

2016

# Systematic Evolution and Study of UAGN Decoding tRNAs in a Genomically Recoded Bacteria

Nanxi Wang

*University of Nebraska-Lincoln, nwang3@unl.edu*

Xin Shang

*University of Nebraska-Lincoln, xshang3@unl.edu*

Ronald Cerny

*University of Nebraska-Lincoln, rcerny1@unl.edu*

Wei Niu

*University of Nebraska-Lincoln, wniu2@unl.edu*

Jiantao Guo

*University of Nebraska-Lincoln, jguo4@unl.edu*

Follow this and additional works at: <http://digitalcommons.unl.edu/chemfacpub>

---

Wang, Nanxi; Shang, Xin; Cerny, Ronald; Niu, Wei; and Guo, Jiantao, "Systematic Evolution and Study of UAGN Decoding tRNAs in a Genomically Recoded Bacteria" (2016). *Faculty Publications -- Chemistry Department*. 95.  
<http://digitalcommons.unl.edu/chemfacpub/95>

This Article is brought to you for free and open access by the Published Research - Department of Chemistry at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications -- Chemistry Department by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

# SCIENTIFIC REPORTS

OPEN

## Systematic Evolution and Study of UAGN Decoding tRNAs in a Genomically Recoded Bacteria

Nanxi Wang<sup>1</sup>, Xin Shang<sup>1</sup>, Ronald Cerny<sup>1</sup>, Wei Niu<sup>2</sup> & Jiantao Guo<sup>1</sup>

Received: 03 December 2015

Accepted: 01 February 2016

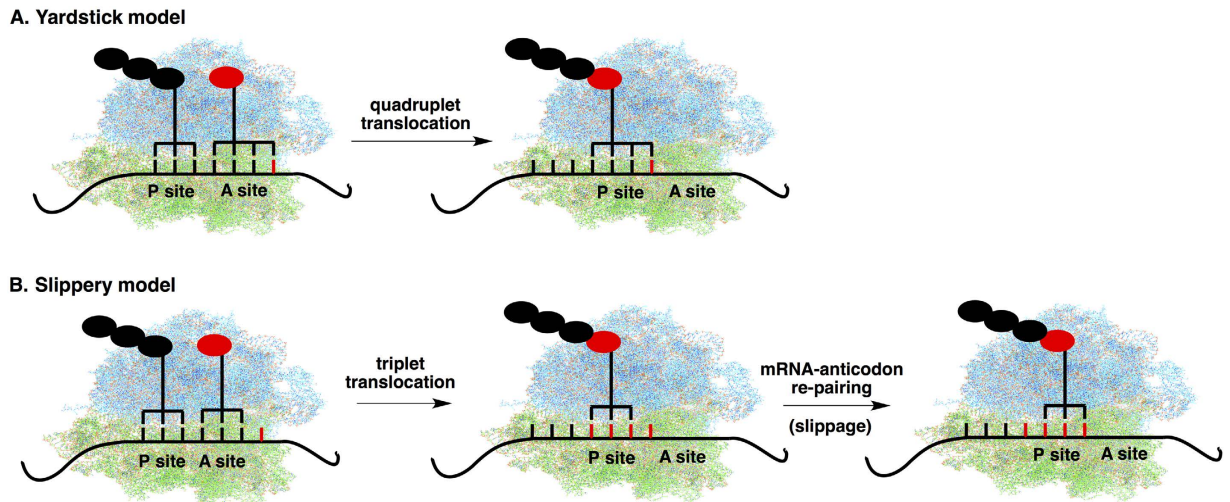
Published: 24 February 2016

We report the first systematic evolution and study of tRNA variants that are able to read a set of UAGN (N = A, G, U, C) codons in a genomically recoded *E. coli* strain that lacks any endogenous in-frame UAGN sequences and release factor 1. Through randomizing bases in anticodon stem-loop followed by a functional selection, we identified tRNA mutants with significantly improved UAGN decoding efficiency, which will augment the current efforts on genetic code expansion through quadruplet decoding. We found that an extended anticodon loop with an extra nucleotide was required for a detectable efficiency in UAGN decoding. We also observed that this crucial extra nucleotide was converged to a U (position 33.5) in all of the top tRNA hits no matter which UAGN codon they suppress. The insertion of U33.5 in the anticodon loop likely causes tRNA distortion and affects anticodon-codon interaction, which induces +1 frameshift in the P site of ribosome. A new model was proposed to explain the observed features of UAGN decoding. Overall, our findings elevate our understanding of the +1 frameshift mechanism and provide a useful guidance for further efforts on the genetic code expansion using a non-canonical quadruplet reading frame.

While the triplet codon is the predominant form of the current genetic code, programmed frameshifts, e.g., +1 frameshift caused by the non-canonical reading of a quadruplet codon, is a natural process. These programmed +1 frameshift events generally involve certain recoding signals that are embedded in the mRNA<sup>1</sup>. On the other hand, tRNA mutants that contain extended anticodon loops (8-base instead of normal 7-base loop) could induce +1 frameshift independent of recoding signals<sup>2–9</sup>. Recent developments in genetic code engineering also demonstrated that quadruplet codons could be used to encode unnatural amino acids (unAAs) under experimental conditions<sup>10–14</sup>. Here we present a systematic study on UAGN (N = A, G, U, C) decoding in order to further expand the current genetic code through suppressing quadruplet codons. Efforts have been made to expand the cellular genetic code with multiple quadruplet codons as an enabling synthetic biology tool for biological investigations<sup>10–14</sup>. Theoretically, a quadruplet codon table provides a maximum of 256 codons, which potentially allows a significant expansion of the current genetic code to facilitate biological studies<sup>15–17</sup> and to eventually enable the ribosomal synthesis of completely artificial biopolymers as new biomaterials. Besides genetic code expansion, we also intend to use UAGN decoding as a model system to study quadruplet codon decoding (+1 frameshift) mechanism.

Two major working models were proposed to explain the mechanism of quadruplet codon decoding (+1 frameshift) with tRNAs bearing extended anticodon loops: (1) the yardstick model<sup>18–20</sup> (Fig. 1A) states that the anticodon loop of tRNA interacts with all four bases of quadruplet codon in the A site of the ribosome, which leads to subsequent quadruplet translocation from the A site to the P site. This model is supported by observations that a number of tRNAs with an extended anticodon loop form apparent Watson-Crick complementarity to their cognate quadruplet codons at all four anticodon positions<sup>4,11,21–25</sup>. A primer extension toeprint assay on tRNA<sub>CCCG</sub> also supported this theory<sup>19</sup>. An NMR study showed that the anticodon stem-loop (ASL) of the tRNA<sub>CCCG</sub> lacked the conserved U-turn motif and could potentially undergo conformational adjustment in order to interact with quadruplet codon<sup>19</sup>. In an altered yardstick model<sup>26</sup>, the complete quadruplet codon-anticodon interaction in the A site is not required. Instead, the extra nucleotide widens the ASL and allows the anticodon nucleotide-34 to interact with either the fourth or the third and fourth codon bases. This model is supported by

<sup>1</sup>Department of Chemistry, University of Nebraska-Lincoln, Lincoln, Nebraska, 68588, United States. <sup>2</sup>Department of Chemical & Biomolecular Engineering, University of Nebraska-Lincoln, Lincoln, Nebraska, 68588, United States. Correspondence and requests for materials should be addressed to W.N. (email: wniu2@unl.edu) or J.G. (email: jguo4@unl.edu)

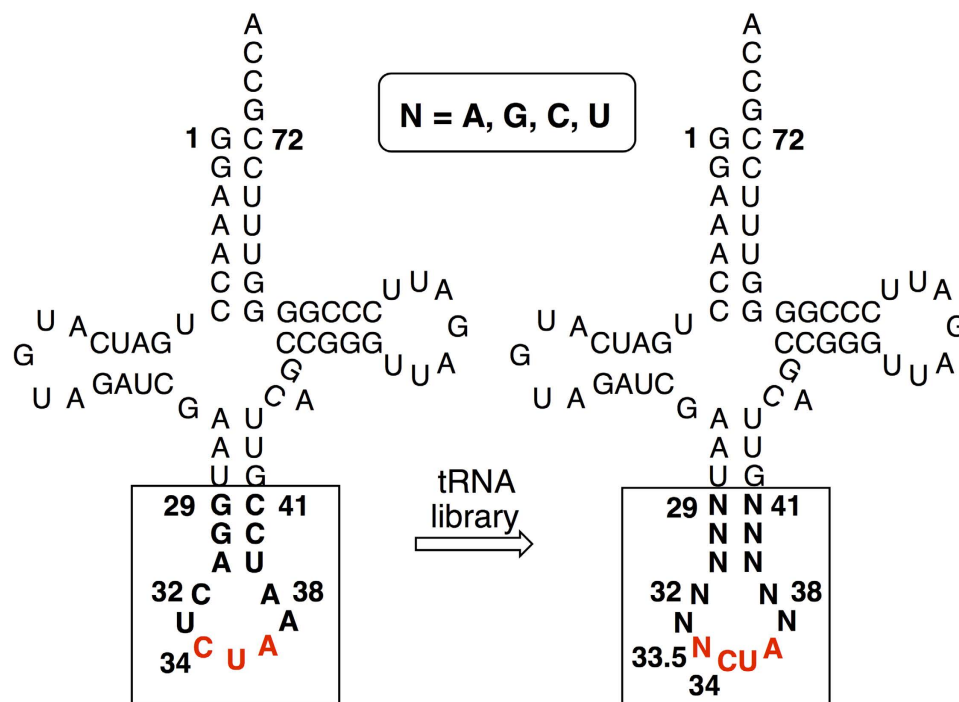


**Figure 1. Hypothetical models for +1 frameshift (quadruplet decoding) with tRNAs containing extended anticodon loop.** (A) The yardstick model features quadruplet anticodon-codon interaction in the A site followed by quadruplet translocation; (B) The Slippery model features normal triplet decoding in the A site, normal triplet translocation, and a slippage in the P site.

reported crystal structures of the 30S ribosomal subunit of *Thermus thermophilus* in complex with tRNAs known to facilitate +1 frameshifting and their cognate mRNA<sup>26</sup>; (2) the slippery model<sup>7,27–29</sup> (Fig. 1B) entails that tRNA makes a normal three-base codon-anticodon interaction in the A site of ribosome and translocation is always triplet. An anticodon-mRNA re-pairing subsequently occurs in the P-site with a slip of the mRNA by one base, which leads to an apparent quadruplet codon decoding. The re-pairing event usually requires that the anticodon binds to a cognate or a near cognate codon in the +1 frame. For example (Fig. S9), a well-studied CCCU suppressor (tRNA<sup>SufA6</sup>)<sup>5,27,30</sup> has a C<sub>34</sub>G<sub>35</sub>G<sub>36</sub> anticodon that forms two G-C base pairs both before and after the re-pairing event, which does not cause significant penalty in binding energy. Evidences suggest that the “ribosomal grip” of the peptidyl-tRNA is pivotal for maintaining the reading frame<sup>31</sup>. The slippery model is also supported by observations that +1 frameshift efficiency was significantly affected by the decoding of the A-site codon following the quadruplet codon, indicating that the frameshift happens in the P site<sup>32–35</sup>. Regardless of the merits of each model, key aspects, such as the codon-anticodon interaction and translocation mechanism, of the quadruplet decoding by tRNAs with extended anticodon loops remain to be resolved.

We<sup>14</sup> and other<sup>36</sup> recently reported a new approach to enhance the efficiency of quadruplet codon decoding, which is based on the engineering of tRNA anticodon stem-loop (ASL). By using this method, here we present the first systematic evolution and study of tRNA variants that decode UAGN codons in a genomically recoded *E. coli* strain that lacks any endogenous in-frame UAGN sequences. Most previous studies on quadruplet codon decoding<sup>10–14</sup> were conducted in hosts that contain a large number of in-frame four base sequences that are identical to the quadruplet codon of interest. An efficient quadruplet decoding system would lead to genome-wide, undesirable frameshifts, which could impair protein synthesis and introduce additional difficulties to the mechanistic investigations on quadruplet codon decoding. In addition, competing recognition of the first three bases of a quadruplet codon by triplet decoding tRNAs or release factors could reduce the efficiency of quadruplet codon decoding and further complicate the study. A recent synthetic biology effort led to the successful construction of an *E. coli* strain, C321.ΔA, that does not express release factor 1 (RF1; decodes the UAG nonsense codon as the termination signal of protein translation) and does not contain endogenous UAG nonsense codon (replaced with ochre UAA nonsense codon)<sup>37</sup>. This genomically recoded *E. coli* strain represents an excellent host for the evolution and study of UAGN decoding by eliminating the pitfalls mentioned above. In fact, a recent study<sup>38</sup> showed that a *Methanocaldococcus jannaschii*-derived tRNA, tRNA<sup>Tyr</sup><sub>UCUA</sub>, could efficiently decode UAGA codons in *E. coli* C321.ΔA by simply replacing the CUA anticodon of *M. jannaschii*-derived amber suppressor tRNA, tRNA<sup>Tyr</sup><sub>CUA</sub>, with UCUA anticodon. Here we report the engineering of the ASL of a *Methanosarcina mazei*-derived tRNA, tRNA<sup>Pyl</sup><sub>CUA</sub>, in order to efficiently decode a set of UAGN (N = A, G, U, C) codons. We chose tRNA<sup>Pyl</sup><sub>CUA</sub> as our model tRNA since its anticodon is unlikely used as an important recognition element by its cognate PylRS according to structural data<sup>39</sup>. We also confirmed this notion by showing that tRNA<sup>Pyl</sup><sub>UCCU</sub>, a derivative of tRNA<sup>Pyl</sup><sub>CUA</sub> in which the anticodon CUA was replaced with UCCU, could be charged with an unAA by a PylRS variant to decode AGGA codon<sup>14</sup>. This structural feature allowed us to engineer the ASL of tRNA<sup>Pyl</sup><sub>CUA</sub> without interfering with the tRNA aminoacylation step. Our work on converting the tRNA<sup>Pyl</sup><sub>CUA</sub> into a quadruplet codon decoding tRNA would augment the current efforts on the genetic incorporation of unAAs by PylRS variants. Since our ASL mutations in tRNA<sup>Pyl</sup> unlikely affect PylRS mutants' specificity towards their unAA substrates, evolved tRNAs can pair with PylRS mutants to incorporate a range of different unAAs.

Our ability to evolve quadruplet decoding tRNAs that are derived from the same ancestor but decode a set of UAGN codons will enable a systematic investigation on the mechanism of quadruplet codon decoding with tRNAs bearing extended anticodon loops. In addition, inferences from this work will likely bring better



**Figure 2.** Clover leaf structures of tRNA<sup>Pyl</sup><sub>CUA</sub> and tRNA<sup>Pyl</sup><sub>NCUA</sub>. Comparing to tRNA<sup>Pyl</sup><sub>CUA</sub>, tRNA<sup>Pyl</sup><sub>NCUA</sub> mutants contain eight rather than seven bases in its anticodon loop. The additional base, N, is numbered as 33.5. Nucleotides in the anticodon stem-loop (ASL) that were randomized are shown in bold.

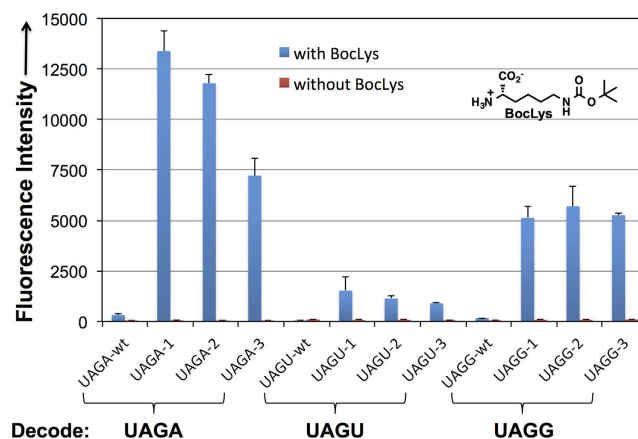
understanding of ribosome function in the future as well, especially how the ribosome maintains the reading frame and its relation to translocation, which remains one of the key questions to be understood in protein translation.

## Results

**tRNA (with an extended anticodon loop) library construction and selection for UAGN (N = A, G, U, C) codon decoding.** An additional and randomized nucleotide, N (N = A, G, U, or C), was inserted between U33 and C34 in the anticodon loop of tRNA<sup>Pyl</sup><sub>CUA</sub>. This extra nucleotide is numbered as 33.5 (Fig. 2) so that the conventional numbering of tRNA molecule at other positions was not changed. Randomization at position 33.5 resulted in a mixture of four tRNAs, including tRNA<sup>Pyl</sup><sub>UCUA</sub>-wt, tRNA<sup>Pyl</sup><sub>CCUA</sub>-wt, tRNA<sup>Pyl</sup><sub>ACUA</sub>-wt, and tRNA<sup>Pyl</sup><sub>GCUA</sub>-wt. The 'wt' designation indicates that these tRNAs have identical sequence as their parent, wild-type tRNA<sup>Pyl</sup><sub>CUA</sub> except the extra nucleotide in the anticodon loop. Each of the four tRNAs now contains an extended (eight rather than seven nucleotides) anticodon loop. We expect that such modification of tRNA can facilitate UAGN decoding since previous work suggested that tRNA mutants containing extended anticodon loops could induce +1 frameshift independent of recoding signals<sup>2-4</sup>.

As mutations in the ASL of tRNA<sup>Pyl</sup><sub>NCUA</sub> (N = A, G, U, or C) likely change the tRNA's interaction with ribosome, mRNA, and other components of the translational machinery, we created a tRNA library (Library-NCUA11, theoretical diversity  $4.2 \times 10^6$ ; actual library size  $5.2 \times 10^6$ ) in which four nucleotides of the anticodon loop (32, 33, 37, 38; Fig. 2) and six nucleotides of the anticodon stem (29, 30, 31, 39, 40, 41; Fig. 2) were completely randomized in tRNA<sup>Pyl</sup><sub>NCUA</sub>. The resulting tRNA library was subsequently subjected to positive selections against UAGA, UAGG, UAGU, and UAGC codons, respectively, in *E. coli* C321.ΔA.exp in order to identify tRNA<sup>Pyl</sup><sub>NCUA</sub> variants that can decode these quadruplet codons. Randomization of the nucleotide at position 33.5 would allow us to select tRNA<sup>Pyl</sup><sub>NCUA</sub> variants against fully matched or fourth-base mismatched quadruplet codons in order to examine the requirement of codon-anticodon pairing in quadruplet codon decoding. Previous data suggest that full quadruplet Watson-Crick base pairing is not absolutely required but indeed results in the highest quadruplet decoding efficiency<sup>4,7,10,22,24,25,40</sup>. With a systematic study and more suitable model system, we expect to gain more reliable and deeper insights into this question.

The selection was conducted in the presence of *N*ε-(tert-butyloxy-carbonyl)-L-lysine (Boc-Lys), BocLysRS (a pyrrolysyl-tRNA synthetase mutant that specifically charges tRNA<sup>Pyl</sup><sub>CUA</sub> with Boc-Lys)<sup>41</sup>, chloramphenicol (concentrations ranging from 34 to 75 μg mL<sup>-1</sup>), and a chloramphenicol acetyltransferase mutant gene containing the quadruplet codon of interest at a permissive site (the codon for Gln98 is replaced with a quadruplet codon). The survivors were further screened with a range of different concentrations of chloramphenicol (50, 75, and 100 μg mL<sup>-1</sup>) in the presence and absence of Boc-Lys. It was considered a hit if the cell grew at 100 μg mL<sup>-1</sup> chloramphenicol in the presence of Boc-Lys, but did not grow at 50 μg mL<sup>-1</sup> chloramphenicol in the absence of



**Figure 3. GFP<sub>UV</sub> fluorescence assays of cells expressing tRNA<sup>Pyl</sup><sub>NCUA</sub> variants.** Fluorescence readings of *E. coli* C321.ΔA cells expressing tRNA<sup>Pyl</sup><sub>NCUA</sub>-wt or the evolved tRNA<sup>Pyl</sup><sub>NCUA</sub> mutants, each coexpressed with BocLysRS and corresponding GFP<sub>UV</sub>-Asn149UAGN. The expressions were conducted either in the presence or in the absence of 5 mM Boc-Lys. Fluorescence intensity was normalized to cell growth. Each data point is the average of triplicate measurements with standard deviation.

Boc-Lys. Multiple hits were obtained for each of the UAGA, UAGG, and UAGU codons. We failed to identify any hit for UAGC codon in a number of attempts. Each hit was designated as UAGN-X, where N represents A, G, or U, and X is the hit number.

**Characterization and validation of evolved tRNAs.** In order to examine the efficiency and fidelity of UAGN codon decoding by the obtained tRNAs, the four quadruplet codons of interest, UAGA, UAGG, UAGU, and UAGC were introduced into GFP<sub>UV</sub> (a variant of green fluorescent protein) to replace the Asn149 codon and yielded GFP<sub>UV</sub>-Asn149UAGA, GFP<sub>UV</sub>-Asn149UAGG, GFP<sub>UV</sub>-Asn149UAGU, and GFP<sub>UV</sub>-Asn149UAGC, respectively. A series of plasmids, pGFP<sub>UV</sub>-NCUA-X (N = A, G, U, or C; X = hit number), were subsequently constructed. Each plasmid contains a tRNA mutant of interest (expressed under the control of a *lpp* promoter) and a GFP<sub>UV</sub> mutant (driven by a T5 promoter) containing corresponding quadruplet codon. As controls, we also constructed four pGFP<sub>UV</sub>-NCUA-wt (N = A, G, U, or C) plasmids in which tRNA<sup>Pyl</sup><sub>NCUA</sub>-wt (tRNA<sup>Pyl</sup><sub>CUA</sub> with an extra base at position 33.5) were used instead of the tRNAs mutants that were obtained from selection.

Plasmids pGFP<sub>UV</sub>-NCUA-X and pGFP<sub>UV</sub>-NCUA-wt were transformed individually into *E. coli* C321.ΔA strain containing plasmid pBK-BocLysRS that carries a copy of BocLysRS-encoding gene. Protein expression was carried out in LB medium supplemented with or without 5 mM Boc-Lys. The Fig. 3 shows the top three tRNA mutants for decoding UAGA, UAGG, and UAGU codons. All mutants exhibited significantly enhanced quadruplet decoding efficiency in the presence of Boc-Lys relative to that of the corresponding tRNA<sup>Pyl</sup><sub>NCUA</sub>-wt. The best UAGA-, UAGG-, and UAGU-decoding tRNA mutant displayed approximately 55, 40, and 9 folds improvement over their corresponding tRNA<sup>Pyl</sup><sub>NCUA</sub>-wt (Fig. 3). Notably, the decoding efficiency of UAGA and UAGG codons was much higher than that of UAGU codon. In addition, fluorescence analyses of *E. coli* cultures also showed that significant amount of full-length GFP<sub>UV</sub> protein was only produced in the presence of Boc-Lys for all of our evolved tRNA<sup>Pyl</sup><sub>NCUA</sub> variants that were examined (Fig. 3). This result indicates that these tRNA mutants can only be charged with Boc-Lys by BocLysRS and cannot be charged with natural amino acids by BocLysRS or any endogenous aminoacyl-tRNA synthetase (aaRSs) in *E. coli*. While no hit was obtained to decode UAGC codon, we still examined the decoding efficiency of tRNA<sup>Pyl</sup><sub>GCUA</sub>-wt against UAGC codon and no detectable decoding was observed.

We further confirmed the incorporation of Boc-Lys into GFP<sub>UV</sub>-Asn149UAGN by mass spectrometry after SDS-PAGE separation and trypsin digest. As shown in the mass spectra (Figs S2, S3 and S4 in the Supporting Information), two mass peaks were observed in each sample. The (M + 2H)<sup>2+</sup> peak corresponds to the peptide fragment (141-LEYNYN**Boc-Lys**VYITADK) from GFP<sub>UV</sub> that contains an intact Boc-Lys residue at position 149. The (M + 3H)<sup>3+</sup> peak corresponds to the peptide fragment that contains a lysine residue at position 149. Since the BocLysRS cannot charge tRNA<sup>Pyl</sup><sub>NCUA</sub> variants with lysine according to both literature report<sup>41</sup> and our data (Fig. S5), the observed peptide that contains lysine at position 149 must be derived from the cleavage of the Boc group under the mass spectrometry conditions. The carbamate cleavage of Boc-Lys was also observed previously with electron spray ionization process in the literature<sup>42,43</sup>. Overall, our data confirmed that there was no undesirable incorporation of natural amino acids. The yields of the mutant GFP<sub>UV</sub> proteins were 54, 9, and 21 mg/L after partial purification by affinity chromatography when tRNA mutants UAGA-1, UAGU-1, and UAGG-2 were used to decode UAGA, UAGU, and UAGG codons, respectively.

**Incorporation of other unAAs using the evolved tRNAs.** Structural data shows the lack of direct interaction between PylRS and the ALS region of tRNA<sup>Pyl</sup><sub>CUA</sub>. Therefore, evolved tRNA<sup>Pyl</sup><sub>NCUA</sub> mutants in theory can be charged by PylRS mutants with unAAs other than Boc-Lys. To demonstrate the general applicability of evolved tRNA<sup>Pyl</sup><sub>NCUA</sub>



Codon	tRNA variants	Positions				
		29–31	32,33	33,5	37, 38	39–41
UAGA	UAGA-wt	G G A	C U	U	A A	U C C
	UAGA-1	G G G	C U	U	A U	C C U
	UAGA-2	U G G	A U	U	A C	C U U
	UAGA-3	C G G	A U	U	A C	C U U
UAGU	UAGU-wt	G G A	C U	A	A A	U C C
	UAGU-1	U G G	C U	U	A U	C U U
	UAGU-2	A G G	C U	U	A U	C U U
	UAGU-3	U G G	A U	U	A C	C U U
UAGG	UAGG-wt	G G A	C U	C	A A	U C C
	UAGG-1	U G G	C U	U	A U	C U U
	UAGG-2	G G G	C U	U	A U	C U U
	UAGG-3	A G G	C U	U	A U	C U U

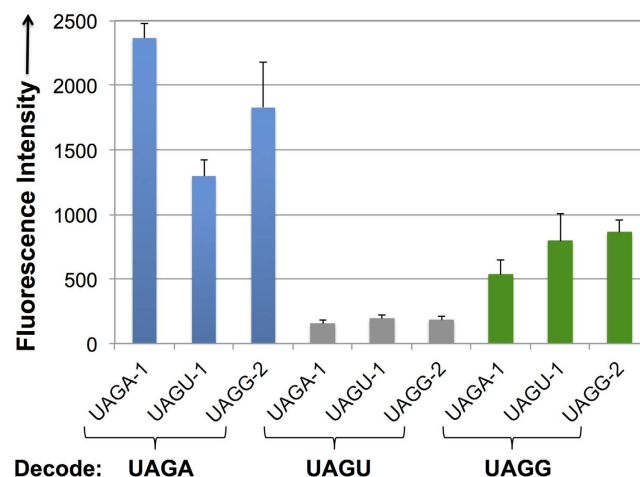
**Table 1. Evolved tRNA<sup>Pyl</sup><sub>NCUA</sub> variants with improved UAGN decoding activity.** Sequences of each tRNA<sup>Pyl</sup><sub>NCUA</sub> variant at randomized positions are listed.

mutants, we randomly selected two PylRS variants for the incorporation of 3'-azibutyl-N-carbamoyl-lysine<sup>44</sup> (AbK) and *o*-nitrobenzyl-oxycarbonyl-Nε-L-lysine<sup>45</sup> (ONBK) in this study. We observed excellent selectivity when tRNA mutants were used to decode UAGN codons (AbK, Fig. S6; ONBK, Fig. S7). The relative incorporation efficiencies of Boc-Lys, Abk, and ONBK in response to either UAGN codons or UAG codon are consistent (Figs 3, S6, S7, and S8). The above results further confirmed that ASL mutations in tRNA<sup>Pyl</sup> do not affect its recognition by PylRS mutants or the synthetases binding specificity towards their unAA substrates. Therefore, the evolved tRNAs can be potentially paired with any PylRS mutants for unAA incorporation, which is a significant augment of the current efforts in genetic code expansion.

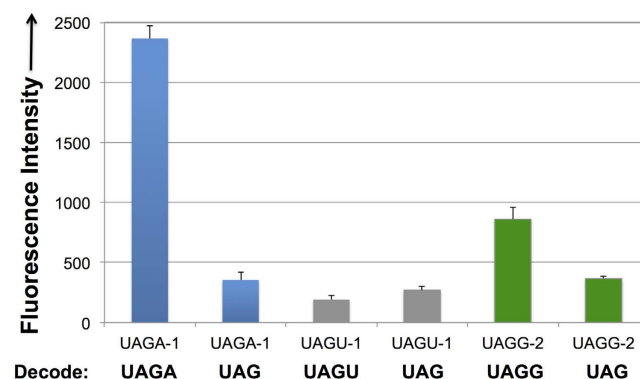
**Analysis of the evolved tRNA mutants.** We did not isolate and directly sequencing tRNA mutants. Instead, genes that encode the evolved tRNA mutants were sequenced. All tRNA-encoding genes were isolated and re-transformed into *E. coli* C321.ΔA for various characterizations throughout the study and consistent data were obtained. Therefore, we assume that no mutations of tRNA occurred during or after transcription and the actual tRNA sequences can be correctly deduced from their encoding genes. The observed mutations of the top three tRNAs from each of UAGA, UAGG, and UAGU selections are shown in Table 1. The most striking observation was that all of these tRNAs contained apparent UCUA (U at 33.5 position) anticodon no matter which UAGN codon they decode. In addition, some tRNAs that were selected against different UAGN codons have the same sequence, such as UAGA-2 and UAGU-3, UAGU-1 and UAGG-1, UAGU-2 and UAGG-3 (Table 1). We subsequently analysed a few less efficient tRNA variants from the selection, three, one, and one additional tRNA sequences were identified from UAGA, UAGU, and UAGG selections, respectively. As shown in Table S1, UAGU-4 and UAGG-4 have the same sequence while all UAGA hits (UAGA-4, UAGA-5, and UAGA-6) are unique. According to our data, all tRNA mutants from UAGU selection could be found from UAGA and/or UAGG selections. A large number of tRNA mutants from UAGU and UAGG selections have sequence overlap. On the other hand, most of tRNA mutants from UAGA selections are unique. Overall, the above results implicate that a fourth base-pairing between anticodon and codon is not essential for UAGN decoding.

The sequencing results also revealed three additional features of the evolved tRNA mutants: (1) no mutation was observed for the two nucleotides that are immediately 5'- or 3'-adjacent (position 33 and 37; Table 1 and Table S1) to the hypothetical four-base anticodon (NCUA). These two positions are likely essential for the function of tRNA<sup>Pyl</sup><sub>NCUA</sub>-derived mutants in *E. coli*; (2) all evolved tRNA mutants contain the same A31G and U39C mutations (Table 1 and Table S1). This led to the replacement of the A31-U39 with a G31-C39 base pair, which likely strengthens the part of the anticodon stem that is directly adjacent to the anticodon loop; (3) most of the evolved tRNA mutants do not have Watson-Crick base pairing at either position 29–41 or position 30–40 (Table 1 and Table S1). These mutations likely lead to structural changes in the anticodon and ASL, which allow a favorable conformational adjustment of tRNA upon its binding to the A site and subsequent translocation to the P and E sites of the ribosome.

**Cross-decoding among UAGN codons.** Three top individual tRNA mutants, UAGA-1, UAGU-1, and UAGG-2, from UAGA, UAGU, and UAGG selection, respectively, were cross-tested against codons that they were not selected against. To simplify the cross-test, a series of plasmids, pGFP<sub>UV</sub>-UAGN-BocLysRS (N = A, G, or U), were constructed. Each plasmid contains a BocLysRS gene (expressed under the control of a *glnS* promoter) and a GFP<sub>UV</sub> mutant containing one of the UAGN (N = A, G, or U) codons. The decoding efficiencies of tRNA mutants toward each UAGN codon are correlated with the intensity of GFP fluorescence. As shown in Fig. 4, while the evolved tRNAs were among the best against the codons they were selected for, significant cross-activities were observed. Again, these results indicate that the fourth base-pairing is not essential for UAGN decoding. In addition, none of the obtained tRNA mutants could efficiently decode UAGC codon (data not shown).



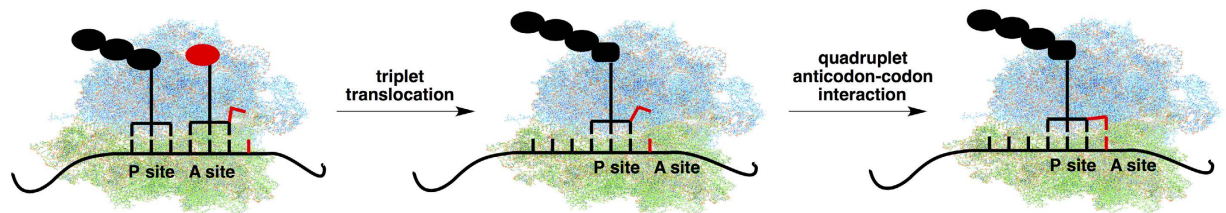
**Figure 4. Cross-decoding among UAGN codons.** Fluorescence readings of *E. coli* C321.ΔA cells expressing the GFP reporter GFP<sub>UV</sub>-Asn149UAGN, BocLysRS (in plasmid pGFP<sub>UV</sub>-UAGN-BocLysRS), and tRNA mutants (in plasmid pBK-UAGN-X; X = hit number) in the presence of 5 mM Boc-Lys. Fluorescence intensity was normalized to cell growth. Each data point is the average of triplicate measurements with standard deviation.



**Figure 5. UAGN and UAG codon decoding efficiency of evolved tRNA variants.** Fluorescence readings of *E. coli* C321.ΔA cells expressing the GFP reporter GFP<sub>UV</sub>-Asn149UAGN (or GFP<sub>UV</sub>-Asn149UAG), BocLysRS (in plasmid pGFP<sub>UV</sub>-UAGN-BocLysRS), and tRNA mutants (in plasmid pBK-UAGN-X; X = hit number) in the presence of 5 mM Boc-Lys. Fluorescence intensity was normalized to cell growth. Each data point is the average of triplicate measurements with standard deviation.

**tRNA (with regular 7-base anticodon loop) library construction and selection for UAGN (N = A, G, U, C) decoding.** Since the fourth base-pairing of anticodon-codon interaction is apparently not essential for UAGN decoding, we next examined if tRNA variants with regular 7-base anticodon loop could decode UAGN codons. To this end, we constructed a tRNA library (Library-CUA10, theoretical diversity  $1.0 \times 10^6$ ; actual library size  $8.7 \times 10^6$ ) based on tRNA<sup>Pyl</sup><sub>CUA</sub> in which four nucleotides of the anticodon loop (32, 33, 37, 38; Fig. 2) and six nucleotides of the anticodon stem (29, 30, 31, 39, 40, 41; Fig. 2) were randomized. The resulting tRNA library was subsequently subjected to positive selections against UAGA, UAGG, UAGU, and UAGC, respectively, employing the same procedure as used in the Library-NCUA11 selection. No hit was obtained from these selections. The result indicates that an extra nucleotide in the anticodon loop is essential for UAGN decoding.

**Cross-decoding against UAG codons.** Three top tRNA mutants, UAGA-1, UAGU-1, and UAGG-2, from UAGA, UAGU, and UAGG selection, respectively, were examined for UAG codon decoding. To do this, a plasmid, pGFP<sub>UV</sub>-UAG-BocLysRS, was constructed and the expression of GFP (contains an amber mutation at position 149) was examined in the presence of UAGA-1, UAGU-1, or UAGG-2. As shown in Fig. 5, UAGA-1 and UAGG-2 displayed significantly higher activity toward UAGA and UAGG codons, respectively, than to UAG codon. On the other hand, UAGU-1 displayed similar activity toward UAGU and UAG codons, which might due to the low decoding efficiency of UAGU-1 against UAGU codon. All three tRNA mutants showed similar and significantly diminished UAG decoding efficiency (~20 fold lower) than that of tRNA<sup>Pyl</sup><sub>CUA</sub>.



**Figure 6. A new model for quadruplet decoding (+1 frameshift).** This model features a triplet translocation followed by a quadruplet anticodon-codon interaction in the P site of ribosome.

## Discussion

This work represents the first example of a systematic evolution and study of tRNAs that decode a quartet of quadruplet codons (UAGN, N = A, U, G, or C) in a genomically recoded host with a clean background. The lack of endogenous UAG nonsense codons eliminates any in-frame UAGN sequences, which rules out the possibility of missing highly efficient UAGN-decoding tRNAs that might be cytotoxic due to undesirable reading-through of stop signals. Utilization of this host strain ensures that the most efficient UAGN-decoding tRNA variants can be identified. We targeted the evolution and study of tRNAs that are derived from the same ancestor, which enables informative and reliable comparison among obtained tRNAs. By introducing mutations into the ASL of parent tRNAs, we identified tRNA<sup>Pyl</sup><sub>NCUA</sub> variants that displayed significant improvement in UAGN decoding efficiency in comparison to their corresponding parents. The best hit (UAGA-1) showed 55-fold improvement in UAGA decoding efficiency than that of tRNA<sup>Pyl</sup><sub>UCUA</sub>-wt. Several sequence convergences were observed among the evolved tRNAs, which likely shift their interaction favorably with other components of the translational machinery for higher efficiency in quadruplet codon decoding. The observed improvement could also be partially attributed to the lack of competition from amber UAG codon-recognizing release factor-1 in the host strain. In addition to Boc-Lys, we showed that the evolved tRNA variants could be used to incorporate AbK and ONBK as well. It is likely that the evolved tRNA variants can be applied to the genetic incorporation of a range of unAAs using PylRS mutants in response to UAGN codons.

Our data suggest that an extra nucleotide in the anticodon loop of tRNA is essential for high-efficiency decoding of UAGN codons within the natural translational machinery. However, the complete four-nucleotide Watson-Crick base-pairing between anticodon and codon is apparently not required for UAGN decoding. In fact, the best tRNA hit that read UAGG and UAGU codons do not have the potential to form four-nucleotide Watson-Crick base-pairing with these codons. In addition, cross activities were observed among tRNA mutants that were selected against different UAGN codons. These results implicated that reading of quadruplet codons is less likely to happen in the A site, which is consistent with the two recent structural investigations on the ASLs of frameshift suppressor tRNAs (tRNA<sup>SufJ</sup> and tRNA<sup>SufA6</sup>) bound to the *Thermus thermophilus* 70S A site<sup>30,46</sup>. These two structural studies suggested that +1 frameshift suppressor tRNAs do not form four-nucleotide Watson-Crick base-pairing with codons. Instead, ASL<sup>SufJ</sup> and ASL<sup>SufA6</sup> form a cognate and a near cognate triplet codon-anticodon interaction in the A site, respectively. It is likely that the evolved tRNA<sup>Pyl</sup><sub>UCUA</sub> variants could engage a similar cognate or near cognate triplet interaction with the first three bases of UAGN sequences. Such cognate or a near cognate triplet codon-anticodon interactions would sufficiently lead to conformational changes of three bases, A1492, A1493, and G530, in 16S rRNA at ribosome A site<sup>47</sup> and allow tRNAs to pass this critical checkpoint for decoding specificity. Since UAGN decoding does not require suppressor tRNA to occupy all four bases of the codon at the decoding center of A site, the overall and apparent quadruplet UAGN codon decoding process may be more accurately described as the induction of +1 frameshift by the evolved tRNA<sup>Pyl</sup><sub>UCUA</sub> variants.

While both ours and others' results indicate that direct quadruplet codon recognition does not happen in the A site, a question remains on how +1 frameshift happens. In the slippery model<sup>7,27–29</sup>, a re-pairing of tRNA to the +1 frame of mRNA (Figs 1B and S9A) occurs after the translocation of tRNA-mRNA complex from A to P site. The tRNA-mRNA re-pairing was suggested to be initiated by the P-site interaction between ribosome and the ASL of tRNA<sup>27</sup>. A recent structural study showed that the insertion of G37.5 in tRNA<sup>SufA6</sup> destabilized the non-canonical U32–A38 base pair, which likely affected EF-G-dependent translocation and therefore facilitated +1 frameshifting<sup>30</sup>. The insertion of an extra nucleotide-33.5 into tRNA<sup>Pyl</sup><sub>NCUA</sub> likely impaired the non-canonical C32–A38 interaction in tRNA<sup>Pyl</sup><sub>NCUA</sub> in a similar way and led to +1 frameshift. In addition, all of our tRNA hits contain either an A38U or an A38C mutation, which likely changed the finely tuned nucleotide identity of the 32–38 pair and influenced the decoding process<sup>48,49</sup>.

Although our data on UAGN decoding are more consistent with the slippery model<sup>7,27–29</sup>, this model cannot explain why a much higher decoding efficiency was observed for UAGA codon by the evolved tRNA<sup>Pyl</sup><sub>UCUA</sub> variants that have potential to form a four-nucleotide Watson-Crick base-pairing with the UAGA codon. In addition, the evolved tRNA<sup>Pyl</sup><sub>UCUA</sub> variants do not have an alternative cognate or near-cognate site for re-pairing in the P site (Fig. S9B), which is a hallmark of the slippery model. Even re-pairing would happen, we would observe that the efficiency of UAGG decoding would be higher than that of UAGA decoding since the re-pairing of the triplet anticodon of tRNA<sup>Pyl</sup><sub>UCUA</sub> is energetically more favored for UAGG. However, we observed the opposite. Here we propose a modified slippery model for UAGN decoding (Fig. 6). The tRNA rearrangement in the P site does not necessarily involve a slippage event. It is possible for a suppressor tRNA to interact with all four bases of a quadruplet sequence in the P site and facilitate +1 frameshifting. In the cases of UAGN decoding, the extra U33.5 in



tRNA<sup>Pyl</sup><sub>UCUA</sub> may engage certain level of interaction with the fourth base of the UAGN sequence, which creates steric interference with the incoming tRNA in the A site and induces +1 frameshift. This hypothesis is consistent with our observations on the efficiencies of UAGN suppression. The U33.5-A interaction in UAGA decoding is the most favorable one and led to the highest decoding efficiency. The U33.5-G interaction in UAGG decoding is also favorable and led to the second highest decoding efficiency. On the other hand, the U33.5-U interaction in UAGU decoding and the U33.5-C interaction in UAGC decoding are less favorable and led to the lowest decoding efficiency. In fact, we did not obtain a hit that is able to suppress UAGC codon. Other mutations in ASL of tRNA<sup>Pyl</sup><sub>UCUA</sub> mutants may augment such interaction between tRNA and UAGN sequences by changing the conformation of engineered tRNAs. While this new model may not be suitable to explain all other +1 frameshifting cases, it provides reasonable molecular details on +1 frameshifting involving UAGN sequences.

Our current data does not pinpoint the reason for the sequence convergence to U at position 33.5 in evolved tRNAs for UAGN suppression. It is possible that only U33.5 could preserve a characteristic U-turn structure of the anticodon loop that is essential for tRNA translocation<sup>50,51</sup>. Either U33 or U33.5 interact with A37 (no mutation) or U38/C38 (A38 in the wild-type) at other side of the anticodon loop. Future structural studies would be needed to obtain more defined answers. Besides conformational factors discussed above, posttranscriptional modifications of tRNA could also play a role. The primary sequence of wild-type and evolved pyrrolysyl tRNAs may not accurately reflect the anticodon base-pairing capability for quadruplet decoding. This is because extensive and chemically diverse nucleoside modifications fine-tune the biophysical and biochemical properties of transcribed tRNA molecules<sup>52,53</sup>. In addition, certain tRNA modifications influence reading frame maintenance of the ribosome and may facilitate +1 frameshift<sup>54,55</sup>. The insertion of U33.5 may impair the normal posttranscriptional modifications of tRNA and induces reading frame shift. Currently, we are investigating the posttranscriptional modification of *M. maize* tRNA in *E. coli* host and the implication in UAGN suppression.

In summary, we have evolved a collection of tRNA mutants that are able to incorporate unAAs in response to quadruplet UAGN sequences. Our work is an augment to the current efforts in genetic code expansion. While, theoretically, a quadruplet codon table provides 256 blank codons, our data suggest that cross-recognition among quadruplet codons may significantly reduce the number of usable quadruplet codons that are intended to be used simultaneously for unAA incorporation. We are currently examining whether we are able to engineer orthogonal tRNA variants for different UAGN sequences in order to truly allow a significant expansion of the genetic code using quadruplet suppression mechanism. The present study also represents a significant step toward the elucidation of the quadruplet suppression mechanism with tRNAs bearing extended anticodon loop. Unlike most of the previous studies that only focused on limited number of +1 frameshift suppressor tRNAs, this study involves a collection of tRNAs that decode a set of UAGN codons. The inferences from this work will likely bring better understanding to how the ribosome maintains the reading frame in the future.

## Methods

**Materials and General Methods.** Boc-Lys was purchased from Bachem. Primers were ordered from Sigma. Restriction enzymes, Antarctic phosphatase (AP) and T4 DNA ligase were purchased from New England Biolabs. KOD hot start DNA polymerase was purchased from EMD Millipore. Protein mass spectrometric data was collected on a Waters Synapt<sup>®</sup> G2 mass spectrometer. Data was processed using Masslynx<sup>™</sup> software (Waters). Standard molecular biology techniques<sup>56</sup> were used throughout. Site-directed mutagenesis was carried out using overlapping PCR. *E. coli* GeneHogs were used for routine cloning and DNA propagation. *E. coli* C321.ΔA.exp (Addgene) and C321.ΔA (Addgene) were used for quadruplet decoding tRNA selection and evaluation. All solutions were prepared in deionized water further treated by Barnstead Nanopure<sup>®</sup> ultrapure water purification system (Thermo Fisher Scientific Inc). Antibiotics were added where appropriate to following final concentrations: ampicillin, 100 mg L<sup>-1</sup>; kanamycin, 50 mg L<sup>-1</sup>; tetracycline, 12.5 mg L<sup>-1</sup>.

**Plasmid construction.** Plasmid pBK-BocLysRS was constructed by inserting BocLysRS-encoding gene behind the constitutive *glnS* promoter (*P<sub>glnS</sub>*) on pBK vector<sup>57</sup>.

Plasmid pREP-BocLysRS-UAGN (for positive selection) was constructed by modifying plasmid pRep-CM12b<sup>58</sup>. Specifically, the UAG codon (position 98) in the chloramphenicol acetyltransferase-encoding gene on pRepCM12b was changed to UAGN by site-directed mutagenesis. The primers that were used for mutagenesis can be found in supporting information. The resulting plasmid, pRepCM12b-UAGN, was digested with *Xba*I and ligated to a DNA fragment containing *P<sub>glnS</sub>*-BocLysRS cassette, which was amplified from template pBK-BocLysRS, to yield plasmid pREP-BocLysRS-UAGN. Individual plasmids with four different quadruplet codons, pREP-BocLysRS-UAGA, pREP-BocLysRS-UAGG, pREP-BocLysRS-UAGU, and pREP-BocLysRS-UAGC, were isolated from single *E. coli* colonies and verified by DNA sequencing.

Plasmids pGFP<sub>UV</sub>-NCUA-wt (N = A, G, U, or C) were obtained by first introducing the desirable UAGN mutation into plasmid pLei-GFP<sub>UV</sub>-Asn149UAG<sup>58</sup> by overlapping PCR. The primers that were used for mutagenesis can be found in supporting information. The resulting plasmid, pGFP<sub>UV</sub>-UAGN, was subsequently modified by replacing the original tRNA with tRNA<sup>Pyl</sup><sub>NCUA</sub>-wt and yielded plasmids, pGFP<sub>UV</sub>-NCUA-wt (N = A, G, U, or C).

Variants of pGFP<sub>UV</sub>-NCUA-X (N = A, G, U, or C; X = hit number) were constructed by replacing the wild-type tRNA<sup>Pyl</sup><sub>NCUA</sub>-wt on pGFP<sub>UV</sub>-NCUA-wt with evolved tRNA mutants.

Plasmids pGFP<sub>UV</sub>-UAGN-BocLysRS (N = A, G, U, or C) were obtained by deleting tRNA fragment and inserting DNA fragment containing *P<sub>glnS</sub>*-BocLysRS cassette into *Pst*I and *Spe*I sites of plasmids pGFP<sub>UV</sub>-UCUA-wt. The *P<sub>glnS</sub>*-BocLysRS cassette was amplified from template pBK-BocLysRS. Individual plasmids, each of which contains a specific UAGN codon of interest, were isolated from single *E. coli* colonies and verified by DNA sequencing.

Plasmids pGFP<sub>UV</sub>-UAG-BocLysRS was obtained by deleting the tRNA fragment and inserting DNA fragment containing *P<sub>glnS</sub>*-BocLysRS cassette into *Pst*I and *Spe*I sites of plasmids pLei-GFP<sub>UV</sub>-Asn149UAG<sup>58</sup>. The *P<sub>glnS</sub>*-BocLysRS cassette was amplified from template pBK-BocLysRS.

**tRNA library construction.** Procedures for library construction were adapted from Noren and Noren<sup>59</sup>. Mutants of tRNA were obtained by overlapping PCR using pBK-mmPylT<sup>45</sup> as the template. Resulting PCR products contain DNA cassettes of *lpp* promoter-tRNA mutant-rrnC terminator. Digestion of PCR products with *Nco*I and *Xho*I followed by ligation between *Nco*I and *Xho*I sites of pBK vector resulted in the tRNA library. The primers that were used for library construction can be found in supporting information.

**Positive selection.** Library DNAs were transformed into *E. coli* C321.ΔA.exp or C321.ΔA electrocompetent cells containing plasmid pREP-BocLysRS-UAGN (N = A, G, U, or C). Transformants were cultivated in LB media containing kanamycin and tetracycline. After 12 h of cultivation, cells were harvested. Based on calculation, a certain number of cells ( $>4.6 \times$  the size of the library) were plated on LB agar containing kanamycin, tetracycline, Boc-Lys (5 mM), and chloramphenicol (concentrations range from  $34 \text{ mg L}^{-1}$  to  $75 \text{ mg L}^{-1}$ ). The selection plates were incubated at 37 °C (C321.ΔA.exp) or 30 °C (C321.ΔA) for 48 h. Selected numbers of single colonies were further screened by replication onto plates with certain concentrations (50, 75, or  $100 \mu\text{g mL}^{-1}$ ) of chloramphenicol in the presence and absence of Boc-Lys. Only the ones that grew in the presence of Boc-Lys and did not grow in the absence of Boc-Lys were selected for further evaluation.

**Fluorescence analysis of bacterial culture.** *E. coli* C321.ΔA strain harboring plasmid pBK-BocLysRS and a pGFP<sub>UV</sub>-NCUA variant (wt or evolved tRNA mutant) was cultured in 1 mL LB media containing ampicillin, kanamycin, and chloramphenicol at 30 °C. After 18 h cultivation, 50 μL culture was sub-cultured in 1 mL LB medium containing ampicillin, kanamycin, chloramphenicol, IPTG (0.1 mM), and 5 mM Boc-Lys (or 5 mM Lys, or none). Following cultivation at 30 °C for an additional 16 h, 1 mL of cell culture were collected, washed, then resuspended in 1 mL of potassium phosphate buffer (50 mM, pH 7.4). The processed cells were directly used for fluorescence and cell density measurements using a Synergy<sup>TM</sup> H1 Hybrid plate reader (BioTek Instruments). The fluorescence of GFP<sub>UV</sub> was monitored at  $\lambda_{\text{Ex}} = 390 \text{ nm}$  and  $\lambda_{\text{Em}} = 510 \text{ nm}$ . The cell density was estimated by measuring the sample absorbance at 600 nm. Values of fluorescence intensity were normalized to cell growth. The incorporations of AbK and ONBK followed the same procedure using previously reported PylRS mutants<sup>44,60</sup>. Reported data are the average of three biological replicates with standard deviations. The cross test among different UAGN codons were conducted using the plasmid combination of pBK-tRNA-hit and pGFP<sub>UV</sub>-UAGN-BocLysRS.

**Protein expression and purification.** *E. coli* C321.ΔA strain harboring plasmid pBK-BocLysRS and a pGFP<sub>UV</sub>-NCUA variant was cultured in 5 mL LB media containing kanamycin and chloramphenicol at 30 °C. After 18 hours of cultivation, 0.5 mL culture was sub-cultured in 50 mL medium. The protein expression was induced at the OD<sub>600nm</sub> of 0.6 by additions of IPTG (0.5 mM) and Boc-Lys (5 mM). Cells were collected by centrifugation at 5 000 g and 4 °C for 15 min. Harvested cells were resuspended in lysis buffer containing potassium phosphate (20 mM, pH 7.4), NaCl (150 mM) and imidazole (10 mM). Cells were subsequently disrupted by sonication. Cellular debris was removed by centrifugation (21 000 g, 30 min, 4 °C). The cell-free lysate was applied to Ni Sepharose 6 Fast Flow resin (GE Healthcare). Protein purification followed manufacturer's instructions. Protein concentrations were determined by Bradford assay (Bio-Rad). Purified protein was isolated by SDS-PAGE and digested with trypsin prior to MS analysis.

## References

- Atkins, J. F. & Bjoerk, G. R. A gripping tale of ribosomal frameshifting: extragenic suppressors of frameshift mutations spotlight P-site realignment. *Microbiol. Mol. Biol. Rev.* **73**, 178–210 (2009).
- Riddle, D. L. & Roth, J. R. Suppressors of frameshift mutations in *Salmonella typhimurium*. *J. Mol. Biol.* **54**, 131–144 (1970).
- Yourn, J. & Kohno, T. Externally suppressible proline quadruplet CCCU. *Science* **175**, 650–652 (1972).
- Riddle, D. L. & Carbon, J. Frameshift suppression: a nucleotide addition in the anticodon of a glycine transfer RNA. *Nat. New Biol.* **242**, 230–234 (1973).
- Sroga, G. E., Nemoto, F., Kuchino, Y. & Bjoerk, G. R. Insertion (sufB) in the anticodon loop or base substitution (sufC) in the anticodon stem of tRNA<sup>Pro2</sup> from *Salmonella typhimurium* induces suppression of frameshift mutations. *Nucleic Acids Res.* **20**, 3463–3469 (1992).
- Cummins, C. M., Donahue, T. F. & Culbertson, M. R. Nucleotide sequence of the SUF2 frameshift suppressor gene of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **79**, 3565–3569 (1982).
- Gaber, R. F. & Culbertson, M. R. Codon recognition during frameshift suppression in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **4**, 2052–2061 (1984).
- Bossi, L. & Roth, J. R. Four-base codons ACCA, ACCU and ACCC are recognized by frameshift suppressor *suff*. *Cell* **25**, 489–496 (1981).
- Bossi, L. & Smith, D. M. Suppressor *suff*: a novel type of tRNA mutant that induces translational frameshifting. *Proc. Natl. Acad. Sci. USA* **81**, 6105–6109 (1984).
- Anderson, J. C. *et al.* An expanded genetic code with a functional quadruplet codon. *Proc. Natl. Acad. Sci. USA* **101**, 7566–7571 (2004).
- Taki, M., Matsushita, J. & Sisido, M. Expanding the genetic code in a mammalian cell line by the introduction of four-base codon/anticodon pairs. *ChemBioChem* **7**, 425–428 (2006).
- Neumann, H., Wang, K., Davis, L., Garcia-Alai, M. & Chin, J. W. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464**, 441–444 (2010).
- Rodriguez, E. A., Lester, H. A. & Dougherty, D. A. *In vivo* incorporation of multiple unnatural amino acids through nonsense and frameshift suppression. *Proc. Natl. Acad. Sci. USA* **103**, 8650–8655 (2006).
- Niu, W., Schultz, P. G. & Guo, J. An expanded genetic code in mammalian cells with a functional quadruplet codon. *ACS Chem. Biol.* **8**, 1640–1645 (2013).
- Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**, 413–444 (2010).
- Chin, J. W. Reprogramming the genetic code. *Science* **336**, 428–429 (2012).
- Niu, W. & Guo, J. Expanding the chemistry of fluorescent protein biosensors through genetic incorporation of unnatural amino acids. *Mol. Biosyst.* **9**, 2961–2970 (2013).
- Curran, J. F. & Yarus, M. Reading frame selection and transfer RNA anticodon loop stacking. *Science* **238**, 1545–1550 (1987).
- Phelps, S. S. *et al.* Translocation of a tRNA with an extended anticodon through the ribosome. *J. Mol. Biol.* **360**, 610–622 (2006).

20. Walker, S. E. & Fredrick, K. Recognition and positioning of mRNA in the ribosome by tRNAs with expanded anticodons. *J. Mol. Biol.* **360**, 599–609 (2006).
21. Bernhardt, H. S. & Tate, W. P. Evidence from glycine transfer RNA of a frozen accident at the dawn of the genetic code. *Biol. Direct* **3**, 53 (2008).
22. Hill, C. W., Combriato, G., Steinhart, W., Riddle, D. L. & Carbon, J. Nucleotide sequence of the GGG-specific glycine transfer ribonucleic acid of *Escherichia coli* and of *Salmonella typhimurium*. *J. Biol. Chem.* **248**, 4252–4262 (1973).
23. Moore, B., Persson, B. C., Nelson, C. C., Gesteland, R. F. & Atkins, J. F. Quadruplet codons: implications for code expansion and the specification of translation step size. *J. Mol. Biol.* **298**, 195–209 (2000).
24. Magliery, T. J., Anderson, J. C. & Schultz, P. G. Expanding the genetic code: selection of efficient suppressors of four-base codons and identification of “shifty” four-base codons with a library approach in *Escherichia coli*. *J. Mol. Biol.* **307**, 755–769 (2001).
25. Anderson, J. C., Magliery, T. J. & Schultz, P. G. Exploring the limits of codon and anticodon size. *Chem. Biol.* **9**, 237–244 (2002).
26. Dunham, C. M. *et al.* Structures of tRNAs with an expanded anticodon loop in the decoding center of the 30S ribosomal subunit. *RNA* **13**, 817–823 (2007).
27. Qian, Q. *et al.* A new model for phenotypic suppression of frameshift mutations by mutant tRNAs. *Mol. Cell* **1**, 471–482 (1998).
28. Pedersen, S. *Escherichia coli* ribosomes translate *in vivo* with variable rate. *EMBO J.* **3**, 2895–2898 (1984).
29. Qian, Q. & Bjork, G. R. Structural alterations far from the anticodon of the tRNA<sup>GAG</sup>Pro of *Salmonella typhimurium* induce +1 frameshifting at the peptidyl-site. *J. Mol. Biol.* **273**, 978–992 (1997).
30. Maehigashi, T., Dunkle, J. A., Miles, S. J. & Dunham, C. M. Structural insights into +1 frameshifting promoted by expanded or modification-deficient anticodon stem loops. *Proc. Natl. Acad. Sci. USA* **111**, 12740–12745 (2014).
31. Naesvall, S. J., Nilsson, K. & Bjoerk, G. R. The ribosomal grip of the peptidyl-tRNA is critical for reading frame maintenance. *J. Mol. Biol.* **385**, 350–367 (2009).
32. Farabaugh, P. J. & Bjork, G. R. How translational accuracy influences reading frame maintenance. *EMBO J.* **18**, 1427–1434 (1999).
33. Gallant, J. *et al.* On the role of the starved codon and the takeoff site in ribosome bypassing in *Escherichia coli*. *J. Mol. Biol.* **342**, 713–724 (2004).
34. Belcourt, M. F. & Farabaugh, P. J. Ribosomal frameshifting in the yeast retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. *Cell* **62**, 339–352 (1990).
35. Farabaugh, P. J. Translational frameshifting: Implications for the mechanism of translational frame maintenance. *Prog. Nucleic Acid Res. Mol. Biol.* **64**, 131–170 (2000).
36. Wang, K. *et al.* Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labeling and FRET. *Nat. Chem.* **6**, 393–403 (2014).
37. Lajoie, M. J. *et al.* Genomically recoded organisms expand biological functions. *Science* **342**, 357–360 (2013).
38. Chatterjee, A., Lajoie, M. J., Xiao, H., Church, G. M. & Schultz, P. G. A bacterial strain with a unique quadruplet codon specifying non-native amino acids. *ChemBioChem* **15**, 1782–1786 (2014).
39. Nozawa, K. *et al.* Pyrrolysyl-tRNA synthetase-tRNA<sup>Pyl</sup> structure reveals the molecular basis of orthogonality. *Nature* **457**, 1163–1167 (2009).
40. Weiss, R. B., Dunn, D. M., Atkins, J. F. & Gesteland, R. F. Ribosomal frameshifting from –2 to +50 nucleotides. *Prog. Nucleic Acid Res. Mol. Biol.* **39**, 159–183 (1990).
41. Yanagisawa, T. *et al.* Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N<sup>ε</sup>-(o-azidobenzoyloxycarbonyl) lysine for site-specific protein modification. *Chem. Biol.* **15**, 1187–1197 (2008).
42. Nguyen, D. P. *et al.* Genetic encoding of photocaged cysteine allows photoactivation of TEV protease in live mammalian cells. *J. Am. Chem. Soc.* **136**, 2240–2243 (2014).
43. Schmied, W. H., Elsasser, S. J., Uttamapinant, C. & Chin, J. W. Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized Pyrrolysyl tRNA Synthetase/tRNA expression and engineered eRF1. *J. Am. Chem. Soc.* **136**, 15577–15583 (2014).
44. Ai, H.-W., Shen, W., Sagi, A., Chen, P. R. & Schultz, P. G. Probing protein-protein interactions with a genetically encoded photocrosslinking amino acid. *ChemBioChem* **12**, 1854–1857 (2011).
45. Chen, P. R. *et al.* A facile system for encoding unnatural amino acids in mammalian cells. *Angew. Chem., Int. Ed.* **48**, 4052–4055 (2009).
46. Fagan, C. E., Maehigashi, T., Dunkle, J. A., Miles, S. J. & Dunham, C. M. Structural insights into translational recoding by frameshift suppressor tRNA<sup>Sup</sup>. *RNA* **20**, 1944–1954 (2014).
47. Steitz, T. A. A structural understanding of the dynamic ribosome machine. *Nat. Rev. Mol. Cell Biol.* **9**, 242–253 (2008).
48. Olejniczak, M. & Uhlenbeck, O. C. tRNA residues that have coevolved with their anticodon to ensure uniform and accurate codon recognition. *Biochimie* **88**, 943–950 (2006).
49. Ledoux, S., Olejniczak, M. & Uhlenbeck, O. C. A sequence element that tunes *Escherichia coli* tRNA<sup>Ala</sup><sub>GCC</sub> to ensure accurate decoding. *Nat. Struct. Mol. Biol.* **16**, 359–364 (2009).
50. Phelps, S. S., Jerinic, O. & Joseph, S. Universally conserved interactions between the ribosome and the anticodon stem-loop of A site tRNA important for translocation. *Mol. Cell* **10**, 799–807 (2002).
51. Phelps, S. S. & Joseph, S. Non-bridging phosphate oxygen atoms within the tRNA anticodon stem-loop are essential for ribosomal A site binding and translocation. *J. Mol. Biol.* **349**, 288–301 (2005).
52. Jackman, J. E. & Alfonzo, J. D. Transfer RNA modifications: nature's combinatorial chemistry playground. *Wiley Interdiscip. Rev.: RNA* **4**, 35–48 (2013).
53. El, Y. B., Bailly, M. & de Crecy-Lagard, V. Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annu. Rev. Genet.* **46**, 69–95 (2012).
54. Bjork, G. R. *et al.* Transfer RNA modification: influence on translational frameshifting and metabolism. *FEBS Lett.* **452**, 47–51 (1999).
55. Yarian, C. *et al.* Accurate translation of the genetic code depends on tRNA modified nucleosides. *J. Biol. Chem.* **277**, 16391–16395 (2002).
56. Sambrook, J. F. & Russell, D. W. Editors. *Molecular cloning: A laboratory manual*, third edition. (Cold Spring Harbor Laboratory Press, 2000).
57. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of *Escherichia coli*. *Science* **292**, 498–500 (2001).
58. Guo, J., Melancon, C. E., III, Lee, H. S., Groff, D. & Schultz, P. G. Evolution of amber suppressor tRNAs for efficient bacterial production of proteins containing nonnatural amino acids. *Angew. Chem., Int. Ed.* **48**, 9148–9151 (2009).
59. Noren, K. A. & Noren, C. J. Construction of high-complexity combinatorial phage display peptide libraries. *Methods* **23**, 169–178 (2001).
60. Lang, K. *et al.* Genetic encoding of bicyclononynes and trans-cyclooctenes for site-specific protein labeling *in vitro* and in live mammalian cells via rapid fluorogenic diels-alder reactions. *J. Am. Chem. Soc.* **134**, 10317–10320 (2012).

## Acknowledgements

This work is supported by the New Faculty Startup Fund to J. G. from the Chemistry Department of University of Nebraska – Lincoln and by Grant 1R01AI11862 (to J. G.) from DHHS-NIH-NIAID. C321.ΔA.exp was a gift from George Church (Addgene plasmid #49018). C321.ΔA was a gift from George Church (Addgene plasmid #48998).

## Author Contributions

N.W. constructed tRNA library and plasmids, performed the selection and characterization of the hits. X.S. synthesized amino acids. R.C. performed mass spectrometry analyses. W.N. designed the study and helped to construct tRNA library and plasmids, and wrote the manuscript. J.G. designed the study and wrote the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Wang, N. *et al.* Systematic Evolution and Study of UAGN Decoding tRNAs in a Genomically Recoded Bacteria. *Sci. Rep.* **6**, 21898; doi: 10.1038/srep21898 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>