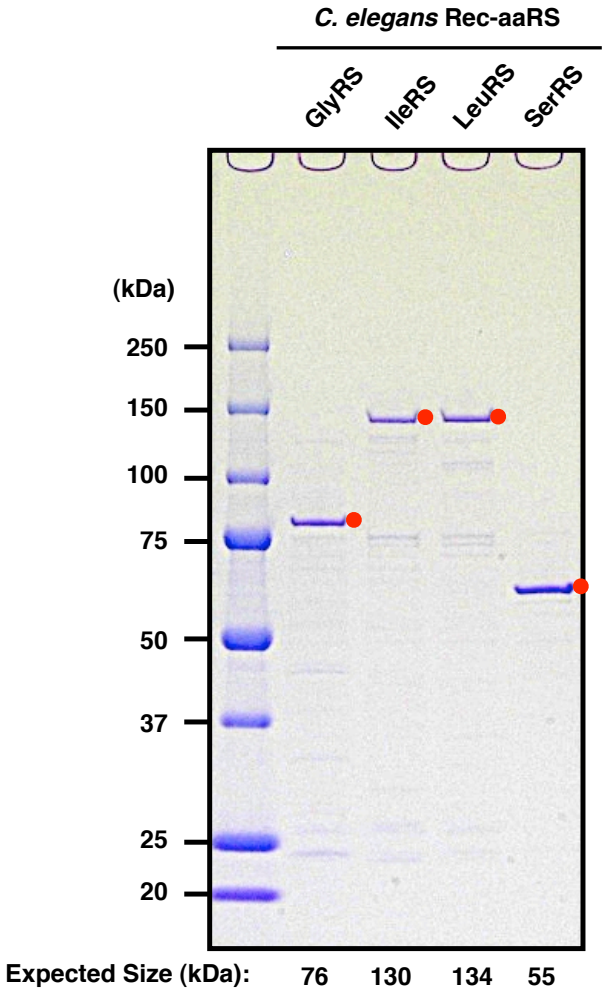


**Figure S3. Purified recombinant aminoacyl-tRNA synthetases from *C. elegans*.** His-tagged recombinant aminoacyl-tRNA synthetase (aaRS) enzymes were purified to near homogeneity using $Ni^{2+}$ agarose resin spin columns followed by gel filtration (see Materials and Methods). The samples were analyzed 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.

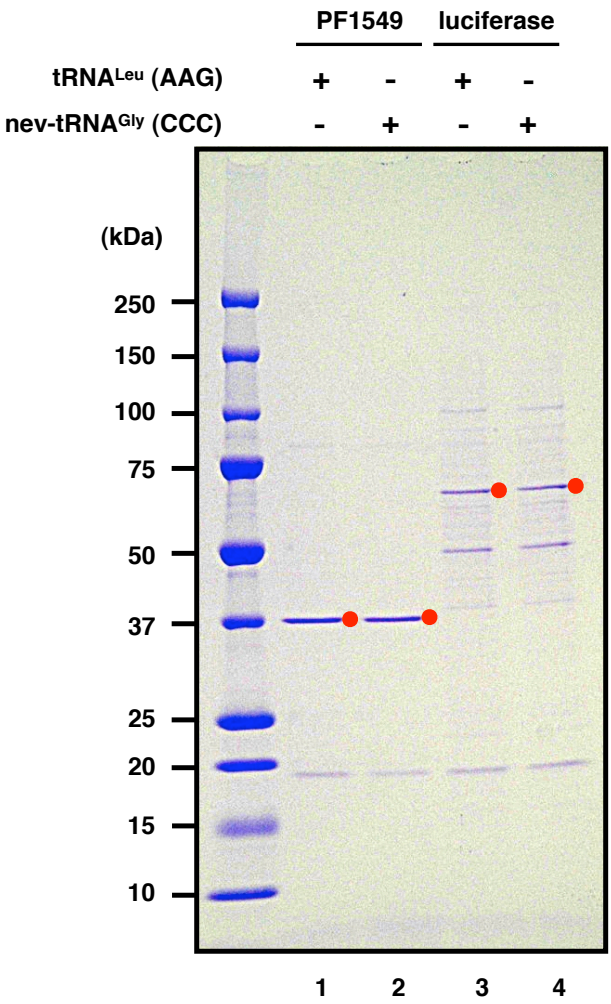


**Figure S4. SDS/PAGE analysis of purified *in vitro*-translated proteins.** His-tagged recombinant proteins synthesized in cell-free systems were purified with a magnetic separator (see Materials and Methods). The samples were analyzed by SDS/PAGE on a 10%-20% gradient gel and stained with Coomassie 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.

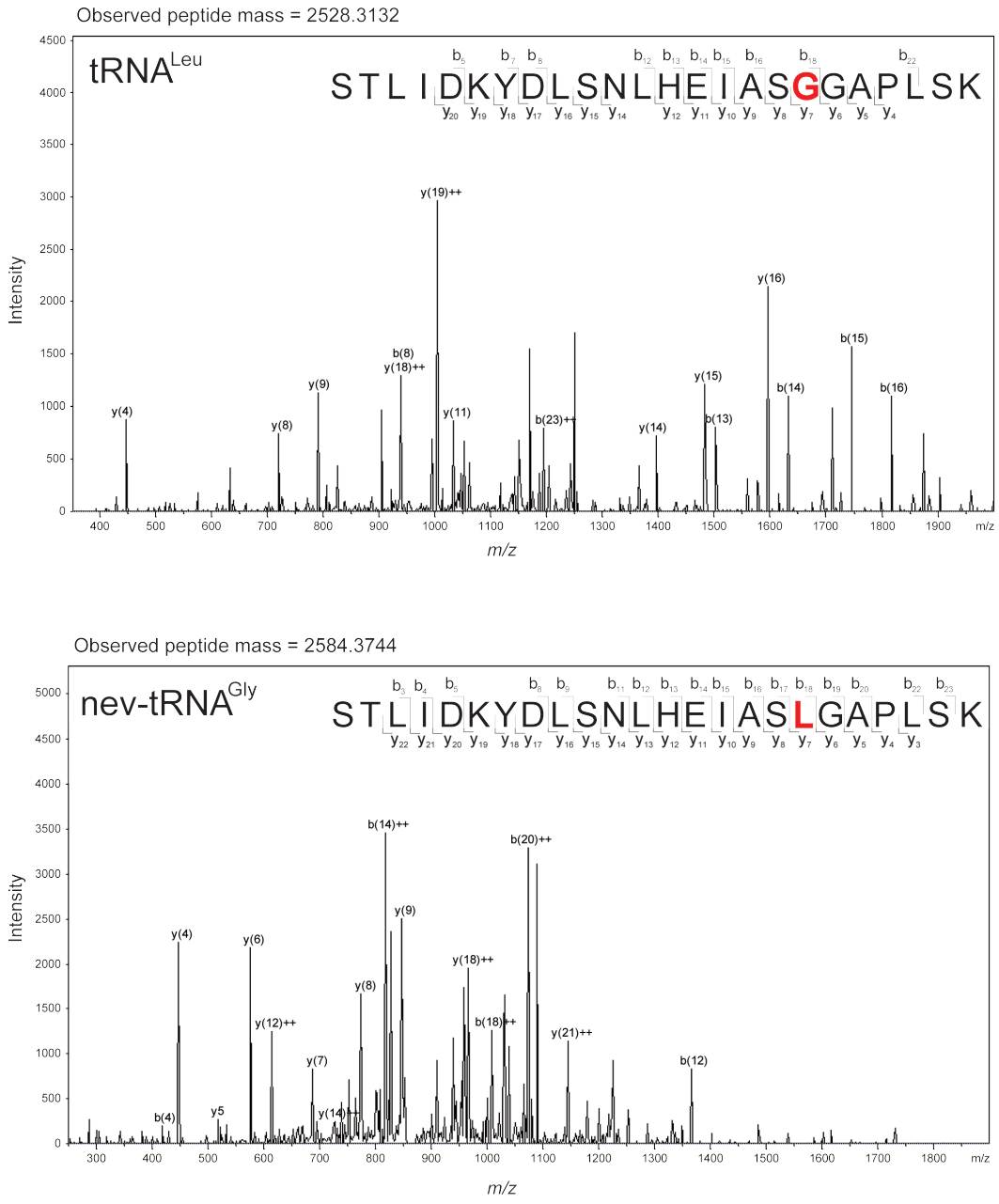


**Figure S5. Comparison of the MS/MS spectra of peptides with amino acid residues arising from the decoding of GGG codons.** Peptides with the sequence STLIDKYDLSNLHEIAS(G/L)GAPLSK from luciferase synthesized in cell-free systems containing $tRNA^{Leu}$ (AAG) or nev-$tRNA^{Gly}$ (CCC) were compared. 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.

| First base of codon | | Second base of codon: T | C | A | 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 S6. 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.

**Dataset S1. List of the predicted nev-tRNAs in nematode genomes**

This table is uploaded separately, as one Excel file.

**Dataset S2. Peptides of PF1549 and luciferase identified in the insect cell-free protein expression system**

This table is uploaded separately, as one Excel file.