Published in final edited form as:

Nat Biotechnol. 2020 August 01; 38(8): 989–999. doi:10.1038/s41587-020-0479-2.

Rapid discovery and evolution of orthogonal aminoacyl-tRNA synthetase–tRNA pairs

Daniele Cervettini¹, Shan Tang¹, Stephen D. Fried^{1,3}, Julian C. W. Willis^{1,4}, Louise F. H. Funke¹, Lucy J. Colwell², Jason W. Chin^{1,2,*}

¹Medical Research Council Laboratory of Molecular Biology, Cambridge, England, UK

²Department of Chemistry, Cambridge University, Cambridge, England, UK

Abstract

A central challenge in expanding the genetic code of cells to incorporate non-canonical amino acids into proteins is the scalable discovery of aminoacyl-tRNA synthetase (aaRS)–tRNA pairs that are orthogonal in their aminoacylation specificity. Here we computationally identify candidate orthogonal tRNAs from millions of sequences and develop a rapid, scalable approach – named tRNA Extension (tREX) – to determine the *in vivo* aminoacylation status of tRNAs. Using tREX, we test 243 candidate tRNAs in *Escherichia coli* and identify 71 orthogonal tRNAs, covering 16 isoacceptor classes, and 23 functional orthogonal tRNA–cognate aaRS pairs. We discover five orthogonal pairs, including three highly active amber suppressors, and evolve new amino acid substrate specificities for two pairs. Finally, we use tREX to characterize a matrix of 64 orthogonal synthetase–orthogonal tRNA specificities. This work expands the number of orthogonal pairs available for genetic code expansion and provides a pipeline for the discovery of additional orthogonal pairs and a foundation for encoding the cellular synthesis of non-canonical biopolymers.

Genetic code expansion enables the cellular synthesis of modified proteins via the co-translational incorporation of non-canonical amino acids^{1, 2}. Orthogonal aaRS–tRNA pairs are crucial to genetic code expansion. These pairs consist of (1) a synthetase that efficiently aminoacylates its cognate tRNA, but minimally aminoacylates endogenous tRNAs in the host organism, and (2) a tRNA that is a substrate for its cognate synthetase but is a poor substrate for endogenous synthetases^{3, 4}. Derivatives of orthogonal pairs that recognize blank codons (most commonly, the amber stop codon) and selectively use non-canonical amino acids (ncAAs) have been used to site-specifically incorporate numerous ncAAs into proteins^{3, 4}. Despite many years of effort, a limited number of orthogonal aaRS–tRNA pairs

*Correspondence: chin@mrc-lmb.cam.ac.uk.

³Present address: Department of Chemistry, Johns Hopkins University, Baltimore, Maryland, USA

⁴Present address: Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

Author contributions

D.C. designed and implemented tREX. D.C., S.T., J.C.W.W. and L.F.H.F performed the tREX screening. D.C. and S.T. performed tRNA and synthetases characterisation and engineering. S.D.F. and D.C. performed tRNA sequence analysis, with initial input for L.J.C. J.W.C. set the direction of research. J.W.C and D.C. wrote the manuscript with input from the other authors.

Competing interests

The authors declare no competing interests.

have been described^{5–21}. The current orthogonal pairs may limit the efficiency and scope of ncAA incorporation, the incorporation of multiple ncAAs and progress towards the encoded cellular synthesis of non-canonical biopolymers¹. The discovery of new orthogonal pairs may enable an increase in efficiency or scope of ncAA incorporation. Moreover, the discovery of additional aaRS–tRNA pairs that are orthogonal with respect to both the host synthetases and tRNAs and each other will provide one key component required for the incorporation of multiple distinct ncAAs and the encoded cellular synthesis of non-canonical biopolymers¹. Genome sequencing and annotation efforts have provided the nucleotide sequences for millions of tRNA genes²², and genome sequences provide a rich source of potential orthogonal pairs. However, previous studies aiming to discover orthogonal pairs have been limited to investigating the properties of one pair or one isoacceptor class^{5–21}; we are not aware of any large-scale effort to identify new orthogonal aaRS–tRNA systems from genomic data.

Here we develop a pipeline (Fig. 1), with associated computational and experimental approaches, to identify orthogonal aaRS–tRNA pairs. We computationally filter 2,799,231 tRNA sequences²² to identify candidate orthogonal tRNAs. We develop a rapid, scalable approach, tREX, to empirically determine the *in vivo* expression and aminoacylation status of tRNAs and use this approach to experimentally test 243 candidate tRNAs. We thereby identify 71 tRNAs, covering 16 isoacceptor classes, which are expressed and orthogonal in *E. coli*. For 23 of these orthogonal tRNAs, we show that the cognate aaRS is functional in *E. coli*, and for five of these pairs, including three highly active amber suppressors, we show that the cognate synthetase is also orthogonal in *E. coli*. Moreover, we evolve new amino acid substrate specificities for two pairs and show that we can incorporate ncAAs with five times greater efficiency when using these new systems. Finally, we characterize a matrix of 64 orthogonal synthetase–orthogonal tRNA specificities and show that the five orthogonal synthetase–orthogonal tRNA pairs reported herein, and three widely used orthogonal pairs, are all orthogonal with respect to each other in their aminoacylation specificity.

Results

Candidate orthogonal tRNAs

Within a cell (and within cellular compartments in eukaryotes), the aaRS–tRNA pairs for each canonical amino acid are orthogonal to each other: each tRNA is aminoacylated by a single endogenous aaRS and is not a substrate to the other 19 synthetases in the cell, and each aaRS specifically recognizes its cognate isoacceptor tRNAs²³. The specificity of aaRS–tRNA recognition is believed to be mediated, in part, by identity elements^{23–26}, which are specific nucleotides within a tRNA sequence (Fig. 2a, Supplementary Table 1).

We scored the extent to which each class of *E. coli* isoacceptor tRNAs contain identity elements recognized by each *E. coli* synthetase using a simple system. For each of the identity elements for a given synthetase, a tRNA received a score of +1 if its nucleotide matched the synthetase's substrate tRNA at that identity element, and a score of -1 otherwise. We note that identity elements do not have equal importance in the recognition of *E. coli* tRNAs by *E. coli* synthetases, and that the relative importance of identity element nucleotide positions in heterologous tRNAs is unknown. The overall score for a tRNA with

respect to each *E. coli* synthetase was the mean of the scores across all the identity elements for the synthetase; thus all scores ranged between -1 and +1. For isoacceptor classes with multiple tRNAs, the comparison was made with each of the isoacceptor tRNAs individually and the scores were then averaged to obtain a single value. This process generated 400 scores: the higher the score, the more similar a tRNA is to the endogenous substrates of the synthetase at the identity elements (Supplementary Table 2).

E. coli tRNAs for amino acids with a single isoacceptor scored +1 for their cognate synthetase (Fig. 2b). However, *E. coli* tRNAs for arginine, glutamine, isoleucine, leucine and serine tRNAs had scores lower than +1, but greater than +0.5; these scores are less than +1 because there are multiple isoacceptors, for these amino acids, which diverge in sequence at the identity elements. Methionine isoacceptors have poor conservation of identity elements between the elongator and the initiator tRNAs and therefore scored less than +0.5 for their cognate synthetase^{22, 23}. Thus, the isoacceptor classes for all elongator tRNAs scored greater than +0.5 for their cognate synthetase. Furthermore, each tRNA had a score of +0.5 or less with the identity elements for each of the 19 non-cognate synthetases. Thus, across all 400 scores, we observed that functional interactions between the tRNAs and the endogenous synthetases correlated with a score of greater than +0.5, and that non-functional interactions of a tRNA with an *E. coli* synthetase correlated with a score of +0.5 or less (Fig. 2b).

tRNAs from heterologous sources that can be used for site-specific incorporation of ncAAs in *E. coli*, including the *Methanocaldococcus jannaschii* (*Mj*)-tRNA^{Tyr} (ref. ³) and the *Methanosarcina barkeri* (*Mb*)-tRNA^{Pyl} (ref. ⁴), are orthogonal with respect to all 20 endogenous aaRSs in *E. coli*. We scored these orthogonal tRNAs against the set of identity elements for each *E. coli* synthetase, generating 40 scores. Each tRNA scored less than +0.5 with every *E. coli* synthetase. Thus, the scoring system we have benchmarked with 400 *E. coli* aaRS-tRNA combinations correctly predicts the orthogonality of these tRNAs on the basis of their identity element scores alone (Fig. 2b, Supplementary Table 2).

Next, we scored the 2,799,231 tRNA sequences from bacteria, archaea, chloroplasts, and bacteriophages present in an established database of aligned tRNA sequences (Supplementary Table 3)²². Our goal was to enrich tRNAs that, on the basis of their identity element nucleotides, may be orthogonal to the 20 *E. coli* aaRSs. We applied our scoring system to the aligned tRNA sequences (Fig. 2a) and discarded any tRNA that had a score of greater than 0.0 with respect to the *E. coli* synthetase for the same amino acid as the heterologous tRNA. This provides a more stringent filter on the identity elements for the *E. coli* synthetase that uses the same amino acid and reduced the number of tRNA sequences for each amino acid by two to four orders of magnitude (Fig. 2c, Supplementary Table 4). We then chose 243 tRNAs that had low scores across the identity elements for all other *E. coli* synthetases and were distributed across isoacceptor classes and derived from organisms in different phyla for experimental characterization (Supplementary Table 5). Of these tRNAs, 86% (208/243) had scores of less than +0.5 for all *E. coli* synthetases (that is, their scores were as low as *E. coli* tRNAs or known orthogonal tRNAs across all identity elements); the remainder of the tRNAs, 14 % (35/243), had a score of 0.0 or less with the *E. coli* synthetase for the same amino acid but had one or two scores for other *E. coli* synthetases that were greater than or equal to +0.5.

A rapid and specific method to measure tRNA aminoacylation

Orthogonal tRNAs are not substrates for endogenous synthetases but can be charged by their cognate aaRSs. To enable the screening of many candidate tRNAs for orthogonality, we wanted to create a rapid screen for tRNA aminoacylation that (1) was independent of the anticodon, (2) was independent of the nature of the amino acid bound to the tRNA of interest and (3) was scalable for the investigation of many tRNAs in parallel.

To develop a technology for screening tRNA aminoacylation, we first designed cyanine-5 (Cy-5)-labelled fluorescent DNA probes. These probes selectively invade the acceptor stem helix of their target tRNA and anneal to their 3'-end. Probes were designed such that annealing to the tRNA would produce a 5'-overhang of single-stranded DNA (Fig. 3a). Addition of a fluorescently labelled DNA probe to RNA extracted from cells expressing *Mb*-tRNA^{Pyl} selectively detected this tRNA and did not detect *E. coli* tRNAs in the absence of *Mb*-tRNA^{Pyl} (Fig. 3b, lanes 3 and 7), confirming the specificity of the probe for the heterologous tRNA.

Next we showed that the fluorescently labelled DNA-tRNA hybrid could be quantitatively extended with the 3'-to-5' exonuclease-deficient Klenow fragment of *E. coli* DNA polymerase I (Klenow exo(-)), and that this extension led to a decreased electrophoretic mobility in native PAGE (Fig. 3b, lanes 7 and 8). We reasoned that, if aminoacylation of the tRNA protects its 3'-end from extension, it might be possible to read out the aminoacylation status of the tRNA via the ratio of extended to non-extended DNA-tRNA hybrid. However, under the pH conditions where the polymerase is active, we observed complete extension of the DNA-tRNA hybrid; this extension was independent of whether the amino acid substrate for PylRS (N^ε-boc-L-lysine, or BocK²⁷) was present in the cells (Fig. 3b, lanes 8 and 12).

To address the problem of coupling the aminoacylation status of the tRNA to extension of the DNA-tRNA hybrid, we took advantage of the well-established procedure for sodium periodate (NaIO₄) oxidation of vicinal diols to dialdehydes^{28, 29}. This reaction selectively oxidizes the 3'-diol of tRNAs, making it unavailable for extension. Treatment of the non-aminoacylated tRNA with NaIO₄ at pH 5, followed by hybridization of the fluorescently labelled DNA probe at pH 7.9 and addition of Klenow exo(-) polymerase, did not lead to extension of the DNA-tRNA hybrid (Fig. 3a and Fig. 3b, lane 10), in line with complete oxidation of the tRNA's 3'-diol. In contrast, when the oxidation, hybridization and extension procedure was repeated on a tRNA extract from cells grown in the presence of BocK, we observed the extension product (Fig. 3a and Fig. 3b, lane 14). We conclude that this procedure allows us to detect aminoacylation of a specific tRNA species. We surmise that this approach works because at low pH the aminoacylation of one of the hydroxyls on the tRNAs 3'-diol is stable and protects it from oxidation³⁰. Raising the pH after oxidation leads to hydrolysis of the labile aminoacyl-ester³¹ and provides the free hydroxyl at the tRNA 3'-end for extension.

Our initial experiments were performed at low levels of aminoacylation. Increasing the concentration of BocK added to cells from 1 to 4 mM led to an increase in the aminoacylated fraction we detected (Fig. 3c). The fraction of aminoacylation determined by our method was similar to that determined by northern blot (Supplementary Fig. 1). Our

method, which we named tRNA Extension (tREX), enables simple and rapid visual screening of tRNA aminoacylation and tRNA orthogonality. tREX is independent of the anticodon of the tRNA, is independent of the identity of the amino acid bound to the tRNA of interest and is scalable to the discovery of many orthogonal tRNAs (see below and Fig. 3).

Scalable identification of orthogonal tRNAs and active cognate synthetases by tREX

We tested 243 tRNAs, selected from our aligned database by using our identity element scoring system and covering all isoacceptor classes, for orthogonality with respect to *E. coli* synthetases (Fig. 3d-m, Supplementary Fig. 2-7, Supplementary Table 5). We cloned each tRNA gene into a high-copy number plasmid under the control of the strong *lpp* promoter and transformed the resulting plasmid into *E. coli*. We then extracted tRNAs, and tested the extract using tREX with a DNA probe complementary to the tRNA under investigation (sequences of the probes are reported in Supplementary Table 5). Specific detection of the tRNA in the extract, but not in extracts from cells lacking the heterologous tRNA gene, demonstrated that the heterologous tRNA was expressed in *E. coli* (Fig. 3d-m).

We created a control for the mobility of the non-aminoacylated tRNA by treating the extract with NaOH, which hydrolyses ester bonds between amino acids and tRNAs, before performing the tREX protocol. We created a control for the mobility of the aminoacylated species by following a variant of the tREX protocol with the NaIO₄ oxidation omitted (Fig. 3b lane 12, Fig. 3d-m). We defined tRNAs as orthogonal in our assay if cells expressing the tRNA produced a band resulting from tREX that co-migrated with the non-aminoacylated control and did not produce a band that co-migrated with the control for the aminoacylated tRNA.

From our screen, we identified 71 tRNAs, covering 16 isoacceptor classes, that were expressed and not measurably aminoacylated by *E. coli* (*Ec*) aaRSs (Supplementary Table 5, Supplementary Fig. 2-7); these tRNAs are orthogonal with respect to *Ec*-aaRSs in our assay. 97 tRNAs were expressed, but showed detectable aminoacylation by endogenous synthetases, and 75 tRNAs were not detected; this indicated that they were either not expressed, were unstable in *E. coli*, or did not hybridize to their cognate probe. Of the 168 tRNAs that we could detect in cells with our probes, 11 of 21 with one or two scores $>+0.5$ were orthogonal, while 60 of 147 with all scores $1+0.5$ were orthogonal.

We cloned codon-optimized genes for 59 of the 71 synthetases under the control of the constitutive *glnS* promoter in a plasmid containing their cognate orthogonal tRNA. tRNA extracts were then prepared from cells transformed with each plasmid and analysed by tREX (Fig. 3d-m, Supplementary Fig. 8-10). Upon coexpression of the cognate aaRS, we detected aminoacylation for 23 of the 59 tRNAs that were orthogonal in *E. coli* (Fig. 3d-m, Supplementary Fig. 8-10). The 23 orthogonal tRNA–active synthetase pairs we have discovered cover 10 of the 20 isoacceptor classes that enforce the canonical genetic code. The remaining synthetases may be poorly expressed or folded, or their cognate tRNA may lack important post-transcriptional modifications.

Orthogonality of synthetases in *E. coli*

Next, we used *in vitro* aminoacylation assays^{9–12, 15} to investigate the orthogonality of 15 of the aaRSs, that were active with their cognate orthogonal tRNAs, with respect to the endogenous tRNAs in *E. coli* (Fig. 4b–f, Supplementary Fig. 11). The 15 aaRSs we chose recognise 9 different amino acids (arginine, aspartic acid, glutamine, glutamic acid, glycine, histidine, isoleucine, proline and tyrosine). The *Mj*-TyrRS–*Mj*-tRNA^{Tyr}_{GUA} pair was characterised as a positive control, as this forms the basis of orthogonal pairs that have been extensively used for ncAA incorporation¹⁵. We observed more rapid aminoacylation by *Mj*-TyrRS of total tRNA from *E. coli* that expressed *Mj*-tRNA^{Tyr}_{GUA} than of total *E. coli* tRNA in the absence of *Mj*-tRNA^{Tyr}_{GUA}. This observation confirms that *Mj*-TyrRS exhibits specificity for its cognate tRNA over *E. coli* tRNAs (Fig. 4a). The *Archaeoglobus fulgidus* tyrosyl-tRNA synthetase (*Af*-TyrRS) and *Aifella pfennigii* histidinyI-tRNA synthetase (*Ap*-HisRS) were also specific for their cognate tRNAs over *E. coli* tRNAs in this assay; as judged by more rapid aminoacylation in the presence of their cognate tRNAs (Fig. 4b, c). The *Caldilinea aerophi* arginyl-tRNA synthetase (*Ca*-ArgRS) and the *Sorangium cellulosum* aspartyl-tRNA synthetase (*Sc*-AspRS) also displayed some specificity for their cognate tRNAs over *E. coli* tRNAs; as judged by more rapid aminoacylation in the presence of their cognate tRNAs (Fig. 4d, e). The other synthetases tested displayed similar increases in aminoacylation with and without the orthogonal tRNAs over the time course of the experiment; we conclude that they exhibit less discrimination between the orthogonal tRNA and *E. coli* tRNAs in this assay (Fig. 4f, Supplementary Fig. 11).

Our data provide support for the orthogonality of the *Af*-TyrRS–*Af*-tRNA^{Tyr}_{GUA} pair, the *Ap*-HisRS–*Ap*-tRNA^{His}_{GUG} pairs, and the *Ca*-ArgRS–*Ca*-tRNA^{Arg}_{GCG} pair, which showed similar aminoacylation selectivity to the *Mj*-TyrRS–*Mj*-tRNA^{Tyr}_{GUA} system. The *in vitro* assay may underestimate orthogonality because it does not capture the competitive aminoacylation of tRAs by their cognate synthetases that occurs in cells. Therefore, it remained possible that other pairs, particularly the *Sc*-AspRS–tRNA^{Asp}_{GUC} pair and the *Ilumatobacter nonamiensis* *In*-GlnRS–*In*-tRNA^{Gln}_{UUG} pair, were orthogonal *in vivo*.

Creating active and orthogonal amber suppressor synthetase–tRNA pairs

Most genetic code expansion experiments incorporate ncAAs in response to the amber stop codon. Therefore, we asked whether amber suppressors could be generated from the pairs we have identified as orthogonal in their aminoacylation specificity (Fig. 5, Supplementary Fig. 12–17). While the parent *Ca*-ArgRS–*Ca*-tRNA^{Arg}_{GCG} and *Ap*-HisRS–*Ap*-tRNA^{His}_{GUG} pairs were orthogonal in their aminoacylation specificity with respect to *E. coli* synthetases and tRNAs, converting the anticodons of *Ca*-tRNA^{Arg} and *Ap*-tRNA^{His} to CUA led to mis-aminoacylation of the resulting tRNAs by *E. coli* aaRSs (Supplementary Fig. 15). Subsequent directed evolution experiments did not lead to amber suppressor derivatives of these pairs.

Converting the anticodons of *In*-tRNA^{Gln}_{UUG} and *Sc*-tRNA^{Asp}_{GUC} to CUA generated tRNAs that were not substrates for either *E. coli* synthetases or the cognate synthetases of the parent tRNAs (Supplementary Fig. 12). By performing rounds of directed evolution, with libraries of mutations in the regions of synthetases that recognise the anticodon of their

cognate tRNAs (Supplementary Fig. 18), and libraries of mutations in the ten nucleotides surrounding the anticodon, we selected amber suppressor derivatives of these pairs (*In*-GlnRS(S9)–*In*-tRNA^{Gln(A1)}_{CUA} and *Sc*-AspRS(C4)–*Sc*-tRNA^{Asp(10)}_{CUA}) that were active and orthogonal (Fig. 5a-d, Supplementary Fig. 12, Supplementary Table 6).

To demonstrate that the active site of *Sc*-AspRS(C4) can be altered to accept new substrates, we identified the residues that interact with the aspartate side chain in the *E. coli* (*Ec*-)AspRS–aspartyl-AMP co-crystal structure³² (Supplementary Fig. 18) and generated a library where these residues (positions 223, 260, and 534 in *Sc*-AspRS(C4)) were randomised to all 20 canonical amino acids. We also mutated Arg536 to the 19 other amino acids to minimize the selection of variants that still used aspartic acid. We performed a positive selection on this library and identified mutants (Lys223Ala, Glu260Lys, Arg536(Gly/Ala/Ser/Asn)), that directed the incorporation of glutamic acid (Fig. 5e, f, Supplementary Table 7). These experiments demonstrated the evolution of the new pair we created for the site-specific incorporation of a new substrate.

Converting the anticodon of *Af*-tRNA^{Tyr}_{GUA} to CUA led to a tRNA that was aminoacylated by endogenous synthetases (Fig. 5g) with either glutamine, glutamic acid or lysine, as indicated by mass spectrometry (MS; Supplementary Fig. 17). To directly determine which amino acids are found on *Af*-tRNA^{Tyr}_{CUA} in *E. coli*, we selectively purified *Af*-tRNA^{Tyr}_{CUA} from cells, under conditions that preserved aminoacylation, by using a complementary biotinylated oligonucleotide. We then deacylated the tRNA and identified the amino acids released by liquid chromatography and mass spectrometry (LC-MS; Supplementary Fig. 17). These experiments identified lysine as the sole amino acid detected, implicating *Ec*-LysRS in the mis-aminoacylation of *Af*-tRNA^{Tyr}_{CUA} in the absence of its cognate synthetase.

Upon co-expression of *Af*-TyrRS with *Af*-tRNA^{Tyr}_{CUA}, the levels of GFP produced by read-through of a GFP^{150TAG} reporter (a sfGFP reporter³³ containing an amber stop codon at position 150) did not increase compared to expression of *Af*-tRNA^{Tyr}_{CUA} alone; however, MS now indicated the incorporation of tyrosine (Supplementary Fig. 17, Supplementary Table 7). These data demonstrate that *Af*-TyrRS outcompetes endogenous synthetases for aminoacylation of *Af*-tRNA^{Tyr}_{CUA}.

Next, we aimed to create a more active and orthogonal *Af*-TyrRS–*Af*-tRNA^{Tyr}_{CUA} pair. To achieve this, we needed to improve the aminoacylation of *Af*-tRNA^{Tyr}_{CUA} by *Af*-TyrRS, and to abolish mis-aminoacylation of *Af*-tRNA^{Tyr}_{CUA} with lysine in the absence of *Af*-TyrRS. We decided to address these challenges by using directed evolution. On the basis of the crystal structure of the closest homologue of *Af*-TyrRS in complex with its cognate tRNA (*Mj*-TyrRS³⁴) we identified Phe274, Leu298 and Asp299 in *Af*-TyrRS (Phe261, Met285 and Asp286 in the *Mj*-TyrRS), which surround position 34 of the tRNA, as targets for site saturation mutagenesis (Supplementary Fig. 18). We created a library of *Af*-TyrRS in which these amino acids were mutated to all 20 canonical amino acids.

To select for active and orthogonal *Af*-TyrRS–*Af*-tRNA^{Tyr}_{CUA} pairs from our library we created a new dual reporter (CAT^{112TAG}-GFP^{67TAG}, contained in the plasmid p15A-

CAT^{112TAG}-GFP^{67TAG}) in which the amber stop codon in the GFP gene replaces the codon for the tyrosine residue that forms the chromophore³⁵. This enables identification of active clones on chloramphenicol and the identification of clones incorporating tyrosine, rather than lysine, on the basis of their fluorescence.

We used the CAT^{112TAG}-GFP^{67TAG} reporter to identify an *Af*-TyrRS mutant, which we named *Af*-TyrRS(G5), with the following amino acid substitutions: Phe274Val, Leu298Gly and Asp299Arg. We co-selected an *Af*-tRNA^{Tyr}_{CUA} mutant (G63T) with *Af*-TyrRS(G5), named *Af*-tRNA^{Tyr(G5)}_{CUA}, that in comparison to *Af*-tRNA^{Tyr}_{CUA} exhibited reduced read-through of the amber codon in the absence of *Af*-TyrRS. This suggested that *Af*-tRNA^{Tyr(G5)}_{CUA} was a poorer substrate for endogenous synthetases. Moreover, the *Af*-TyrRS(G5)–*Af*-tRNA^{Tyr(G5)}_{CUA} pair was substantially more active than the *Af*-TyrRS–*Af*-tRNA^{Tyr}_{CUA} pair from which it was derived (Fig. 5g).

To further improve the orthogonality of *Af*-tRNA^{Tyr(G5)}_{CUA}, we generated a library of mutants that varied nucleotides at positions where *Af*-tRNA^{Tyr(G5)}_{CUA} has the same sequence as the *Ec*-tRNA^{Lys}, but differs from *Mj*-tRNA^{Tyr}_{CUA}, which is orthogonal in *E. coli* (namely U7•A66, C11•G24, U44 and G45) (Fig. 5h). After selection with the CAT^{112TAG}-GFP^{67TAG} reporter, we identified a new mutant (*Af*-tRNA^{Tyr(22)}_{CUA}) that was less active with endogenous synthetases than *Af*-tRNA^{Tyr(G5)}_{CUA}, but more active with *Af*-TyrRS(G5) than the tRNA from which it was derived (Fig. 5g).

Finally, to optimise the interaction between *Af*-TyrRS(G5) and the amber suppressor tRNA, we generated a library that randomised the five nucleotides on either side of the anticodon in *Af*-tRNA^{Tyr(22)}_{CUA} (Fig. 5h). After selection with the CAT^{112TAG}-GFP^{67TAG} reporter we identified a new tRNA (*Af*-tRNA^{Tyr(A01)}_{CUA}) with the following mutations: G29C•G41C, G30U•C40A, A31U•U39A, G37A. *Af*-tRNA^{Tyr(A01)}_{CUA} did not lead to read through of the amber stop codon, suggesting that it is not aminoacylated by the *Ec*-LysRS or any other endogenous synthetases. Moreover, the *Af*-TyrRS(G5)–*Af*-tRNA^{Tyr(A01)}_{CUA} pair was approximately 6-times more active than the original *Af*-TyrRS–*Af*-tRNA^{Tyr}_{CUA} pair (Fig. 5g). Remarkably, this pair was also approximately 5-times more active than the *Mj*-TyrRS–*Mj*-tRNA^{Tyr}_{CUA} pair commonly used for genetic code expansion² (Fig. 5g) and produced wild-type levels of GFP when used to suppress a TAG codon in the GFP gene (Supplementary Table 6).

Efficient ncAA incorporation with the evolved *Af*-TyrRS– *Af*-tRNA^{Tyr}_{CUA} pair

Because the *Af*-TyrRS(G5)–*Af*-tRNA^{Tyr(A01)}_{CUA} pair was orthogonal and exceptionally active, we investigated the directed evolution of this pair for efficient ncAA incorporation. On the basis of the crystal structure of *Af*-TyrRS³⁶, we generated a library that mutated residues Tyr36, His74, Gln116, Asp165 and Ile166 in *Af*-TyrRS(G5) (Supplementary Fig. 19). Our library mutated 4 of the 5 positions to all 20 amino acids, while at position 165 we excluded negatively charged amino acids from the library to minimize recognition of tyrosine.

To identify mutants of *Af*-TyrRS(G5) that could direct the incorporation of ncAAs, we created a new triple reporter in which the 3'-end of the GFP^{67TAG} gene was fused in frame to

the gene encoding the far-red fluorescent protein E2Crimson³⁷ (plasmid p15A-CAT^{112TAG}-GFP^{67TAG}E2Crimson) (Fig. 6a). This new reporter also maintained the CAT^{112TAG} gene to enable selection for active synthetases on chloramphenicol. Cells containing the *Mj*-TyrRS-*Mj*-tRNA^{Tyr}_{CUA} pair and this reporter exhibited green and red fluorescence (because installing tyrosine in response to the amber codon in GFP at position 67 enables chromophore formation of GFP and production of E2Crimson) and chloramphenicol resistance (Fig. 6a). In contrast, cells containing the *Mj*-*p*B(OH)₂PheRS-*Mj*-tRNA^{Tyr}_{CUA} pair³⁸ in the presence of *p*-borono-L-phenylalanine and the triple reporter exhibit red fluorescence and chloramphenicol resistance, but no green fluorescence (as amino acids other than tyrosine at position 67 in GFP do not form the canonical chromophore). The differential fluorescence profile of this reporter allows the identification of active synthetases that incorporate amino acids other than tyrosine.

We performed chloramphenicol selections using the *Af*-TyrRS(G5) library with the cognate *Af*-tRNA^{Tyr(A01)}_{CUA}, our new triple reporter, and several ncAAs (*p*-iodo-L-phenylalanine, *p*-azido-phenylalanine and *O*-methyl-L-tyrosine). We identified surviving clones exhibiting red fluorescence but little or no visible green fluorescence, and thereby rapidly identified *Af*-TyrRS(G5) mutants that incorporated *O*-methyl-L-tyrosine (Tyr36Ile, His74Leu, Gln116Asn, Asp165Phe and Ile166His), *p*-iodo-L-phenylalanine (Tyr36Ile, His74Leu, Gln116Glu, Asp165Thr and Ile166Gly; an additional Leu69Met substitution was subsequently identified by selection on a library generated by error prone PCR), and *p*-azido-L-phenylalanine (Tyr36Thr, His74Leu, Gln116Glu, Asp165Thr and Ile166Gly; an additional Asn190Lys substitution was subsequently identified by selection on a library generated by error prone PCR).

The evolved *Af*-TyrRS(G5)-*Af*-tRNA^{Tyr(A01)}_{CUA} variants incorporate their cognate ncAAs with high specificity (Fig. 6). *O*-methyl-L-Tyr was incorporated with an efficiency similar to that of the previously described pair derived from *Mj*-TyrRS-*Mj*-tRNA^{Tyr}_{CUA} (ref. 39). However, *p*-iodo-L-phenylalanine and *p*-azido-L-phenylalanine were incorporated with efficiencies similar to that of the parent *Af*-TyrRS(G5)-*Af*-tRNA^{Tyr(A01)}_{CUA} pair with tyrosine (Fig. 5g, Fig. 6), and with much higher efficiencies than seen with the previously described derivatives of the *Mj*-TyrRS-*Mj*-tRNA^{Tyr}_{CUA} pair for these ncAAs^{40, 41} (Fig. 6, Supplementary Table 6).

Eight mutually orthogonal aaRS-tRNA pairs

To assess whether the 5 orthogonal pairs developed in this work (*Ca*-ArgRS-*Ca*-tRNA^{Arg}_{GCG}, *Sc*-AspRS(C4)-*Sc*-tRNA^{Asp(10)}_{CUA}, *In*-GlnRS(S9)-*In*-tRNA^{Gln(A1)}_{CUA}, *Ap*-HisRS-*Ap*-tRNA^{His}_{GUG}, *Af*-TyrRS(G5)-*Af*-tRNA^{Tyr(A1)}_{CUA}) are orthogonal in their aminoacylation specificity with respect to each other and with respect to three previously described orthogonal pairs (*Ma*-PylRS-*Ma*-tRNA^{Pyl(6)}_{CUA} (ref. 20), *Mb*-PylRS-*Mb*-tRNA^{Pyl}_{CUA}, *Mm*-SepRS(v1.0)-*Mm*-tRNA^{Sep(v2.0)}_{CUA} (ref. 17, 42)) we performed tREX on all 64 combinations of these 8 orthogonal synthetases and 8 orthogonal tRNAs (Fig. 7). Our data demonstrate that all 8 pairs are orthogonal in their aminoacylation specificity with respect to *E. coli* pairs and to each other.

Discussion

We have developed a pipeline for the discovery of aaRS–tRNA pairs that are orthogonal in their aminoacylation specificity. Our method uses a simple identity element-based scoring system, which we benchmark on both *E. coli* tRNAs and known orthogonal tRNAs, to enrich for potentially orthogonal tRNAs. tRNAs have complex and variable three dimensional structures, and the contribution of *E. coli* identity elements to the recognition of heterologous tRNAs by *E. coli* synthetases may be modulated by other structural features; these include variable loops and additional identity elements or antideterminants, which lie outside the *E. coli* identity element positions²³. Future work may include our developing understanding of these factors to refine and improve orthogonal tRNA prediction.

Central to our approach is the tREX method we report for determining the aminoacylation status of specific tRNAs. Previous methods for measuring aminoacylation either (1) relied on northern blots to detect changes in tRNA mobility upon aminoacylation (this is exceptionally slow and low throughput, e.g.: approximately 1 d per gel, and some amino acids do not lead to sufficient mobility shifts to make this generally feasible)^{43–45} or (2) converted the anticodon of the tRNA to CUA to generate an amber suppressor that could be used in stop codon read-through assays^{7, 9}. However, for most tRNAs this simple modification of the anticodon affects the aminoacylation of the tRNA by its cognate synthetase, rendering the pair inactive (or weakly active), and/or leads to mis-aminoacylation by other synthetases; this is because the anticodon is recognized by the cognate synthetases for 17 of the 20 canonical amino acids. Our method allows us to rapidly measure the aminoacylation and orthogonality of tRNAs with their native anticodons, thereby uncoupling the ability to determine the aminoacylation status of a tRNA from the mutation of its anticodon. Moreover, our approach works for tRNAs aminoacylated with any amino acid. Thus, our method allows us to discover tRNAs that are orthogonal with respect to endogenous synthetases; these tRNAs may be subsequently evolved to read new codons, as exemplified herein by the generation of amber suppressors. We have demonstrated the broad utility and scalability of tREX by determining the aminoacylation status of 243 tRNAs and 59 tRNA–cognate synthetase combinations. The experimental data we have generated on tRNA orthogonality may enable the creation of improved computational predictions, which both weight the contributions of identity elements²³ and factor in the complex contributions of other parts of the tRNA molecule to orthogonality²³.

By using our pipeline, we have discovered 71 orthogonal tRNAs; thus, 29% of the tRNAs identified by our computational filter are expressed and orthogonal in *E. coli* (42% of tRNAs that we can detect in cells are orthogonal). We identified 5 aaRS–tRNA pairs that exhibit orthogonality in their aminoacylation specificity with respect to *E. coli* synthetases and tRNAs. For three of these pairs we generated amber suppressor derivatives, demonstrating that they can be altered to decode different codons. For two of these pairs we explicitly showed that we could alter the specificity of the active site to recognize new substrates, and for the evolved *Af*-TyrRS–*Af*-tRNA^{Tyr}_{CUA} pair we showed that we could incorporate several ncAAs with an efficiency 5 times greater than that mediated by derivatives of the *Mj*-TyrRS–*Mj*-tRNA^{Tyr}_{CUA} pair, which is commonly used for genetic code expansion. As several of the pairs we have developed recognize chemical structures that are distinct from

the aromatic amino acids and lysine derivatives that are common substrates for the currently prevalent genetic code expansion systems^{1, 2}, we anticipate that these pairs may enable an expansion in the range of ncAA chemistries that can be incorporated into proteins.

We demonstrated that 8 orthogonal pairs are mutually orthogonal in their aminoacylation specificity. To use these pairs together to encode up to eight distinct ncAAs in a single protein, it will be necessary to alter the pairs to decode distinct blank codons and alter the active sites of the pairs to selectively recognize distinct ncAA substrates.

Strategies analogous to those we have used here for creating amber suppressors may prove useful for redirecting these pairs to new codons. These codons may include (in addition to stop codons) quadruplet codons⁴⁶, codons containing non-natural bases⁴⁷ and sense codons in organisms with compressed genetic codes⁴⁸.

Because the majority of the new aaRS–tRNA pairs naturally recognize chemically divergent structures, it may be relatively simple to generate derivatives of each pair that recognize one non-natural substrate but do not recognize the non-natural substrates of the other pairs. Moreover, we have previously reported a scalable strategy for the explicit selection of aaRS variants that are mutually orthogonal in their non-natural substrate specificity⁴²; extensions of this approach should prove useful for generating active sites that selectively use one ncAA, but specifically exclude other substrates. We anticipate that combining the advances reported herein, with advances in generating blank codons^{46–48} and progress toward expanding the chemical scope of cellular ribosomes⁴⁹ may enable the encoded cellular synthesis of non-canonical biopolymers.

Online Methods

Materials

Arabinose, antibiotics, IPTG, liquefied phenol, chloroform, *p*-iodo-L-phenylalanine (CAS 24250-85-9, cat. No. I8757-1G), and *O*-methyl-L-tyrosine (CAS 6230-11-1, cat. No. 158259-1G) were purchased from Sigma-Aldrich. *p*-azido-L-phenylalanine (CAS 33173-53-4, cat. No. 4020250.0005) was purchased from Bachem. All ¹⁴C-labelled radiochemicals (arginine: cat. No. MC137; aspartic acid: cat. No. MC139; glutamic acid: cat. No. MC 156; glutamine: at. No MC1124; glycine: cat. No. MC163; isoleucine: cat. No. MC174; proline: cat. No. MC263; tyrosine: cat. No. MC275) and ³H-labelled histidine (cat. No. MT905) were purchased from Moravek Inc.

tRNA alignment

tRNA sequences were downloaded from tRNA-DB-CE together with the information about the secondary structure of the candidate tRNAs, which constitutes the basis for the alignment performed (<http://trna.ie.niigata-u.ac.jp/cgi-bin/trnadb/index.cgi>). The sequences were sorted on the basis of their isoacceptor classes. The database contains the following structural information:

1→7	Acceptor stem, first strand
8→9	Unpaired bases
10→13	D arm, first strand
14→21	D loop
22→25	D arm, second strand
26	Unpaired base
27→31	Anticodon stem, first strand
32→38	Anticodon loop
39→43	Anticodon stem, second strand
44→48	Variable loop
49→53	TΨC stem, first strand
54→60	TΨC loop
61→65	TΨC stem, second strand
66→72	Acceptor stem, second strand
73	Discriminator base

To convert this information into an alignment of tRNAs composed of canonical positions 1 to 76, the following procedure was performed: (1) for tRNAs with an 8-nucleotide acceptor stem, the first 7 base pairs (bp) were assigned to positions 1→7/66→72. The only *E. coli* tRNA with such a feature, tRNA^{Sec}, aligned in such a manner to the tRNA^{Ser} sequences, with which it shares the recognition by the seryl-tRNA synthetase; (2) if the D arm is formed of 3 bp, positions 13 and 22 were assigned to the first and last nucleotides of the D loop; (3) the D loop presents a very high degree of variability, and commonly lacks the two distinctive guanines at positions 18 and 19 found in *E. coli* tRNAs. For these positions, the alignment was manually generated as shown in Supplementary Table 3; (4) as the variable loop can have a very broad range of lengths, the positions 44→48 were assigned in the following order: 44→48→45→47→46. For loops longer than 5 nucleotides, the first two positions were numbered 44 and 45, while the last three were numbered 46→48; (5) positions 74→76 always have a CCA sequence, either because the tRNA gene natively has this sequence, or because the tRNA processing enzymes edit the original sequence and *a posteriori* add it. The sequence alignment of *E. coli* tRNAs is shown in Supplementary Table 2. Fifteen candidate tRNA sequences (Supplementary Table 4; Arg_01 to Arg_05, Arg_07, Arg_09, Arg_10, Tyr_04, Ser_2682, Ser_5752, Leu_12217, Leu_13688, Leu_10255 and Leu_8506) were manually aligned before the automatic method (described above) was developed. The scores for these tRNAs (Supplementary Table 5) are from the automatic alignment.

Computational analysis

Candidate tRNAs, aligned as described above, were compared individually to each of the distinct 48 *E. coli* isoacceptor tRNAs, which were also aligned as described above (Supplementary Table 2 –*E. coli* **Isoacceptors Alignment**). For each candidate tRNA (tRNA^X) and *E. coli* isoacceptor (for example, *E. coli* tRNA^{Ala}_{GGC}), the nucleotides of the candidate tRNA^X at the positions of the identity elements for the *E. coli* isoacceptor (for

example, positions 2, 3, 4, 20, 69, 70, 71 and 73 for *E. coli* alanine isoacceptors) were scored. Each position of tRNA^X was scored +1 if it matched the nucleotide at that position of the *E. coli* isoacceptor (for example, *E. coli* tRNA^{Ala}_{GGC}); otherwise it was scored -1. The final score for this pairwise comparison (tRNA^X to *E. coli* tRNA^{Ala}_{GGC}) was the average of the individual scores across the positions of the identity elements. The score of tRNA^X for *E. coli* AlaRS was the average score of tRNA^X for each of the *E. coli* tRNA^{Ala} isoacceptors (e.g. *E. coli* tRNA^{Ala}_{GGC} and tRNA^{Ala}_{TGC}). The complete list of identity elements for *E. coli* aaRSs used in this work is shown in Supplementary Table 1 (refs. ^{23–26}). After the scoring procedure, tRNAs with a score for the same *E. coli* aaRS as their isoacceptor class greater than 0.0 were filtered out. A subset of the remaining tRNAs were chosen for experimental characterisation (Supplementary Table 5). tRNAs for selenocysteine and natural suppressor tRNAs were not considered for experimental characterisation.

Plasmid generation

Our standard expression plasmid for tRNA/synthetase pairs consisted of the pBR322 origin of replication (high copy number); the constitutive *E. coli glnS* promoter controlling the expression of the aminoacyl-tRNA synthetase of interest followed by a spacer; the strong constitutive *E. coli lpp* promoter controlling the expression of the tRNA of interest and an *rnc* terminator following the tRNA. All plasmids were constructed starting from the plasmid containing PylRS and tRNA^{Pyl} from *M. barkeri* and removing the coding sequence of the synthetase from the start ATG codon to the last codon of the coding sequence (CDS; the TAA stop codon was left in place) by NEB HiFi DNA assembly (Cat. No. E2621L). The remaining plasmid containing only a tRNA was used as a template in which tRNA^{Pyl} was systematically replaced by the tRNA under investigation (tRNA^X). These plasmids were called pKW-(tRNA^X); for example, pKW-*Af*tRNA^{Tyr}. For tRNAs that were investigated together with their cognate synthetase, the synthetase was introduced where the PylRS CDS was located. These plasmids were called pRST-(tRNA^X)-aaRS (for example, pRST-*Af*-tRNA^{Tyr}-*Af*TyrRS). All tRNA libraries were generated on pKW plasmids, and all synthetase libraries were generated on pRST plasmids.

The reporter plasmid p15A-CAT^{112TAG}-GFP^{150TAG} contained the p15A origin of replication and a constitutively expressed tetracycline resistance gene, in addition to the constitutive positive selection marker CAT^{112TAG} (a chloramphenicol acetyltransferase gene interrupted by an amber stop codon at position 112) and the fluorescent reporter GFP^{150TAG} (the sfGFP gene containing an amber stop codon at position 150) under the control of the L-arabinose-inducible P_{BAD} promoter. This plasmid was used for selections, unless otherwise stated. The p15A-CAT^{112TAG}-GFP^{67TAG} reporter plasmid was derived from the p15A-CAT^{112TAG}-GFP^{150TAG} plasmid by mutating codon 150 to AAC (asparagine codon) and codon 67 to TAG. The p15A-CAT^{112TAG}-GFP^{67TAG}E2Crimson plasmid was derived from the p15A-CAT^{112TAG}-GFP^{67TAG} plasmid by fusing the GFP gene in frame with a linker and the E2crimson gene.

tRNA extraction

The tRNA extraction described below was adapted from a previous method⁵⁰. A pre-culture of the desired strain was incubated overnight at 37°C with shaking at 220 r.p.m. 2 ml of pre-

culture was diluted into 50 ml of pre-warmed rich medium (2xYT) and the cells were incubated at 37°C with shaking at 220 r.p.m. for 1–2h. At OD₆₀₀ = 0.5–1.0, cells were pelleted (5 min at 5,000 RCF). The cell pellet was resuspended in 800 µl of buffer D (NaOAc 50 mM pH 5, NaCl 150 mM, MgCl₂ 10 mM, EDTA 0.1 mM) and transferred to a 2 ml tube. Cells were pelleted (2 min at 5,000 RCF) and the pellet was resuspended in 450 µl of buffer D. 50 µl of liquefied phenol (90% phenol in water from Sigma Aldrich, cat. No. P9346) was added to the cells and the lysis was performed by head-over-tail rotation (15 r.p.m., 15 min, room temperature). The lysed cells were pelleted (25 min, >20,000 RCF, 4°C). ~500 µl of supernatant containing the RNAs was recovered and transferred to a clean tube. 500 µl of chloroform was added to the solution. The tube was thoroughly vortexed for 1 min until a cloudy emulsion was formed. The emulsion was separated by centrifugation (1 min at >20,000 RCF). The top layer containing the tRNAs was recovered (~480 µl). This solution was either frozen at –20°C or immediately processed with periodate as follows.

tREX probe design

The tREX probes were DNA probes composed of two sections. The 3'-section of the probes was the reverse complement of the sequence of the tRNA of interest between canonical position 45 (e.g: from the second nucleotide of the variable loop) and canonical position 76 (e.g: to the 3'-CCA end of the tRNA). The 5'-section of the probe was a poly A sequence whose length scaled with the length of the 3'-section as follows:

Length of the 3'-section	30 to 31	32 to 33	34 to 36	37 to 38	39 to 41	42 to 43	44 to 46	47 to 48	49 to 51
Number of As	12	13	14	15	16	17	18	19	20

Probes were purchased from Sigma-Aldrich as PAGE-purified oligonucleotides and were labelled at the 3'-OH with the Cy5 dye.

tREX protocol

The tRNA extract was aliquotted into 3 136-µl samples (A, B and C). Sample A (positive control for full extension) was brought to 160 µl with buffer D and the tRNAs were precipitated with 375 µl of absolute ethanol. After incubation (1h, 10°C), the sample was centrifuged (25 min at >16,100 RCF), the supernatant removed and the pellet was dried. The pellet was then resuspended in buffer D to a final concentration of 1 µg/µl, as measured by nanodrop.

Sample B (negative control for no extension) was deacylated by adding NaOH (8 µl, 300 mM, 42°C, 1h), then neutralized by addition of NaOAc (8 µl, 3M, pH 5) and oxidized with NaIO₄ (8 µl, 100 mM, 1h, 10°C). Next, 375 µl of absolute ethanol was added and the tRNAs were precipitated and resuspended as for sample A.

Sample C was brought to 152 µl with buffer D, then oxidized with NaIO₄ (8 µl, 100 mM, 1h, 10°C). Next, the tRNAs were ethanol precipitated with 375 µl of absolute ethanol and resuspended as for samples A and B.

The enzymatic reactions were assembled as follows: 2 µl of tRNA (1 µg/µl, from sample A, B or C, plus a control for probe specificity); 1 µl of dNTPs (10 mM each), 5 µl of NEBuffer 2.1, 1 µl of Cy5-labelled DNA probe 1 µM, 40.5 µl of ddH₂O. The control for probe specificity was tRNA extract from untransformed *E. coli* DH10b treated as sample C. The reactions were annealed in a thermocycler using the following settings: 95°C for 1 min, 70°C for 2 min, 50°C for 2 min, then 4°C. After annealing, 0.5 µl of Klenow Fragment (3'→5' exo -; NEB M0212S) was added to each reaction. The samples were incubated at 37°C for 20 min.

Next, each reaction was mixed with 50 µl of loading dye (8 M urea and 0.04% Orange G) and 10 µl was run on an 8% acrylamide (19:1) gel (1x TBE, 200 V, approximately 45 min). Gels were imaged on an RGB Typhoon Imager (635 nm).

Northern blot analysis of aminoacylated tRNAs

tRNA extracts containing the tRNA of interest were ethanol-precipitated (no oxidation was performed) and then resuspended in buffer D to a final concentration of 1 µg/µl, as assessed by nanodrop. 2 µg of the tRNA sample was run for 4.5 h at 4°C on a 15 cm long acidic urea PAGE (6.5% acrylamide 19:1, 100 mM NaOAc pH 5, 8 M urea, 0.1% TEMED and 0.1% (NH₄)₂S₂O₈) using 300 mM NaOAc pH 5 as running buffer and 12 W of constant power (~140 V, ~85 mA). The composition of the loading dye (2x) was 100 mM NaOAc pH 5, 8 M urea, 0.1% xylene cyanol FF and 0.1% bromophenol blue.

After electrophoresis, the gel was stained with SYBR Gold to identify the position of tRNAs. An appropriate section of the gel was cut out and transferred onto nylon membrane (Ambion® BrightStar®-Plus Positively Charged Nylon Membrane) using the iBlot™ DNA Transfer Stack for the iBlot® Dry Blotting System (the membrane contained in the transfer stack is replaced). The tRNAs were cross-linked to the membrane (Stratalinker® UV Crosslinker 2400), which was later immersed for 20 min in Ambion® ULTRAhyb®-Oligo buffer for blocking. The biotinylated DNA probe was then added to a final concentration of 3 ng/µl and hybridised overnight at 37°C (tRNA^{Pro} probe: 5'-TGGCGGAAACCCCGGGAATCTAACCCGGCT-3'). Excess probe was washed off using 2xSSC buffer and then blotting was then performed with the Thermo Scientific Pierce Chemiluminescent Nucleic Acid Detection Module.

Synthetase purification

Synthetases were cloned in a pET expression vector under the control of a T7 promoter. To allow for a one-step affinity purification, a StrepTag followed by a TEV cleavage site was both added after the initial methionine (e.g: MWSHPQFEKSENLYFQG[...], where the StrepTag sequence is underlined and the TEV cleavage site is in italics.).

Synthetase expression plasmids were transformed into *E. coli* BL21 (DE3) cells. One liter of cell culture was grown in 2xYT after overnight expression at 20°C with shaking at 220 r.p.m. and induction at OD₆₀₀ = 0.5 with 0.2 mM IPTG. Cells were harvested at 5,000 RCF for 10 min, resuspended in 2 volumes of wash buffer containing cComplete™ Protease Inhibitor Cocktail (Roche) and sonicated. The debris were sedimented at 18,000 RCF and the supernatant was mixed with 1 mg of avidin and loaded onto a StrepTrap HP column

(GE) for affinity purification of synthetase protein. The column was washed with wash buffer until the UV trace at 280 nm of the measured flow-through returned to baseline, and the column was then equilibrated in elution buffer minus desthiobiotin before the protein of interest was eluted with elution buffer and concentrated. The compositions of the buffers were as follows: Wash Buffer: 50 mM sodium phosphate buffer pH 8, 250 mM NaCl, 50 mM KCl, 2 mM MgCl₂, 1 mM DTT; Elution Buffer: 20 mM sodium phosphate buffer pH 7.2, 100 mM KCl, 2 mM MgCl₂, 1 mM DTT, 1.5 mM desthiobiotin.

***In vitro* aminoacylation**

¹⁴C-labelled amino acids were purchased from Moravek Inc. (500 µl solution, 100 µCi/ml, variable specific concentration). Each amino acid was then diluted with cold amino acid to a final concentration of 75 µCi/ml and a specific activity of 60 mCi/mmol (1.25 mM, Amino Acid Stock 5x). In the case of histidine, the amino acid stock 5x was prepared by diluting the ³H-labelled amino acid, purchased from the same supplier at a concentration of 1.0 mCi/mL, to a final concentration of 0.75 mCi/ml and a specific activity of 600 mCi/mmol (1.25 mM). The reaction was performed using Aminoacylation Buffer (10x, 200 mM HEPES pH 7.2, 500 mM KCl and 100 mM MgCl₂). tRNA stock was prepared at 2 µg/µl, while the Enzyme Stock was prepared at concentrations between 4 and 12.5 µM. Before the reaction is started, Whatman glass microfiber filters GF/B (21 mm, CAT No. 1821-021) were wetted with 250 µl of cold 5% trichloroacetic acid (TCA). For each enzyme, a 2x Master Mix (2xMM) was assembled with 32 µl of Aminoacylation Buffer 10x, 12.8 µl of DTT 20 mM, 64 µl of Amino Acid Stock 5x, 3.84 µl of Enzyme Stock and 15.36 µl of Water.

30 µl of the 2xMM was mixed with 37.5 µl of tRNA Stock 2x. 9 µl was spotted on a wet glass filter as time 0 min. The reaction was started by adding 6.5 µl of ATP 10 mM, then 10 µl aliquots were taken at time 0.5, 1, 2, 3 and 5 min and spotted on filters. The filters were transferred to a vacuum filtration apparatus and washed with 1 ml of 5% TCA followed by 4 ml of 1% TCA and 1 ml of 70% ethanol to remove the unbound amino acid. The filters were then transferred to the appropriate vial for liquid scintillation counting. To measure total counts, 10 µl of each reaction was transferred to a dry filter and directly transferred to a vial for liquid scintillation counting without any washing.

tRNA purification and amino acid analysis

A biotinylated DNA probe (100 µM) was added to the tRNA solution to a final concentration of 0.1 –1 µM. The mixture was heated to 75°C for 10 min to anneal the probe to the target tRNA. The extract was then incubated on ice and the DNA –tRNA hybrid captured on high-capacity streptavidin-coated agarose (twice the volume of 100 µM biotinylated probe used for annealing was added). The solution was shaken (30 min, 4°C), then the beads were recovered and transferred to a 1 ml empty chromatographic spin column (Bio-rad, cat. No. 7326207). The beads were washed 5 times with 100 mM ammonium acetate pH 5, then 3 times with 20 mM ammonium acetate pH 5.

To hydrolyse the amino acid from tRNA, 200 µl of 20 mM ammonium carbonate pH 9.6 was added and the solution incubated (20 min, 37°C) before collection. To maximize recovery of

the amino acid, the beads were further washed with 200 µl of water, 200 µl of 20 mM ammonium bicarbonate pH 3 (x2), then 200 µl of water (x2). All the washes were combined and lyophilized until dry. The amino acid was dissolved in 40 µl of water:methanol (40:60) solution, and the solution was centrifuged (30 min, 21,000 RCF, 4°C) to sediment precipitation. The top 20 µl of the solution was transferred into a 250 µl glass insert (Agilent) for MS analysis.

All amino acid samples were analysed on an Agilent 1260 Infinity equipped with an Agilent 6130 Quadrupole LC –MS unit. A HILIC-Z column (4.6 x 150 mm) equipped with a guard column (Agilent) with 0.5 ml/min flow rate was used to elute amino acids. Buffer A (10 mM ammonium formate in water) and buffer B (acetonitrile:H₂O 9:1 (v/v) solution with 10 mM ammonium formate) were used for RP –HPLC. 5 µl of amino acid-containing solution was injected and eluted using a gradient of 100% buffer B to 70% buffer B in buffer A over 10 min at 30°C. The mass spectrometer was set to selected-ion monitoring (SIM) mode. Pure amino acids, purchased from Sigma-Aldrich, were run as standards for comparison.

GFP expression and mass spectrometry

The gene encoding sfGFP was contained in a plasmid with the p15A origin of replication under the control of the L-arabinose-inducible P_{BAD} promoter. The resulting protein had the following sequence, with the asterisk at position 150 corresponding to the amber stop codon used for ncAA incorporation:

```

1  MYSKGEELFT  GVPILVELD  GDVNGHKFSV  RGEGEDATN  GKLTCLKFICT  TGKLPVPWPT
61  LVTTLTYGVQ  CFSRYPDHMK  RHDFFKSAMP  EGYVQERTIS  FKDDGTYKTR  AEVKFEGDTL
121 VNRIELKGID  FKEDGNILGH  KLEYNFNSH*  VYITADKQKN  GIKANFKIRH  NVEDGSVQLA
181 DHYQQNTPIG  DGPVLLPDNH  YLSTQSVLSK  DPNEKRDHNV  LLEFVTAAGI  THGMDELYKG
241 SHHHHHH-

```

E. coli DH10b cells harbouring this sfGFP-containing plasmid were transformed by heat shock (42°C, 50 s, 150 ng of DNA, 25 µl of cells) with a plasmid that constitutively expressed tRNA –synthetase pair (pRST). The plasmid contained the pMB1 origin of replication. After transformation, cells were recovered for 1 h (1 ml SOC medium, 37°C, 850 r.p.m.). After recovery, the transformation was diluted 100-fold in 2xYT containing antibiotics for selection of transformants, 0.2% L-arabinose and the appropriate ncAA (2 mM), when required. After overnight expression, 5 to 50 ml of cells were collected by centrifugation (5,000 RCF, 10 min), the supernatant was discarded and the pellet was resuspended in 1 ml of lysis buffer (BugBuster™ Protein Extraction Reagent supplemented with 20 mM Tris pH 8.0, 500 mM NaCl and 40 mM imidazole) and incubated for 15 min on head-over-tail rotation at room temperature. The crude lysate was centrifuged at >16,100 RCF for 20 min at 4°C, the supernatant was recovered and 50 –100 µL of a 1:1 slurry of Ni-NTA agarose beads (Qiagen) was added. The beads were recovered by centrifugation at 200 RCF and washed 5 times with 1 ml of wash buffer (20 mM Tris pH 8.0, 500 mM NaCl and 40 mM imidazole), then sfGFP was eluted using 50 µl of elution buffer (50 mM Tris pH 8.0, 50 mM NaCl, 300 mM imidazole).

Mass spectra of all protein samples were acquired on an Agilent 1200 LC –MS system equipped with a 6130 Quadrupole spectrometer. A Phenomenex Jupiter C4 column (150×2 mm, 5 µm) was used to elute proteins. Buffer A (0.2% formic acid in H₂O) and buffer B (0.2% formic acid in acetonitrile (MeCN)) were used for RP –HPLC. Mass spectra were acquired in the positive mode and analysed with the MS Chemstation software (Agilent Technologies). The deconvolution program provided in the software was used to obtain the entire mass spectra.

GFP expression for fluorescence quantification

E. coli DH10b cells containing the gene encoding sfGFP with an in-frame amber stop codon at position 150 (see above) were transformed with a pWK or pRST plasmid by heat shock. Recovery medium was added (1 ml, SOC medium), and the transformation was divided into 3 aliquots. After recovery (1 h, 37°C, 850 r.p.m.), 50 µl of each transformation was added to a different well of a 24 deep well plate (Riplate®SW 24, PP, 10 ml) containing 5 ml of 2xYT medium supplemented with the appropriate antibiotics for selection, 0.2% arabinose to induce expression of GFP and 2 mM of ncAA, as necessary. Cells were incubated at 37°C (22 h, 220 rpm), then 500 µl of medium was transferred to clear-bottom 24-well plates. The GFP fluorescence in this standard volume was measured with the PHERAstar FS (BMG Labtech) plate reader using the optic module with an excitation wavelength of 485 nm and an emission wavelength of 520 nm. The gain was set to 8 (arbitrary units). Measurements were performed on each sample grown in individual wells. In each graph, individual data points are shown together with the standard deviation.

Site-saturation mutagenesis

Site saturation mutagenesis on aaRSs or tRNAs was carried out by enzymatic inverse PCR (eiPCR)⁵¹. Primers were designed to contain the BsaI or SapI restriction site. PCRs were performed with Q5® Hot Start High-Fidelity DNA polymerase (NEB) or alternatively with PrimeSTAR® GXL DNA polymerase (TaKaRs). Libraries were produced to contain >2x10⁸ individual transformants. Sanger sequencing of the library pool and/or a small number (fewer than ten) of individual clones was used to confirm that the libraries were randomized at the desired positions; these experiments do not allow quantitative assessment of potential biases in the library sequences.

Mutagenesis by error-prone PCR

The gene encoding the synthetase targeted for evolution was amplified by PCR using primer immediately flanking the CDS. The DNA obtained was subsequently purified and used as a template for random mutagenesis by error-prone PCR, which was carried out with the GeneMorph II Random Mutagenesis kit (Agilent Technologies). Amplification was performed according to the manufacturer's instructions to achieve a medium mutation frequency (4.5-9 mutations/kbp). 30 cycles were performed with 250 ng of initial target DNA. Random mutagenesis libraries were generated by assembly of the mutagenised synthetase with the plasmid backbone, which was amplified with Q5® Hot Start High-Fidelity DNA Polymerase (NEB, Cat. No. M0493S). Assembly was performed for 1 h at 50°C with NEBuilder HiFi DNA Assembly Master Mix, followed by transformation into homemade electrocompetent cells. The transformation efficiency was >3x10⁷.

aaRS library selections

pRST plasmids containing the synthetase of interest together with its cognate tRNA were used to build the libraries. Ten aliquots of 1 µg of library DNA were transformed into ten aliquots of 100 µl of homemade electrocompetent *E. coli* DH10b cells containing an appropriate reporter plasmid (for example, p15A-CAT^{112TAG}-GFP^{150TAG}, p15A-CAT^{112TAG}-GFP^{67TAG} or p15A-CAT^{112TAG}-GFP^{67TAG}E2Crimson). To ensure high electroporation efficiency, electrocompetent cells were freshly prepared by diluting an overnight preculture of the appropriate cells 25-fold in 2xYT medium containing the appropriate antibiotics. Cells were grown to OD₆₀₀ = 0.6–0.8. Cells were then collected and resuspended in ice-cold 10% glycerol in water; this procedure was repeated 3 times. Finally, the cells were collected and the cell pellet was resuspended with 10% glycerol in water (2 times the pellet volume). After electroporation, cells were recovered (1 h at 37°C, SOC medium) and then were incubated overnight (37°C, 1 liter of 2xYT, 220 r.p.m.) in medium contained the appropriate antibiotics (commonly, tetracycline to select for the reporter plasmid and spectinomycin to select for the library plasmid). An appropriate dilution of this medium was plated on agar plates to ensure the transformation had yielded >5x10⁸ individual transformants (for example, >500 colonies on the 10⁻⁶ dilution plate). The overnight culture of library-containing transformants was then diluted 25-fold in 2xYT and allowed to grow to OD₆₀₀ = 1; if necessary, the appropriate ncAA was added to a concentration of 2 mM. When OD₆₀₀ = 1 was reached, 2 ml of cells were plated on 25 x 25 cm² agar plates containing the appropriate antibiotics to select for the markers on the plasmids, together with chloramphenicol, 0.2% arabinose to induce expression of the fluorescent markers (e.g. GFP or E2Crimson), and the appropriate ncAA (2 mM) if needed. The chloramphenicol concentration was chosen in accordance with the desired level of activity of the synthetase (range: 25–500 mg/l). The plates were incubated at 37°C overnight. For selections using the p15A-CAT^{112TAG}-GFP^{150TAG} or p15A-CAT^{112TAG}-GFP^{67TAG} reporter plasmid, surviving colonies were illuminated with blue light (485 nm) to test for GFP production. For selections using the p15A-CAT^{112TAG}-GFP^{67TAG}E2Crimson reporter plasmid, surviving colonies were illuminated with blue light (485 nm) to test for GFP production and with red light (635 nm) to test for E2Crimson production. Individual colonies displaying the correct fluorescence pattern were grown and characterised, or the surviving cells were pooled together and the DNA was extracted and transformed again for another round of selection or for a round of activity screening.

tRNA library selections for improved orthogonality and activity

tRNA selections were performed with a two-plasmid system. A pKW plasmid encoding the tRNA was used to build the library via site-saturation mutagenesis. A second plasmid was a modified version of the appropriate reporter plasmid (for example, p15A-CAT^{112TAG}-GFP^{150TAG} or p15A-CAT^{112TAG}-GFP^{67TAG}), which additionally encoded the synthetase.

1 µg of library DNA was transformed into homemade electrocompetent *E. coli* DH10b cells (0.1 ml) containing the appropriate synthetase-encoding reporter plasmid. After recovery (1 h at 37°C, SOC medium 220 r.p.m.), the cells were incubated overnight (37°C, 100 ml of 2xYT, 220 rpm). The medium contained the appropriate antibiotics (commonly, tetracycline to select for the reporter and spectinomycin to select for the library plasmid). An appropriate

dilution (for example, 10^{-5}) of the transformation was plated on agar plated to ensure the transformation efficiency was enough to yield $>5 \times 10^7$ individual transformants. The overnight culture was then diluted 25-fold in 2xYT and allowed to grow to $OD_{600} = 1$. When $OD_{600} = 1$ was reached, 2 ml of cells were plated on 25 x 25 cm² agar plates containing the appropriate antibiotics to select for the markers on the plasmids, together with chloramphenicol and 0.2% arabinose to induce expression of GFP. The chloramphenicol concentration was chosen in accordance with the desired level of activity of the synthetase (range: 25-500 mg/l). The plates were incubated at 37°C overnight. Surviving colonies were illuminated with blue light to test for GFP production. DNA was extracted either from individual GFP-positive colonies, or from GFP-positive colonies pooled together.

The reporter plasmid was selectively digested with an appropriate restriction enzyme and T5 exonuclease treatment. The now isolated tRNA library plasmid was purified again and transformed into cells containing the appropriate selection plasmid (for example, p15A-CAT^{112TAG}-GFP^{150TAG} or p15A-CAT^{112TAG}-GFP^{67TAG}) lacking the cognate synthetase, and the transformants were plated on agar plates containing 0.2% arabinose for GFP screening in the absence of chloramphenicol. GFP-negative colonies contained a variant of the tRNA that was selectively aminoacylated to a desired level in the presence of the synthetase and not aminoacylated in its absence, thus being more active and orthogonal.

Statistics and reproducibility

The number of replicates and types of replicates performed are described in the legend to each figure. Individual data points are shown and, where relevant, the mean $\pm$ s.d. is shown; this information is provided in each figure legend. No statistical tests were needed or performed.

Life sciences reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the Medical Research Council (MRC), UK (MC_U105181009 and MC_UP_A024_1008), and an ERC Advanced Grant SGCR, all to J.W.C. We thank W. Schmied, W. Robertson and R. Hegde for helpful discussions.

Data availability

The tREX screening data used in this study are available in Supplementary Fig. 2–10. Supplementary Fig. 2–7: tREX screening for tRNA orthogonality in *E. coli*. Supplementary Fig. 8–10: tREX screening for aaRS activity on their cognate tRNA. Supplementary Table 4: complete list of tRNAs generated by the filter described. Supplementary Table 5: tRNAs which were selected for experimental investigation including tRNA accession number on

tRNA-DB-CE and cognate synthetase accession number on NCBI Protein, sequence of corresponding tREX probe. All other datasets and material generated or analysed in this study are available from the corresponding author upon reasonable request.

References

1. Chin JW. Expanding and reprogramming the genetic code. *Nature*. 2017; 550:53–60. [PubMed: 28980641]
2. Young DD, Schultz PG. Playing with the Molecules of Life. *ACS Chem Biol*. 2018; 13:854–870. [PubMed: 29345901]
3. Liu CC, Schultz PG. Adding new chemistries to the genetic code. *Annu Rev Biochem*. 2010; 79:413–444. [PubMed: 20307192]
4. Chin JW. Expanding and reprogramming the genetic code of cells and animals. *Annu Rev Biochem*. 2014; 83:379–408. [PubMed: 24555827]
5. Edwards H, Schimmel P. An *E coli* aminoacyl-tRNA synthetase can substitute for yeast mitochondrial enzyme function in vivo. *Cell*. 1987; 51:643–649. [PubMed: 3315228]
6. Chin JW, et al. An expanded eukaryotic genetic code. *Science*. 2003; 301:964–967. [PubMed: 12920298]
7. Wang L, Magliery TJ, Liu DR, Schultz PG. A New Functional Suppressor tRNA/Aminoacyl-tRNA Synthetase Pair for the in Vivo Incorporation of Unnatural Amino Acids into Proteins. *Journal of the American Chemical Society*. 2000; 122:5010–5011.
8. Ambrogelly A, et al. Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. *Proceedings of the National Academy of Sciences*. 2007; 104:3141–3146.
9. Park H-S, et al. Expanding the Genetic Code of *Escherichia coli* with Phosphoserine. *Science*. 2011; 333:1151–1154. [PubMed: 21868676]
10. Santoro SW, Anderson JC, Lakshman V, Schultz PG. An archaeobacteria - derived glutamyl - tRNA synthetase and tRNA pair for unnatural amino acid mutagenesis of proteins in *Escherichia coli*. *Nucleic Acids Research*. 2003; 31:6700–6709. [PubMed: 14627803]
11. Pastrnak M, Magliery TJ, Schultz PG. A New Orthogonal Suppressor tRNA/Aminoacyl-tRNA Synthetase Pair for Evolving an Organism with an Expanded Genetic Code. *Helvetica Chimica Acta*. 2000; 83:2277–2286.
12. Anderson JC, Schultz PG. Adaptation of an Orthogonal Archaeal Leucyl-tRNA and Synthetase Pair for Four-base, Amber, and Opal Suppression. *Biochemistry*. 2003; 42:9598–9608. [PubMed: 12911301]
13. Anderson JC, et al. An expanded genetic code with a functional quadruplet codon. *Proc Natl Acad Sci U S A*. 2004; 101:7566–7571. [PubMed: 15138302]
14. Chatterjee A, Xiao H, Schultz PG. Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in *Escherichia coli*. *Proceedings of the National Academy of Sciences*. 2012; 109:14841–14846.
15. Steer BA, Schimmel P. Major Anticodon-binding Region Missing from an Archaeobacterial tRNA Synthetase. *Journal of Biological Chemistry*. 1999; 274:35601–35606.
16. Neumann H, Peak-Chew SY, Chin JW. Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat Chem Biol*. 2008; 4:232–234. [PubMed: 18278036]
17. Rogerson DT, et al. Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. *Nat Chem Biol*. 2015; 11:496–503. [PubMed: 26030730]
18. Hughes RA, Ellington AD. Rational design of an orthogonal tryptophanyl nonsense suppressor tRNA. *Nucleic Acids Res*. 2010; 38:6813–6830. [PubMed: 20571084]
19. Chatterjee A, Xiao H, Yang PY, Soundararajan G, Schultz PG. A tryptophanyl-tRNA synthetase/tRNA pair for unnatural amino acid mutagenesis in *E. coli*. *Angew Chem Int Ed Engl*. 2013; 52:5106–5109. [PubMed: 23554007]
20. Willis JCW, Chin JW. Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs. *Nat Chem*. 2018; 10:831–837. [PubMed: 29807989]

21. Wu N, Deiters A, Cropp TA, King D, Schultz PG. A genetically encoded photocaged amino acid. *J Am Chem Soc.* 2004; 126:14306–14307. [PubMed: 15521721]
22. Abe T, et al. tRNADB-CE: tRNA gene database well-timed in the era of big sequence data. *Frontiers in Genetics.* 2014; 5
23. Giegé R, Sissler M, Florentz C. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Research.* 1998; 26:5017–5035. [PubMed: 9801296]
24. Aldinger CA, Leisinger AK, Igloi GL. The influence of identity elements on the aminoacylation of tRNA(Arg) by plant and *Escherichia coli* arginyl-tRNA synthetases. *FEBS J.* 2012; 279:3622–3638. [PubMed: 22831759]
25. Larkin DC, Williams AM, Martinis SA, Fox GE. Identification of essential domains for *Escherichia coli* tRNA(leu) aminoacylation and amino acid editing using minimalist RNA molecules. *Nucleic Acids Res.* 2002; 30:2103–2113. [PubMed: 12000830]
26. Quinn CL, Tao N, Schimmel P. Species-specific microhelix aminoacylation by a eukaryotic pathogen tRNA synthetase dependent on a single base pair. *Biochemistry.* 1995; 34:12489–12495. [PubMed: 7547995]
27. Nguyen DP, Garcia Alai MM, Kapadnis PB, Neumann H, Chin JW. Genetically encoding N(epsilon)-methyl-L-lysine in recombinant histones. *J Am Chem Soc.* 2009; 131:14194–14195. [PubMed: 19772323]
28. Zamecnik PC, Stephenson ML, Scott JF. PARTIAL PURIFICATION OF SOLUBLE RNA. *Proc Natl Acad Sci U S A.* 1960; 46:811–822. [PubMed: 16590676]
29. Rizzino AA, Freundlich M. Estimation of in vivo aminoacylation by periodate oxidation: tRNA alterations and iodate inhibition. *Analytical Biochemistry.* 1975; 66:446–449. [PubMed: 166571]
30. Lodemann E, Niedenthal I, Wacker A. Influence of pH on the stability of some aminoacyl transfer ribonucleic acids and their elution pattern in chromatography on columns of methylated albumin adsorbed on kieselguhr. *Z Naturforsch B.* 1970; 25:845–848. [PubMed: 4394218]
31. Hentzen D, Mandel P, Garel J-P. Relation between aminoacyl-tRNA stability and the fixed amino acid. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis.* 1972; 281:228–232.
32. Eiler S, Dock-Bregeon A, Moulinier L, Thierry JC, Moras D. Synthesis of aspartyl-tRNA(Asp) in *Escherichia coli*—a snapshot of the second step. *EMBO J.* 1999; 18:6532–6541. [PubMed: 10562565]
33. Pedelacq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol.* 2006; 24:79–88. [PubMed: 16369541]
34. Kobayashi T, et al. Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nat Struct Biol.* 2003; 10:425–432. [PubMed: 12754495]
35. Wang L, Xie J, Deniz AA, Schultz PG. Unnatural Amino Acid Mutagenesis of Green Fluorescent Protein. *The Journal of Organic Chemistry.* 2003; 68:174–176. [PubMed: 12515477]
36. Kuratani M, et al. Crystal structures of tyrosyl-tRNA synthetases from *Archaea*. *J Mol Biol.* 2006; 355:395–408. [PubMed: 16325203]
37. Strack RL, et al. A rapidly maturing far-red derivative of DsRed-Express2 for whole-cell labeling. *Biochemistry.* 2009; 48:8279–8281. [PubMed: 19658435]
38. Brustad E, et al. A genetically encoded boronate-containing amino acid. *Angew Chem Int Ed Engl.* 2008; 47:8220–8223. [PubMed: 18816552]
39. Wang L, Brock A, Herberich B, Schultz PG. Expanding the genetic code of *Escherichia coli*. *Science.* 2001; 292:498–500. [PubMed: 11313494]
40. Xie J, et al. The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nat Biotechnol.* 2004; 22:1297–1301. [PubMed: 15378068]
41. Chin JW, et al. Addition of p-azido-L-phenylalanine to the genetic code of *Escherichia coli*. *J Am Chem Soc.* 2002; 124:9026–9027. [PubMed: 12148987]
42. Zhang MS, et al. Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. *Nat Methods.* 2017; 14:729–736. [PubMed: 28553966]

43. Varshney U, Lee CP, RajBhandary UL. Direct analysis of aminoacylation levels of tRNAs in vivo. Application to studying recognition of *Escherichia coli* initiator tRNA mutants by glutamyl-tRNA synthetase. *J Biol Chem.* 1991; 266:24712–24718. [PubMed: 1761566]
44. Stenum TS, Sorensen MA, Svenningsen SL. Quantification of the Abundance and Charging Levels of Transfer RNAs in *Escherichia coli*. *J Vis Exp.* 2017
45. Walker SE, Fredrick K. Preparation and evaluation of acylated tRNAs. *Methods.* 2008; 44:81–86. [PubMed: 18241790]
46. Neumann H, Wang K, Davis L, Garcia-Alai M, Chin JW. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature.* 2010; 464:441–444. [PubMed: 20154731]
47. Zhang Y, et al. A semi-synthetic organism that stores and retrieves increased genetic information. *Nature.* 2017; 551:644–647. [PubMed: 29189780]
48. Fredens J, et al. Total synthesis of *Escherichia coli* with a recoded genome. *Nature.* 2019; 569:514–518. [PubMed: 31092918]
49. Schmied WH, et al. Controlling orthogonal ribosome subunit interactions enables evolution of new function. *Nature.* 2018; 564:444–448. [PubMed: 30518861]
50. Brubaker LH, McCorquodale DJ. The Preparation of Amino Acid-Transfer Ribonucleic Acid from *Escherichia coli* by Direct Phenol Extraction of Intact Cells. *Biochim Biophys Acta.* 1963; 76:48–53. [PubMed: 14068560]
51. Stemmer WP, Morris SK. Enzymatic inverse PCR: a restriction site independent, single-fragment method for high-efficiency, site-directed mutagenesis. *Biotechniques.* 1992; 13:214–220. [PubMed: 1327007]

Fig. 1. Pipeline for identifying orthogonal aaRS–tRNA pairs.

(1) A list of candidate orthogonal tRNAs is identified computationally. (2) Candidate orthogonal tRNAs are expressed in *E. coli*. Orthogonal tRNAs that are both expressed and not aminoacylated by endogenous synthetases, are identified. (3) Cognate synthetases that aminoacylate the orthogonal tRNAs when coexpressed in *E. coli* are identified. (4) The orthogonality of the cognate synthetases with respect to *E. coli* tRNAs (black) is determined.

Fig. 2. Developing and applying a computational filter to identify candidate orthogonal tRNAs.

a) Secondary structure diagram for tRNAs showing the canonical numbering scheme.

Conserved nucleotide numbering is indicated. The identity elements for synthetases that

recognize each amino acid are colour-coded as indicated to the left of the secondary

structure diagram. **b)** Top: application of our scoring scheme to *E. coli* aaRSs and tRNAs.

The score for each *E. coli* synthetase with each set of isoacceptor tRNAs is shown in the 20

x 20 matrix. Bottom: the score for each *E. coli* synthetase with two commonly used

orthogonal tRNAs. **c)** Filtering tRNAs from bacteria, archaea, mitochondria, chloroplasts

and phage present in the database tRNA-DB-CE database. The number of tRNA sequences

for the indicated amino acid, as of March 2017, is represented by the grey bars. Removing

tRNAs that scored more than 0.0 for the *E. coli* synthetase for the same amino acid reduced

the number of candidates for experimental investigation by orders of magnitude (black bars).

Fig. 3. tREX enables rapid determination of tRNA aminoacylation status.

a) A total tRNA extract from cells contains a tRNA of interest (light blue). The fate of the tRNA of interest differs depending on whether it is aminoacylated with an amino acid (yellow hexagon) or not. Total tRNA is treated with NaIO₄ at pH 5 to selectively oxidize the diol on the ribose at the 3'-end of non-aminoacylated tRNAs (black bar). Aminoacylation is maintained at this pH and aminoacylated tRNAs are not oxidized. A DNA probe specific for the tRNA of interest (grey) with a fluorophore attached (red star) anneals to the tRNA at pH 7.9 (I: probe annealing); at this higher pH, tRNAs are deacylated. Next, the exo(-) Klenow

fragment of *E. coli* DNA polymerase I is added to the hybridization solution (II: enzymatic extension). tRNAs that were aminoacylated before oxidation are extended by the polymerase using the probe as a template, whereas oxidized tRNA, which were never aminoacylated, cannot be extended. **b)** Validation of tREX on *Mb*-tRNA^{Pyl} (tRNA^{Pyl}). tRNA^{Pyl} probe, the Cy5-labelled DNA probe complementary to tRNA^{Pyl}, was added to all reactions. *E. coli* tRNAs are present in all lanes containing a tRNA extract. PylRS + tRNA^{Pyl} indicates whether the cells from which the tRNAs were extracted contained these components. NaIO₄ indicates whether the periodate oxidation was performed. BocK indicates whether the ncAA substrate (N^ε-boc-L-lysine) was added to cells from which tRNAs were subsequently extracted. Klenow exo(−) indicates whether this polymerase was added after annealing of the probe to the tRNA. The gel is native PAGE stained with SYBR Gold (green). The Cy5 on the probe is visualized in magenta. The ladder is O'RangeRuler 5 bp DNA Ladder by Thermo Scientific. This experiment was repeated in two independent replicates with similar results. **c)** Lanes 1–7 were created by mixing deacylated and oxidised tRNA^{Pyl} (0% equivalents) with non-oxidised tRNA^{Pyl} (100% equivalents) and subjecting the mixture to enzymatic extension. The percentage of non-oxidised tRNA is indicated above each lane. These lanes provide standards for the equivalent percentages of aminoacylation. Lanes 8–10 show the tREX experiment from cells expressing the orthogonal PylRS–tRNA^{Pyl} as a function of BocK concentration; BocK is an amino acid substrate for PylRS. Cy5 fluorescence is visualized. This experiment was repeated in two independent replicates with similar results. **d)** Investigation of the expression and orthogonality of *Capnocytophaga sp. oral taxon 329 str. F0087* (Cs)-tRNA^{Arg} CCG and the activity of Cs-ArgRS in *E. coli*. Lane 1 shows a control for probe specificity with respect to *E. coli* tRNAs. Lanes 2 and 3 show references for the electrophoretic mobility of the extended or native tRNAs obtained by not oxidizing the extract (lane 2) or by oxidation after total deacylation with NaOH (lane 3). Lane 4 shows the tREX experiment for the Cs-ArgRS–Cs-tRNA^{Arg} CCG pair, and lane 5 shows the tREX experiment for Cs-tRNA^{Arg} CCG. Cy5 fluorescence is visualized. The fluorescence pattern in lanes 2–5 likely results from the probe binding to partially degraded or processed forms of the target tRNA, as well as to full-length tRNA. **e)** As in **(d)** but with the *Variovorax paradoxus S110* AspRS–tRNA^{Asp} GUC. **f)** As in **(d)** but with the *Calothrix sp. PCC 7507* CysRS–tRNA^{Cys} GCA. **g)** As in **(d)** but with the *Ilumatobacter nonamiensis YM16-303* GlnRS–tRNA^{Gln} UUG. **h)** As in **(d)** but with the *Streptomyces davawensis JCM 4913* GluRS–tRNA^{Glu} CUC. **i)** As in **(d)** but with the *Bacteroides vulgatus ATCC 8482* GlyRS–tRNA^{Gly} CCC. **j)** As in **(d)** but with the *Afifella pfennigi DSM 17143* HisRS–tRNA^{His} GUG. **k)** As in **(d)** but with the *Salinispora arenicola CNS673* IleRS–tRNA^{Ile} GAU. **l)** As in **(d)** but with the *Coprobacillus sp. D7* ProRS–tRNA^{Pro} UGG. **m)** As in **(d)** but with the *Archaeoglobus fulgidus DSM 4304* TyrRS–tRNA^{Tyr} GUA. The experiments shown in **(d)** to **(m)** were repeated in two independent replicates with similar results.

Fig. 4. In vitro aminoacylation assays.

a) Purified *Mj*-TyrRS (150 nM) was incubated with tRNA extract from either *E. coli* (grey dot, duplicates) or from *E. coli* expressing *Mj*-tRNA^{Tyr}_{GUA} (black dots, duplicates). Incorporation of ¹⁴C-labelled tyrosine was initiated by the addition of ATP and monitored over 5 min. **b)** As in **a)**, but with *A. fulgidus* DSM 4304 TyrRS (150 nM)-tRNA^{Tyr}_{GUA} (black dots, duplicates). **c)** As in **a)**, but with *A. pfennigii* DSM 17143 HisRS (100 nM)-tRNA^{His}_{GUG} and ³H-labelled histidine. **d)** As in **a)**, but with *C. aerophila* DSM 14535 = NBRC 104270 ArgRS (100 nM)-tRNA^{Arg}_{GCG} and ¹⁴C-labelled arginine. **e)** As in **a)**, but with *S. cellulorum* So ce56 AspRS (50 nM)-tRNA^{Asp}_{GUC} and ¹⁴C-labelled aspartic acid. **f)** As in **a)**, but with *I. nonamiensis* YM16-303 GlnRS (100 nM)-tRNA^{Gln}_{UUG} and ¹⁴C-labelled glutamine.

Fig. 5. Directed evolution creates active and orthogonal *In*-GlnRS-*In*-tRNA, *Sc*-AspRS-*Sc*-tRNA^{Asp} and *Af*-TyrRS-*Af*-tRNA pairs for amber suppression and altered amino acid specificity.

a) Evolution of *In*-GlnRS allowed the selection of a mutant, *In*-GlnRS(S9), with activity toward *In*-tRNA^{Gln} CUA. After evolution of the synthetase, a variant of the tRNA was selected, *In*-tRNA^{Gln(A1)} CUA, which substantially increased the amber suppression efficiency of the pair. Quantification of GFP production was performed on three independent cell cultures, individual data points are shown as dots, bars represent the mean and the error bars represent the s.d. No statistical test was needed. AU: arbitrary units. **b)** *In vitro* aminoacylation of purified *In*-GlnRS(S9) (100 nM) incubated with tRNA extracts from either *E. coli* (grey dots, duplicates) or from *E. coli* expressing *In*-tRNA^{Gln(A1)} CUA (black dots, duplicates). Incorporation of ¹⁴C-labelled glutamine was initiated by the addition of ATP and monitored over 5 min. The lack of aminoacylation on *E. coli* tRNAs suggests orthogonality of the evolved synthetase. **c)** Evolution of *Sc*-AspRS allowed the selection of a mutant, *Sc*-AspRS(C4), with activity toward *Sc*-tRNA^{Asp} CUA. After evolution of the synthetase, a variant of the tRNA was selected, *Sc*-tRNA^{Asp(10)} CUA, which substantially increased the amber suppression efficiency of the pair. Quantification of GFP production was performed on three independent cell cultures; individual data points are shown as dots,

bars represent the mean and the error bars represent the s.d. No statistical test was needed. **d)** *In vitro* aminoacylation by purified *Sc*-AspRS(C4) (50 nM) incubated with tRNA extracts from either *E. coli* (grey dots, duplicates) or from *E. coli* expressing *Sc*-tRNA^{Asp(10)}_{CUA} (black dots, duplicates). Incorporation of ¹⁴C-labelled aspartic acid was initiated by the addition of ATP and monitored over 5 min. The minimal aminoacylation of *E. coli* tRNAs suggests good orthogonality of the evolved synthetase. **e)** Production of GFP from GFP^{150TAG} by *Sc*-tRNA^{Asp(10)}_{CUA} alone or together with *Sc*-AspRS(C4Glu), an evolved variant of *Sc*-AspRS(C4) incorporating glutamic acid. Quantification of GFP production was performed on three independent cell cultures; individual data points are shown as dots, bars represent the mean and the error bars represent the s.d. No statistical test was needed. **f)** Electrospray ionisation (ESI)–MS of GFP purified from cells containing GFP^{150TAG}, *Sc*-tRNA^{Asp(10)}_{CUA} and the *Sc*-AspRS(C4Glu) confirming that the evolved pair incorporates glutamic acid. Measured mass: 27.842 kDa; expected mass: 27842. kDa. Measured mass (–Met) 27.713 kDa; expected mass (–Met): 27.711 kDa. This experiment was repeated on two independent protein samples with similar results. **g)** Production of GFP from GFP^{150TAG} by the indicated tRNA or synthetase–tRNA pair for different intermediates in the evolution of the *Af*-TyrRS–*Af*-tRNA^{Tyr}_{CUA} pair. The *Mj*-TyrRS–*Mj*-tRNA^{Tyr}_{CUA} is shown as a reference. Quantification of GFP production was performed on three independent cell cultures; individual data points are shown as dots, bars represent the mean and the error bars represent the s.d. No statistical test was needed. **h)** Mutations in the *Af*-tRNA^{Tyr} body selected during each round of directed evolution. *Af*-tRNA^{Tyr}_{CUA} is the anticodon transplant. *Af*-tRNA^{Tyr(G5)}_{CUA}, has a G63T mutation shown in red; the residues targeted for mutagenesis in the first library are shown in orange on this structure. *Af*-tRNA^{Tyr(22)}_{CUA} has alterations at the residues are shown in red; residues that were targeted for mutagenesis but did not change are in orange, and residues surrounding the anticodon targeted for mutation in the next library are in green. *Af*-tRNA^{Tyr(A01)}_{CUA} has alterations at the residues are shown in red; residues that were targeted for mutagenesis but did not change are in orange or green.

Fig. 6. Efficient incorporation of ncAAs, via evolution of the *Af*-TyrRS(G5)-*Af*-tRNA^{Tyr}(A01) CUA pair.

a) Schematic of the GFP^{67TAG}-E2Crimson fusion protein in the triple reporter. This reporter was used to select for mutant TyrRSs that were active in the presence of an ncAA, but did not efficiently incorporate tyrosine; such synthetases are candidates for effective ncAA incorporation. The data show the fluorescence generated by the reporter with the indicated synthetase-tRNA-amino acid combinations; cells were resuspended in phosphate-buffered saline, excited at the indicated wavelength, and imaged on a phosphorimager. This experiment was repeated on two independent cell cultures with similar results. WT: wild type. **b)** Characterization of an evolved derivative of the *Af*-TyrRS(G5)-*Af*-tRNA^{Tyr}(A01) CUA pair directing the incorporation of *O*-methyl-L-tyrosine. The activity of this pair is compared to that of the derivative of the *Mj*-TyrRS-*Mj*-tRNA^{Tyr}_{CUA} pair previously reported for incorporating this ncAA. The data show fluorescence resulting from production of GFP from GFP^{150TAG} by the indicated aaRS-tRNA_{CUA} pair. Quantification of GFP production was performed on three independent cell cultures; individual data points are shown as dots, bars represent the mean and the error bars represent the s.d. No statistical test was needed. **c)** ESI-MS of GFP purified from cells containing GFP^{150TAG}, *Af*-tRNA^{Tyr}(A01) CUA and the *Af*-*O*-methyl-TyrRS confirms incorporation of *O*-methyl-L-tyrosine. Measured mass: 27.889 kDa; expected mass: 27.890 kDa. Measured mass (-Met) 27.757 kDa; expected mass (-Met): 27.759 kDa. This experiment was repeated on two independent protein samples with similar results. **d)** As in **b)** but with the *Af*-*p*-iodo-PheRS. Quantification of GFP production was performed on three independent cell cultures; individual data points are shown as dots, bars represent the mean and the error bars represent the s.d. No statistical test was needed. **e)** As in **c)** but with the *Af*-*p*-iodo-PheRS, confirmed incorporation of *p*-iodo-L-phenylalanine. Measured mass: 27.986 kDa; expected mass: 27.986 kDa. Measured mass (-Met) 27.853 kDa; expected mass (-Met): 27.855 kDa. This experiment was repeated on two independent protein samples with similar results. **f)** As in

b) but with the *Af-p*-azido-PheRS. Quantification of GFP production was performed on three independent cell cultures, individual data points are shown as dots, bars represent the mean and the error bars represent the s.d. No statistical test was needed. **g)** As in **c)** but with the *Af-p*-azido-PheRS, confirming incorporation of *p*-azido-L-phenylalanine. Measured mass: 27.898 kDa; expected mass: 27.901 kDa. This experiment was repeated on two independent protein samples with similar results.

Fig. 7. Mutual orthogonality of eight aaRS–tRNA pairs.

tREX experiments were performed for all 64 combinations of 8 aaRSs (*Ca*-ArgRS, *Sc*-AspRS(C4), *In*-GlnRS(S9), *Ap*-HisRS, *Ma*-PylRS, *Mb*-PylRS, *Mm*-SepRS^(v1.0), and *Af*-TyrRS(G5)) and 8 tRNAs (*Ca*-tRNA^{Arg}_{GCG}, *Sc*-tRNA^{Asp(10)}_{CUA}, *In*-tRNA^{Gln(A1)}_{CUA}, *Ap*-tRNA^{His}_{GUG}, *Ma*-tRNA^{Pyl(6)}_{CUA}, *Mb*-tRNA^{Pyl}_{CUA}, *Mm*-tRNA^{Sep(v2.0)}_{CUA} and *Af*-tRNA^{Tyr(A1)}_{CUA}). The first column provides a control for potential nonspecific binding of the probe to *E. coli* tRNAs, while the second and third columns represent controls for the electrophoretic mobility of the fully extended and not extended (native) bands, respectively. Plasmids containing *Mm*-SepRS^(v1.0) were transformed in *>serB E. coli* DH10b, which have high intracellular concentration of phosphoserine. Cells harbouring *Ma*-PylRS or *Mb*-PylRS were grown in the presence of 2 mM Bock. Controls are repeated in the last three lanes. This experiment was performed once.