

Table S1. The usage of +4U in 100 non-pseudo, non-hypothetical HEGs and LEGs for each stop codon, ranked by CAI, in 25 bacterial species. Together with the number of 30S and 50S ribosomal protein genes in each species. The reference HEGs to compute CAI are the same 40 ribosomal protein genes used for .ITE reference files.

SPECIES NAME	ACCESSION	N.R. <sup>a</sup>	UAA <sup>b</sup>		UAG		UGA	
			P <sub>U.HEG</sub>	P <sub>U.LEG</sub>	P <sub>U.HEG</sub>	P <sub>U.LEG</sub>	P <sub>U.HEG</sub>	P <sub>U.LEG</sub>
<i>Bacillus anthracis</i>	NC_005945	56	0.77	0.23	0.37	0.29	0.32	0.24
<i>Bacillus subtilis</i>	NC_000964	56	0.65	0.24	0.19	0.10	0.44	0.34
<i>Bartonella henselae</i>	NC_005956	55	0.25	0.35	0.21	0.19	0.39	0.35
<i>Staphylococcus aureus</i>	NC_002758	53	0.78	0.35	0.34	0.32	0.49	0.343
<i>Listeria monocytogenes</i>	NC_003210	57	0.52	0.29	0.17	0.23	0.36	0.36
<i>Streptococcus pyrogenes</i>	NC_002737	51	0.61	0.23	0.34	0.24	0.51	0.49
<i>Lactococcus lactis</i>	NC_002662	57	0.64	0.17	0.36	0.34	0.32	0.32
<i>Leptospira interrogans</i>	NC_005823	54	0.37	0.26	0.30	0.34	0.41	0.33
<i>Deinococcus deserti</i>	NC_002937	51	0.17	0.14	0.12	0.11	0.12	0.17
<i>Bacteriodes thetaiotaomicron</i>	NC_004663	55	0.62	0.29	0.34	0.21	0.46	0.28
<i>Escherichia coli</i>	NC_000913	57	0.62	0.3	0.36	0.24	0.72	0.30
<i>Salmonella enterica</i>	NC_003197	59	0.60	0.35	0.36	0.25	0.69	0.30
<i>Yersinia pestis</i>	NC_003143	56	0.63	0.39	0.33	0.31	0.65	0.45
<i>Shewanella oneidensis</i>	NC_004347	62	0.71	0.27	0.31	0.27	0.49	0.328

<i>Neisseria meningitidis</i>	NC_003112	57	0.52	0.21	0.33	0.32	0.59	0.27
<i>Legionella pneumophila</i>	NC_002942	54	0.37	0.39	0.29	0.29	0.42	0.26
<i>Acidithiobacillus ferrooxidans</i>	NC_011761	56	0.15	0.30	0.15	0.20	0.24	0.31
<i>Campylobacter jejuni</i>	NC_002163	54	0.45	0.30	0.29	0.28	0.34	0.43
<i>Desulfovibrio vulgaris</i>	NC_002937	56	0.19	0.18	0.14	0.11	0.37	0.21
<i>Mycobacterium tuberculosis</i>	NC_000962	58	0.17	0.14	0.16	0.17	0.20	0.24
<i>Microcystis aeruginosa</i>	NC_010296.1	53	0.39	0.19	0.34	0.37	0.35	0.31
<i>Helicobacter pylori</i>	NC_000915	54	0.21	0.20	0.22	0.34	0.41	0.45
<i>Pseudomonas aeruginosa</i>	NC_002516	58	0.21	0.22	0.11	0.13	0.37	0.23
<i>Shigella flexneri</i>	NC_004337	53	0.62	0.17	0.47	0.20	0.63	0.37
<i>Synechocystis sp.</i>	NC_017277	52	0.32	0.26	0.35	0.18	0.38	0.33

---

a Number of 30S and 50S ribosomal protein genes.

b The +4U usage in 100 UAA-ending HEGs and LEGs (highest and lowest CAI scores, respectively).

## FIGURE LEGENDS

**Fig. S1.** Relationship between  $I_{TE}$  and usage of termination signals (stop codons and +4 bases), in *E. coli*. All non-pseudo, non-hypothetical CDSs were ranked by  $I_{TE}$  and binned into 10 sets. Twenty sequences in the 5' UTR were randomly shuffled using Sequence Manipulation Suite: Shuffle DNA (Stothard 2000). The stop codon usage and +4 base usage was obtained in each set. Stop codon usage ( $P_{UAA}$ ,  $P_{UAG}$ ,  $P_{UGA}$ ) is represented by solid lines; +4 base usage ( $P_A$ ,  $P_C$ ,  $P_G$ ,  $P_U$ ) is represented by dotted lines.

**Fig. S2.** Relationship between nc\_tRNA abundance and +4U usage, represented by linear regression between 100 UAA-ending HEGs (highest  $I_{TE}$  scores) and abundance of UAG nc\_tRNAs with a single mismatch at A) the first stop codon site, B) the second stop codon site, C) the third stop codon site, and abundance of UGA nc\_tRNAs with a single mismatch at D) the first stop codon site, E) the second stop codon site, F) the third stop codon site.

**Fig. S3.** The +4 base usage in all non-pseudo, non-hypothetical CDSs in 19 bacterial species (Table 1).

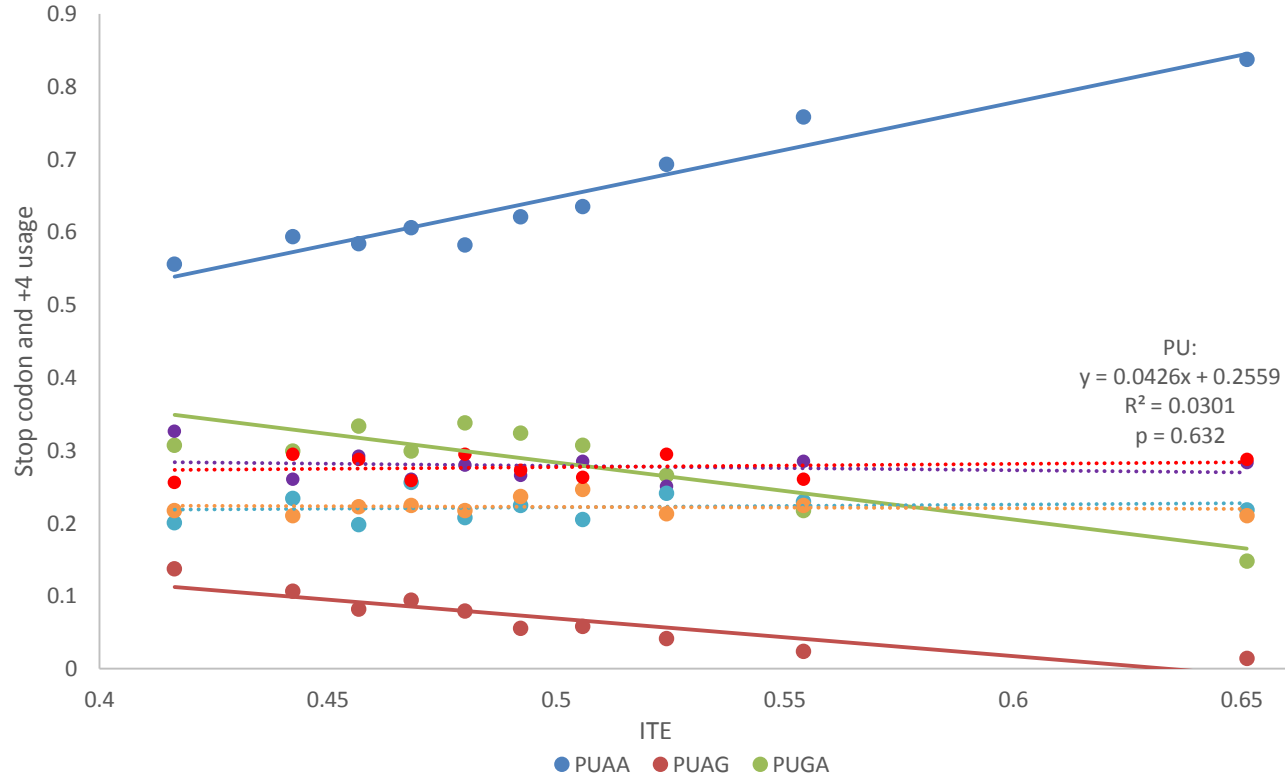


Fig. S1

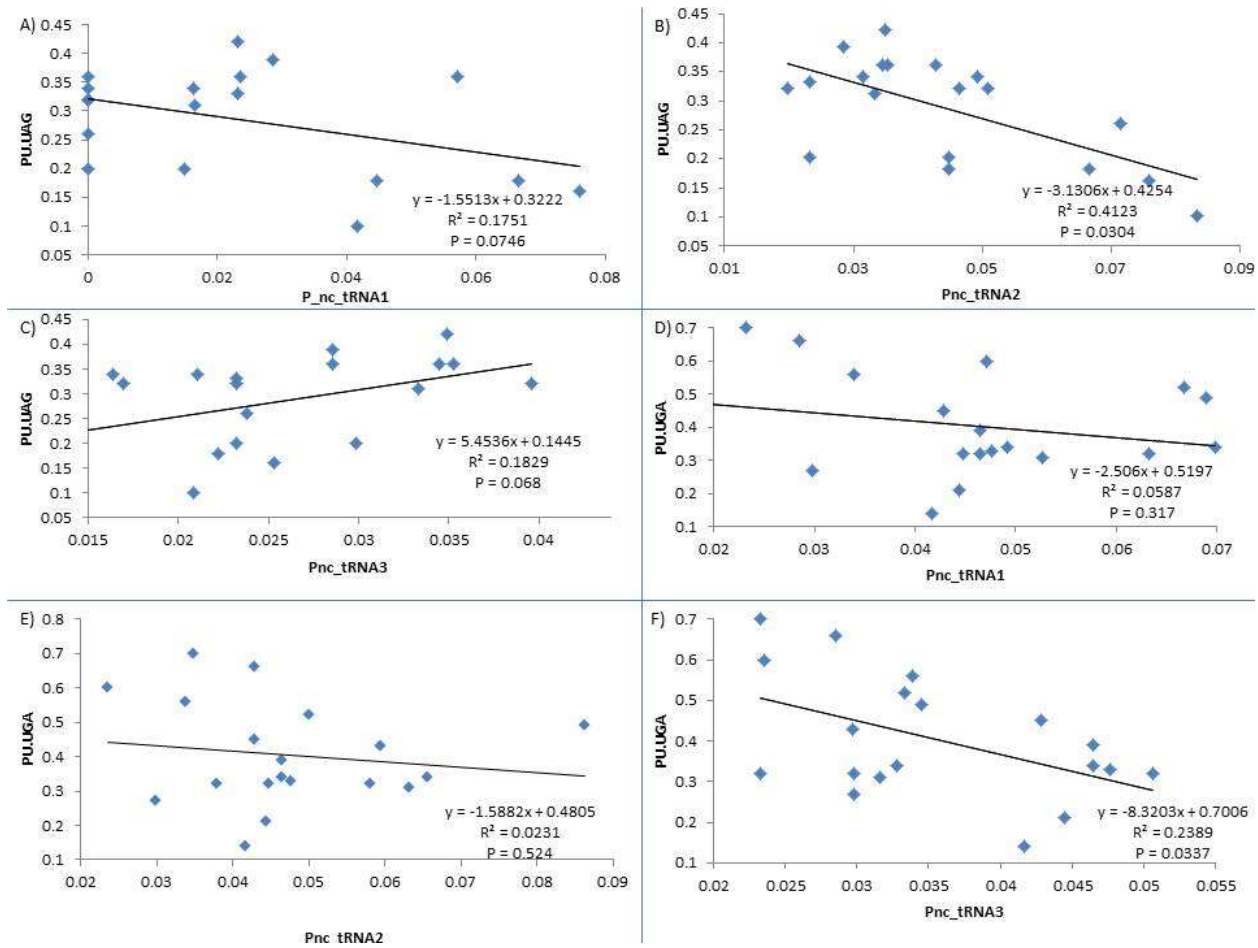


Fig. S2

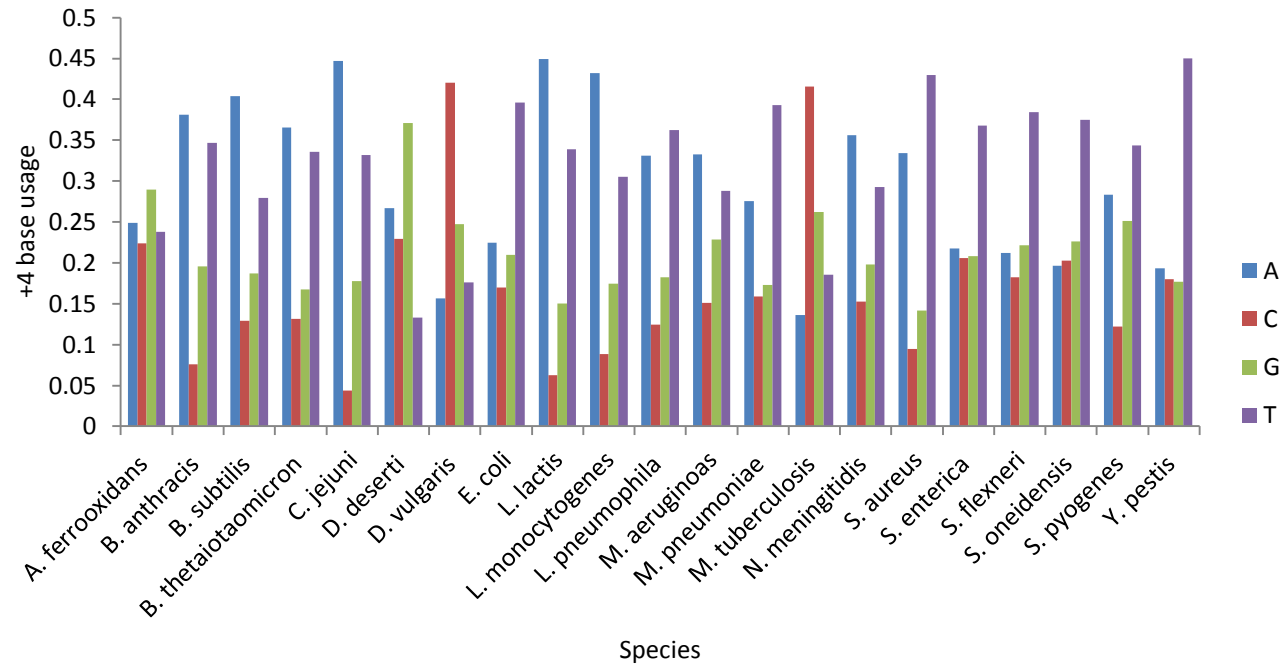


Fig. S3