Article

Engineering a genomically recoded organism with one stop codon

Michael W. Grome[1,2], Michael T. A. Nguyen[1,2,6], Daniel W. Moonan[2,3,6], Kyle Mohler[2,4,6], Kebron Gurara[1,2,6], Shenqi Wang[1,2], Colin Hemez[1,2,5], Benjamin J. Stenton[1,2], Yunteng Cao[1,2], Felix Radford[1,2], Maya Kornaj[1,2,4], Jaymin Patel[1,2], Maisha Prome[1,2], Svetlana Rogulina[2,4], David Sozanski[1,2], Jesse Tordoff[1,2], Jesse Rinehart[2,4 ✉] & Farren J. Isaacs[1,2,5 ✉]

The genetic code is conserved across all domains of life, yet exceptions have revealed variations in codon assignments and associated translation factors[1–3]. Inspired by this natural malleability, synthetic approaches have demonstrated whole-genome replacement of synonymous codons to construct genomically recoded organisms (GROs)[4,5] with alternative genetic codes. However, no efforts have fully leveraged translation factor plasticity and codon degeneracy to compress translation function to a single codon and assess the possibility of a non-degenerate code. Here we describe construction and characterization of Ochre, a GRO that fully compresses a translational function into a single codon. We replaced 1,195 TGA stop codons with the synonymous TAA in ΔTAG *Escherichia coli* C321.ΔA[4]. We then engineered release factor 2 (RF2) and tRNA[Trp] to mitigate native UGA recognition, translationally isolating four codons for non-degenerate functions. Ochre thus utilizes UAA as the sole stop codon, with UGG encoding tryptophan and UAG and UGA reassigned for multi-site incorporation of two distinct non-standard amino acids into single proteins with more than 99% accuracy. Ochre fully compresses degenerate stop codons into a single codon and represents an important step toward a 64-codon non-degenerate code that will enable precise production of multi-functional synthetic proteins with unnatural encoded chemistries and broad utility in biotechnology and biotherapeutics.

The standard genetic code contains 64 codons that degenerately encode 20 natural amino acids and translation termination. Organismal variations in codon recognition and inclusion of natural non-standard amino acids (nsAAs) have given rise to alternative genetic codes[1–3]. Cognate and near-cognate codon recognition by tRNAs can vary widely across domains and between species as tRNAs are gained, lost, mutated or reassigned to recognize different codons—for example, via variations in anticodon wobble and base modifications (such as inosine)[1]. Deviations in code may render codons unassigned[6], encoding alternative canonical amino acids[2,3] or nsAAs[7,8], or serving as noncanonical stop codons (such as UUA and UCA)[7].

Alternative codon configurations and codon redundancy highlight the plasticity of the genetic code, motivating genome-scale efforts to construct GROs—organisms constructed with alternative genetic codes whereby redundant codons are reassigned to new functions. GROs possess several advantageous phenotypes, including for enhanced phage resistance[8–10], biocontainment strategies[11,12] and efficient site-specific incorporation of nsAAs to imbue novel protein properties[13,14]. The first constructed GRO was an *E. coli* in which all instances of TAG stop codons were replaced with synonymous TAA codons, followed by deletion of release factor 1 (RF1)[4,15]. The resulting strain, C321.ΔA, released TAG as an unassigned codon amenable for reassignment.

Recent efforts have commenced to reassign sense codons in *E. coli*[5,16] and other organisms[17–19].

Whole-genome recoding inspires unresolved questions regarding the malleability of the genetic code and whether a minimal non-degenerate genome is feasible. By assigning unused codons to new functions, such genomes could expand in vivo protein design with unnatural biochemistries, conferring valuable properties for biomaterials and drugs (for example, immunity to proteases). Achieving such goals requires engineering essential translation factors (tRNAs and release factors) for defined single-codon specificities (codon exclusivity), decoupling codons from translational crosstalk (for example, wobble effect or near-cognate suppression) and eliminating functional redundancy among synonymous codons. Single-codon specificities found in nature demonstrate the feasibility of engineering exclusivity—for example, AUG specificity for tRNA[Met] enabled by discriminatory post-transcriptional modifications of AUG-specific tRNA[Met](CAU) and AUA-specific tRNA[Ile](CAU)[20]. Eliminating similar modifications has mitigated anticodon wobble in native tRNAs, rendering them codon exclusive[21,22]. Engineering essential translation factors presents unique challenges related to their cryptic alternative functions that could be revealed by wide-scale genomic recoding efforts.

[1]Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT, USA. [2]Systems Biology Institute, Yale University, West Haven, CT, USA. [3]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA. [4]Department of Cellular and Molecular Physiology, Yale University, New Haven, CT, USA. [5]Department of Biomedical Engineering, Yale University, New Haven, CT, USA. [6]These authors contributed equally: Michael T. A. Nguyen, Daniel W. Moonan, Kyle Mohler, Kebron Gurara. ✉e-mail: jesse.rinehart@yale.edu; farren.isaacs@yale.edu
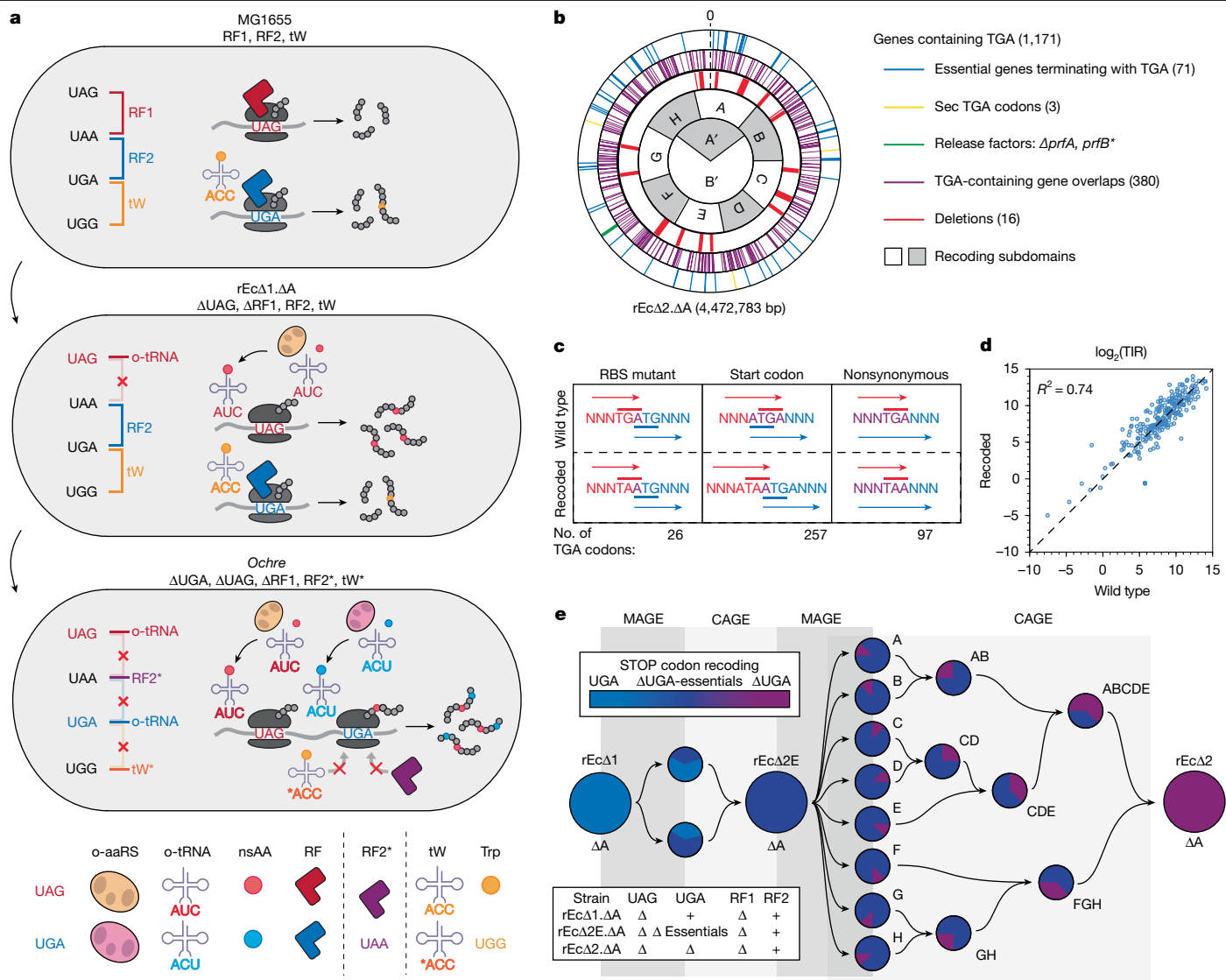
Fig. 1 | Design and construction of genomically recoded *E. coli* with a single stop codon. a, Cellular schematic summary of translational decoding of the three canonical stop codons—UAG, UGA and UAA—and the tryptophan-encoding UGG codon alongside their native cognate translation factors across three strains of *E. coli* after iterative rounds of recoding: MG1655 (wild type), rEcΔ1.ΔA (ΔUAG, ΔRF1) and Ochre (ΔUAG, ΔRF1, RF2*, tW*). Asterisks denote translation factors engineered to discriminate against UGA. Codons with multiple iterations of genomic recoding from wild-type *E. coli* to rEcΔ2.ΔA.B3.tW*, which is a ΔTAG-ΔTGA recoded strain containing a mutant RF2 and a mutant tRNA^Trp both engineered to discriminate against UGA. OTS components (o-aaRS and o-tRNA) in recoded strains enable reassignment of UAG and UGA to sense codons incorporating nsAAs. RF, release factor. **b**, Diagram of ΔTAG-ΔTGA recoded genome (rEcΔ2.ΔA) depicting relative locations of important designed mutagenic events in the genome-wide removal of TGA. Innermost sections of circles depict genomic subdivisions for targeted recoding of rEcΔ1.ΔA to acquire rEcΔ2E.ΔA (A' and B') and subsequently to acquire fully recoded rEcΔ2.ΔA.B1 (A–H). Sec, selenocysteine. **c**, Top, representations of common terminal TGA codons (red) overlapping genes (blue). Bottom, common mutational means of resolving codon conversions with minimal genetic deviation. **d**, Predicted translation initiation rates (TIRs) for genes overlapped by genes terminating with TGA shown in **b** recoded from TGA to TAA plotted against the predicted translation initiation rates for wild-type overlapped genes. The diagonal corresponds to recoding events that do not alter the expected translation rate. R^2 is the Pearson correlation coefficient. **e**, Schematic overview depicting iterative steps of MAGE-mediated genomic recoding and CAGE-mediated hierarchical assembly of recoded subdomains into a final fully recoded strain (rEcΔ2) lacking terminal TGA codons.

Here we describe Ochre, a GRO that fully compresses a redundant codon functionality into a single codon, liberating two essential stop codons for reassignment. Specifically, synonymous replacement of 1,195 instances of the stop codon TGA alongside ΔTAG[4], combined with engineering of essential translation factors (RF2 and tRNA^Trp) to attenuate native UGA recognition, enabled multi-site incorporation of two distinct nsAAs at UAG and UGA with greater than 99% accuracy within single proteins. Although past recoding efforts have repurposed stop and sense codons via deletion of associated translation machineries (RF1[4] and tRNAs[5], respectively), a genome reduced to a

single non-degenerate stop codon had not been achieved owing to challenges in engineering translation factors to precisely tune codon specificities. Our efforts disentangled translational crosstalk within the stop codon block, rendering four codons non-degenerate, with each serving a unique function: UAA as the sole stop codon, UGG encoding Trp and UGA and UAG liberated to encode two unique nsAAs (Fig. 1a). Distinct from prior efforts[4,5], we demonstrate the importance of compressing redundant codons and disentangling translational crosstalk to fully liberate codons for precise nsAA reassignments, establishing a benchmark for future recoding. Ochre provides a novel context

Article

to interrogate translation and cryptic codon functions and to evolve new translation machinery with applications in biocontainment[11,12], genetic isolation[8–10] and biomanufacturing of proteins with synthetic chemistries[13,14,23].

Constructing a ΔTAG/ΔTGA recoded strain

The stop codon block consists of four translationally linked codons: TAG, TGA, TAA and TGG. In C321.ΔA[4], TAG is freed as an open coding channel, leaving TAA (stop), TGG (Trp) and TGA (stop and near-cognate Trp). With only 71 essential genes among 1,216 total open reading frames (ORFs) containing TGA in the MG1655 genome (Extended Data Table 1), TGA is the second rarest codon. This scarcity and translational redundancy makes TGA an attractive target alongside TAG. Additionally, dual orthogonal translation systems (OTSs) expressing orthogonal aminoacyl-tRNA synthetases (o-aaRSs) and tRNAs (o-tRNAs) targeting UAG and UGA have been deployed in unrecoded *E. coli* to incorporate two distinct nsAAs[24], but their efficacy was limited by competition with native translation factors.

There are four primary open questions for our recoding effort: (1) whether TGA codons are essential; (2) whether redundant codon functions can be compressed into one codon; (3) whether release factors can attain single-codon specificity; and (4) whether codon suppression can be mitigated without deletion of translation factors. Release factors[7,25] and tRNAs[26,27] with altered specificities suggest it is feasible to engineer attenuated UGA recognition, mitigating translation factor competition to free TGA alongside TAG for reassignment. Addressing these questions would enable construction of a strain lacking both TAG and TGA codons that, when complemented by a noncompetitive mutant release factor and tRNA^Trp, could be reassigned as sense codons. Although previous recoding efforts enabled nsAA incorporation at two open serine codons[9], unresolved translational crosstalk (for example, o-tRNA wobble[28]) is likely to result in misincorporations. We hypothesize that resolving crosstalk at UGA within a ΔTGA recoded strain will enable precise, site-specific dual incorporation of nsAAs at both UAG and UGA and establish functional codon exclusivity (non-degeneracy) in the stop codon block.

To construct a ΔTAG, ΔTGA recoded strain (rEcΔ2.ΔA), we used the ΔTAG progenitor C321.ΔA[4], which we refer to as rEcΔ1.ΔA. All TGA codons, with few exceptions, occur at the end of a gene to terminate translation. Out of the 1,216 annotated ORFs that contain TGA, 1,171 are annotated genes and 45 are listed as pseudogenes. To reduce recoding efforts, 76 non-essential genes and 3 pseudogenes containing TGA were removed with 16 targeted genomic deletions[29] (Fig. 1b). The remaining 1,134 terminal TGA codons (1,092 genes and 42 pseudogenes) were targeted for conversion to TAA via multiplex automated genomic engineering (MAGE)[30] (Supplementary Table 1). Three formate dehydrogenase genes (*fdhF*, *fdoG* and *fdnG*) containing internal TGA codons were not targeted for recoding owing to their role in encoding selenocysteine[31]. Four distinct oligonucleotide designs were used to convert codons. One strategy introduced single-nucleotide substitutions to 833 non-overlapping ORFs and 3 refactoring strategies targeted 380 overlapping ORFs in which such substitutions might affect neighbouring gene expression (Fig. 1c and Supplementary Methods). These latter strategies resulted in changes to more than 300 overlapping coding sequences (Supplementary Table 2) and few predicted disruptions to translation initiation rates (Fig. 1d and Supplementary Table 3).

Construction of ΔTGA rEcΔ2.ΔA constituted two major phases (Fig. 1e), each utilizing iterative cycles of MAGE[30] concurrently targeting distinct genomic subdomains within clonal progenitor strains, followed by conjugative assembly genome engineering (CAGE)[15] to hierarchically assemble recoded subdomains into a final strain. TGA-to-TAA conversions were confirmed via whole-genome sequencing (WGS) after each assembly. In phase 1, rEcΔ2E.ΔA was constructed from rEcΔ1.ΔA by

converting 71 essential genes terminating with TGA split between 2 distinct genomic subdomains (A′ and B′) among two clones (Supplementary Table 4). An additional 44 proximal genes terminating with TGA were converted and 3 genes terminating with TGA were replaced with selectable markers. Recoded subdomains A′ and B′ were assembled via CAGE into rEcΔ2E.ΔA (Supplementary Table 5).

Phase 2 constructed the final ΔTGA recoded rEcΔ2.ΔA. Another 1,012 ORFs terminating with TGA (980 genes and 35 pseudogenes) were converted via MAGE (Supplementary Table 1 and 6) and 229 non-essential ORFs (72 genes terminating with TGA and 3 pseudogenes terminating with TGA)[29] were deleted via marker placements (Extended Data Fig. 1 and Supplementary Table 7). TGA sites were divided among eight rEcΔ2E.ΔA clones and targeted concurrently across distinct genomic subdomains (A–H) split among clones (Fig. 1b and Supplementary Table 8). Preserving chemotaxis and motility, non-essential genomic regions were identified[29] and those with a high density of genes terminating with TGA were deleted, yielding 16 genomic deletions. Unwanted markers were deleted via *tolC* displacements[15]. The non-essential glycerate kinase I gene *garK* was recalcitrant to conversion and was inactivated by a frameshift nonsense mutation (Supplementary Table 9). Recoded subdomains were hierarchically assembled via CAGE into the final rEcΔ2.ΔA (Extended Data Fig. 2 and Supplementary Table 10). Out of 1,216 predicted TGA codons in *E. coli*, 1,195 were abolished, leaving 10 predicted pseudogenes, 8 insertion sequences and 3 internal selenocysteine codons unrecoded (Extended Data Table 1). WGS data and breseq analysis[32] were used to assess background mutations[33]. More than 700 unintended background mutations accumulated in our final recoded strain (Supplementary Tables 11–13), although this was likely to be an overestimate (Supplementary Methods). To mitigate detrimental phenotypes arising during construction, cell colonies were routinely measured for maximum OD$_{600}$ (MaxOD) and doubling time in lysogeny broth (LB) and minimal (M9) media. The fastest growing strains with highest MaxOD were selected to progress for further recoding. When auxotrophies arose, conjugal backcrossing to parental strains (Supplementary Methods and Supplementary Fig. 2) and targeted reversions of implicated mutations (Supplementary Methods and Supplementary Tables 14a and 15) were used to restore prototrophy.

Engineering mutant RF2

Engineering UAA-specific translation termination is a first step in decoupling UGA translational crosstalk and liberating UAG and UGA for efficient reassignment. RF1, encoded by *prfA*, terminates translation at UAG and UAA, whereas RF2, encoded by *prfB*, terminates translation at UGA and UAA. RF1 was deleted within ancestral rEcΔ1.ΔA[4], confirming that RF1 is non-essential without TAG codons and that RF2 can serve as the sole release factor. After ΔTGA recoding, we tested whether RF1 could independently support translation termination. Unlike RF1, RF2 serves additional roles beyond termination, including post-peptidyl transfer quality control with release factor 3 (RF3)[34] and ArfA-cooperative ribosomal rescue from 'non-stop' mRNAs[35]. Our inability to delete RF2 in the presence of RF1 suggests that RF1 cannot functionally compensate for loss of RF2. The capacity of RF2 to act as a sole release factor made it our target for engineering specificity for UAA. Thus, we sought to identify mutations that attenuate RF2-mediated termination at UGA while retaining essential functions. Using MAGE, five primary mutations were introduced to the *prfB* gene: (1) T246A, to restore wild-type release factor methylation[36,37]; (2) terminal TGA-to-TAA conversion; (3) internal TGA-to-TAA for autoregulatory maintenance[38,39]; (4) S205P for UAA-specific termination activity[40]; and (5) E170K charge flip[41,42] to enable strain viability (Fig. 2a,b and Supplementary Table 14b). Two in vivo assays measured the effect of these mutations on the ability of RF2 to: (1) outcompete cognate/near-cognate suppression in fluorescent reporter translation; and (2) inhibit phage infection. Derived from K-12 strains, rEcΔ1.ΔA contains
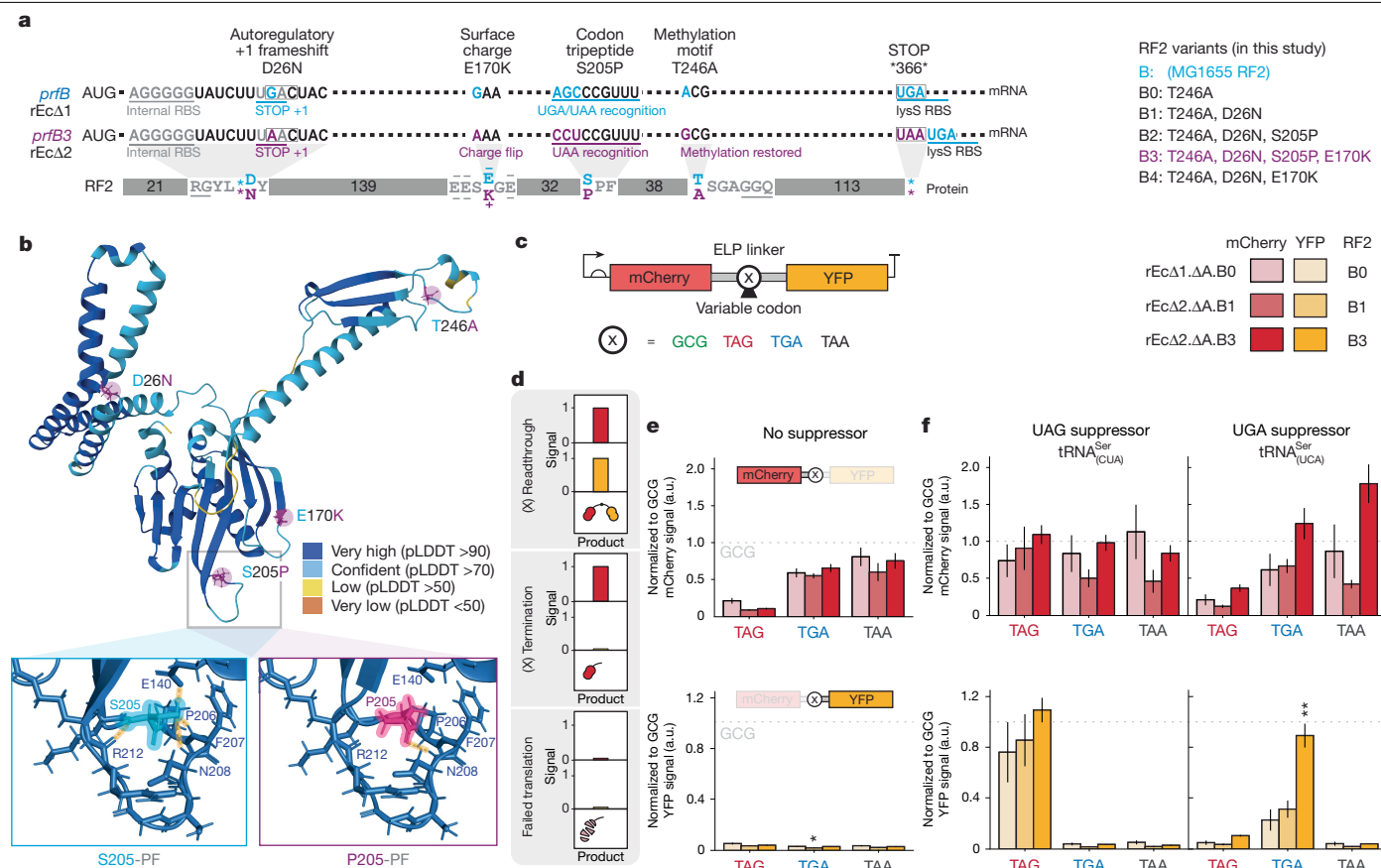
Fig. 2 | Engineering RF2 to attenuate UGA termination. a, Left, comparative mRNA transcript and protein diagrams displaying key nucleotides and amino acids in wild-type RF2.B (cyan) and mutant RF2.B3 (purple). Right, RF2 variants used in this study. **b**, Top, AlphaFold[48] structural prediction of RF2.B3 with residue deviations from wild-type RF2 marked in purple. Bottom, magnified views of wild-type RF2 SPF (left) and mutant RF2 PPF (right) codon recognition loops, highlighting interactions with residue 205. AlphaFold[48] structural prediction suggests that wild-type S205 presents up to four hydrogen bonds within the RF2 codon recognition loop, whereas P205 presents one. **c**, Schematic representation of mCherry–YFP fluorescent reporter used to assess release activity at three target codons (X is UAG, UGA or UAA) relative to a sense codon control (GCG) in a peptide linker between mCherry and YFP. Release activity at codon X prevents downstream YFP expression and

fluorescence without compromising mCherry, whereas stalling and degradation leads to loss of mCherry and YFP signals. High YFP fluorescence suggests readthrough. **d**, Schematic representation of expected fluorescent outcomes resulting from the mCherry–YFP readthrough assay. **e,f**, mCherry (top) and YFP fluorescence (bottom) expressed as a function of translation termination or readthrough at target codon X during reporter translation in variant strains with release factors. Fluorescence is normalized to GCG codon construct fluorescence in absence (**e**) or presence (**f**) of UAG- or UGA-suppressing supD tRNAs. Colour scheme in **d**–**f** as indicated in **c**. a.u., arbitrary units. Data are mean ± s.e.m. P values for comparison with rEcΔ1.ΔA by unpaired t-tests (n = 3 biological replicates). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.

an A246T mutation in RF2 that hinders termination efficiency compared with wild-type *E. coli*[42]. We introduced T246A to restore RF2 methylation by PrmC, generating variant RF2.B0 and improving strain fitness[42]. To make RF2.B0 translation compatible with attenuated UGA termination, we converted both the internal autoregulatory TGA[38,39] and terminal TGA to TAA, generating variant RF2.B1 (Supplementary Table 5). The *prfB* mRNA, encoding RF2, contains an internal ribosome binding site (RBS) upstream of a conditionally frameshifting CUU-UGA sequence that regulates RF2 translation based on cellular concentrations[38,39]. Although the CUU-UGA sequence is highly conserved in *prfB* mRNA, exceptions such as *Chlorobium tepidum* instead have CUU-UAA[38], suggesting permissibility. TGA-to-TAA conversion introduced a D26N mutation with no significant growth defects in rEcΔ2.ΔA. B1 compared with B0 (Extended Data Fig. 3a). Downstream, the *prfB* terminal TGA overlaps the RBS for *lysS*, encoding lysine-tRNA synthetase. To preserve *lysS* translation, we inserted TAA directly upstream of the terminal TGA (Fig. 2a). In prokaryotic translation, release factor codon specificity is predominantly conferred by helix α5, specific to U at position 1 of the stop codon, and a 10 or more amino acid codon recognition loop whose structure and sequence discriminates between the second

and third codon bases[43–45]. Alternative release factor codon specificity is well precedented in prokaryotic systems. Single point mutations in RF2 alter codon specificity (for example, E167K or F207T), expanding recognition to all three stop codons, creating 'omnipotent' release factors[41]. In addition, some mitochondrial release factors terminate at canonical sense codons (for example, AGA or AGG)[7]. Various works have interrogated codon specificity via tripeptide 'anticodon' motifs located within recognition loops: PXT in RF1 (where X is A or V) or SPF in RF2[43,46,47]. Crystallography implicates hydrogen bonding by the threonine in PXT and the serine in SPF in discrimination at the second base of the codon[44], although molecular dynamics simulations dispute this[45].

We hypothesized that one characterized *E. coli* RF2 mutation (S205P) in the tripeptide SPF might endow UAA specificity within rEcΔ2.ΔA[46]. Structural prediction by AlphaFold[48] (Fig. 2b) suggests that wild-type S205 presents up to 4 hydrogen bonds within the codon recognition loop of RF2, whereas P205 presents 1, probably altering recognition loop stability. Previous characterizations of S205P evaluated release factor codon recognition in vivo and in vitro[41,46]. Using a chimeric RF1–RF2 backbone, Ito et. al. recapitulated wild-type RF2 termination with SPF[46]. Introducing S205P (forming chimeric PPF) attenuated

a

Phage	λ
TAG	4
TAA	29
TGA	40

	MG1655	rEcΔ1.ΔA	rEcΔ2.A.B0	rEcΔ2.ΔA.B0	rEcΔ2.ΔA.B3	rEcΔ2.A.B3
TAG	+	–	–	–	–	–
TGA	+	+	–	–	–	+
RF1	+	–	+	–	–	+
RF2 26	D	D	D	D	N*	N*
RF2 170	E	E	E	E	K*	K*
RF2 205	S	S	S	S	P*	P*
RF2	B	B	B0	B0	B3	B3

b

Phage	μ
TAG	0
TAA	32
TGA	23

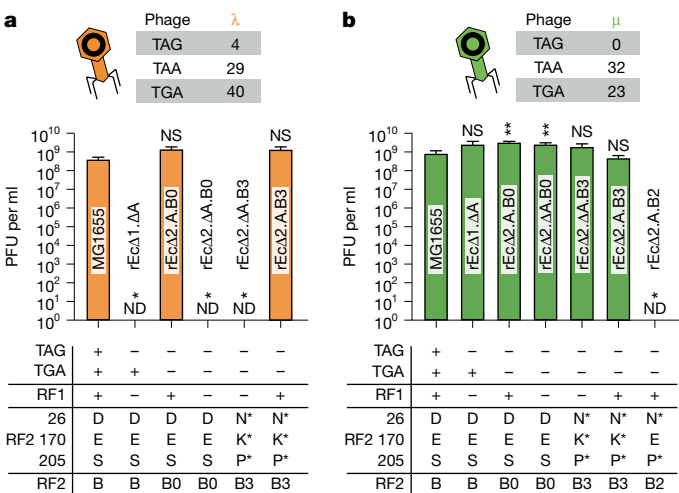|	MG1655	rEcΔ1.ΔA	rEcΔ2.A.B0	rEcΔ2.ΔA.B0	rEcΔ2.ΔA.B3	rEcΔ2.ΔA.B3	rEcΔ2.A.B2
TAG	+	–	–	–	–	–	–
TGA	+	+	–	–	–	–	+
RF1	+	–	+	–	–	–	+
RF2 26	D	D	D	D	N*	N*	N*
RF2 170	E	E	E	E	K*	K*	E
RF2 205	S	S	S	S	P*	P*	P*
RF2	B	B	B0	B0	B3	B3	B2

Fig. 3 | Characterization of release factor variants using phage assays. a, RF1 and RF2 variant strains were challenged with λ phage to assess phage infectivity with genes terminating with TAG, TAA and TGA. PFU, plaque-forming units. **b**, Release factor variant strains were challenged with μ phage to assess phage infectivity with genes terminating with TAA and TGA only. Data are mean ± 95% confidence interval for $n = 3$ biological replicates. Unpaired t-test for comparison with *E. coli* MG1655. NS, not significant; ND, not detected (lower than the measurable detection limit).

recognition of UGA in vitro without compromising UAA termination. Nonetheless, expression in *E. coli* did not complement native RF2 knockout[46]. A second study introduced S205P into wild-type RF2[41]. Unsurprisingly, supplemental expression in *E. coli* led to growth interference and cell nonviability. Given these results, we hypothesized that unlike wild-type *E. coli*, which is dependent on native UGA termination, our ΔTGA rEcΔ2.ΔA.B1 strain could uniquely tolerate RF2 S205P. Using rEcΔ2.ΔA.B1, we attempted to introduce S205P into genomic RF2.B1 to generate variant RF2.B2, but did not recover viable clones without *prfA*, compensatory mutations in translation regulatory protein RF3 (*prfC*) (Supplementary Table 14b), or episomal RF2.B2 expression with growth defects (Extended Data Fig. 4 and Supplementary Discussion 1). To enable S205P within a genomic ΔRF1/RF2.B1 background and maintain cellular fitness, we used MAGE oligonucleotides to introduce mutations across RF2.B1 concurrent with S205P (Supplementary Table 16). Targeted mutations were identified through *E. coli* RF1 and RF2 sequence alignments with RF2s from low-TGA-frequency bacterial genomes[49] that are hypothesized to encode RF2 variants with poorer UGA recognition. Residues were cross-referenced with the literature to avoid conferring recognition of UAG[41,42,46]. Viable S205P strains arose only in co-occurrence with E170K (Extended Data Fig. 3b). E170K is a previously characterized charge flip mutation that compensates growth interference from S205P[41] and independently arose from rEcΔ1.ΔA adaptive evolution[42]. Release factor complementation assays revealed that residue charge flips close to E170—such as E167K—altered codon recognition, unlike E170K[41]. Introduction of S205P and E170K into RF2.B1 generated RF2.B3 (Fig. 2a,b and Supplementary Tables 14, 16 and 17). To examine the essentiality of E170K in our rEcΔ2.ΔA.B3 strain, we used MAGE to establish a K170E reversion but did not yield viable clones in absence of wild-type S205 or RF1, confirming the essentiality of E170K to support viability of the RF2 S205P mutation.

Characterizing mutant RF2 termination

To assess whether RF2.B3 abolishes recognition of UGA, we conducted an in vivo assay measuring release activity of RF2 variants at all three stop codons. Using validated methods[50], this assay evaluates the ability of a release factor to terminate translation of an mCherry gene ending in a target codon upstream of a YFP gene (Fig. 2c, Extended Data Fig. 5b–d and Supplementary Table 18). If no termination occurs at the target codon, translation reads through mCherry to YFP, expressing both proteins. If termination occurs, only mCherry is expressed, yielding low YFP fluorescence. Figure 2d–f presents fluorescence of mCherry and YFP expression as a function of termination at a target codon, normalized to fluorescent expression from GCG (positive control). The results revealed low YFP expression at all stop codons, regardless of release factor status (Fig. 2e). Previous work showed that rEcΔ1.ΔA genes terminating with TAG are degraded via transfer messenger RNA (tmRNA) tagging, reducing UAG codon readthrough and GFP expression[51]. This is consistent with low expression levels observed for the mCherry gene terminating with UAG and downstream YFP. For more robust readouts, we used modified UAG- and UGA-suppressor serine tRNAs (supD)[52] that precisely base pair with their respective codons to examine direct codon competition with release factor variants, effective for assessing codon reassignment (Fig. 2f). Results with mutant rEcΔ2.ΔA. B3 suggest termination at UAA, approximately 42-fold enhanced readthrough at UAG and 45-fold enhanced readthrough at UGA relative to no suppressors, leading to 11-fold and 8-fold increases in UGA YFP fluorescence compared with wild-type RF2 strains rEcΔ1.ΔA.B0 and rEcΔ2.ΔA.B1, respectively. These results demonstrate that UGA recognition by RF2.B3 is significantly attenuated compared with RF2. B0, as RF2.B3 is outcompeted by UGA-specific suppressor tRNAs. To assess whether S205P (not E170K) causes attenuation, we reverted RF2.B3 P205 back to serine, generating variant RF2.B4. We repeated the in vivo mCherry–YFP readthrough assays without suppressors to assess rEcΔ2.A.B2 (RF2(S205P)/RF1+) alongside isogenic rEcΔ2.ΔA.B4 (RF2(E170K)/ΔRF1) and two wild-type RF2/ΔRF1 controls: rEcΔ2.ΔA. B0 and rEcΔ1.ΔA.B0 (Extended Data Fig. 5e). The high mCherry and low YFP fluorescence for UAG from rEcΔ2.A.B2 is indicative of RF1 terminating at UAG. For UGA, results reveal low YFP fluorescence from both RF2.B0 and RF2.B4 variants, suggesting little effect from E170K and D26N on UGA termination. However, S205P in the absence of E170K (rEcΔ2.A.B2) results in high YFP fluorescence, suggesting significant UGA readthrough from native suppression and possibly abolition of RF2 termination. This result stands in stark contrast to RF2. B3 (S205P with E170K) without supD suppressors (Fig. 2e), revealing S205P to be uniquely responsible for attenuation of UGA termination. These data suggest that E170K restores S205P termination at UAA sufficient for cell viability and partially restores termination at UGA, although significantly attenuated, to enable robust supD suppression (Fig. 2f).

The effects of individual RF2 mutations on translation are further evidenced by phage assays assessing genetic orthogonality[8–10] (Supplementary Discussion 2). We challenged strains with infection from two bacteriophages: Lambda (λ) (containing TAG, TGA and TAA) and Mu (μ) (containing TGA and TAA only) (Fig. 3 and Extended Data Fig. 6). Whereas all strains lacking RF1 retained resistance to λ, as previously described[51] (Fig. 3a), only rEcΔ2.A.B2 (RF2(S205P)/RF1+) displayed resistance to μ (Fig. 3b). These results suggest that: (1) UGA termination may be abolished in rEcΔ2.A.B2 and thus is not necessary for ΔTGA strain viability; and (2) that E170K in rEcΔ2.A.B3 restores partial UGA termination function in alignment with readthrough data.

Eliminating UGA suppression by tRNA^Trp

We established proteomics and reporter protein assays to examine how UGA is decoded in rEcΔ2.ΔA.B3. Mass spectrometry revealed marked Trp suppression comprising around 80% of residues at UGA, indicating significant suppression by tRNA^Trp (Fig. 4 and Supplementary Table 19) as the *E. coli* tRNA^Trp (Ec-tRNA^Trp) anticodon loop wobble pairs with UGA[53]. To reduce native Trp suppression and disentangle
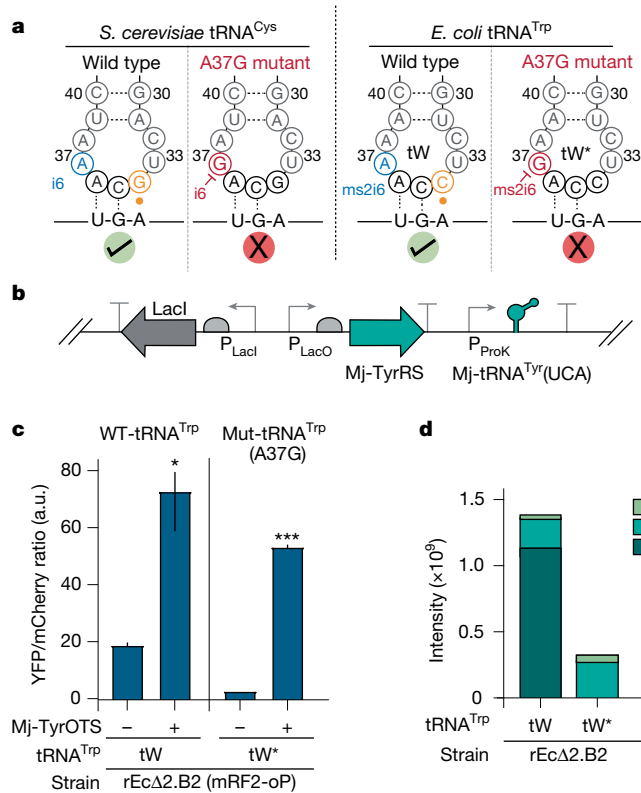
Fig. 4 | Engineering tRNA^Trp to mitigate UGA suppression. a, Schematics of wild-type and mutant (A37G) tRNA^Cys(GCA) anticodon loops in *S. cerevisiae* and wild-type and mutant (A37G) tRNA^Trp(CCA) anticodon loops in *E. coli*, depicting loss of anticodon modification and codon recognition in mutant forms. **b**, Schematic representation of Mj-TyrOTS expression plasmid: IPTG-inducible *M. jannaschii* tyrosyl-tRNA synthetase (Mj-TyrRS) and accompanying constitutively expressed tRNA^Tyr(UCA). **c**, Mj-TyrOTS-enabled suppression of UGA within a mCherry–YFP fluorescent reporter expressed in rEcΔ2.B2 strains with wild-type or mutant tRNA^Trp (tW*). Data are mean ± s.e.m. *P* values for comparison with no Mj-TyrOTS and no Ara (inducer) conditions for each strain by unpaired *t*-tests (*n* = 3 biological replicates). **d**, Stacked mass spectrometry data depicting identities of incorporated residues at UGA from MS-READ analysis of rEcΔ2.B2 strains with wild-type or mutant tRNA^Trp.

translational crosstalk between TGA and TGG, we modified *trpT*, which encodes tRNA^Trp(CCA), to abolish UGA recognition. Modifications at positions 34 and 37 of tRNA anticodon loops have been demonstrated to dictate codon recognition fidelity across kingdoms[22,26,54]. Mutations that knock out A37-modifying enzymes have been shown to abolish UGA suppression from *Saccharomyces cerevisiae* tRNA^Cys(GCA) (Sc-tRNA^Cys)[22] and strongly attenuate suppression (by 98.5% compared to the wild type) by Ec-tRNA^Trp (ref. 21). We compared the sequence surrounding the anticodon region of Ec-tRNA^Trp to that of Sc-tRNA^Cys and identified a conserved A36-A37-A38 sequence in the anticodon loop of both tRNAs that is required for modification of A37 to *N*⁶-(isopentenyl) adenosine (i6A37) in both prokaryotes and eukaryotes[26]. On the basis of these observations, we hypothesized that the UGA suppression activity of Ec-tRNA^Trp is linked to A37 and that an A37G mutation, which disables ms2i6A37 modification, could abolish UGA suppression in *E. coli*. To test this hypothesis, we mutated A37 in native Ec-tRNA^Trp to G37 (tW*) to disrupt the A36-A37-A38 substrate motif and prevent base modification (Fig. 4a), generating rEcΔ2.ΔA.B3.tW* (Ochre). We used our mCherry–YFP fluorescent assays with an isopropyl β-ᴅ-1-thiogalactopyranoside (IPTG)-inducible *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase and proK constitutively expressed tRNA^Tyr(UCA) pair (Mj-TyrOTS) designed to encode Tyr at UGA to
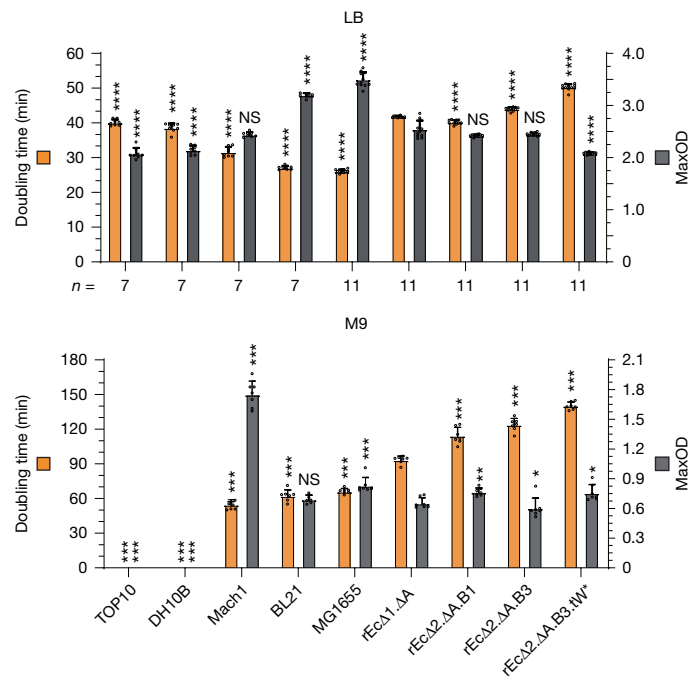
evaluate OTS activity and native tRNA^Trp suppression at UGA (Fig. 4b). We observed a significant reduction in YFP/mCherry expression in rEcΔ2.ΔA.B3.tW* compared with rEcΔ2.ΔA.B3 (wild-type tRNA^Trp), suggesting a reduction of native tRNA^Trp suppression (Fig. 4c). To confirm decreased Trp suppression in rEcΔ2.ΔA.B3.tW*, we used mass spectrometry reporter for exact amino acid decoding (MS-READ) to assay amino acid incorporation at UGA in wild-type and A37G tRNA^Trp strains containing Mj-TyrOTS[13,55]. Mass spectrometry analysis confirmed Mj-TyrOTS-mediated Tyr incorporation at UGA codons, and unexpectedly, a small amount (less than 1%) of Cys incorporation, possibly due to a parallel third base wobble to Sc-tRNA^Cys[22]. Whereas Trp was the most abundant amino acid at UGA in rEcΔ2.ΔA.B3 cells with wild-type tRNA^Trp, it was completely absent in rEcΔ2.ΔA.B3.tW* (Fig. 4d and Supplementary Table 19). These results confirm that a single base substitution in Ec-tRNA^Trp preserves native Trp decoding while eliminating UGA suppression, removing the wobble effect and establishing an open UGA codon for reassignment. With the combination of RF1 deletion and attenuated UGA recognition of RF2.B3 and tW*, four codons that naturally display translational crosstalk (UAG-open, UGA-open, UAA-stop and UGG-Trp) can be rendered translationally isolated and functionally exclusive (Fig. 1a).

Characterizing growth of recoded strains

The cumulative effects of whole-genome TGA replacement and RF2–tRNA^Trp translation engineering can be seen when comparing our rEcΔ2.ΔA.B3.tW* with ancestral rEcΔ1.ΔA in fitness (Extended Data Fig. 7) and morphology (Extended Data Fig. 8). To assess the effects on cellular fitness, we measured doubling time and MaxOD of each strain in LB and M9 media (Fig. 5). Doubling times for rEcΔ2.ΔA.B3.tW* increased by approximately 50% in M9 and 20% in LB, whereas MaxOD increased by around 16% in M9 and decreased by around 18% in LB. Further analysis dissecting the individual effects of ΔTGA recoding and translation factor mutations on growth revealed that RF2 S205P was

Fig. 5 | Growth of recoded strains. MaxOD and doubling time of recoded and laboratory *E. coli* strains grown in nutrient-rich LB (top) and nutrient-poor M9 (bottom) liquid media. Data are mean ± 95% confidence interval for *n* = 11 (LB) or 7 (M9) biological replicates. *P* values for comparison with rEcΔ2.ΔA.B1 by Mann–Whitney *U*-test.

Article

a primary contributing factor to fitness impairments (Extended Data Fig. 3a and Supplementary Discussion 3). These data reveal general growth impairment relative to ancestral rEcΔ1.ΔA, but with minor improvement in M9 MaxOD. However, the similarity in growth of rEcΔ2.ΔA.B1 with commercial strains TOP10 and DH10B in LB (Fig. 5) incentivizes further tuning of translation factors to improve growth and fitness.

nsAA dual incorporation into proteins

To assess the practical implications of disentangling translational crosstalk to liberate codons, we assessed the ability of rEcΔ2.ΔA. B3 and rEcΔ2.ΔA.B3.tW* (Ochre) to incorporate two distinct nsAAs at UAG and UGA in reporter proteins. Previous recoding of TAG and RF1 deletion in rEcΔ1.ΔA demonstrated multi-site incorporation of nsAAs in elastin-like polypeptides (ELPs) with more than 95% accuracy[13]. To reassign both UAG and UGA codons, we constructed a dual-OTS fluorescent reporter plasmid for incorporating *para*-acetyl-L-phenylalanine (pAcF) at UGA and N^ε-Boc-L-lysine (BocK) at UAG (Fig. 6 and Extended Data Fig. 9a). This plasmid contains a bicistronic cassette with L-arabinose-inducible (pBAD) o-aaRSs paired with constitutively active (proK) o-tRNAs assigned to each codon: pAzFRS.2.t1[13] paired with anticodon-modified *M. jannaschii* tRNA(UCA) (Mj-tRNA(UCA)) targeting UGA, and chPylRS[56] paired with optimized *PylT* targeting UAG. o-tRNAs were expressed with a valX linker for dual expression. For a reporter, we expressed a Tet repressor (TetR)–anhydrotetracycline (aTc)-inducible ELP–GFP[13] with the ELP sequence modified for MS-READ[55]. Seven ELP–GFP constructs were tested to assess nsAA encoding at UAG, UGA, or both within peptide linkers: (1) 1×TAG; (2) 3×TAG; (3) 1×TGA; (4) 3×TGA; (5) 1×(TGA-TAG); (6) 3×(TGA-TAG); and (7) 3×TAC as a control construct for native incorporation (Fig. 6a, Extended Data Fig. 9b and Supplementary Tables 18 and 20). We tested four strains to compare single and dual incorporation: rEcΔ1.ΔA.B0, rEcΔ2.ΔA.B1, rEcΔ2.ΔA.B3 and rEcΔ2.ΔA.B3.tW* (Fig. 6b–d).

We first analysed single nsAA incorporation at one and three codons. Without RF1 present, we expected minimal differences between strains incorporating BocK at UAG. Indeed, induction of OTS with expression of 3×TAG or 1×TAG and BocK demonstrated little appreciable difference in normalized fluorescence between strains (Fig. 6b and Extended Data Fig. 9c). As expected, induction of 3×TAG and OTS with pAcF presented around fourfold (rEcΔ1.ΔA), sevenfold (rEcΔ2.ΔA.B1), ninefold (rEcΔ2.ΔA.B3) and ninefold (rEcΔ2.ΔA.B3.tW*) increased fluorescence relative to no nsAAs (Fig. 6b). Expression of 1×TAG expression demonstrated a similar trend with higher normalized fluorescence and approximately twofold change (Extended Data Fig. 9c). These results reveal no negative effect on nsAA incorporation efficiencies at UAG resulting from ΔTGA recoding with or without translation engineering.

With translation engineering, we hypothesized that pAcF incorporation at UGA would significantly increase over wild-type RF2 variants. As expected, induction of 3×TGA and OTS with pAcF presented approximately threefold (rEcΔ1.ΔA), threefold (rEcΔ2.ΔA.B1), eightfold (rEcΔ2.ΔA.B3) and sevenfold (rEcΔ2.ΔA.B3.tW*) increased fluorescence relative to no nsAAs (Fig. 6c). Meanwhile, 1×UGA fluorescence fold changes were similar across all strains, resulting from increased fluorescence in the absence of nsAAs (Extended Data Fig. 9d). This effect probably resulted from previously observed mischarging of Mj-tRNA by pAzRS.2.t1[13]. Overall, these data reveal that TGA reassignment was not enabled by ΔTGA recoding alone, highlighting the role of RF2 and tRNA^Trp engineering in mitigating codon recognition to liberate and reassign UGA.

Finally, we assessed dual incorporation of BocK and pAcF at UAG and UGA, respectively. Induction of 3×(TGA-TAG) and OTS with BocK and pAcF presented approximately 2-fold (rEcΔ1.ΔA), 7-fold (rEcΔ2.ΔA. B1), 20-fold (rEcΔ2.ΔA.B3) and 17-fold (rEcΔ2.ΔA.B3.tW*) increased

fluorescence relative to no nsAAs (Fig. 6d). Expression of 1×TAG resulted in similar relative fluorescence fold changes, but with rEcΔ1.ΔA matching rEcΔ2.ΔA (Extended Data Fig. 9e). Demonstrably, the attenuation of UGA competition from native translation factors enabled UAG and UGA stop codon reassignment for multi-site encoding of two distinct nsAAs.

To assess translational accuracy, we used mass spectrometry to assay BocK and pAcF incorporation at six codons (three TGA and three TAG) in a modified MS-READ reporter peptide designed for unambiguous identification of two MS-READ reporter segments (Fig. 6e and Supplementary Table 21). Tandem mass spectra for each reporter segment confirmed incorporation of 3×BocK and 3×pAcF at all intended positions. Product ion spectra were more clearly differentiated by a strong 100-Da neutral loss signature at BocK sites (Fig. 6f). In rEcΔ2.ΔA.B3, peptide intensities for reporter segments revealed Trp incorporation at UGA, competing with pAcF, and Lys, Gln and Tyr incorporation at UAG, consistent with previous observations when reassigning UAG[13] (Fig. 6g,h), providing approximately 93% on-target incorporations. By contrast, results with rEcΔ2.ΔA.B3.tW* were completely free of Trp contamination and accompanied by a marked reduction in Lys, Gln and Tyr incorporation at UAG, yielding more than 99% on-target incorporation at all sites (Fig. 6g,h). This result demonstrates the effect of the tRNA^Trp A37G mutation in enhancing incorporation efficiency and mitigating competition from native UGA suppressors. These conclusions are validated by western blot assays dissecting translation at individual codons (Extended Data Fig. 9f and Supplementary Discussion 4). To evaluate the applicability of our expression chassis, we measured purified 3×(TGA-TAG) ELP–GFP containing BocK and pAcF dual incorporations, obtaining yields of approximately 2 mg l^−1 from Ochre, around 21-fold higher than yields from unrecoded BL21 (Extended Data Fig. 9g,h and Supplementary Discussion 5). Together, these results emphasize the utility of our work and the importance of mitigating native codon competition to facilitate nsAA incorporation into proteins with high fidelity.

Discussion

Here we describe an integrated genomic, biomolecular and protein engineering effort to disentangle translational crosstalk within the stop codon block, rendering four codons (UAG, UAA, UGA and UGG) functionally non-degenerate, thereby compressing the stop function to UAA while opening UAG and UGA for reassignment. This was achieved by replacing all genomic stop codons with synonymous TAA, coupled with engineering of tRNA^Trp and RF2 to mitigate UGA recognition. Deploying compatible OTSs, the resulting strain, Ochre, enabled UAG and UGA reassignment for multi-site incorporation of two distinct nsAAs within proteins with more than 99% accuracy. This result highlights the efficacy of integrating translation factor engineering with genomic recoding to more comprehensively compress, release and accurately reassign codons. As recoding efforts delete cognate translation factors to free codons[4,9], underlying webs of natively competing translation factors may be revealed, serving as a major barrier to comprehensive recoding. Our work overcomes this by engineering translation factors to disentangle crosstalk, which is essential to completely liberate codons, offering a new lens for recoding efforts across domains of life.

Codon recognition was a central focus of the study. Our efforts affirm that translation factor decoding is a delicate balance, involving highly tuned structural interactions, concentration kinetics and competition for codons (Supplementary Discussion 1), with RF2 decoding influenced by residues beyond the recognition loop (for example, E170K). Future recoding efforts will require engineering codon recognition. As in tRNA^Trp, natural anticodon modifications offer promise for future codon isolation. Modifications may be added or removed to widen or narrow codon specificity. Alternatively, ribosomal components[57,58] or RF3 may provide other levers for
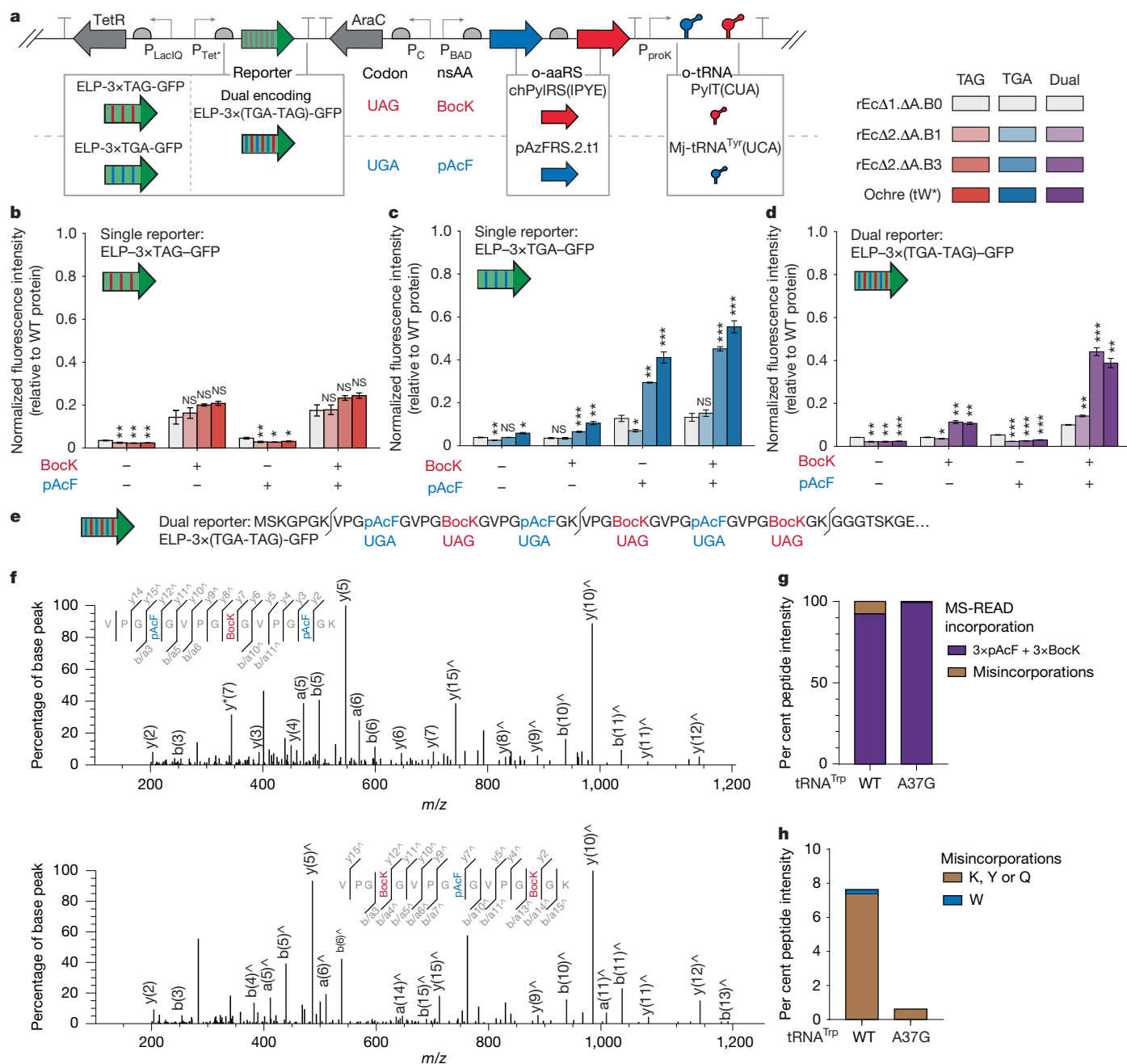
Fig. 6 | Single and dual incorporation of nsAAs at UAG and UGA in ELP–GFP proteins. a, Left, schematic depicting the dual-OTS and fluorescent reporter expression system. L-arabinose-inducible o-aaRSs and accompanying constitutively expressed UAG- and UGA-suppressing o-tRNAs coordinate charging and incorporation of the nsAAs BocK and pAcF at UAG and UGA, respectively, in one of three aTc-inducible fluorescent ELP–GFP reporters. Each reporter contains three instances of TAG, TGA or TAG and TGA within a central ELP–GFP linker to study single or dual nsAA encoding. Right, key representing nsAA incorporation at UAG, UGA or both within four variant recoded strains. **b–d**, Fluorescent expression of ELP–GFP reporter containing 3 TAG codons (**b**), 3 TGA codons (**c**) or 3 TAG and 3 TGA codons (**d**) in variant strains with or without 10 mM BocK and/or 1 mM pAcF added to the growth medium, along

with 0.05% (w/v) L-arabinose and 100 ng ml⁻¹ aTc for induction. Colour scheme as shown in the key in **a**. Fluorescence is normalized to optical density, then to 3×TAC control construct fluorescence/optical density. Data are mean ± 95% confidence interval for *n* = 3 biological replicates. *P* values for comparison with rEcΔ1.ΔA by unpaired *t*-tests (*n* = 3). WT, wild type. **e**, Diagrammatic depiction of dual reporter ELP–3×(TGA-TAG)–GFP with TAG (red) and TGA (blue) stop codon positions and associated nsAAs. **f**, Mass spectrometry data showing incorporation of BocK and pAcF in ELP–3×(TGA-TAG)–GFP reporter proteins expressed in rEcΔ2.ΔA.B3. ^ indicates BocK sites. **g**, MS-READ peptide intensity for expected BocK incorporation at 3×UAG and pAcF at 3×UGA, and misincorporations identified in all examined digested peptides. **h**, MS-READ peptide intensity of misincorporations from **g**.

modulating decoding. Given the limited sampling of translation factor mutations in this study (Supplementary Table 14b), there exists broader opportunity to engineer factor functionalities in ways that are uniquely enabled by GROs—for example: (1) directed evolution of translation machineries may foster kinetics that are impermissible

in other contexts; (2) additional release factor engineering to abolish UGA decoding may enable Ochre to further obstruct horizontal transfer of genetic elements[8–10] or improve protein yields; or (3) engineered noncanonical stop codons may enable bidirectional genetic isolation[7].

Article

This study provides novel insights into fundamental biology. Spontaneous mutations that arise in RF3 (Supplemental Table 14b) may help to elucidate regulatory mechanisms of translation. The replacement of TGA, decoupling it from stop function, provides insight into codon redundancy and cryptic codon functions removed during codon compression. Differential suppression of UGA codons by tRNAs in times of stress can alter cellular transcription patterns[50,59], demonstrating nonsynonymous stop codon functions. Other recoding efforts in *E. coli* have uncovered expression-regulating functions that are unique to AGG/AGA codons[60]. Future recoding efforts should take note of alternative functions of translation factors or codons, especially as these efforts reveal them. The effects of ΔTGA recoding on cellular responses in Ochre have yet to be investigated—this approach may have potential for improving fitness and investigating cryptic codon functions.

Our study provides insights into the science of genetic codes and translation by utilizing a synthetic biology approach that recodes genomes and repurposes the function of translation components in an integrated framework. More broadly, this study helps to establish design rules to probe and engineer genomes alongside translation factors to achieve a deeper and predictive understanding of the universal translation machinery and the canonical genetic code, setting the stage for engineering genomes with non-degenerate genetic codes. This research also offers valuable applications in safeguarding genetically modified organisms with enhanced genetic isolation[8,10,51], biocontainment[11,12] and expanded capacity for producing entirely new classes of synthetic proteins, biomaterials or therapeutic agents with diverse chemistries[13,14,23].

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-024-08501-x.

1. Shulgina, Y. & Eddy, S. R. A computational screen for alternative genetic codes in over 250,000 genomes. *eLife* **10**, e71402 (2021).
2. Ivanova, N. N. et al. Stop codon reassignments in the wild. *Science* **344**, 909–913 (2014).
3. Lekomtsev, S. A. Non-standard genetic codes and translation termination. *Mol. Biol.* **41**, 964–972 (2007).
4. Lajoie, M. J. et al. Genomically recoded organisms expand biological functions. *Science* **342**, 357–360 (2013).
5. Fredens, J. et al. Total synthesis of *Escherichia coli* with a recoded genome. *Nature* **569**, 514–518 (2019).
6. Young, R. E. & Purton, S. Codon reassignment to facilitate genetic engineering and biocontainment in the chloroplast of *Chlamydomonas reinhardtii*. *Plant Biotechnol. J.* **14**, 1251–1260 (2016).
7. Sengupta, S., Yang, X. & Higgs, P. G. The mechanisms of codon reassignments in mitochondrial genetic codes. *J. Mol. Evol.* **64**, 662–688 (2007).
8. Nyerges, A. et al. A swapped genetic code prevents viral infections and gene transfer. *Nature* **615**, 720–727 (2023).
9. Robertson, W. E. et al. Sense codon reassignment enables viral resistance and encoded polymer synthesis. *Science* **372**, 1057–1062 (2021).
10. Ma, N. J. & Isaacs, F. J. Genomic recoding broadly obstructs the propagation of horizontally transferred genetic elements. *Cell Syst.* **3**, 199–207 (2016).
11. Mandell, D. J. et al. Biocontainment of genetically modified organisms by synthetic protein design. *Nature* **518**, 55–60 (2015).
12. Rovner, A. J. et al. Recoded organisms engineered to depend on synthetic amino acids. *Nature* **518**, 89–93 (2015).
13. Amiram, M. et al. Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids. *Nat. Biotechnol.* **33**, 1272–1279 (2015).
14. Jin, X., Park, O. J. & Hong, S. H. Incorporation of non-standard amino acids into proteins: challenges, recent achievements, and emerging applications. *Appl. Microbiol. Biotechnol.* **103**, 2947–2958 (2019).
15. Isaacs, F. J. et al. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333**, 348–353 (2011).
16. Ostrov, N. et al. Design, synthesis, and testing toward a 57-codon genome. *Science* **353**, 819–822 (2016).
17. Lau, Y. H. et al. Large-scale recoding of a bacterial genome by iterative recombineering of synthetic DNA. *Nucleic Acids Res.* **45**, 6971–6980 (2017).
18. Zhao, Y. et al. Debugging and consolidating multiple synthetic chromosomes reveals combinatorial genetic interactions. *Cell* **186**, 5220–5236.e5216 (2023).
19. Richardson, S. M. et al. Design of a synthetic yeast genome. *Science* **355**, 1040–1044 (2017).
20. Soma, A. et al. An RNA-modifying enzyme that governs both the codon and amino acid specificities of isoleucine tRNA. *Mol. Cell* **12**, 689–698 (2003).
21. Petrullo, L. A., Gallagher, P. J. & Elseviers, D. The role of 2-methylthio-N6-isopentenyladenosine in readthrough and suppression of nonsense codons in *Escherichia coli. Mol. Genet. Genomics* **190**, 289–294 (1983).
22. Blanchet, S. et al. Deciphering the reading of the genetic code by near-cognate tRNA. *Proc. Natl Acad. Sci. USA* **115**, 3018–3023 (2018).
23. Vanderschuren, K. et al. Tuning protein half-life in mouse using sequence-defined biopolymers functionalized with lipids. *Proc. Natl Acad. Sci. USA* **119**, e2103099119 (2022).
24. Zheng, Y., Gilgenast, M. J., Hauc, S. & Chatterjee, A. Capturing post-translational modification-triggered protein–protein interactions using dual noncanonical amino acid mutagenesis. *ACS Chem. Biol.* **13**, 1137–1141 (2018).
25. McGowan, J. et al. Identification of a non-canonical ciliate nuclear genetic code where UAA and UAG code for different amino acids. *PLoS Genet.* **19**, e1010913 (2023).
26. Atkin, A. L., Riazi, M. A., Greer, C. L., Roy, K. L. & Bell, J. B. The functional analysis of nonsense suppressors derived from in vitro engineered *Saccharomyces cerevisiae* tRNA(Trp) genes. *Gene* **134**, 57–65 (1993).
27. Albers, S. et al. Repurposing tRNAs for nonsense suppression. *Nat. Commun.* **12**, 3850 (2021).
28. dos Reis, M., Savva, R. & Wernisch, L. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.* **32**, 5036–5044 (2004).
29. Kolisnychenko, V. et al. Engineering a reduced *Escherichia coli* genome. *Genome Res.* **12**, 640–647 (2002).
30. Wang, H. H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).
31. Wright, D. E. & O'Donoghue, P. Biosynthesis, engineering, and delivery of selenoproteins. *Int. J. Mol. Sci.* **25**, 223 (2023).
32. Deatherage, D. E. & Barrick, J. E. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods Mol. Biol.* **1151**, 165–188 (2014).
33. Ellis, H. M., Yu, D., DiTizio, T. & Court, D. L. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl Acad. Sci. USA* **98**, 6742–6746 (2001).
34. Petropoulos, A. D., McDonald, M. E., Green, R. & Zaher, H. S. Distinct roles for release factor 1 and release factor 2 in translational quality control. *J. Biol. Chem.* **289**, 17589–17596 (2014).
35. Shimizu, Y. ArfA recruits RF2 into stalled ribosomes. *J. Mol. Biol.* **423**, 624–631 (2012).
36. Mora, L., Heurgue-Hamard, V., de Zamaroczy, M., Kervestin, S. & Buckingham, R. H. Methylation of bacterial release factors RF1 and RF2 is required for normal translation termination in vivo. *J. Biol. Chem.* **282**, 35638–35645 (2007).
37. Uno, M., Ito, K. & Nakamura, Y. Functional specificity of amino acid at position 246 in the tRNA mimicry domain of bacterial release factor 2. *Biochimie* **78**, 935–943 (1996).
38. Baranov, P. V., Gesteland, R. F. & Atkins, J. F. Release factor 2 frameshifting sites in different bacteria. *EMBO Rep.* **3**, 373–377 (2002).
39. Craigen, W. J. & Caskey, C. T. Expression of peptide chain release factor 2 requires high-efficiency frameshift. *Nature* **322**, 273–275 (1986).
40. Nakamura, Y. & Ito, K. A tripeptide discriminator for stop codon recognition. *FEBS Lett.* **514**, 30–33 (2002).
41. Uno, M., Ito, K. & Nakamura, Y. Polypeptide release at sense and noncognate stop codons by localized charge-exchange alterations in translational release factors. *Proc. Natl Acad. Sci. USA* **99**, 1819–1824 (2002).
42. Wannier, T. M. et al. Adaptive evolution of genomically recoded *Escherichia coli. Proc. Natl Acad. Sci. USA* **115**, 3090–3095 (2018).
43. Young, D. J., Edgar, C. D., Poole, E. S. & Tate, W. P. The codon specificity of eubacterial release factors is determined by the sequence and size of the recognition loop. *RNA* **16**, 1623–1633 (2010).
44. Korostelev, A. et al. Crystal structure of a translation termination complex formed with release factor RF2. *Proc. Natl Acad. Sci. USA* **105**, 19684–19689 (2008).
45. Sund, J., Ander, M. & Aqvist, J. Principles of stop-codon reading on the ribosome. *Nature* **465**, 947–950 (2010).
46. Ito, K., Uno, M. & Nakamura, Y. A tripeptide 'anticodon' deciphers stop codons in messenger RNA. *Nature* **403**, 680–684 (2000).
47. Mora, L., Zavialov, A., Ehrenberg, M. & Buckingham, R. H. Stop codon recognition and interactions with peptide release factor RF3 of truncated and chimeric RF1 and RF2 from *Escherichia coli. Mol. Microbiol.* **50**, 1467–1476 (2003).
48. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
49. Wei, Y., Wang, J. & Xia, X. Coevolution between stop codon usage and release factors in bacterial species. *Mol. Biol. Evol.* **33**, 2357–2367 (2016).
50. Fan, Y. et al. Heterogeneity of stop codon readthrough in single bacterial cells and implications for population fitness. *Mol. Cell* **67**, 826–836.e825 (2017).
51. Ma, N. J., Hemez, C. F., Barber, K. W., Rinehart, J. & Isaacs, F. J. Organisms with alternative genetic codes resolve unassigned codons via mistranslation and ribosomal rescue. *eLife* **7**, e34878 (2018).
52. Steege, D. A. A nucleotide change in the anticodon of an *Escherichia coli* serine transfer RNA results in supD-amber suppression. *Nucleic Acids Res.* **11**, 3823–3832 (1983).
53. Hirsh, D. & Gold, L. Translation of the UGA triplet in vitro by tryptophan transfer RNA's. *J. Mol. Biol.* **58**, 459–468 (1971).
54. Lamichhane, T. N., Blewett, N. H. & Maraia, R. J. Plasticity and diversity of tRNA anticodon determinants of substrate recognition by eukaryotic A37 isopentenyltransferases. *RNA* **17**, 1846–1857 (2011).
55. Mohler, K. et al. MS-READ: quantitative measurement of amino acid incorporation. *Biochim. Biophys. Acta* **1861**, 3081–3088 (2017).

56. Bryson, D. I. et al. Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat. Chem. Biol.* **13**, 1253–1260 (2017).
57. Sato, H., Ito, K. & Nakamura, Y. Ribosomal protein L11 mutations in two functional domains equally affect release factors 1 and 2 activity. *Mol. Microbiol.* **60**, 108–120 (2006).
58. Goringer, H. U., Hijazi, K. A., Murgola, E. J. & Dahlberg, A. E. Mutations in 16S rRNA that affect UGA (stop codon)-directed translation termination. *Proc. Natl Acad. Sci. USA* **88**, 6603–6607 (1991).
59. Romero Romero, M. L. et al. Environment modulates protein heterogeneity through transcriptional and translational stop codon readthrough. *Nat. Commun.* **15**, 4446 (2024).
60. Napolitano, M. G. et al. Emergent rules for codon choice elucidated by editing rare arginine codons in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **113**, E5588–E5597 (2016).

Article

Methods

DNA

Oligonucleotides were purchased from Integrated DNA Technologies with standard purification and desalting. See Supplementary Data for all oligonucleotides and DNA constructs.

Media

Unless otherwise stated, all cultures were grown in LB-Lennox medium (10 g l^{-1} bacto tryptone, 5 g l^{-1} sodium chloride, 5 g l^{-1} yeast extract). LB agar plates were composed of LB plus 15 g l^{-1} bacto agar. M9 minimal medium (12.8 g l^{-1} Na$_2$HPO$_4$, 3 g l^{-1} KH$_2$PO$_4$, 1 g l^{-1} NH$_4$Cl, 0.5 g l^{-1} NaCl, 3 mg l^{-1} CaCl$_2$) was adjusted to pH 7.5 with 10 M NaOH. For phage experiments, Tryptone-KCl (TK) liquid medium comprised of 10 g l^{-1} Tryptone, 5 g l^{-1} KCl and 0.5 ml l^{-1} of 1 M CaCl$_2$. TK sloppy agar contained 10 g l^{-1} Tryptone, 5 g l^{-1} NaCl and 7 g l^{-1} agar. TK bottom agar contained 10 g l^{-1} Tryptone, 2.5 g l^{-1} NaCl, 2.5 g l^{-1} KCl, 0.5 ml l^{-1} of 1M CaCl$_2$ and 10 g l^{-1} agar. Super optimal broth with catabolite repression (SOC) liquid medium contained (20 g l^{-1} tryptone, 5 g l^{-1} yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl$_2$, 10 mM MgSO$_4$ and 20 mM glucose). For protein yield quantification, terrific broth medium (TB) contained 11.8 g l^{-1} tryptone, 23.6 g l^{-1} yeast extract, 9.4 g l^{-1} K$_2$HPO$_4$, 2.2 g l^{-1} KH$_2$PO$_4$ and 0.8% v/v glycerol.

Selective agents. ColE1 was expressed in strain JC411 and purified as previously described[61]. All other selective agents were purchased commercially: carbenicillin (50 μg ml^{-1}), chloramphenicol (40 μg ml^{-1}), gentamycin (10 μg ml^{-1}), kanamycin (30 μg ml^{-1}), sodium dodecyl sulfate (SDS) (0.005% w/v), spectinomycin (190 μg ml^{-1}), tetracycline (15 μg ml^{-1}), hygromycin (150 μg ml^{-1}), zeocin (10 μg ml^{-1}) and colicin E1 (ColE1; ~10 μg ml^{-1}).

Promoter inducers. Plasmid-induced pORTMAGE[62] and Recombineering[63]: L-arabinose (0.2% w/v, pBAD promoter), rhamnose (0.15% w/v, *rhaB* promoter). Plasmid maintenance was carried out with tetracycline (5–10 ng ml^{-1}, *tetA* selectable marker).

OTS expression for nsAA incorporation assays: L-arabinose (0.05% w/v, pBAD promoter) and anhydrotetracycline (100 ng ml^{-1}, pL *tetO* promoter).

Non-standard amino acids. BocK was purchased from Chem-Impex (00363) and dissolved in LB or TB to a final concentration of 10 mM. pAcF was purchased from Chem-Impex (24756), dissolved in sterile water to a concentration of 50 mM and used at a final concentration of 1 mM.

Strains

All strains were built on rEcΔ1.ΔA-C321.ΔA[4] (MG1655 dnaG.Q576A. exoX.K28TAA.ΔtolC.ΔprfA.ΔmutS::zeo.Δ (ybhB-bioAB)::[λcI857 N(cro-ea59)::*tetR-bla*].12B.tolQRA); with all instances of TAG codons converted to TAA and 1,195 naturally occurring TGA codons.

Selectable marker preparation

Selectable markers were amplified by PCR (40 μl per reaction) performed using Kapa HiFi HotStart ReadyMix according to the manufacturer's protocols with annealing at 61 °C. Primers were designed via Benchling primer creation software, confirmed with Kun's oligonucleotide T_m calculator (http://arep.med.harvard.edu/kzhang/cgi-bin/myOligoTm.cgi). PCR products were purified using a Qiagen PCR purification kit, eluted in 30 μl dH$_2$O and quantified either using an Eppendorf BioPhotometer plus or NanoDrop ND1000 spectrophotometer. For analysis, amplicons were run on a 1.5% agarose gel stained with ethidium bromide to confirm the expected band sizes.

TGA-to-TAA recoding

The starting strain was a standard rEcΔ1.ΔA (C321.ΔA) with a zeocin marker cassette replacing a chloramphenicol marker at the *mutS* locus (Supplementary Table 10). The *ybgC-tolQRA* locus was duplicated between the *ycgV* and *ychF* genes as previously described[64] to increase the fidelity of ColE1 selections. The previously reported mutations to *dnaG* and *exoX* to improve MAGE efficiency were also present in the starting strain[65]. All genes containing TGA codons were identified from WGS (Supplementary Table 1; see Methods section 'WGS and analysis' below). Eliminating TGA codons required a combined strategy of MAGE-mediated conversions of TGA to TAA and non-essential gene deletions (Supplementary Table 7). The oligonucleotides were designed as previously described[30], ordered from IDT and grouped into pools up to 11 oligonucleotides per pool at 10–12 μM total DNA per pool, regardless of number of oligonucleotides (N) (Supplementary Table 1). These pools were used for $N + 1$ rounds of MAGE before cultures were plated. MAGE was performed as previously described[30]. Forty-seven colonies were picked and screened using Multiplex allele-specific colony PCR (MASC-PCR), as previously described[15]. Simultaneously, picked colonies were grown to confluence, diluted 1:15 into 150 μl of M9 minimal medium, then inoculated 1:100 into 150 μl LB and M9 to assess growth phenotypes (MaxOD, doubling time and lag time) (see 'Fitness analysis'). The colony containing the largest number of lowest frequency conversions and minimal deviations in growth fitness was chosen for subsequent rounds of MAGE.

Native RBS predictions. Predicted RBS translation rates of genes overlapped by TGA stop codons before and after recoding to TAA were calculated in alignment with previous protocols[36].

MAGE and λ-Red-mediated recombination. MAGE, pORTMAGE and λ-Red-mediated recombination were performed as previously described[15,62]. Cells were transferred to 0.1 cm cuvettes, electroporated (BioRad GenePulser, 1.78 kV, 200 Ω, 25 μF) and immediately resuspended into 3 ml LB (MAGE) or 1 ml SOC medium (dsDNA), grown at 34 °C, 225 rpm. For *tolC* and *galK* negative selections, cultures were recovered for at least 7 h to allow complete protein turnover before exposure to ColE1 and 2-deoxygalactose, respectively. Once all strains were conjugated into a final strain (see 'CAGE assembly'), genomically integrated λ-Red was displaced by *tolC* negative selection. To convert background mutations and remaining TGA codons we employed a plasmid-based single stranded DNA recombineering approach. L-arabinose-inducible λ-Red-derived recombineering machinery was expressed on a temperature-curable plasmid. For double stranded DNA (dsDNA) recombination, rhamnose-inducible episomal λ-Red-derived recombineering machinery was also expressed on a temperature-curable plasmid containing L-arabinose-inducible *ccdA* antitoxin for negative selections as previous described[63] for both marker removal and release factor placement/displacement. Both λ-Red plasmids contained *tetA* selectable markers. To cure plasmids, strains were incubated with λ-Red induction at 37 °C overnight in 3 ml LB, plated onto solid medium, incubated at 42 °C for 2–4 h, then incubated at 37 °C until colonies were visible. Colonies were picked into 150 μl LB, incubated for 3 h, then transferred 10 μl into LB with tetracycline to identify colonies lacking plasmids.

Genotyping. MASC-PCR was used to simultaneously detect up to 11 TGA-to-TAA conversions or background mutation reversions, as previously described with TAG conversions[15]. All primers are designed with a 61 °C annealing temperature, according to Kun's oligonucleotide T_m calculator, while avoiding 3′ end binding in secondary structure, according to Benchling's primer secondary structure prediction function. Each reaction consisted of KAPA 2 G Fast Multiplex ReadyMix (Kapa Biosystems, KK5802), 2 μl of template DNA and 0.2 μM of each

primer for each 10 µl reaction. MASC-PCR results were run on 2.2% agarose gels with ethidium bromide staining. After λ-Red-mediated recombination or conjugation, colony PCR was used to confirm the presence and absence of selectable markers at desired positions. Colony PCR (10 µl per reaction; annealing at 61 °C (Kun's oligonucleotide T_m calculator)) was performed using Kapa 2G Fast HotStart ReadyMix following manufacturer's protocols. Results were analysed on a 1.5% agarose gel stained with ethidium bromide. Sanger sequencing was performed by Genewiz.

CAGE assembly. Conjugative assembly genome engineering (CAGE) was performed using the protocol as previously described[15]. Deviating from previous protocols, the donor strain had two positive markers— a spectinomycin marker and a gentamycin marker, each flanking the recoded region, in addition to the kanamycin resistance origin of transfer (kan^R-$oriT$) cassette set ~3–5 kb upstream (Extended Data Fig. 2 and Supplementary Table 10). The donor strain also contained a modified RK24 plasmid in which all genes that end in a TAG codon were recoded to TAA[10]. Cell spots were rinsed twice with 500 µl LB and collected to give a 1 ml 10^{-2} dilution of conjugated cells, followed by an additional 10-fold dilution in a new tube. 50 µl of 10^{-2} and 10^{-3} dilutions were each plated onto LB agar plate with appropriate antibiotics to select for both the recipient background selectable markers and the donor region selectable markers. Forty-seven candidate colonies were grown in a 96-well format and screened for desired genotypes via PCR (to confirm presence and absence of selectable markers) and MASC-PCR (to confirm the presence of interspersed desired codon replacements). For large genomic transfers (>1 Mb), final strains were sequence verified through WGS (see 'Genotyping'). The resulting strain contained the three selectable markers, all of which were subsequently deleted via dsDNA λ-Red tolC-mediated selection–counterselection (via SDS or ColE1 selection, respectively), or maintained for the next conjugation[64].

Genomic deletions. To reduce the overall number of genes requiring ΔTGA recoding via MAGE, 16 multigenic regions containing a total of 229 genes (3 later categorized as pseudogenes) were identified for deletion based on previous work by the Blattner laboratory[29] (Extended Data Fig. 1). Each targeted deletion site (up to ~34 kb in size) was displaced via tolC selectable marker displacement and counterselection for subsequent marker removal in accordance with previous protocols[64]. ColE1 selections on solid medium were performed as previously described[15,64]. For pre-selection, 5 µl of recovered cultures were inoculated into 150 µl LB with either carbenicillin (control) or ColE1 with vancomycin (64 µg ml⁻¹), in triplicate, in a 96-well plate and incubated at 34 °C for 16 h to monitor growth along the progenitor strain with tolC. Strains exhibiting growth in both carbenicillin and ColE1 with vancomycin were considered positive for tolC deletion and subsequently plated onto LB agar plates with ColE1 for monoclonal colony selection and PCR screened to confirm the loss of tolC. Subsequent MASC analyses allowed for selection of strains with both deletion events and TGA conversions to reduce the total number of MAGE cycles. After every tolC placement and gene displacement, strain growth curves were analysed to assess gene deletion impact on strain phenotype (see 'Fitness analysis'). Major growth impairments were assessed. Additional major deletions (>50 bp) resulted either from mutagenesis of highly repetitive noncoding genes, active transposable elements or scarring from conjugations.

WGS and analysis. Whole genomes were isolated using the Qiagen DNeasy Blood and Tissue isolation kit. For high fidelity genome analysis performed after conjugations, raw reads were acquired via short-read 150-bp (50× coverage) paired-end Illumina sequencing data were collected with Hiseq 4000 with libraries prepared by the Yale Center for Genome Analysis. Long-read (<25 kb, 50× coverage) data were prepared by Pacific Biosciences Single Molecule Real-Time (SMRT) Analysis.

For rapid whole-genome analysis, raw reads were acquired via Plasmidsaurus Oxford nanopore standard bacterial WGS (30× coverage) via bacterial pellet submissions suspended in DNA/RNA Shield. To perform analyses on raw reads, latest versions of breseq 0.38.1 computation pipeline were employed for aligning sequence reads to rEc1.ΔA (C321.ΔA) reference genome, run using Ubuntu LTS on Windows Subsystem for Linux, in accordance with https://barricklab.org[37]. Summary .html outputs provided comprehensive lists of mismatches, indels and missing or novel junctions to monitor TGA conversion progress, deletions and background mutagenesis. Mutations were then organized by type in Excel (see Supplementary Tables).

AlphaFold structure of RF2.B3
The 3D structure of RF2.B3 in Fig. 2b was acquired through Benchling AlphaFold[48] online prediction from amino acid sequence submission (Supplementary Table 17) and spatially oriented for view of primary reaction sites.

Liquid selection complementation
In triplicate, strains harbouring prfB variant expression plasmids (as described in Extended Data Fig. 4 and Supplementary Discussion 1) were electroporated with a kanamycin resistance (kan^R) dsDNA cassette for displacement of genomic prfB and recovered in varying concentrations of vanillic acid inducer for 3 h at 37 °C to allow episomal expression of prfB variants. Strains were then diluted 1:50 into selective LB medium 25 µg ml⁻¹ kanamycin and incubated with shaking at 37 °C, monitoring absorbance at 600 nm in a BioTek Synergy H1 plate reader (Agilent). After 36 h of growth, OD_{600} for each strain was reported as viability after complementation. Individual colonies were isolated by plating knockout strains on selective medium after recovery and PCR screened to confirm kan^R displacements of genomic prfB. Colonies that tested positive with PCR screen were confirmed by WGS (Plasmidsaurus) to not harbour genomic prfB or secondary mutations, and whole-plasmid sequencing (Plasmidsaurus) of their prfB-containing plasmid to confirm variant.

Readthrough fluorescence assay
We used a dual fluorescence mCherry–YFP reporter to test stop codon readthrough as previously reported[66]. This reporter was constitutively expressed from a low copy backbone (p15A) to maximize dynamic range for detecting post-transcriptional fluctuations. Each evaluated strain was transformed with four variants of the reporter (three stop codons and a sense codon positive control GCG) and plated on selective medium for overnight growth. Colonies were picked in triplicate into 150 µl cultures of LB with appropriate selection in 96-well plates, grown with shaking at 225 rpm in 37 °C for 18 h after which timepoints measurements were taken for mCherry (excitation: 585 nm, emission: 635 nm, gain: 100), YFP (excitation: 500 nm, emission: 541 nm, gain: 80) and OD_{600} (absorbance: 600 nm) in a BioTek Synergy H1 plate reader (Agilent). To assess cognate amber or opal suppression, a plasmid bearing supD[52] with CUA or UCA anticodon was co-transformed with the dual fluorescence reporter plasmid. Fluorescence values of all strain and plasmid conditions were normalized to OD_{600} before dividing by the positive control described above to arrive at fractional mCherry and YFP signals reported.

Bacteriophage assays
For all phage experiments, growth was carried out in TK at 37 °C; infection and propagation occurred in TK sloppy agar poured onto solid TK bottom agar incubated at 37 °C. TK sloppy agar mixes were maintained at 45 °C before use.

Phage propagation. For propagation, *E. coli* MG1655 was grown to mid-log phase in 3 ml of LB. Two-hundred microlitres of bacteria was added to 3 ml TK sloppy agar. Immediately following, 50 µl of phage

Article

(10 to 100-fold dilutions) was added directly from refrigerated stock into the sloppy-bacterial culture. Three millilitres of sloppy bacteria phage culture was poured onto solid TK bottom agar plates, dispersed evenly, then left at room temperature to solidify. Plates were then incubated at 37 °C for ~16 h to permit lysis to proceed to completion. The entire sloppy agar was collected and centrifuged (12,000g, 2 min) and 3 ml of supernatant was filtered with 0.22-µm filter column to remove bacteria.

Phage titration. Bacteria strains were grown overnight at 37 °C until OD_{600} reached 2–3. Twenty microlitres of bacteria was added to 3 ml sloppy agar. Three millilitres of sloppy bacteria culture was poured onto solid TK bottom agar plates. After incubated at room temperature for 15 min, 3 µl of the phage dilutions (10^1- to 10^8-fold dilutions) were dropped on the surface of the solidified sloppy agar. Once the drops dried, plates were incubated overnight at 37 °C. Visible plaques were counted for the individual drops. Titres (in PFU per ml) were calculated according to PFU per ml = $N \times \frac{1}{DF} \times \frac{1}{V}$, where N is the number of plaques, DF is the phage dilution factor and V is the volume of phage dilution pipetted on the plate.

Fitness analysis

Kinetic growth curves. Kinetic growth (OD_{600}) curves were obtained via monitoring strain growth within a BioTek Synergy HT plate reader. Each strain was grown in triplicate within 96-well plates, 150 µl LB and 150 µl M9, within 96-well flat-bottom plates, incubated at 34 °C for intermediate λ-Red+ strains and 37 °C for final strains, 225 rpm, for 16–36 h and absorbance at 600 nm was read at 10-min intervals. In preparation, strains were grown to confluence in LB, diluted 1:15 into 150 µl of M9 minimal medium and inoculated 1:100 into 150 µl LB and M9. Auxotrophic strains revealed no growth in M9.

OD_{600} calibration. The absorbance obtained by the BioTek Synergy HT plate reader was recalibrated to OD_{600} (absorbance at 600 nm through 1 cm pathway) using a standard curve $y = 2.746x + 1.878x^2$. The OD_{600} of an overnight LB culture of MG1655 was measured with a Biochrom Libra S4 Spectrophotometer at 600 nm wavelength in a semi-micro cuvette (1 cm pathway) after 1:10 dilution of culture into 1 ml LB. To generate the calibration curve, a series of cultures with OD_{600} ranging from 0 to 6 were prepared by diluting in LB medium. These cultures were then measured by the BioTek Synergy HT plate reader, with the same settings as growth cultures. The average values were then fitted to a polynomial standard curve for recalibration. The effects of medium evaporation in plate wells are not considered.

Doubling time and MaxOD calculation. The recalibrated growth curve was used to calculate the doubling time and MaxOD. Linear fitting of $log_2 OD_{600}$ was performed using a sliding-window method[67], where the window size is 50 min for rich medium and 100 min for M9 minimal medium in the early log phase, respectively. The doubling time was calculated as the reciprocal of the slope. MaxOD was obtained within the 36 h growth period. Individual biological sample $n \geq 4$.

nsAA incorporation assays

Plasmid construction. Gene fragments for aaRS and tRNA were synthesized by Twist Bioscience and cloned into expression vectors by Golden Gate assembly. Plasmids were sequence verified by whole-plasmid sequencing (Plasmidsaurus or Quintarabio). All cloning was made in Mach1 (Thermo Fisher, C862003).

Incorporation of BocK and pAcF into proteins. Recoded strains were transformed with OTS-reporter plasmids by standard electroporation protocols. Electroporated strains were recovered in 2 ml LB or SOC for at least 2 h before plating onto LB agar plates with kanamycin (50 µg ml⁻¹) and incubated at 37 °C overnight. Three single colonies from each plate were picked and grown in 800 µl LB supplemented with kanamycin (50 µg ml⁻¹) in a 96 deep-well plate sealed with a Breathe-Easy film (Sigma-Aldrich) and incubated at 37 °C with shaking at 220 rpm for 20–24 h. For nsAA incorporation at UAG and/or UGA, after overnight growth the cultures were back-diluted 1:50 onto a clear-bottom black 96-well plate (Costar) in a total of 150 µl of LB supplemented with kanamycin (50 µg ml⁻¹), aTc (100 ng ml⁻¹), L-arabinose (0.05% w/v), 1 mM pAcF and/or 10 mM BocK. Cell growth (absorbance at OD_{600}) and GFP fluorescence (excitation 485 nm, emission 525, gain 70, bottom measurement) were measured in a BioTek Synergy H1 plate reader (Agilent) for 24 h at 10 min intervals with linear shaking. Data were analysed with a custom Python script. All reported GFP fluorescence was normalized to ELP–3×TAC–GFP control fluorescence within each corresponding condition.

Expression of ELP–3×(TGA-TAG)–GFP with nsAAs for mass spectrometry. Single colonies of rEcΔ2.ΔA.B3 and rEcΔ2.ΔA.B3.tW* (Ochre) containing the dual-OTS dual reporter plasmid were inoculated in 2 ml LB with 50 µg ml⁻¹ kanamycin in a 14 ml falcon tube overnight at 37 °C with shaking at 220 rpm. After overnight growth the cultures were diluted 1:100 in 25 ml LB-kanamycin and grown until OD_{600} at 0.6, where the cultures were supplemented with 1 mM pAcF and 10 mM BocK and induced with 100 ng ml⁻¹ aTc and 0.05% w/v L-arabinose. The cultures were grown at 37 °C overnight, after which the cells were collected by centrifugation at 3,200g for 20 min in a 50 ml centrifuge tube and stored at −80 °C until protein purification.

Protein mass spectrometry

ELP–GFP reporter protein purification. Frozen *E. coli* cell pellets were thawed on ice and pellets were lysed by sonication with lysis buffer consisting of 50 mM Tris-HCl (pH 7.4, 23 °C), 500 mM NaCl, 0.5 mM EGTA, 1 mM DTT, 10% glycerol, 50 mM NaF and 1 mM Na_3O_4V. The extract was clarified with two rounds of centrifugation performed for 20 min at 4 °C and 14,000g. Cell-free extracts were applied to Ni-NTA metal affinity resin and purified according to the manufacturer's instructions. Wash buffers contained 50 mM Tris pH 7.5, 500 mM NaCl, 0.5 mM EGTA, 1 mM DTT, 50 mM NaF, 1 mM Na_3VO_4 and 10 mM imidazole. Proteins were eluted with a wash buffer containing 250 mM imidazole. Eluted protein was subjected to 4 rounds of buffer exchange (20 mM Tris pH 8.0 and 100 mM NaCl) and concentrated using a 30 kDa molecular weight cut-off spin filter (Amicon).

Protein digestion and mass spectrometry. Affinity purified, buffer exchanged ELP–GFP reporter protein, or whole cell lysates, were digested and analysed by mass spectrometry as described previously with some modifications[55,68,69]. ELP–GFP reporter protein (5–10 µg) was diluted with water and 20% SDS for a final volume of 115 µl and final concentration of 1% SDS. Samples were denatured for 15 min at 55 °C in a heat block. Reduction and alkylation of cysteines was performed with TCEP and 2-chloroacetamide (CAM) using a final TCEP and CAM concentration of 10 mM and 44 mM respectively. The reduction-alkylation reaction proceeded for 20 min at 55 °C. 6 µl of 50 mg ml⁻¹ SP3 beads (Speed Bead, Cytiva) pre-washed and resuspended with water were added to samples for a final working volume of 134 µl. Binding of protein to the beads was induced by adding 150 µl of 100% ethanol. The binding mixture was incubated in a ThermoMixer Eppendorf at 24 °C for 10 min at 1,400 rpm. After binding, the beads were magnetized on a magnetic rack and supernatants were removed. This was followed by three rounds of bead washes with 500 µl of 80% ethanol per wash. All traces of 80% ethanol were removed after the last wash. Beads in each sample were resuspended with 50 µl of digestion solution containing 0.4 µg of sequencing grade trypsin (Promega) in 50 mM triethylammonium bicarbonate (TEAB) buffer (Sigma). Digests were incubated for 16 h at 37 °C in a ThermoMixer at 1,400 rpm. Beads were magnetized and 50 µl supernatants were moved to fresh tubes. Beads were

resuspended in 50 µl of 50 mM TEAB buffer and incubated for 5 min at 37 °C in a ThermoMixer at 1,400 rpm to maximize peptide recovery. Beads were magnetized again and the two 50 µl supernatants were combined. Peptides were dried in a vacuum centrifuge at room temperature. Dried peptides were reconstituted in 2/98 acetonitrile/water with 0.1% formic acid and analysed by LC–MS/MS. LC–MS/MS was performed using a Vanquish Neo UHPLC system (Thermo) and an Orbitrap Eclipse Tribrid Mass Spectrometer (Thermo). The analytical column employed was a 75 µm inner diameter, fused silica capillary tube (Molex) packed in-house to a length of 15 cm with 1.9 µm ReproSil-Pur 120 Å C18-AQ (Dr. Maisch) using methanol as the packing solvent. Column was attached to a PepSep Spray Adapter with a fused silica emitter (Bruker). Peptide separation was achieved using mixtures of 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B) with a 41-min gradient; 0/5, 30/30, 39/45, 40/55, 41/100 (time (min)/B (%), linear ramping between steps). The gradient was performed with a flowrate of 300 nl min^{-1}. At least one blank injection (5 µl 2% B) was performed between samples to eliminate peptide carryover on the analytical column. One-hundred nanomoles of trypsin-digested BSA and 100 ng of trypsin-digested HeLa protein standard were run periodically between samples as quality control standards. The mass spectrometer was operated with the following parameters: (MS1) 60,000 orbitrap resolution, 250% normalized AGC target, 50 ms maximum injection time, 300–1,400 m/z scan range; (data dependent-MS2) Ion Trap detector, 200% normalized AGC target, 13 ms maximum injection time, top 10 mode, 1.2 m/z isolation window, 30% normalized HCD collision energy, 40 s dynamic exclusion. Data were searched using MaxQuant version 1.6.10.43 with deamidation (NQ), oxidation (M) and phospho (STY) as variable modifications and carbamidomethyl (C) as a fixed modification with up to 3 missed cleavages, 5 amino acids minimum length and 1% false discovery rate against a modified Uniprot *E. coli* database containing custom MS-READ reporter proteins. MS-READ search results were analysed using MaxQuant and Perseus version 1.6.2.2

Protein yield quantification

Protein expression. Precultures were inoculated from fresh colonies picked from selective LB agar plates (LB plates + antibiotics at 37 °C for 24 h) into 2 ml TB with 50 µg ml^{-1} kanamycin and grown overnight at 37 °C with shaking at 220 rpm. Overnight precultures were diluted to OD_{600} of 0.1 in 20 ml TB with 50 µg ml^{-1} kanamycin and grown until $OD_{600} = 0.6$–0.8. pAcF was added to a final concentration of 1 mM. Reporter and OTS expression was induced with 100 ng ml^{-1} aTc and 0.05% arabinose. Cells were collected 8 h post-induction by centrifugation at 4,000g and the supernatant was removed. Cell pellets were resuspended in 1 ml wash buffer (50 mM Tris-HCl, 150 mM NaCl, pH 7.5), transferred to 1.7 ml tubes and centrifuged at 17,000g. The wash buffer was then removed and the cell pellet stored at −80 °C until further use.

Lysis and extraction. Pellets were resuspended in sufficient volume of wash buffer to normalize OD_{600} to 10. 1 ml of cells were lysed using Sigma-Aldrich 1× BugBuster (from 10× stock) and centrifuged at 16,000g to isolate soluble fraction.

Quantifying GFP concentration. GFP quantification methods were adapted from previous work[70]. An ELP–sfGFP standard with an N-Terminal 6×His tag (containing no in-frame stop codon and cloned in a pET28a vector) was expressed in BL21. The ELP–sfGFP standard was purified using a Ni-NTA column and the protein was quantified using a Pierce BCA (Bicinchoninic Acid) Protein Assay kit (Thermo Fisher) using established methods[68]. A GFP standard curve was produced by spiking the purified GFP standard into lysate prepared from Ochre that contained no GFP. GFP was spiked in at 28, 18.6, 12.5, 8.3, 5.5, 3.7, 2.5, 1.6, 1.1, 0.7, 0.5 and 0.3 µg ml^{-1} to generate triplicate, linear standard GFP curves. Lysates of different GFP expressing strains were prepared in a similar fashion and dilutions were fit to the linear range of the standard

curve. Fluorescence of clarified lysates was measured in an H1M BioTek plate reader (excitation: 485 nm, emission : 510 nm, gain: 60) alongside the GFP standard. Fluorescence values were converted to µg ml^{-1} and used to calculate total yields of expressed protein.

Statistical analysis

Statistical significance was typically generated from two-tailed unpaired t-tests. P values for mCherry–YFP readthrough assays (Fig. 2e,f and Extended Data Fig. 5e) were calculated using Welch's unpaired t-test by comparison to rEcΔ1.ΔA within each paired codon and suppressor condition. Error bars display s.e.m., n = 3. P values for phage infection assays (Fig. 3) and tRNA^{Trp} UGA suppression assays (Fig. 4c,d) were derived from unpaired t-tests by comparison to *E. coli* MG1655 within each graph. Error bars display 95% confidence interval, n = 3. P values for tRNA^{Trp} UGA suppression assays (Fig. 4c) were derived from unpaired t-tests in relation to no Mj-TyrOTS plasmid within each strain. Error bars display 95% confidence interval, n = 3. P values for strain doubling times and maximum OD_{600} were calculated using Mann–Whitney U-tests comparing strain values with those of rEcΔ1.ΔA within each graph. Error bars display 95% confidence interval, n = 7 or 11 replicates (Fig. 5), n = 8 (Extended Data Fig. 3, LB) or n = 4 (Extended Data Fig. 3, 2×YT and TB) replicates. P values in ELP–GFP fluorescence assays assessing nsAA incorporation efficiencies (Fig. 6b–d and Extended Data Fig. 9c–e) were derived from unpaired t-tests comparing data to rEcΔ1.ΔA within each nsAA condition. Error bars display 95% confidence interval, n = 3. P values for RF2 complementation fluorescent protein production (Extended Data Fig. 4f and Supplementary Fig. 3a) compare fluorescence of GFP expressed from RF2 B2(P205) complemented strains to fluorescence of B1(wild type (S205)) strains with a similar genotype and vanillic acid concentration using unpaired t-tests (n = 3). Error bars display s.e.m. P values comparing RF2.B1 and B2 TGA readthrough (Supplementary Fig. 3b) for various vanillic acid concentrations compared to 0.1 µM vanillic acid within each strain release factor variant were calculated by Welch's t-tests (n = 3). Error bars display s.e.m. Calculations and graphs were generated using Graph Pad Prism (v.10.1.0). Figures were prepared using Adobe Illustrator 2023 (v.28.2). All measurements were taken from distinct samples except when measured repeatedly over time to collect time courses.

Reporting summary

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

Data availability

All data generated or analysed during this study are included in this published article (and its supplementary information files). The genome sequence for *E. coli* Ochre is available on GenBank (accession number CP173712). Any additional information needed will be available from the corresponding authors upon reasonable request. Source data are provided with this paper.

61. Schwartz, S. A. & Helinski, D. R. Purification and characterization of colicin E1. *J. Biol. Chem.* **246**, 6318–6327 (1971).
62. Nyerges, A. et al. A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. *Proc. Natl Acad. Sci. USA* **113**, 2502–2507 (2016).
63. Wang, H. et al. Improved seamless mutagenesis by recombineering using ccdB for counterselection. *Nucleic Acids Res.* **42**, e37 (2014).
64. Gregg, C. J. et al. Rational optimization of tolC as a powerful dual selectable marker for genome engineering. *Nucleic Acids Res.* **42**, 4779–4790 (2014).
65. Lajoie, M. J., Gregg, C. J., Mosberg, J. A., Washington, G. C. & Church, G. M. Manipulating replisome dynamics to enhance lambda Red-mediated multiplex genome engineering. *Nucleic Acids Res.* **40**, e170 (2012).
66. Zhang, H. et al. Metabolic stress promotes stop-codon readthrough and phenotypic heterogeneity. *Proc. Natl Acad. Sci. USA* **117**, 22167–22172 (2020).
67. Kuznetsov, G. et al. Optimizing complex phenotypes through model-guided multiplex genome engineering. *Genome Biol.* **18**, 100 (2017).

68. Mohler, K., Moen, J. M., Rogulina, S. & Rinehart, J. System-wide optimization of an orthogonal translation system with enhanced biological tolerance. *Mol. Syst. Biol.* **19**, e10591 (2023).
69. Hughes, C. S. et al. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat. Protoc.* **14**, 68–85 (2019).
70. Csibra, E. & Stan, G. B. Absolute protein quantification using fluorescence measurements with FPCountR. *Nat. Commun.* **13**, 6600 (2022).

Author contributions M.W.G. performed experiments on recoding, translation engineering and strain characterization, analysed data and wrote the manuscript. M.T.A.N. contributed to strain characterization and encoding nsAAs into protein. D.W.M. performed experiments on recoding. K.M. contributed to strain and translation engineering. K.G. contributed to strain characterization of translation factor function. S.W. performed strain characterization of phage resistance. C.H. contributed to recoding effort. B.J.S. contributed to recoding effort. Y.C. contributed to strain characterization of growth fitness and protein synthesis. F.R. contributed to strain characterization of encoding nsAAs into protein. M.K. performed proteomics experiments to validate encoding of nsAAs in proteins. J.P. designed oligonucleotides for recoding and screening. M.P. contributed to strain characterization of growth. S.R. contributed to characterization of translation factor function and protein analytics. D.S. contributed to recoding effort. J.T. contributed to recoding effort. J.R. served as co-principal investigator and participated in the design and analysis of experiments and manuscript writing. F.J.I. served as co-principal investigator, conceived the project and participated in the design and analysis of experiments and manuscript writing.

Competing interests Yale University has filed a patent (PCT 63/663,593) related to this study. F.J.I. is a co-founder of, scientific adviser for and has equity in Pearl Bio. J.R. is a scientific advisor for and has an equity in Pearl Bio. J.R. is a co-founder of and has equity in Kapis Biosciences. All other authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-08501-x.
Correspondence and requests for materials should be addressed to Jesse Rinehart or Farren J. Isaacs.
Peer review information *Nature* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.
Reprints and permissions information is available at http://www.nature.com/reprints.
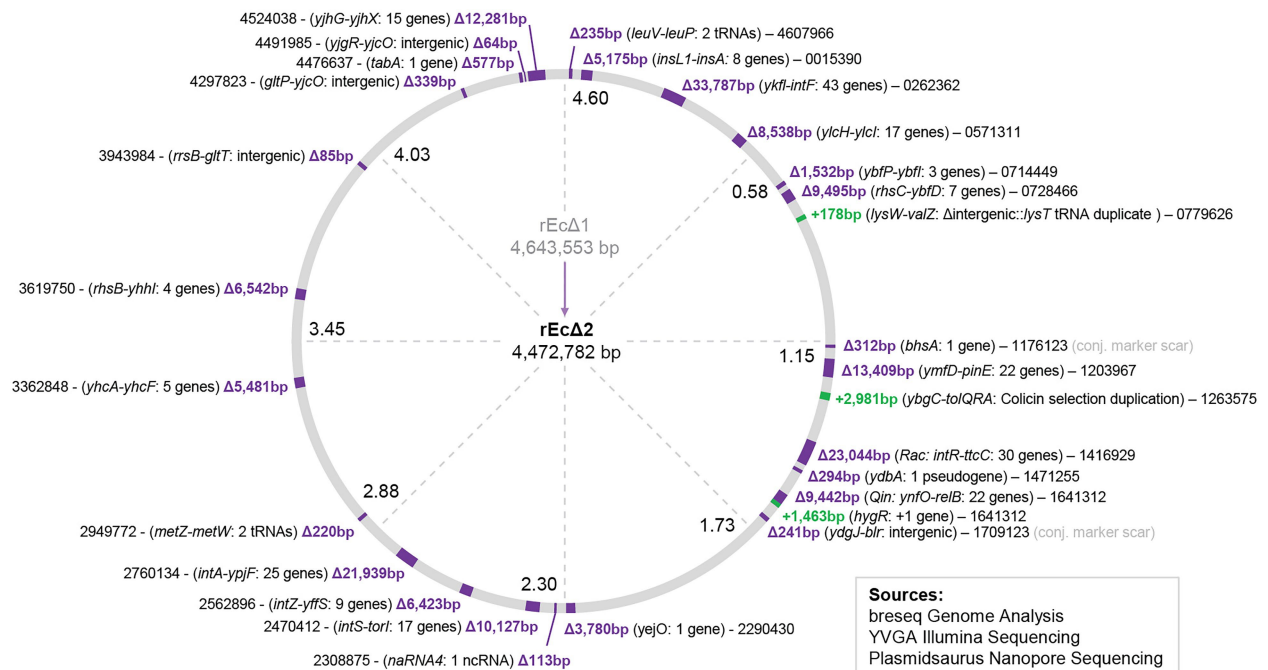
Extended Data Fig. 1 | Major genomic deletions & insertions (>50bp). Figure displays all insertions and deletions greater than 50 bp identified through breseq[32] compared to progenitor strain rEcΔ1.ΔA2 as a reference genome. Details on deliberate mutations can be found in Supplementary Table 7.

	Parents		Offspring Variants						
	D	EF	DEF	1	2	3	4	5	6
LB-OD600	0.74	0.73	0.64	0.62	0.66	0.67	0.69	0.64	0.64
DT (min)	104	85	79	82	84	81	94	82	85
M9-OD	0.30	0.43	0.38	0.36	0.23	0.37	0.28	0.39	0.36
DT (min)	393	306	279	298	305	293	295	272	280

Raw OD600 values not corrected for OD or DT calculations used for relative strain comparisons
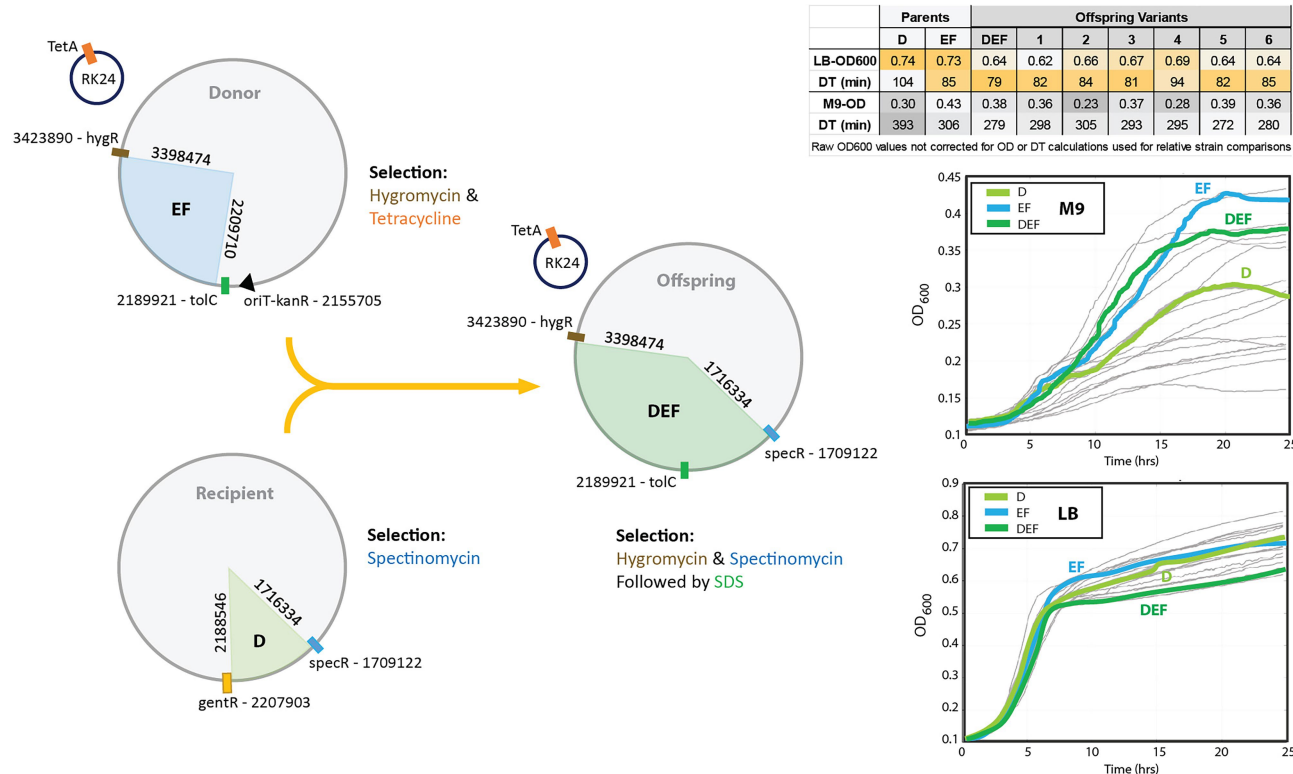
Extended Data Fig. 2 | Example conjugation assembly of recoded regions. Diagrammatic sample of CAGE-mediated assembly of partially recoded ΔTGA regions within rEcΔ2E.ΔA(EF) & (D). Selectable markers are depicted by rectangles (brown = hyg^R; green = $tolC$; blue = $spec^R$; yellow = $gent^R$; orange = $tetA$). RK24 plasmid and an $oriT$-kan^R (Origin of Transfer ($oriT$) adjacent to a kanamycin selectable marker (kan^R) – depicted by black triangle) are present in all donor strains. Point of $oriT$-kan^R triangle conveys counter-clockwise directional orientation. Selectable marker (e.g. $tolC$) is placed 3–5 kb downstream of the $oriT$. Two selecta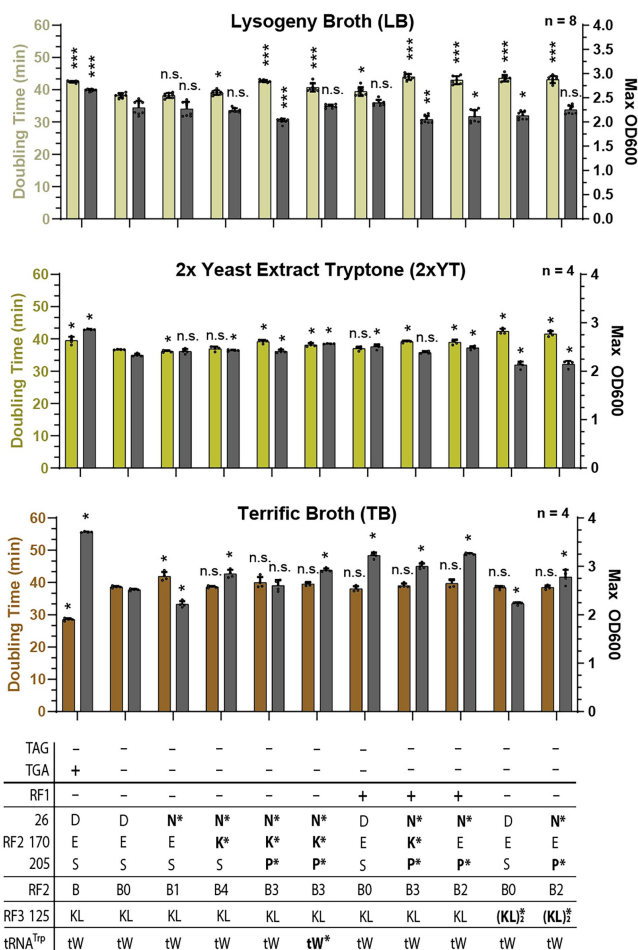ble markers (Supplementary Table 10) are placed flanking recoded regions to select for and preserve these regions within final strains. To select for multiple recoded regions with greater than two antibiotics, subsequent rounds of selection with additional antibiotics are performed. Marker $gent^R$ was overwritten by conjugal transfer and homologous recombination, while kan^R is excluded by its position upstream of $oriT$. Offspring strain is confirmed via MASC-PCR of known recoded sites, followed by whole genome sequencing. Graphs and table to the right depict comparative kinetic growth curve data (OD$_{600}$ over time in hours) in nutrient rich (LB) and minimal media (M9) for donor (cyan), recipient (light green), and offspring (dark green) strains. Grey lines correspond to alternative offspring strains resulting from the same conjugation experiment, revealing a diversity of phenotypic results. Non-selected strains displayed poor growth or undesired genotypes (deleterious background mutations or unrecoded TGA codons). Graph displays raw plate reader OD600 readings and DT calculations (n = 1) per strain: parental strains (D & EF) and offspring (DEF & samples 1–6). Figure data are actual and representative of typical experimental results.
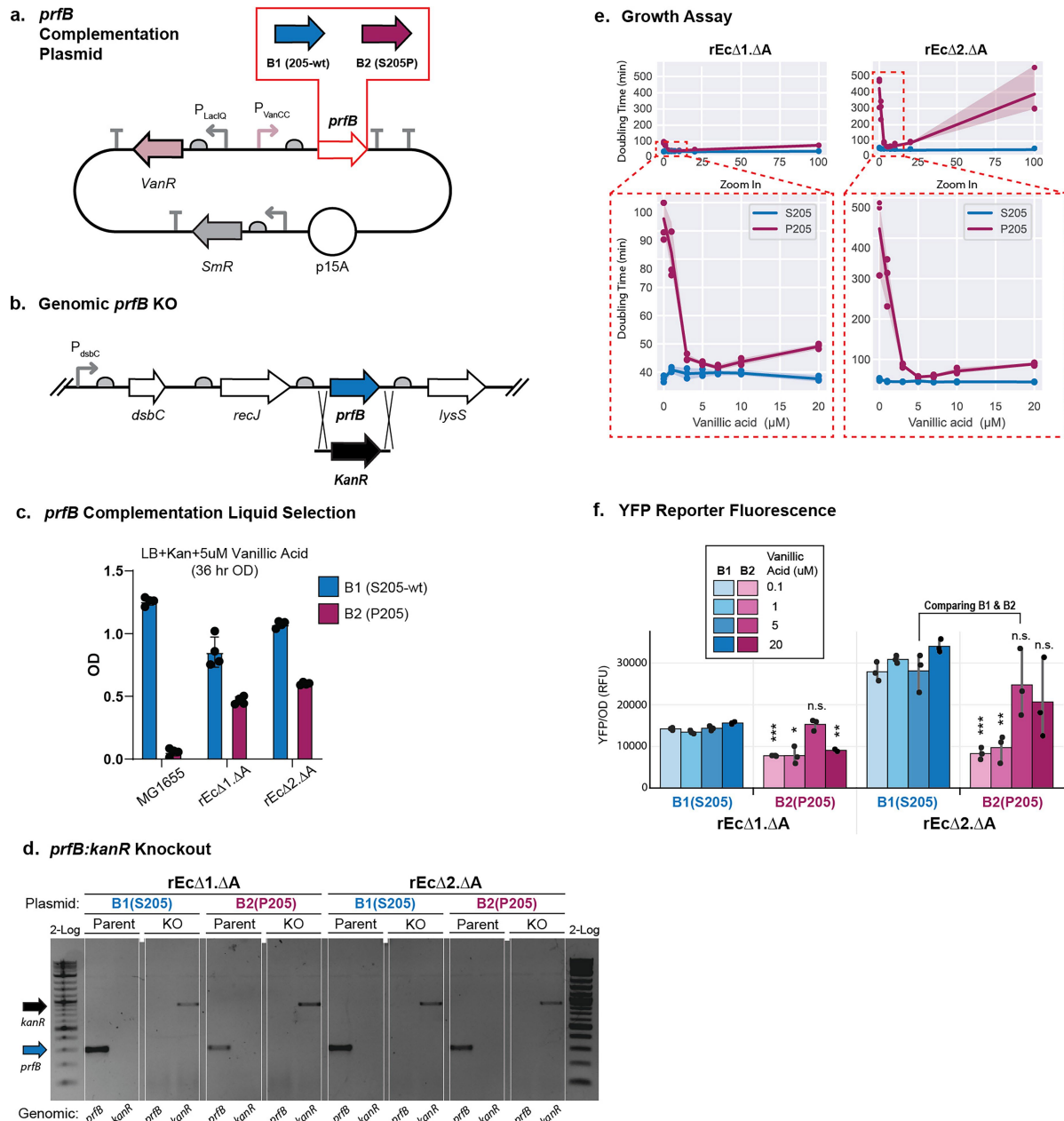
Extended Data Fig. 3 | Release factor (RF) mutagenesis

a.

TAG	–	–	–	–	–	–	–	–	–	–	–
TGA	+	–	–	–	–	–	–	–	–	–	–
RF1	–	–	–	–	–	–	+	+	+	–	–
RF2 26	D	D	N*	N*	N*	N*	D	N*	N*	D	N*
RF2 170	E	E	E	K*	K*	K*	E	K*	E	E	E
205	S	S	S	S	P*	P*	S	P*	P*	S	P*
RF2	B	B0	B1	B4	B3	B3	B0	B3	B2	B0	B2
RF3 125	KL	KL	KL	KL	KL	KL	KL	KL	KL	(KL)‡*	(KL)‡*
tRNA^Trp	tW	tW	tW	tW	tW	tW*	tW	tW	tW	tW	tW

b.

	prfB Mutations					
	E149L	S205P	RR116LL	ED311RR	E170K	
Mutation Freq. / Strain Wells	4.79%	1.06%	6.38%	3.19%	29.79%	# Mutations / Strain
rEcΔ2.ΔA	0	0	0	0	0	0
A05	1	0	0	1	1	3
A09	0	0	0	0	1	1
B11	0	0	1	0	1	2
C01	0	0	1	0	0	1
D09	0	1	0	0	1	2
E07	0	0	0	1	1	2
F05.1	1	0	0	0	0	1
F05.2	1	0	0	1	0	2
H04.1	0	0	1	0	0	1
H04.2	1	0	1	0	0	2
G06	1	0	1	0	1	3

N = 94 Total Strains

Extended Data Fig. 3 | Release factor (RF) mutagenesis. a, RF-variant phenotypic growth data, MaxOD and doubling time (hours), for recoded strains with release factor variants after 36 h incubation at 37 °C in variable media: lysogeny broth (LB), 2x yeast tryptone (2xYT), and terrific broth (TB). Errors bars display confidence interval 0.95 for growth replicates n = 4. Statistical significance is displayed compared to rEcΔ2.ΔA.B0. Error bars display confidence interval 0.95 of n = 4 (2xYT and TB) or n = 8 (LB) biological replicates. Mann-Whitney U tests, * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.001, n.s. = not significant. **b**, Table of RF2 S205P oligo mutagenesis and frequencies. Six MAGE cycles with nine oligos targeting *prfB* within rEcΔ2.ΔA.B0 produced eleven unique genotypes. The final MAGE cycle culture was streaked onto LB agar and grown overnight at 34 °C. Colonies were picked into a 96-well plate and PCR screened via MASC primers. Conversion frequencies are listed. One well yielded the desired S205P mutation alongside E170K. Following the same protocol, multiple attempts to revert E170K, leaving S205P alone, failed. 1 = presence of mutant band in MASC gel; 0 = presence of WT band. Strains from wells F5 and H4 presented both Mut and WT genotypes, but resolved into two distinct genotypes, designated .1 and .2, when streaked onto agar plates and rescreened. Mutations not listed as columns had 0 conversion frequency.
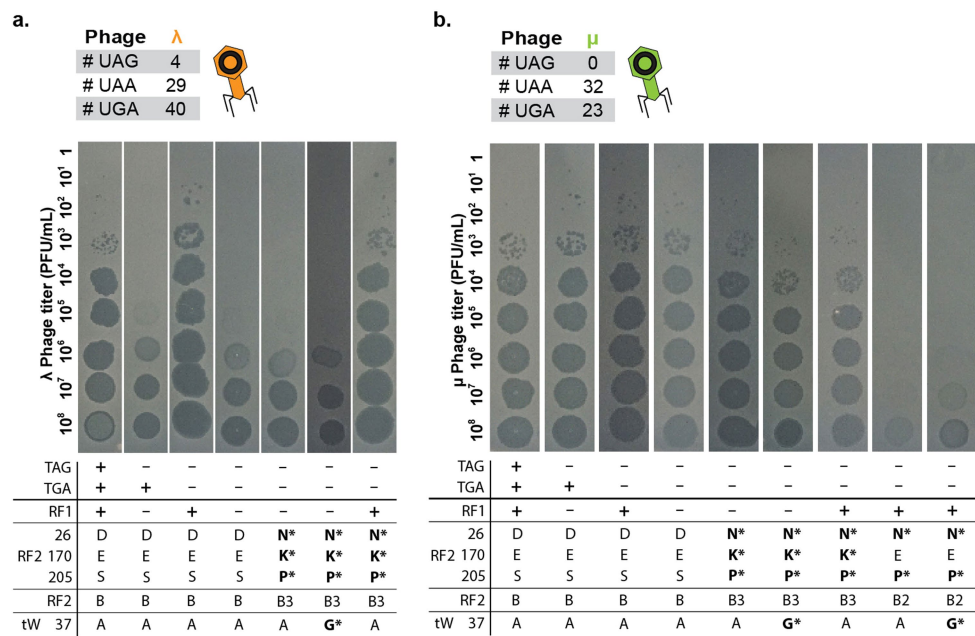
Article

Extended Data Fig. 4 | RF2 S205P complementation. a, Schematic of *prfB* expression plasmid harbouring either a wildtype (Ser) or mutant (Pro) 205 residue (RF2.B1 or B2, respectively) under the control of vanillic acid (VA)-inducible promoter pVanCC. **b**, Schematic of kanamycin resistance (*kan^R*) dsDNA cassette displacing genomic *prfB* while maintaining native overlapping RBS of the downstream gene *lysS*. **c**, Complementation assay: measured OD of variant strains containing *prfB* expression plasmid (RF2.B1 or B2) induced at 5 µM Van recovered in liquid LB after *kan^R*-mediated displacement of genomic *prfB*. **d**, PCR screen of isolated colonies (n = 4) confirming displacement of genomic *prfB* with *kan^R* in the presence of plasmid-expressed RF2.B1 and

B2 variants. For gel source data, see Supplementary Fig. 1. 2-Log ladder used as reference. Positive *prfB*-knockouts were not recovered from MG1655. **e**, Effect of RF2.B1 and RF2.B2 plasmid induction (0.1–20 µM Van) on **e**, doubling time and **f**, fluorescent protein production in absence of genomic *prfB*. Statistical significance compares RF2 B2(mut-P205) and B1(wt-S205)-mediated fluorescent expression within the same strains at the same corresponding Van concentration. Unpaired t-tests (n = 3 biological replicates), * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.001. Error bars display standard errors of the mean.
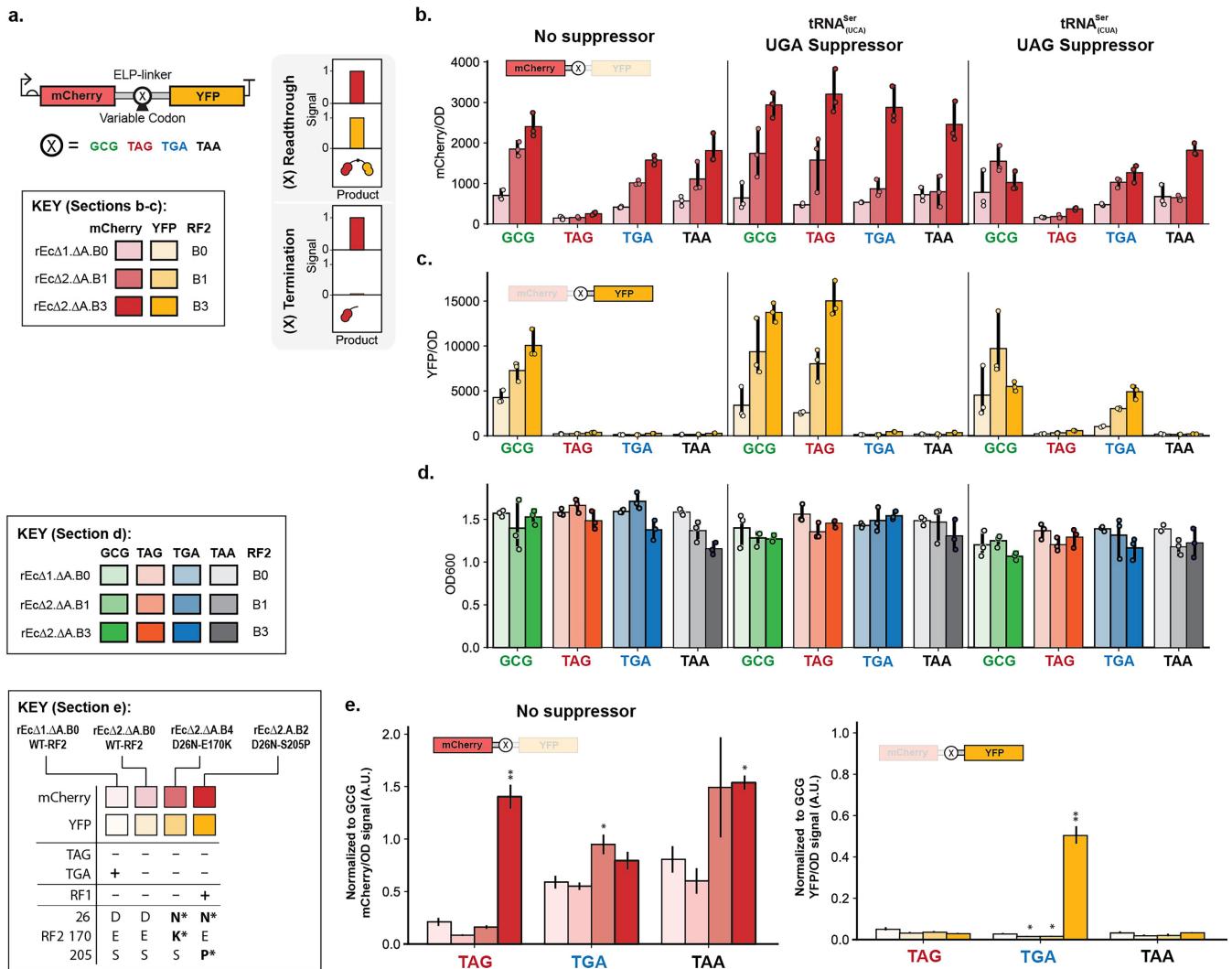
Extended Data Fig. 5 | Growth and fluorescence of strains with readthrough reporter. a, Schematic representation of YFP:mCherry fluorescent reporter used to assess release activity at the three stop codons (X = UAG, UGA, or UAA) relative to sense codon (GCG) within a peptide linker between mCherry and YFP. Release activity at codon X prevents downstream YFP expression and fluorescence without compromising mCherry. Meanwhile, stalling and degradation results in loss of mCherry and YFP fluorescence. High YFP fluorescence suggests readthrough. A colour key for sections b-c is located below. Schematic representation of expected fluorescent outcomes resulting from the mCherry-YFP readthrough assay is to the right. **b-d**, OD-normalized **b**, mCherry (top) and **c**, YFP (middle) fluorescence (corresponds to Fig. 2e,f), along with **d**, strain OD at time of measurement (bottom) resulting from YFP:mCherry fluorescent assay presenting readthrough and Ser suppression

at GCG, UAG, UGA, and UAA in the absence and presence of UAG- or UGA-suppressing SupD tRNAs. Error bars display standard errors of the mean, n = 3 biological replicates. **e**, Comparison of RF variant stop codon readthrough: RF2.B2 (S205P) and RF2.B4 (E170K) with WT-RF2.B0). Normalized mCherry and YFP readthrough fluorescence (same assay as Fig. 2d – "No Suppressor") interrogating the impacts of E170K or S205P mutations in RF2.B3 release activity. mCherry (left) and YFP (right) fluorescence resulting from YFP:mCherry fluorescent assay presenting translation termination or native suppression (readthrough) at GCG, UAG, UGA, and UAA, normalized to GCG (+ control) mCherry or YFP fluorescence, respectively. Statistical significance is displayed compared to rEcΔ1.ΔA. Unpaired t-tests (n = 3), * = p-value < 0.05, ** = p-value < 0.01. Error bars display standard errors of the mean, n = 3 biological replicates.

a.

Phage	λ
# UAG	4
# UAA	29
# UGA	40

b.

Phage	μ
# UAG	0
# UAA	32
# UGA	23

λ Phage titer (PFU/mL): 10^1 10^2 10^3 10^4 10^5 10^6 10^7 10^8

TAG	+	–	–	–	–	–	–
TGA	+	+	–	–	–	–	–
RF1	+	–	+	–	–	–	+
RF2 26	D	D	D	D	**N***	**N***	**N***
RF2 170	E	E	E	E	**K***	**K***	**K***
RF2 205	S	S	S	S	**P***	**P***	**P***
RF2	B	B	B	B	B3	B3	B3
tW 37	A	A	A	A	A	**G***	A

μ Phage titer (PFU/mL): 10^1 10^2 10^3 10^4 10^5 10^6 10^7 10^8

TAG	+	–	–	–	–	–	–	–	–
TGA	+	+	–	–	–	–	–	–	–
RF1	+	–	+	–	–	–	+	+	+
RF2 26	D	D	D	D	**N***	**N***	**N***	**N***	**N***
RF2 170	E	E	E	E	**K***	**K***	**K***	E	E
RF2 205	S	S	S	S	**P***	**P***	**P***	**P***	**P***
RF2	B	B	B	B	B3	B3	B3	B2	B2
tW 37	A	A	A	A	A	**G***	A	A	**G***

Extended Data Fig. 6 | Plaque images from relative titre spotting assays.
a, Close-up images from viral relative titre assays for λ phage. **b**, Close-up images from viral relative titer assays for μ phage. Spots with phage λ or μ at high multiplicity of infection are clear of bacterial growth for all strains.
* = mutant variant. Experiments were conducted with 3-4 biological replicates. For plate source data, see Supplementary Fig. 1.

a.

Nutrient Rich Media (LB)
Kinetic Log

Legend:
- MG1655
- rEcΔ1.ΔA (C321.ΔA)
- rEcΔ2.ΔA.B1
- rEcΔ2.ΔA.B3
- rEcΔ2.ΔA.B3.tW* (Ochre)

n = 11, mean & SD

Minimal Media (M9)
Kinetic Log

Legend:
- MG1655
- rEcΔ1.ΔA (C321.ΔA)
- rEcΔ2.ΔA.B1
- rEcΔ2.ΔA.B3
- rEcΔ2.ΔA.B3.tW* (Ochre)

n = 8, mean & SD

Time (hrs:min:sec)

b.

Nutrient Rich Media (LB)
Kinetic Log

Legend:
- MG1655
- rEcΔ1.ΔA (C321.ΔA)
- rEcΔ2.ΔA.B1
- rEcΔ2.ΔA.B3
- rEcΔ2.ΔA.B3.tW* (Ochre)

n = 11, mean & SD

Minimal Media (M9)
Kinetic Log

Legend:
- MG1655
- rEcΔ1.ΔA (C321.ΔA)
- rEcΔ2.ΔA.B1
- rEcΔ2.ΔA.B3
- rEcΔ2.ΔA.B3.tW* (Ochre)

n = 8, mean & SD

Time (hrs:min:sec)

Extended Data Fig. 7 | Recoded strain kinetic growth curves and phenotypic comparisons with conventional E. coli strains. a-b, Measured OD_{600} kinetic growth of Fig. 5 recoded strains over 36-hours at 37 °C in nutrient rich (LB) and minimal (M9) media, displaying OD_{600} in **a**, log (ln) scale displaying mean and standard deviation, and **b**, linear scale with mean and standard deviation error bars (n = 11 LB, n = 7 M9).
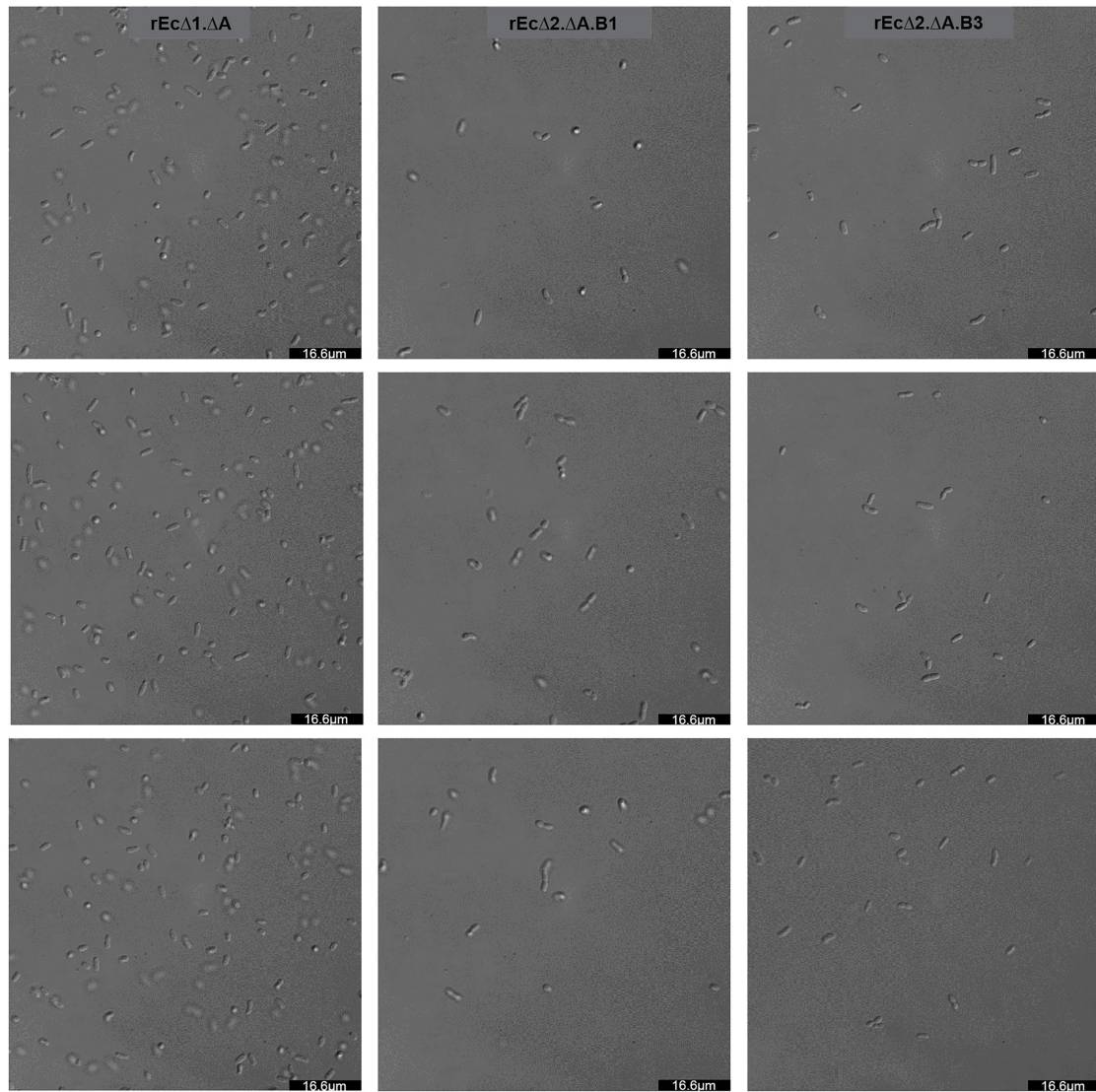
Extended Data Fig. 8 | Strain morphology. Fully recoded strain cell morphology with wildtype and mutant RF2 imaged via differential interference contrast (DIC) brightfield Leica DMi8 S Wide-Field microscopy using a 100x oil immersion objective, 1.6x internal lens. Recoding and RF2 mutations do not cause cell aggregation or a filamentous phenotype, which are indicators of cell stress. Scale bars are 16.6 μm. Three images are presented in columns for each strain.
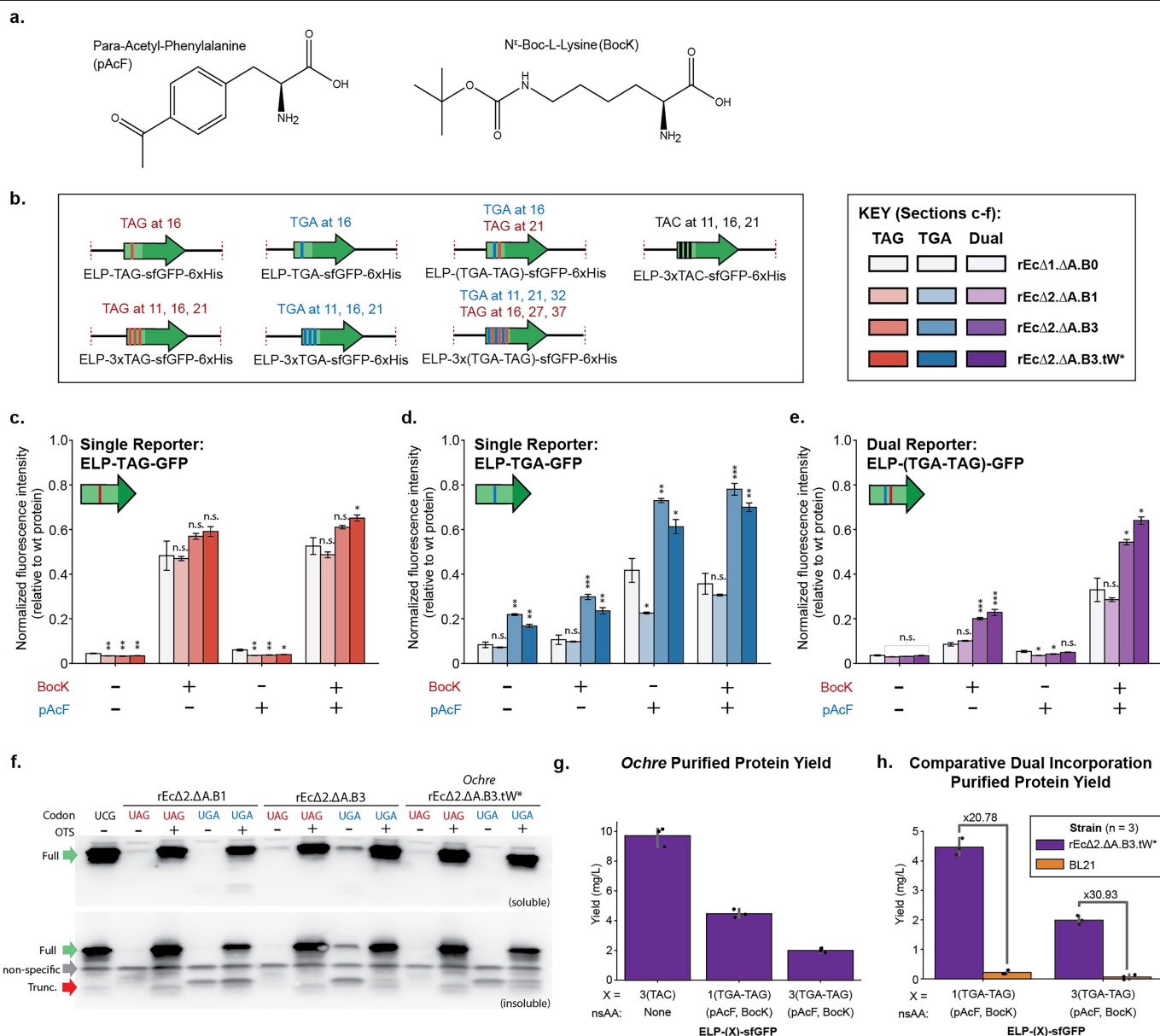
Extended Data Fig. 9 | Further characterization of single and dual-incorporation at UAG and UGA. a, Nonstandard amino acid chemical structures of BocK & pAcF: non-standard amino acids (nsAAs) Nε-Boc-L-lysine (BocK) and para-acetyl-L-phenylalanine (pAcF) incorporated at UAG and UGA, respectively. **b**, Schematic representations of histidine-tagged elastin-like polypeptide super-folder green fluorescent protein (ELP-sfGFP-6xHis) reporters expressed within Fig. 6a expression plasmid to assess nsAA incorporation. To the right is a colour key for sections c-e. **c-e**, Fluorescent expression of ELP-GFP reporter containing: **c**, 1 TAG codon, **d**, 1 TGA codon, or **e**, 1 TAG and 1 TGA codons within variant strains in the absence or presence of one or both: BocK and pAcF. Statistical significance is displayed compared to rEcΔ1.ΔA. Error bars display confidence interval 0.95 of n = 3 biological replicates. Unpaired t-tests (n = 3), * = p-value < 0.05, ** = p-value < 0.01,

*** = p-value < 0.001. **f**, Western blot analysis of GFP expression with and without nsAA incorporation at UAG (BocK) and UGA (pAcF) at position 151, with UCG as a control (UAG/UGA/UCG @151 in GFP). Soluble (top) and insoluble (bottom) fractions display purified full-length GFPs (green arrow), non-specific peptides (grey arrow), and truncated GFPs @151 (red arrow) resulting from RF2 and tRNA^Trp(CCA) mutations within recoded rEcΔ2.ΔA variant strains. For gel source data, see Supplementary Fig. 1. **g-h**, Purified protein yields of ELP-GFP expressed within final recoded strain rEcΔ2.ΔA.B3.tW* resulting from dual incorporation of nsAAs at 1x and 3x UAG (BocK) and UGA (pAcF) codon pairs compared to **g**, 3xTAC ELP-GFP control expression employing native translation machinery, and **h**, dual incorporation expression within standard E. coli protein expression strain BL21. Error bars display confidence interval 0.95 of n = 3 biological replicates.

Article

Extended Data Table 1 | TGA conversion summary

UGA Recoding		Additional Notes	
Total Gene SNPs	**844**	**Total Overlapped ORFs Mutated**	**302**
G→A	389	Genes	294
C→T	455	Noncoding RNAs	8
Total Gene Indels	**239**		
+TAA	117		
+TTA	120		
+TAAT	1		
Δ1 bp	1		
Recoded Pseudogenes	**32**		
G→A	12		
C→T	14		
+TAA	3		
+TTA	3		
Total Recoded TGA ORFs	**1115**		
Gene Knockout (*garK*) ΔG→AT	**1**		
Deleted TGA ORFs	**79**		
Genes	76		
Pseudogenes	3		
Total Gene TGAs Abolished	**1160**		
Total Terminal TGA Genes	1168		
Total Essential TGA Genes	71		
Total TGA Pseudogenes Abolished	**35**		
Total TGA Pseudogenes	45		
Total ORF TGAs Abolished	**1195**		
Total TGA ORFs	1216		
Selenoproteins	**3**		
Unrecoded ins Elements	**8**		
Unrecoded Pseudogenes	**10**		

Genes and pseudogenes containing TGA codons and those deliberately converted to TAA, deleted, or deactivated (i.e. *garK*). Release factor 2 gene *prfB* contains 2 TGAs converted to TAA, counted as one converted gene. Over 300 open reading frames (ORFs) overlapping TGA codons were intentionally mutated as a result of targeted TGA-to-TAA conversions via MAGE. Twenty-one ORFs remain unrecoded, including 3 selenoproteins that contain UGA-specific SECIS elements, 8 transposons, and 10 predicted pseudogenes leftover from extra and unnecessary TGA conversion targeting. See Supplementary Table 1 for full list of TGA conversions. Unrecoded ORFs wither terminal TGAs include: *agaA, agaW, pbl, peaD, yrdE, yeeP, ylbl, yjhZ, yibV, yncI, insQ, insI1(2-3), insE1(1-5)*.

Corresponding author(s): Farren Isaacs, Jesse Rinehart

Last updated by author(s): Nov 10, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	*Provide a description of all commercial, open source and custom code used to collect the data in this study, specifying the version used OR state that no software was used.*
Data analysis	Breseq 0.38.1 (Ubuntu LTS on Windows Subsystem for Linux), Python v3.10.3, GraphPad Prism 10

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data generated or analyzed during this study are included in this published article (and its supplementary information files). Any additional information needed will be available from the corresponding authors upon reasonable request.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Reporting on race, ethnicity, or other socially relevant groupings	*Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status). Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.) Please provide details about how you controlled for confounding variables in your analyses.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were determined by number of colonies or culture wells tested
Data exclusions	No data excluded
Replication	Experimental measures were taken in triplicate
Randomization	Groups were defined by particular features, e.g., each strain containing a different protein-of-interest variant. Strains were compared in isogenic background except when otherwise stated.
Blinding	Blinding was not relevant to this study. Research measured responses to molecular changes and reported quantitative outputs.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒ ☐	Antibodies
☒ ☐	Eukaryotic cell lines
☒ ☐	Palaeontology and archaeology
☒ ☐	Animals and other organisms
☒ ☐	Clinical data
☒ ☐	Dual use research of concern
☒ ☐	Plants

Methods

n/a	Involved in the study
☒ ☐	ChIP-seq
☒ ☐	Flow cytometry
☒ ☐	MRI-based neuroimaging

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*