nature portfolio

https://doi.org/10.1038/s41586-024-08501-x

Supplementary information

Engineering a genomically recoded organism with one stop codon

In the format provided by the authors and unedited

Supplementary Information - Engineering a genomically recoded organism with one stop codon

Michael W. Grome[1,2], Michael T. A. Nguyen[1,2^], Daniel W Moonan[2,3^], Kyle Mohler[2,4^], Kebron Gurara[1,2^], Shenqi Wang[1,2], Colin Hemez[1,2,5], Benjamin 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*]

[1] Department of Molecular, Cellular & Developmental Biology, Yale University, New Haven, CT 06520, USA
[2] Systems Biology Institute, Yale University, West Haven, CT 06516, USA
[3] Department of Molecular Biophysics & Biochemistry, Yale University, New Haven, CT 06520, USA
[4] Department of Cellular & Molecular Physiology, Yale University, New Haven, CT 06520, USA
[5] Department of Biomedical Engineering, Yale University, New Haven, CT 06520, USA

[^] These authors contributed equally to this study
*Correspondence and requests for materials should be addressed to:
Jesse Rinehart (jesse.rinehart@yale.edu) - 203.737.3144;
Farren Isaacs (farren.isaacs@yale.edu) - 203.737.3156

Supplementary Information – Table of Contents

PDF File: Supplementary Information

Excel File: Supplementary Tables

Excel file contains 21 supplementary tables providing information on genomic mutations, DNA oligos, important gene sequences, nomenclature guides, and mass spectrometry results.

Table of Contents: Delineated by tabs

Supplementary Table 1 | TGA Conversions & MAGE Oligos. List of all instances of predicted TGA opal-codon-containing genes, along with pseudogenes with TGAs abolished. Table includes gene names, relative loci within rEcΔ1.ΔA reference genome, overlapping genes, mutations made to convert TGA to TAA, gene orientation, breseq[32] reference database annotations, and MAGE oligos used to introduce codon conversions - 0 signifies absence, 1 signifies presence. '→' = + sense strand, '←' = - sense strand. NOTE: some mismatching gene names are synonyms.

Supplementary Table 2 | Overlapping ORF Mutations Resulting from TGA Conversions. 294 overlapping ORFs were identified via breseq[32] out of 380 expected. 8 were ncRNAs not originally available in annotations. The remaining mutations are likely included within other tables obscured by annotations. '→' = + sense strand, '←' = - sense strand - position relative to rEcΔ1.ΔA reference genome.

Supplementary Table 3 | Native RBS Predictions. Comparative calculated mRNA free energy (ΔG) of wildtype (WT) and modified (mut - TGA to TAA) ribosomal binding sites (RBS) for 249 overlapping genes and predicted translation initiation rates (TIR) to assess impact on downstream gene expression.

Supplementary Table 4 | Essentials Strain (rEcΔ2E.ΔA) TGA Conversions. List of TGA-ending essential genes and additional converted genes within essentials recoded strain, including sense strand, replicore, loci relative to rEcΔ1.ΔA reference genome, overlapping genes, MAGE oligos used for TGA conversions - 0 signifies absence, 1 signifies presence.

Supplementary Table 5 | Strain Nomenclature & Genotypes. Green text refers to WT conditions, while red refers to mutated conditions. Blue text indicates key strains examined in this study.

Supplementary Table 6 | Converted TGA Pseudogene List. Top table contains list of pseudogenes ending in TGA converted to TAA; bottom table contains list of pseudogenes unconverted. '→' = + sense strand, '←' = - sense strand - position relative to rEcΔ1.ΔA reference genome.

Supplementary Table 7 | Major Intentional Deletions. Large genomic deletions intentionally created via tolC-mediated displacement (selection & counterselection). Most deleted regions pertained to cryptic prophages. All sites were listed as unessential and deleted by the Blattner research lab[29]. ORF names listed in blue are TGA-ending genes. ORF names listed in green are TGA-ending pseudogenes. Brackets '[]' indicate partial ORF deletions - position relative to rEcΔ1.ΔA reference genome.

Supplementary Table 8 | Genomic Subdivisions & Conjugative Intermediates. The genome of rEcΔ1.ΔA is subdivided into two subdomains (A' & B'). A' contains 35 essential genes, while B' contains 36 essential genes. Two clones of rEcΔ1.ΔA were categorized as rEcΔ1.ΔA(A') and rEcΔ1.ΔA(B'). Essential TGA-terminating genes within each region were targeted for recoding within each respectively named strain. During MAGE, additional oligos targeting adjacent TGA-terminating genes were included, resulting in 40 conversions within region A' in rEcΔ1.ΔA(A') and 68 conversions in B' within rEcΔ1.ΔA(B'). Resulting partially recoded genomes were combined via CAGE, resulting in rEcΔ2E.ΔA. Eight copies of rEcΔ2E.ΔA were each assigned a particular genomic subdomain (A through H). All TGA-terminating genes within each subdomain was similarly targeted within each respective strain. Once nearly all genes are converted, genomes are hierarchically assembled via CAGE until a fully recoded strain rEcΔ2.ΔA is achieved. Actual numbers of TGA ORFs recoded after conjugations vary as recoded codons are lost via genomic crossovers during conjugations and subsequently recoded again throughout the CAGE assembly process. NOTE: number of ORFs includes pseudogenes converted, but not ins elements or pseudogenes unconverted.

Supplementary Table 9 | *garK* MAGE Oligos. MAGE oligos (v1-v5) used to mutate *garK*, along with mutagenesis details and MASC PCR screening primers to interrogate desired SNPs. '→' = + sense strand, '←' = - sense strand

Supplementary Table 10 | Genomic Selectable Marker Sequences. Sequences include a promoter, open-reading frame, and terminator. Promoter and terminator for kan^R and chloramphenicol acetyltransferase (*CAT*) are the same. MASC PCR primers amplify internally to

each marker and were pooled into single Multiplex PCR reactions to interrogate marker presence and absence per strain. Antibiotic concentrations differ between rEcΔ1 & rEcΔ2.

Supplementary Table 11 | Background Indel List. Supplementary Table 11: Background Indel List. Complete list of background indels categorized by breseq[32] alignment of Ochre whole genome sequencing raw reads with a rEcΔ1.ΔA annotated reference. - '→' = + sense strand, '←' = - sense strand. Tables exclude whole gene deletions, & heavy mutagenic events. List may contain some intentional mutations from overlapping gene conversions.

Supplementary Table 12 | Background SNP List. Complete list of background SNPs categorized by breseq[32] alignment of Ochre whole genome sequencing raw reads with a rEcΔ1.ΔA annotated reference. - '→' = + sense strand, '←' = - sense strand - position relative to rEcΔ1.ΔA reference genome. Tables exclude intentional SNPs and those within proximity of heavy mutagenic events. Nonsynonymous, synonymous, and nonsense categorizations of SNPs were not provided for pseudogenes.

Supplementary Table 13 | Major Background Mutagenic Events. Corresponds to Extended Data Fig 2. Purple regions depict net change in sequence number (indels). Yellow regions depict translation machinery. Mutations resulted from either targeted tolC-mediated deletion, unintended background mutagenesis, or conjugative scarring by marker placement. Sequences are provided for translation machinery alignments - position relative to rEcΔ1.ΔA reference genome.

Supplementary Table 14 | Key Mutations. Three tables present key mutations identified or created within this study, divided into three sub-tables pertaining to (A) mutations involved in auxotrophy/prototrophy, (B) translation machinery mutations, and (C) reverted mutations that impact cellular mechanisms or fitness - MASC PCR primers are included to interrogate desired SNPs. '→' = + sense strand, '←' = - sense strand - position relative to rEcΔ1.ΔA reference genome

Supplementary Table 15 | Growth Data - Reverting Mutations Implicated in Auxotrophy. Growth data (doubling time – DT; maximum cellular density at OD600 – MaxOD) in nutrient rich media (LB) and minimal media (M9) before and after reversion of acquired mutations back to wildtype sequences. Mutation reversions in rEcΔ2E.ΔA(AB) & (C) did not alleviate auxotrophic phenotypes. Identified mutations in *purL*, *cysU*, & *hisI* (see Supplementary Table 14A – Key Mutations) each individually conveyed auxotrophy. Only when all three genes were reverted to wildtype sequences was prototrophy (growth in M9) restored. Reverting the mutation in *argG* alone revealed variable growth effects in a strain dependent manner (data not shown). For strain notation, reference Supplementary Table 8 – Genomic Subdivisions & Conjugative Intermediates). Ancestral rEcΔ1.ΔA strain and optimization mutations serve as a control.

Supplementary Table 16 | RF2 Mutagenesis Oligo Pool. Sites selected were preliminary oligos to sample various residue swaps identified through RF1/RF2 sequence comparisons, E coli RF2 with RF2 sequences from low frequency TGA bacteria (e.g. S. aureus), and eliminating residues identified in literature to confer codon infidelity - MASC PCR primers are included to

interrogate desired SNPs. '→' = + sense strand, '←' = - sense strand - position relative to rEcΔ1.ΔA reference genome. Mutations made via MAGE into rEcΔ2.ΔA.B1

Supplementary Table 17 | RF2 Variant Sequences. Nucleotide (gene: *prfB*) and amino acid (protein: release factor 2) sequences for wildtype B-wt and B3. Discrepancies between sequences are highlighted and in bold. Tripeptide sequences are underlined.

Supplementary Table 18 | Fluorescent Reporter Sequences (ELP-GFP, YFP-mCherry). Sequences include open-reading frame 1, linker sequence, and open-reading frame 2.

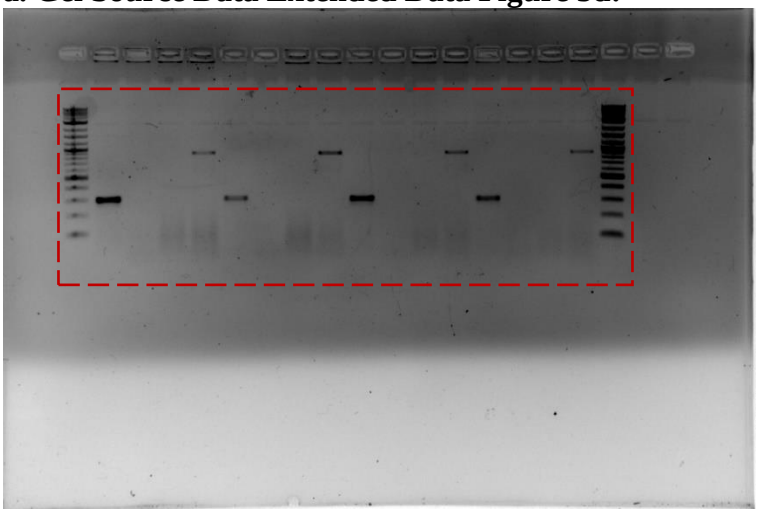
Supplementary Table 19 | Mass Spec Peptide Data for tRNATrp Suppression. Mass spectrometry data corresponding to Figure 4d - displays peptide sequences and data regarding amino acid incorporations.

Supplementary Table 20 | Raw OTS Data - GFP Fluorescence and OD. Raw GFP fluorescence and OD readings from Figure 6 and Extended Data Figure 10 data. L-arabinose inducible aaRSs and accompanying constitutively expressed UAG- and UGA-suppressing tRNAs coordinate charging and incorporation of nsAAs Bock and pAcF at UAG and UGA, respectively, within one of three aTc-inducible fluorescent ELP-GFP reporters. Each reporter contains either 1 or 3 instances of TAG, TGA, or TAG and TGA within a central ELP-GFP linker to study single or dual nsAA encoding expressed within variant strains in the absence or presence of one or both: BocK and pAcF. Samples marked "MD" represent "missing data," or data dropped due to inconsistent bacterial overgrowth.

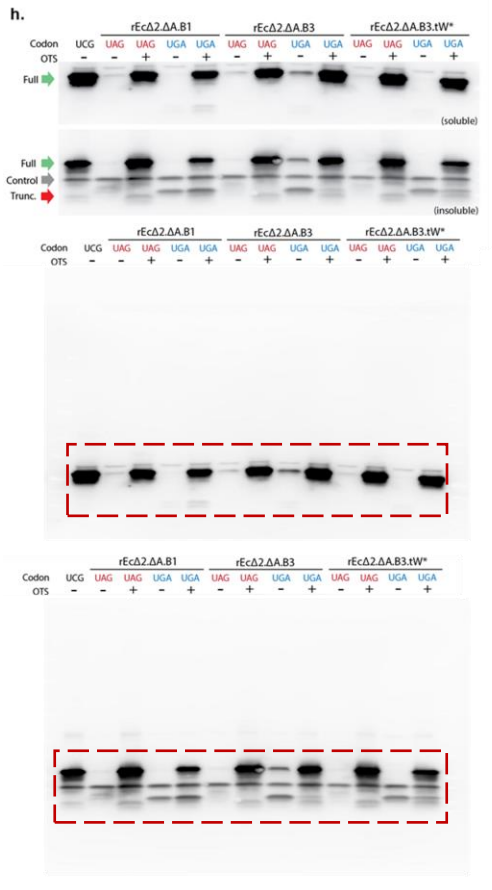
Supplementary Table 21 | Mass Spec Peptide Data for BocK/pAcF OTS. Mass spectrometry data corresponding to Figure 6e-h - displays peptide sequences and data regarding non-standard amino acid and off-target amino acid incorporations.

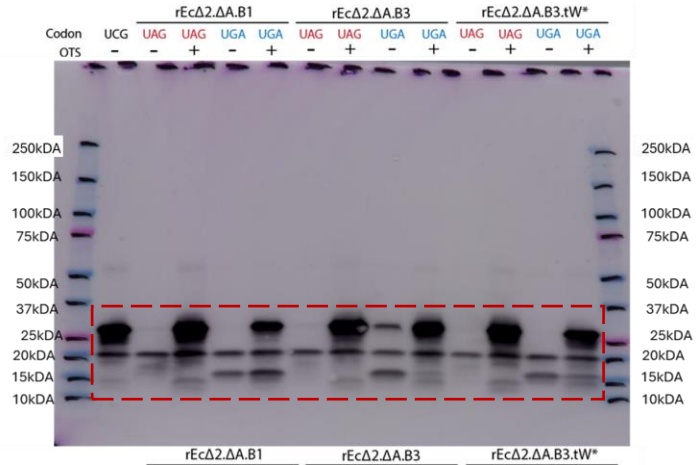
Supplementary Fig. 1 | Gel Source Data. Uncropped images of electrophoresis gels, western blots, and phage spotting agar plates shown in this study.

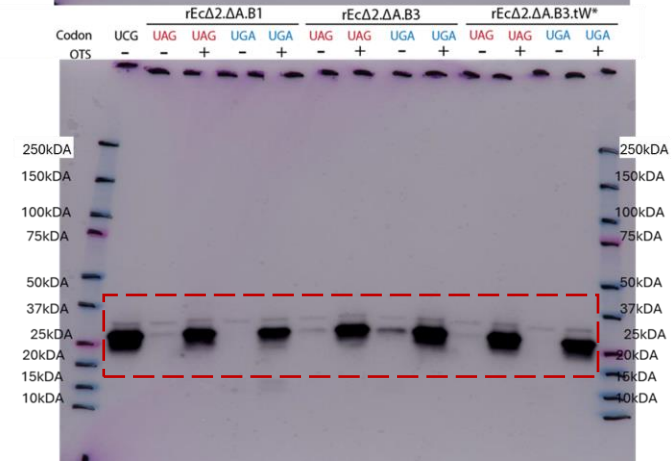
a. Gel Source Data Extended Data Figure 5d:



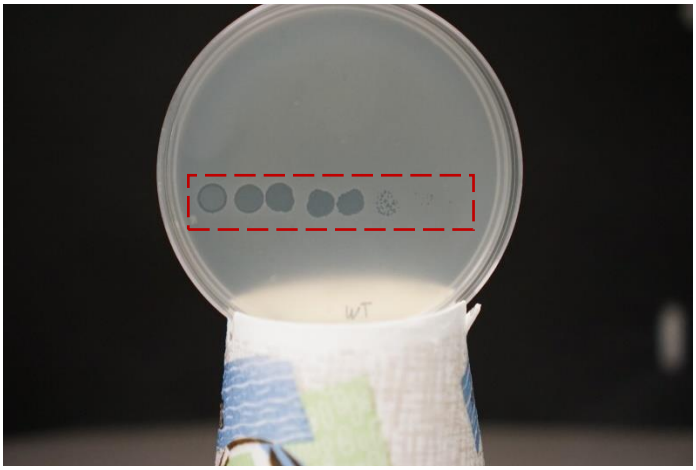
b. Gel Source Data Extended Data Figure 10h:



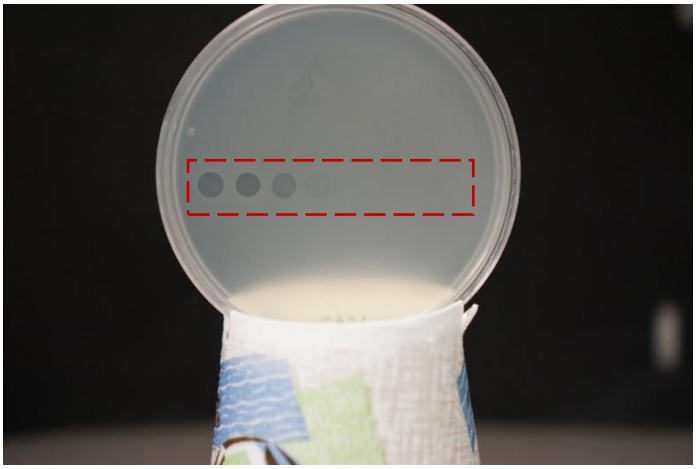




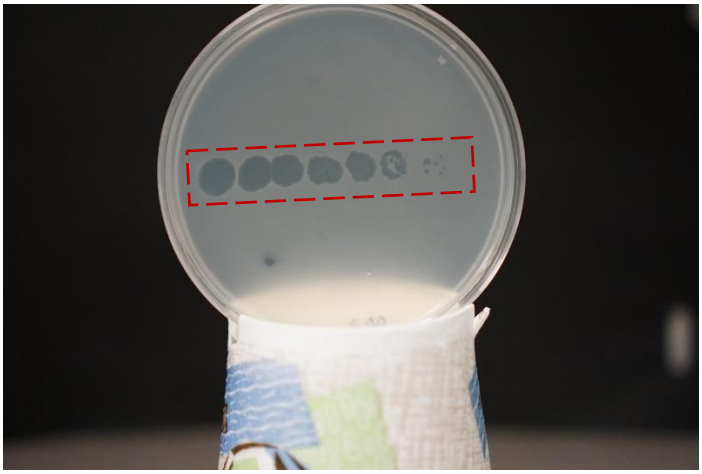
c. Gel Source Data Extended Data Figure 7:
Lambda – MG1655



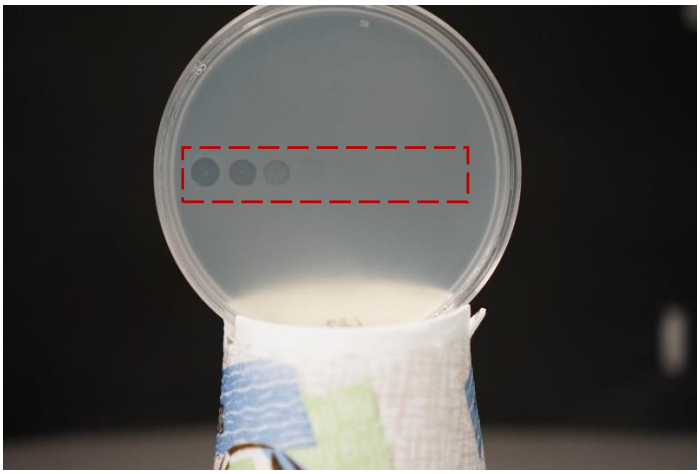
Lambda – C321



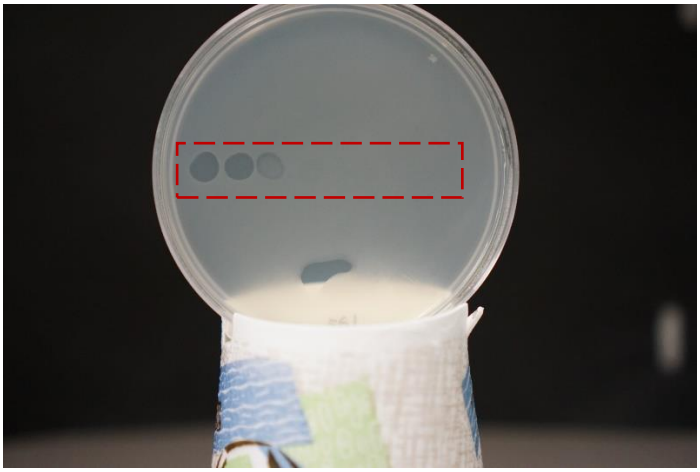
Lambda – rEcΔ2



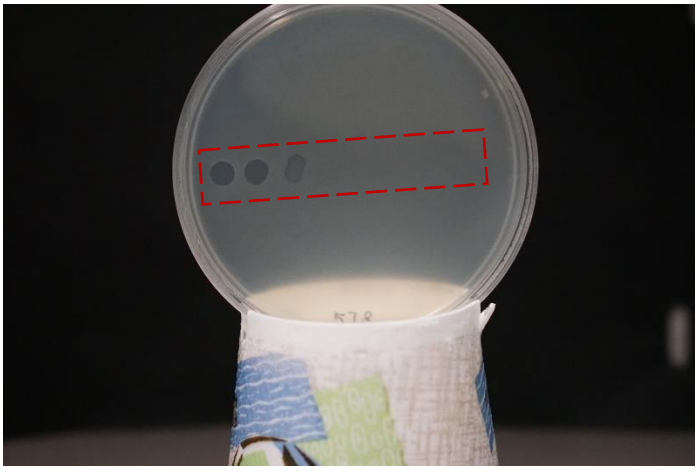
Lambda - rEcΔ2.ΔA



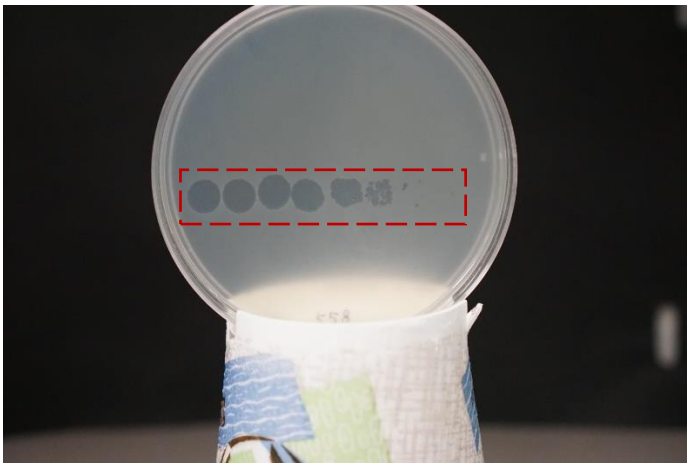
Lambda - rEcΔ2.ΔA.B3



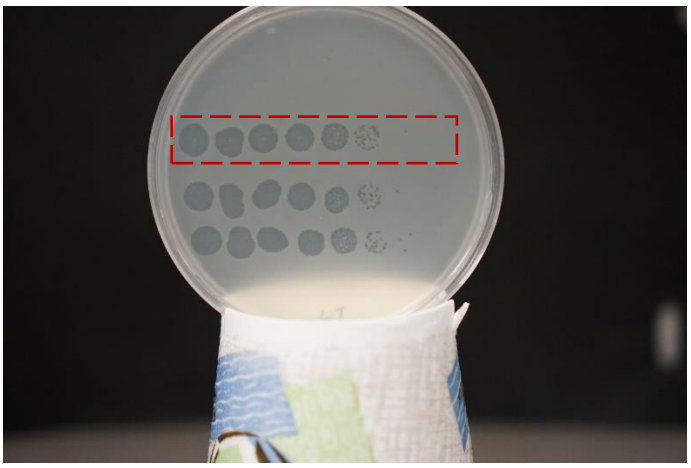
Lambda - rEcΔ2.ΔA.B3.tW*



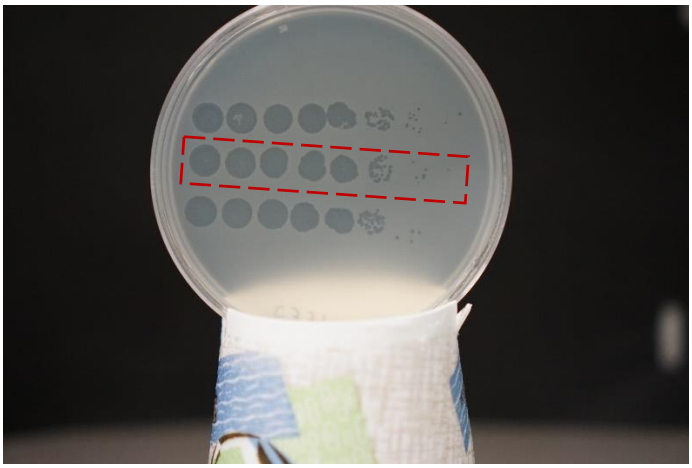
Lambda – rEcΔ2.A.B3



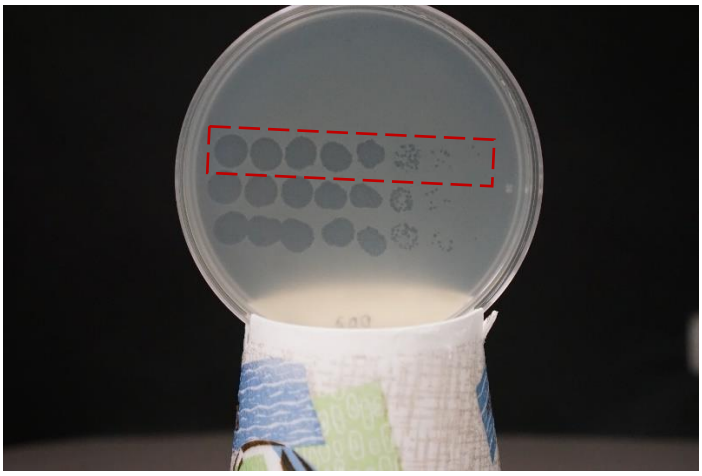
Mu – MG1655



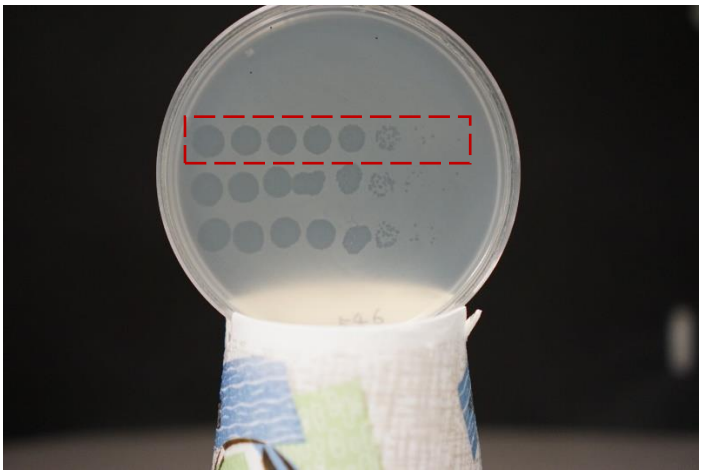
Mu – C321



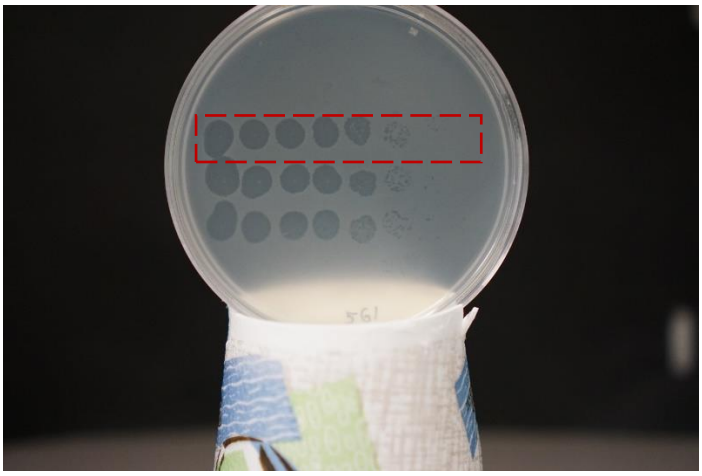
Mu - rEcΔ2.A



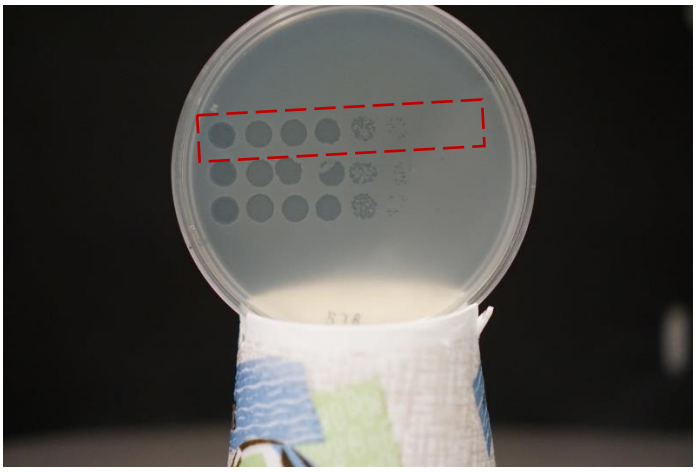
Mu - rEcΔ2.ΔA



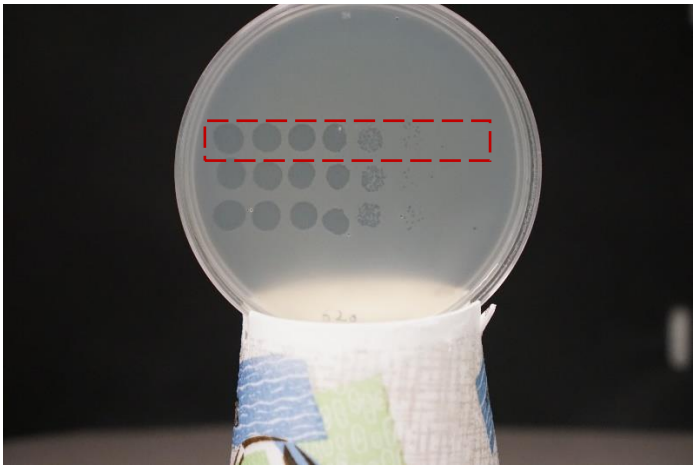
Mu - rEcΔ2.ΔA.B3



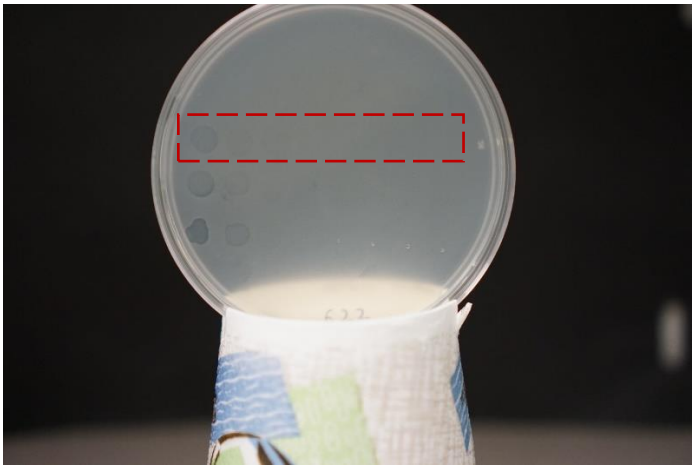
Mu - rEcΔ2.ΔA.B3.tW*



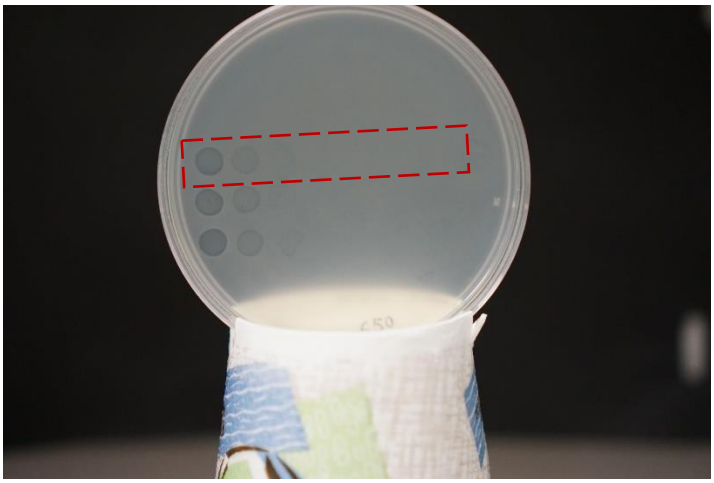
Mu - rEcΔ2.A.B3



Mu - rEcΔ2.A.B2



Mu - rEcΔ2.A.B2.tW*



Supplementary Methods

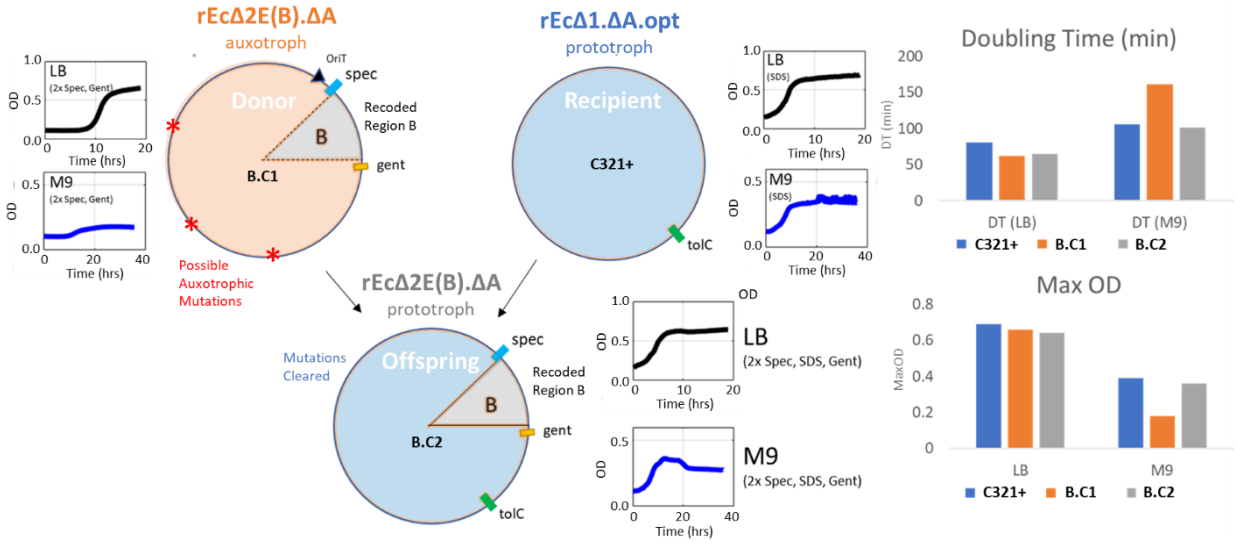
MAGE Oligo Design for Overlapping Genes: Although most oligos were designed to introduce a single-nucleotide substitution to convert TGA to TAA within non-overlapping ORFs, 380 terminal TGA codons overlap a neighboring ORF, and most cannot be converted in this way without potentially impacting expression of neighboring genes. This overlap can fall into one of three categories (**Fig. 1c**). In the first category (26 instances), the A of TGA and ATG is shared, and thus mutating G to A in the TGA codon does not alter the ATG codon. In the second category (257 instances), both the T and G are shared between the ATG and TGA codons. Therefore, recoding TGA to TAA disrupts the start codon, which could preclude expression of the overlapped gene. To avoid this problem, the insertion of three bases – TAA – preserves both a new TAA codon and the ATG start codon, albeit with the ribosome binding site (RBS) shifted by three nucleotides from the start codon. The third category of overlap (97 instances) encompasses all instances where two genes overlap extensively (3+ nucleotides shared), introducing a G to A conversion that generally results in nonsynonymous mutations within overlapping genes (**Supplementary Table 2)**. Since spacing of the RBS from the start codon or RBS-overlapping mutations can impact translation initiation rates[71], we used the RBS calculator[72] to predict translation initiation rates for both the wild type and recoded sequences (**Supplementary Table 3)** and found little predicted deviation in translation initiation upon recoding (**Fig. 1d**).

Background Mutations: Upon completion of the final *Ochre* genome, background mutations were acquired across strain lineages, comprising 193 indels (including 129 frameshifts; **Supplementary Table 11**) and 591 single nucleotide polymorphisms (SNPs), which include 170 synonymous SNPs and 302 nonsynonymous SNPs resulting in 13 nonsense mutations (**Supplementary Table 12**). In addition, longread sequencing revealed various complex mutations, including the loss of 4 tRNAs (LeuP, LeuQ, MetZ, MetW), duplication of a mutant LysT, and instances of rRNA mutagenesis (**Supplementary Table 13**). It is likely that total mutations acquired are overrepresented by breseq[32] outputs presented in supplementary tables. For example, breseq frequently separates single mutation events (e.g., +TAA insertion) as multiple mutation events (e.g., indels and SNPs).

Reverting Background Mutations & Auxotrophic Phenotypes:
When MAGE mutagenesis is performed within multiple strains to be conjugated together, off-target genomic mutations can arise that offer little phenotypic impairment alone, but when combined with other mutations through conjugation can result in synthetic deleterious mutations with phenotypic impairments. When auxotrophies arose, particularly after conjugations where a negative phenotype affects all resultant offspring, we employed two methods of prototrophy restoration: (1) conjugative backcrossing with ancestral rEcΔ1.ΔA (C321), or (2) targeted mutation reversions via MAGE.

(1) Conjugative Backcrossing with rEcΔ2.ΔA: Early in this research, when only an eighth or quarter of a genome is being recoded, there was high probability of mutations arising within the unrecoded background, representing a majority of the genomic regions. To address these, strains can be conjugatively backcrossed with ancestral rEcΔ2.ΔA via CAGE (**Supplementary Fig. 2**), in accordance with previous protocols[15,73]. This method was used routinely to "clean up" strains prior to major conjugations and ensured genomic background outside of recoded regions were

genotypically parental and unmutated, thereby reducing probability of synthetic lethality leading to conjugation failure.



Supplementary Fig. 2 | Conjugative backcross to restore prototrophy. CAGE was employed to revert unknown auxotrophy-inducing mutations within the background of partially recoded strains. Recoded regions (e.g. region B) were conjugatively transferred into a rEcΔ1.ΔA optimized parental strain to clear unknown auxotrophy mutations. Kinetic growth curves and growth data (Max OD & Doubling Times) are displayed for each parental strain and the resulting conjugative offspring, revealing an improvement in M9 growth after backcrossing (Note: OD readings are raw, untransformed plate reader values). Selectable markers are depicted by rectangles (blue = *specR*; yellow = *gentR*; green = *tolC*). RK24 plasmid and an *OriT-kanR* (Origin of Transfer (*OriT*) adjacent to a kanamycin selectable marker (*kanR*) – depicted by black triangle) are present in all donor strains. Point of *OriT-kanR* triangle conveys clockwise directional orientation. Selectable marker (e.g. *specR*) is placed 3-5kb downstream of the *OriT*. Two selectable markers (Supplementary Table 9) are placed flanking recoded regions to select for and preserve these regions within final strains. Excess donor genomic material (outside of recoded regions) may contain deleterious mutations. A third marker is placed in the recipient strain and selected via antibiotic to limit transfer to mostly the desired recoded region. Antibiotic selection and PCR amplification of marker sequences are used to confirm transfer of genomic regions. MASC PCR of select recoded sites within the recoded region are used to confirm preservation of recoded genes. The resulting strain is moved forward for additional recoding or conjugation with other partially recoded strains.

(2) MAGE-Targeted Reversion of Off-Target Mutations: Early in this research, growth impairment mutations would arise within recoded regions, resulting in poor conjugated offspring phenotypes that could not be addressed with backcrossing. When multiple recoded regions were assembled, an accumulation of background mutations within recoded regions led to deleterious phenotypes in growth and antibiotic resistances, as well as synthetic auxotrophies (i.e. an inability to grow in M9 minimal media supplemented with glucose). In one instance, we were unable to conjugate together partially recoded strains. Literature searches suggested that knockout mutations within genes *fliK* and *kefB*, responsible for flagellar assembly and potassium transporter, respectively, were implicated in disrupting conjugal ability[74] (**Supplementary Table**

13C). MAGE-mediated reversions of these mutations restored strain conjugative capacities. Employing breseq[32] WGS analysis of recoded strains along their lineages, we identified various non-synonymous mutations within genomes that correlated with undesirable phenotypes (**Supplementary Table 13A**). The resulting list of mutations was examined through ecocyc.org database, uniprot.org mutation lists, and relevant literature to identify potentially phenotypically relevant mutations. Once a narrowed list of probable causative mutations was achieved, MAGE oligos were used to revert targeted mutations. Growth curves or fluorescent assays were then used to assess changes in phenotypes resulting from mutation reversions. One demonstrative example of this method was the presence of an auxotrophy that arose post-conjugation between two partially recoded strains. Each strain was backcrossed with rEcΔ2.ΔA prior to conjugation, yet auxotrophy was not resolved. Via breseq analysis and ecocyc.org growth evidence, we identified a mutation in *purL* suspected to lead to auxotrophy. The gene *purL* is involved in purine biosynthesis, located within a partially TGA-recoded genomic region (**Supplementary Table 1**). Once the *purL* mutation was reverted to WT via MAGE, the prototrophic phenotype was restored (**Supplementary Table 14**).

Supplementary Discussion 1 | *Mutant Release Factor 2 Complementation Assays*

RF2 S205P Complementation rEcΔ2.ΔA RF2-S205P (RF2.B2) is Inviable with Genomic Expression: Release factor 2 (RF2), expressed from gene *prfB*, recognizes and terminates translation at UGA and UAA codons. RF2 termination ability is dependent on multiple factors that regulate RF2 expression and function, including the structure of its codon recognition loop[46], methylation of its peptidyl transfer arm by methyltransferase PrmC[36], internal autoregulatory sequences[39], and additional translation factors like release factor 3 (RF3)[34]. Our work seeks to generate an RF2 variant with codon recognition limited to the UAA codon to open UGA for reassignment.

We first modified the *prfB* internal autoregulatory stop codon from UGA to UAA to make RF2 expression compatible with attenuated UGA recognition and termination (**Fig. 2a**). This introduced a D26N mutation in RF2, generating a new variant (RF2.B1). We then identified a mutation in literature, S205P, that we hypothesized would abolish UGA recognition while preserving that of UAA[46]. Within our rEcΔ2.ΔA.B1 strain, we attempted to introduce the S205P mutation into genomic RF2 B1(wt-S205) to generate B2(mut-P205) yet failed to recover viable clones. Reintroduction of *prfA* (release factor 1: RF1) into its native genomic locus enabled viable production of the S205P mutation (**Supplementary Table 5**; **Extended Data Fig. 4**), and in combination abolished UGA termination (**Extended Data Fig. 6e & 7b**), revealing successful elimination of UGA-codon dependence within our strains. However, once B2(mut-P205) was established, subsequent *prfA* deletion was only achieved when accompanied by a spontaneous P205S reversion, or via acquisition of a spontaneous compensatory mutation in *prfC* (RF3) **(Supplementary Table 14b**): a protein that monitors peptide quality control in translation alongside accelerating RF1/RF2-mediated termination and release from the ribosome[34,75]. This RF3 mutation was only revertible when mutant P205 was again reverted to wildtype Serine, or when *prfA* was reintroduced, reinforcing the lethality of S205P. This supplementary discussion assesses three questions: (1) can B2(mut-P205) complement knockouts of genomic *prfB* in rEcΔ2.ΔA and ancestral strains? (2) Can rEcΔ2.ΔA.B2 viability be rescued with elevated B2(mut-P205) expression? (3) What impact could S205P have on RF2 function?

rEcΔ2.ΔA RF2-S205P (RF2.B2) is Viable with Plasmid Overexpression*:* Given the conversion of *prfB* internal TGA to TAA, RF2 expression may be regulated in part by RF1 termination, reducing RF2.B2 concentrations and contributing to inviability. To assess the viability of S205P (RF2.B2) in absence of RF1, we conducted a complementation assay by transforming rEcΔ2.ΔA.B1 and ancestral strains (rEcΔ1.ΔA and MG1655) with a vanillic acid-inducible expression plasmid containing one of two *prfB* variants: B1(wt-S205) and B2(mut-P205) (**Extended Data Fig. 5a).** Genomic *prfB* was subsequently deleted by displacement with a kanamycin resistance (*kanR*) antibiotic selectable marker at varying concentrations of vanillic acid (Van), revealing *prfB* deletion was strain dependent (**Extended Data Fig. 5b-d).** Overexpression of B2(mut-P205) was able to complement Δ*prfB* in both rEcΔ1.ΔA and rEcΔ2.ΔA but not MG1655 (**Extended Data Fig. 5b**). We found that a 50x higher concentration of inducer (5 µM vs 0.1 µM Van) was required for sufficient expression of B2(mut-P205) relative to B1(wt-S205) to maintain cell viability and restore fitness (e.g. doubling time and fluorescent protein production) in rEcΔ1.ΔA and rEcΔ2.ΔA (**Extended Data Fig. 5e-f)**. These data suggest that S205P compromises essential RF2 function that can be compensated by elevated expression. RF2 has two cited essential functions: translation termination and ribosomal rescue[34]. Given the viability of genomic B2(mut-P205) in the presence of RF1, which does not participate in ribosomal rescue[34], the compromised essential RF2 function is likely translation termination. As stated in the main text, AlphaFold[41] structural prediction (**Fig. 2b**) suggests wild-type S205 presents up to 4 hydrogen bonds within the codon recognition loop of RF2, while mutant P205 presents one, likely altering codon recognition loop stability. The fact that genomic B2(mut-P205) complemented by RF1 abolished UGA recognition (**Extended Data Fig. 6e & 7b**) in rEcΔ2.A.B2 while enabling viability suggests that a failure in UAA termination caused by S205P was responsible for inviability. Meanwhile, complementation data suggests B2(mut-P205) retains some degree of UAA and UGA termination function sufficient to complement *prfB* knockouts in both ΔUGA rEcΔ2.ΔA and UGA+ rEcΔ1.ΔA with elevated plasmid expression (**Extended Data Fig. 5c**).

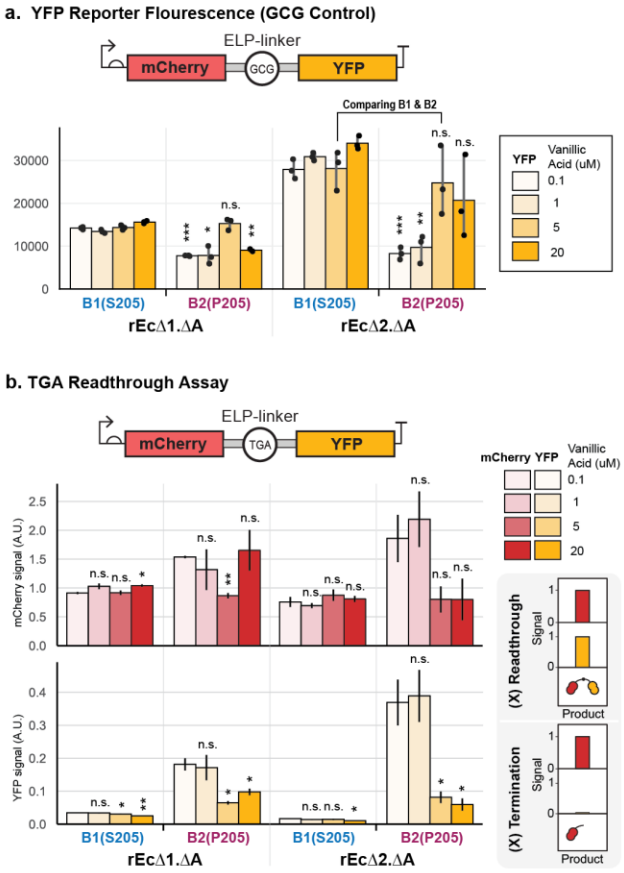
Attenuated UGA Termination by RF2-S205P (B2) is Compensated by Elevated Expression: To assess the degree to which B2(mut-P205) expression retains UAA and UGA recognition at different levels of induction, we conducted an *in vivo* codon readthrough assay that reports on release activity of B1(wt-S205) and B2(mut-P205) at stop codon TGA and a sense codon control (GCG). Derived from established methods[52], this assay determines an RF's ability to terminate translation of an mCherry gene ending in a target codon upstream of a YFP gene terminating in TAA (**Supplementary Fig. 3**). If no termination occurs at the target codon, translation reads through mCherry to YFP, expressing both proteins. However, if successful termination occurs, only mCherry is expressed, yielding low YFP signal.

While titrating B2(mut-P205) expression shows a clear minimum for doubling time at 5 µM Van induction for both rEcΔ1.ΔA(Δ*prfB)* and rEcΔ2.ΔA*(*Δ*prfB)* (**Extended Data Fig. 5d**), fluorescent protein expression (GCG codon – no termination) also presents a maximum fluoresence at 5 µM Van for both strains (**Extended Data Fig. 5e, Supplementary Fig. 3a**). Meanwhile, UGA readthrough, measured as YFP signals from mCherry-[UGA]-YFP normalized to mCherry-[GCG]-YFP at similar culture conditions (**Supplementary Fig. 3b**), significantly decreases going from 1 µM to 5 µM Van. These data suggest low B2(mut-P205) UGA termination function at < 5 µM Van induction levels, and a restoration in UGA termination at

higher expression levels of induction, demonstrating a tradeoff between cellular fitness (optimal at 5 μM Van) and UGA readthrough (optimal < 5 μM Van).

UAA Termination by RF2-S205P (B2) is Tightly Regulated: The ability of RF1 to compensate B2(mut-P205) for rEcΔ2 viability suggests B2(mut-P205) is compromised in UAA termination. Unsurprisingly, both low induction (<5 μM Van) and high induction (>5 μM Van) of B2(mut-P205) expression is detrimental to cell fitness (**Extended Data Fig. 5e-f**). Low induction offers the cell insufficient B2(mut-P205) concentrations to permit efficient termination. Meanwhile, high induction disrupts RF2-mediated autoregulation of expression to also hinder efficient termination. RF2 expression is regulated via RF2-mediated recognition and termination at a stop codon internal to *prfB*, prematurely cleaving RF2 peptides when concentrations are high[39]. If RF2 is compromised in its ability to terminate at this stop codon, expression will rise, increasing RF2 concentrations[36]. Methylation of RF2 by methyltransferace PrmC enables efficient termination. If RF2 levels outpace PrmC methylation kinetics, RF2s become hypomethylated, further lowering termination efficiencies and limiting cell growth[36]. The sensitivity in growth and protein production to varying expression of B2(mut-P205) but not B1(wt-S205) (**Extended Data Fig. 5e-f**) suggests the negative feedback mechanism of B2(mut-P205), which relies on internal UAA termination (**Fig. 2a**), is compromised by inefficient UAA termination resulting from S205P compounded by hypomethylation. This is supported by high B2(mut-P205) expression resulting in more pronounced changes in doubling times for rEcΔ2.ΔA than rEcΔ1.ΔA (**Extended Data Fig. 5e**), as the *prmC* gene in rEcΔ1.ΔA contains a ribosomal binding site (RBS) predicted to have an 11-fold increased transcription rate than that found in rEcΔ2.ΔA (DeNovoDNA.com).

Summary: Our findings reveal that mutating or replacing wt-RF2 is nuanced. RF2 function is tightly controlled under multiple mechanisms of regulation, including RF2 codon loop stability, methylation, premature termination, and RF3. Genomic recoding, through the eliminaiton of native TGA codons, has allowed for complementation of *prfB* knockouts with a mutation that is not tolerated in MG1655 (i.e. S205P). However, both rEcΔ1.ΔA and rEcΔ2.ΔA only permit *prfB* knockout and maintain cellular fitness (doubling times and fluorescent protein production) when the expression level of B2(mut-P205) is narrowly constrained. These results confirm that recoding permits RF2 mutations with expanded functionality, but also highlights that RF2 function needs to be very specific and tightly regulated to maintain cellular fitness. Moreover, complementation with B2(mut-P205) also supports that more complex mutations within RF2 (i.e. S205P & E170K) may be required to achieve a rEcΔ2.ΔA strain with the most desirable traits without compensatory mutations in other genes (e.g. RF3).



Supplementary Fig. 3 | S205P Complementation Readthrough Assay. a, Recolored representation of Extended Data Fig. 5f, including a schematic representation of control YFP-[GCG]-mCherry fluorescent reporter used to assess fluorescent protein production from rEcΔ1.ΔA and rEcΔ2.ΔA episomally expressing B1(wt-S205) or B2(mut-P205) at varying levels of Van induction in absence of genomic prfB. Statistical significance compares B2(mut-P205) and B1(wt-S205)- mediated fluorescent expression within the same strains at the same corresponding Van concentration. Welch's unpaired t-tests (n = 3), * = p-value < 0.05, ** = p-value < 0.01. Error bars display standard errors of the mean. **b,** Schematic representation of YFP-[TGA]-mCherry fluorescent reporter used to assess termination at UGA. High YFP fluorescence suggests readthrough. Release activity at UGA prevents downstream YFP expression and fluorescence without compromising mCherry, while stalling and degradation leads to loss of mCherry and YFP signal. A color key and schematic representation of expected fluorescent outcomes resulting from the mCherry-YFP readthrough assay are displayed to the right. Fluorescence in (b) is normalized to GCG codon construct fluorescence from (a). Statistical significance is displayed compared to 0.1 µM Van within each strain RF variant strain. Welch's unpaired t-tests (n = 3), * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.001. Error bars display standard errors of the mean.

Supplementary Discussion 2 | *Characterizing mutant RF2 with phage*

To further examine impacts of RF2 engineering on protein translation, we conducted phage infection assays that have been used to assess the impact of genomic recoding and translation factor deletions in establishing genetic code orthogonality[8-10]. Hosts that lacked machinery for decoding essential phage codons mistranslated viral proteins, halting propagation and conferring resistance to phage. Our prior work revealed Lambda (λ) phage genes containing TAG codons are targeted for degradation via tmRNA tagging when expressed in a strain lacking RF1 recognition of the UAG codon, preventing phage infection[10,51]. We hypothesize that abolition of UAG and UGA termination will confer strains with resistance to phage that employ essential genes ending in TAG or TGA. RF-variant strains and predecessors were challenged with two genotypically distinct bacteriophages: λ (containing TAG, TGA, and TAA) and Mu (μ) (containing TGA and TAA only). Strains incubated with serially diluted phage were used to quantify plaque-forming units per milliliter (PFU/ml), presenting no visible plaques in resistant strains and reduced plaque sizes in partially resistant strains (**Extended Data Fig. 6)**. Results revealed that all strain variants lacking RF1 were resistant to λ (**Fig. 3a**), retaining the genetic orthogonality of rEcΔ1.ΔA conferred by deletion of RF1. To specifically examine the impact of UGA decoding alone, we challenged strains with μ (lacking TAG). If RF2 decoding of UGA is abolished, we hypothesize a strain would be resistant to μ infection. Indeed, rEcΔ2.A.B2, expressing RF2.B2 with RF1, demonstrates robust resistance to μ (**Fig. 3b**), suggesting abolition of UGA termination. However, all other RF2 variants, with or without RF1, were susceptible to μ infection, suggesting sufficient UGA decoding to enable phage translation, consistent with readthrough data (**Fig. 2e**). Taken together, these phage assays suggest (i) robust RF2 decoding of UGA is not necessary for rEcΔ2 viability, as demonstrated by rEcΔ2.A.B2 resistance to μ, and (ii) S205P functionally abolishes UGA termination, which is partially restored in the presence of E170K, consistent with readthrough results.

Supplementary Discussion 3 | *Assessing Individual Impacts on Growth*

To decouple the fitness impacts of genome-scale TGA replacements from that of translation engineering, we compared growth data from ΔTGA-recoded rEcΔ2.ΔA.B1 to ancestral rEcΔ1.ΔA, then to RF-engineered rEcΔ2.ΔA.B3 and $tRNA^{Trp}$-engineered rEcΔ2.ΔA.B3.tW*. Comparing rEcΔ2.ΔA.B1 with ancestral rEcΔ1.ΔA revealed that genome-scale TGA replacements resulted in media-specific divergences in fitness. In M9, MaxOD increased by ~17% while DT slowed by ~23%. In LB, MaxOD decreased by ~4% while growth rates improved, decreasing DT by ~5%. These results demonstrate both growth impairments and improvements after ΔTGA recoding, depending on which metric (MaxOD or DT) or media (LB or M9) are considered.

To assess the impacts of cumulative RF2 engineering, we compared rEcΔ2.ΔA.B3 to ancestral rEcΔ2.ΔA.B1. DTs slow in M9 by ~8% and in LB by ~9%, while MaxOD decreased in M9 by ~21% and remained stable in LB. Further analysis dissecting the impacts of individual RF2.B3 mutations on growth revealed S205P to be the primary RF2 mutation contributing to fitness impairments (**Extended Data Fig. 3a**). Combining RF2.B3 with engineering of $tRNA^{Trp}$ (rEcΔ2.ΔA.B3.tW*) resulted in slower DTs in M9 (~15%) and LB (~25%) compared to rEcΔ2.ΔA.B1, while decreasing LB MaxOD by ~14%. Meanwhile, tW* restored M9 MaxOD to that of rEcΔ2.ΔA.B1. We hypothesize the fitness impairments from tW* result from compromised cognate UGG recognition[56]. Taken together, these results reveal that fitness is

impaired by both ΔTGA recoding and cumulative translation engineering of RF2 exacerbated by tW* (**Extended Data Fig. 3a**).

When comparing DT and MaxOD of rEcΔ1.ΔA and rEcΔ2.ΔA independent of translation engineering (**Fig. 5, Extended Data Fig. 8**), we examined the accumulated effects of TGA codon replacement, deliberate genomic deletions (**Supplementary Tables 7**), TGA-overlapping gene mutations (**Supplementary Table 2**), and unintended background mutations acquired through selections (e.g., rRNAs), MAGE (e.g., off-targets), and conjugations (e.g., combing mutations) (**Supplementary Tables 11, 12, and 13, Extended Data Fig. 1**). While deletions and ΔTGA recoding play some role in cell fitness, most phenotypic disparities may have arisen from synthetic effects of accumulated unintended background mutations. This factor establishes opportunities for targeted and evolved fitness improvements[42,67], as well as analysis into the impacts of uncharacterized mutations on individual protein expression, function, and associated pathways.

To dissect the impacts of individual RF mutations on growth, RF-variants in isogenic rEcΔ2 backgrounds (**Supplementary Table 14b**), were assessed for growth fitness (DT and MaxOD) in diverse laboratory media (**Extended Data Fig. 4a**), but focusing primarily on growth in LB. All RF variants tested were derived from RF2-T246A (RF2.B0). The D26N conversion in RF2.B0, generating RF2.B1, resulted from the recoding of *prfB* internal autoregulatory TGA to TAA. RF2.B1 presented no significant change in LB DT or MaxOD compared to RF2.B0. Introducing E170K atop D26N, generating RF2.B4, presented no significant change in LB DT from RF2.B0, while decreasing MaxOD by ~2%. The introduction of S205P atop D26N and E170K, generating RF2.B3, presented greater shifts in LB fitness, slowing LB DT from RF2.B0 by ~12%, while decreasing MaxOD by ~12%.

To analyze the impact of S205P independent of E170K, RF2 variants were assessed in $RF1^+$ backgrounds. Expressing RF1 alongside RF2.B0 ($RF1^+$/B0) improved growth rates compared to ΔRF1/B0, decreasing DT in LB by ~3%, while not significantly altering MaxOD. Introducing D26N and S205P into B0 generated $RF1^+$/B2, decoupling S205P from E170K. Compared to $RF1^+$/B0, $RF1^+$/B2 presented an ~9% slower DT in LB and ~12% reduction in MaxOD. Meanwhile, introducing E170K to generate $RF1^+$/B3 presented ~12% slower DTs in LB and ~14% reductions in MaxOD compared to $RF1^+$/B0. These data suggest fitness impairments in rEcΔ2.ΔA.B3 result from cumulative effects of RF2 mutations, with S205P being a key contributor to changes in DT and MaxOD in LB.

Supplementary Discussion 4 | *Western Blot Analysis of Single Codon Translation*

To dissect translation at individual codons, western blots examining the products of single codon incorporations in GFP at position 151 (**Extended Data Fig. 9f**) were conducted in accordance with established protocols[4]. These results reveal improved full protein expression from nsAA incorporations at UGA within translationally engineered rEcΔ2.ΔA.B3 and rEcΔ2.ΔA.B3.tW* strains relative to ancestral rEcΔ2.ΔA.B1. GFP expression in the absence of OTSs demonstrated reduced protein truncation (i.e., termination at UGA) with mutant RF2 B3 strains compared to rEcΔ2.ΔA.B1. Meanwhile, the appearance of full protein expression (i.e., suppression at UGA) in absence of OTS within rEcΔ2.ΔA.B3 and subsequent loss within rEcΔ2.ΔA.B3.tW* demonstrates the impact of RF2 engineering in mitigating termination (enabling native Trp suppression) and of the $tRNA^{Trp}$ A37G mutation in mitigating this native UGA Trp suppression. Notably, minor remaining truncated protein identified in mutant RF2 strain blots under UGA

likely limits UAG/UGA-containing recombinant protein yields, though not accuracy of incorporations (**Fig. 6g**). Further evolution of RF2 to fully abolish UGA recognition may alleviate this limitation, enhancing UGA-UAG-containing recombinant protein yields and enabling both μ and λ phage resistance in *Ochre* (rEcΔ2.ΔA.B3.tW*).

Supplementary Discussion 5 | *Dual-Incorporation Protein Yields*

To evaluate applicability of our protein expression chassis, we measured purified protein yields derived from dual-incorporation constructs within rEcΔ2.ΔA.B3.tW* (**Extended Data Fig. 9g-h**). Dual incorporation of pAcF and BocK at a single codon pair UGA-UAG, respectively, presented ~2.3-fold greater yield (4.5 mg/L) than corresponding triple UGA-UAG codon pair constructs (2 mg/L), and ~2.2-fold lower yield than the 3xTAC sense codon control (9.7 mg/L) employing native translation machinery (**Extended Data Figure 9g**). Comparing these dual incorporation yields from rEcΔ2.ΔA.B3.tW* to an unrecoded conventional protein expression strain BL21, conveys the usefulness of this chassis in expressing nsAA-incorporated recombinant proteins. rEcΔ2.ΔA.B3.tW* ELP-GFP expression presented ~21-fold and ~31-fold enhanced protein expression for single and triple UGA-UAG dual-nsAA incorporations, respectively, relative to BL21 (**Extended Data Figure 9h**).

References:

71 Curry, K. A. & Tomich, C. S. Effect of ribosome binding site on gene expression in Escherichia coli. *DNA* **7**, 173-179, doi:10.1089/dna.1988.7.173 (1988).

72 Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* **27**, 946-950, doi:10.1038/nbt.1568 (2009).

73 Ma, N. J., Moonan, D. W. & Isaacs, F. J. Precise manipulation of bacterial chromosomes by conjugative assembly genome engineering. *Nat Protoc* **9**, 2285-2300, doi:10.1038/nprot.2014.081 (2014).

74 Van Wonterghem, L. *et al.* Genome-Wide Association Study Reveals Host Factors Affecting Conjugation in Escherichia coli. *Microorganisms* **10**, doi:10.3390/microorganisms10030608 (2022).

75 Zaher, H. S. & Green, R. A primary role for release factor 3 in quality control during translation elongation in Escherichia coli. *Cell* **147**, 396-408, doi:10.1016/j.cell.2011.08.045 (2011).