www.sciencemag.org/cgi/content/full/342/6156/357/DC1

Supplementary Materials for

Genomically Recoded Organisms Expand Biological Functions

Marc J. Lajoie, Alexis J. Rovner, Daniel B. Goodman, Hans-Rudolf Aerni, Adrian D. Haimovich, Gleb Kuznetsov, Jaron A. Mercer, Harris H. Wang, Peter A. Carr, Joshua A. Mosberg, Nadin Rohland, Peter G. Schultz, Joseph M. Jacobson, Jesse Rinehart, George M. Church,* Farren J. Isaacs*

*Corresponding author. E-mail: farren.isaacs@yale.edu (F.J.I.); gchurch@genetics.med.harvard.edu (G.M.C.)

Published 18 October 2013, *Science* **342**, 357 (2013)
DOI: 10.1126/science.1241459

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/cgi/content/full/342/6156/357/DC1)

Table of contents for Supplementary Online Text

Supplemental Figures

Supplemental Tables: (‡ indicates supplementary tables included as separate files)

A. Materials and Methods

All DNA oligonucleotides were purchased with standard purification and desalting from Integrated DNA Technologies (Table S19). Unless otherwise stated, all cultures were grown in LB-Lennox medium (LB^L, 10 g/L bacto tryptone, 5 g/L sodium chloride, 5 g/L yeast extract) with pH adjusted to 7.45 using 10 M NaOH. LB^L agar plates were LB^L plus 15 g/L bacto agar. Top agar was LB^L plus 7.5 g/L bacto agar. MacConkey agar was prepared using BD Difco™ MacConkey agar base according to the manufacturer's protocols. M9 medium (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 1 g/L NH_4Cl, 0.5 g/L NaCl, 3 mg/L $CaCl_2$) and M63 medium (2 g/L $(NH_4)_2SO_4$, 13.6 g KH_2PO_4, 0.5 mg $FeSO_4 \cdot 7H_2O$) were adjusted to pH 7 with 10 M NaOH and KOH, respectively. Both minimal media were supplemented with 1 mM $MgSO_4 \cdot 7H_2O$, 0.083 nM thiamine, 0.25 μg/L D-biotin, and 0.2% w/v carbon source (galactose, glycerol, or glucose).

The following selective agents were used: carbenicillin (50 μg/mL), chloramphenicol (20 μg /mL), kanamycin (30 μg/mL), spectinomycin (95 μg/mL), tetracycline (12 μg/mL), zeocin (10 μg/mL), gentamycin (5 μg/mL), SDS (0.005% w/v), Colicin E1 (ColE1; ~10 μg/mL), and 2-deoxygalactose (2-DOG; 0.2%). ColE1 was expressed in strain JC411 and purified as previously described (*26*). All other selective agents were obtained commercially.

The following inducers were used at the specified concentrations unless otherwise indicated: anhydrotetracycline (30 ng/μL), L-arabinose (0.2% w/v).

p-acetyl-L-phenylalanine (pAcF) was purchased from PepTech (# AL624-2) and used at a final concentration of 1 mM. O-phospho-L-serine (Sep) was purchased from Sigma Aldrich (# P0878-25G) and used at a final concentration of 2 mM.

Strains

All strains were based on EcNR2 (*11*) (*Escherichia coli* MG1655 Δ*mutS*::*cat* Δ(*ybhB-bioAB*)::[λcI857 N(*cro-ea59*)::*tetR-bla*]). Strains C321 [strain 48999 (www.addgene.org/48999)] and C321.ΔA [strain 48998 (www.addgene.org/48998)] are available from addgene.

Selectable marker preparation

Selectable markers were prepared using primers described in Table S19. PCR reactions (50 μL per reaction) were performed using Kapa HiFi HotStart ReadyMix according to the manufacturer's protocols with annealing at 62 °C. PCR products were purified using the Qiagen PCR purification kit, eluted in 30 μL of dH_2O, quantitated using a NanoDrop™ ND1000 spectrophotometer, and analyzed on a 1% agarose gel with ethidium bromide staining to confirm that the expected band was present and pure.

MAGE and λ Red-mediated recombination

MAGE (*13*), CoS-MAGE (*14*), and λ Red-mediated recombination (*27*) were performed as previously described. Briefly, an overnight culture was diluted 100-fold into 3 mL LB^L plus antibiotics and grown at 30 °C in a rotator drum until mid-log growth was achieved (OD_{600} ~0.4-0.6). Lambda Red was induced in a shaking water bath (42 °C, 300 rpm, 15 minutes), then induced culture tubes were cooled rapidly in an ice slurry for at least two minutes. Electrocompetent cells were prepared at 4 °C by pelleting 1 mL of culture (centrifuge at 16,000

rcf for 20 seconds) and washing the cell pellet twice with 1 mL ice cold deionized water (dH_2O). Electrocompetent pellets were resuspended in 50 µL of dH_2O containing the desired DNA. For MAGE oligos, no more than 5 µM (0.5 µM of each oligo) was used. For CoS-MAGE, no more than 5.5 µM (0.5 µM of each oligo including the co-selection oligo) was used. For dsDNA PCR products, 50 ng was used. Cells were transferred to 0.1 cm cuvettes, electroporated (BioRad GenePulser™, 1.78 kV, 200 Ω, 25 µF), and then immediately resuspended in 3 mL LB^L (MAGE and CoS-MAGE) or 1.5 mL LB^L (dsDNA). Recovery cultures were grown at 30 °C in a rotator drum. For continued MAGE cycling, cultures were recovered to mid-log phase before being induced for the next cycle. To isolate monoclonal colonies, cultures were recovered for at least 3 hours (MAGE and CoS-MAGE) or 1 hour (dsDNA) before plating on selective media. For *tolC* and *galK* negative selections, cultures were recovered for at least 7 hours to allow complete protein turnover before exposure to ColE1 and 2-deoxygalactose, respectively.

CAGE

CAGE was performed as previously described (*11*). Briefly, conjugants were grown to late-log phase in all relevant antibiotics (including tetracycline in the donor culture to select for the presence of conjugal plasmid pRK24 (*28*)). At mid-log growth, 2 mL of each culture was transferred to a 2 mL microcentrifuge tube and pelleted (5000 rcf, 5 minutes). Cultures were washed twice with LB^L to remove antibiotics, then the pellets were resuspended in 100 µL LB^L. Donor (10 µL) and recipient (90 µL) samples were mixed by gentle pipetting and then spotted onto a pre-warmed LB^L agar plate (6 x 10 µL and 2 x 20 µL spots). Conjugation proceeded at 30 °C without agitation for 1 – 24 hours. Conjugated cells were resuspended off of the LB^L agar plate using 750 µL liquid LB^L, and then 3 µL of the resuspended conjugation was inoculated into 3 mL of liquid LB^L containing the appropriate selective agents. The population with the correct resistance phenotype was then subjected to ColE1 negative selection to eliminate cells that retained *tolC*.

Each round of conjugation, genotyping, and strain manipulation required a minimum of 5 days to complete. On day 1, the conjugation and positive selections were performed. On day 2, the population of cells exhibiting the desired resistance phenotype was subjected to a ColE1 selection to eliminate candidates that retained *tolC*. The ColE1-resistant population was then spread onto plates to isolate monoclonal colonies. On day 3, candidate colonies were grown in a 96-well format and screened for the desired genotypes via PCR (to confirm loss of *tolC*) and MASC-PCR (to confirm the presence of the desired codon replacements). On day 4, *tolC* or *kanR*-oriT was recombined directly into one of the positive markers, and recombinants were plated on LB^L plates containing SDS or kanamycin, respectively. On day 5, candidate colonies were grown in liquid LB^L containing SDS or kanamycin and used as PCR template to confirm successful replacement of positive selection markers with *tolC* or *kanR*-oriT. These strains were ready for the next conjugation.

Positive/Negative selections

Positive selection for tolC: TolC provides robust resistance to SDS (0.005% w/v) in LB^L (both liquid and LB^L agar).

Negative selection for tolC: After *tolC* was removed via λ Red-mediated recombination or conjugation, cultures were recovered for at least 7 hours prior to ColE1 selection. This was

enough time for the recombination to proceed and for complete protein turnover in the recombinants (*i.e.* residual TolC protein no longer present). ColE1 selections were performed as previously described (*11*). Briefly, pre-selection cultures were grown to mid-log phase (OD_{600} ~0.4), then diluted 100-fold into 150 µL of LB^L and LB^L + ColE1. Once growth was detected, monoclonal colonies were isolated on non-selective plates and PCR screened to confirm the loss of *tolC*.

Positive selection for galK: GalK is necessary for growth on galactose (0.2% w/v) as a sole carbon source. It is important to thoroughly wash the cells with M9 media to remove residual carbon sources prior to selection in M63 + galactose (both liquid and M63 agar). Noble agar must be used, since Bacto agar may contain contaminants that can be used as alternative carbon sources.

Negative selection for galK: After *galK* was removed via λ Red-mediated recombination or conjugation, cultures were recovered for at least 7 hours prior to 2-DOG selection. This was enough time for the recombination to proceed and for complete protein turnover in the recombinants (*i.e.* residual GalK protein no longer present). 2-DOG selections were performed as previously described (*29*). Briefly, pre-selection cultures were grown to mid-log phase (OD_{600} ~0.4), washed three times in M9 medium to remove residual nutrients from LB^L, and then inoculated into M63 + 0.2% glycerol and M63 + 0.2% glycerol + 0.2% 2-DOG. Once growth was detected, monoclonal colonies were isolated on non-selective plates (LB^L agar or MacConkey agar) and PCR screened to confirm the loss of *galK*. When possible, colonies were streaked onto MacConkey + 0.2% galactose indicator plates (white colonies are Gal- and red colonies are Gal+) prior to PCR screening, but MacConkey media is toxic to strains that do not express TolC, which provides resistance to bile salts. We also found that 2-DOG selection was effective in LB^L, but PCR screening was important because LB^L + 2-DOG selection was less stringent.

Screening for galK and malK: Cultures were diluted and plated for single colonies on MacConkey agar + 0.2% galactose (galK) or MacConkey agar + 0.2% maltose (malK) indicator plates (white colonies are Gal- or Mal-, and red colonies are Gal+ or Mal+). The genotypes were confirmed *via* PCR.

Genotyping

After λ Red-mediated recombination or conjugation, colony PCR was used to confirm the presence or absence of selectable markers at desired positions. Colony PCR (10 µL per reaction) was performed using Kapa 2G Fast HotStart ReadyMix according to the manufacturer's protocols with annealing at 56 °C. Results were analyzed on a 1% agarose gel with ethidium bromide staining.

Multiplex allele-specific colony PCR (MASC-PCR) was used to simultaneously detect up to 10 UAG→UAA conversions as previously described (*11*). Briefly, each allele was interrogated by two separate PCRs to detect the UAG/UAA status. The two reactions shared the same reverse primer but used different forward primers whose 3′ ends annealed to the SNP being assayed. Amplification only in the wt-detecting PCR indicated a UAG allele, whereas amplification only in the mutant-detecting PCR indicated a UAA allele. Each primer set produced a unique

amplicon size corresponding to its target allele (100, 150, 200, 250, 300, 400, 500, 600, 700 and 850 bp). Template was prepared by growing monoclonal colonies to late-log phase in 150 µl LB^L and then diluting 2 µl of culture into 100 µl dH_2O. Initially, we used Qiagen Multiplex PCR kit, but KAPA 2G Fast Multiplex Ready Mix produced cleaner, more even amplification across our target amplicon size ranges. Therefore, typical MASC-PCR reactions contained KAPA 2G Fast Multiplex ReadyMix (Kapa Biosystems, # KK5802) and 10X Kapa dye in a final volume of 10 µl, including 2 µl of template and 0.2 µM of each primer. PCR activation occurred at 95°C (3 min), followed by 27 cycles of 95°C (15 sec), 63–67°C (30 sec; annealing temperature was optimized for each set of MASC-PCR primers), and 72°C (70 sec). The final extension was at 72°C (5 min). MASC-PCR results were analyzed on 1.5% agarose gels with ethidium bromide staining to ensure adequate band resolution.

Sanger sequencing was performed by Genewiz or Eton Bioscience, Inc.

Genomic DNA for whole genome sequencing was prepared using a Qiagen Genomic DNA purification kit or by simultaneously lysing raw culture and shearing genomic DNA using a Covaris E210 AFA Ultrasonication machine. Illumina libraries were prepared as previously described (*30*). Each strain was barcoded with a unique 6 bp barcode for pooling. Up to 16 strains were pooled for sequencing on a single HiSeq lane, and up to 4 genomes were pooled for sequencing on a single MiSeq lane. Whole genome sequencing was performed using Illumina HiSeq or MiSeq systems. The HiSeq samples were sequenced with paired end 50 bp or 100 bp reads, and the MiSeq samples were sequenced with paired end 150 bp reads.

Sequencing analysis

In order to analyze the sequencing data from 68 distinct genomes, we developed a software pipeline that connects several modular tools and custom scripts for analysis and visualization. The goal of our pipeline was to identify SNPs and structural variants relative to the reference genome *E. coli* K-12 MG1655 (U00096.2, GI:48994873). Note that we use the term SNP to mean any small mismatches or indels identified by Freebayes (<22 bp). We use the term structural variant to refer to large insertions detected by Breakdancer and Pindel, deletions, or other significant junction events (confirmed variants of size 170 bp and 776 bp in C321.ΔA).

FASTQ conversion to SAM/BAM: FASTQ reads were split using individual genome barcodes with the FASTX toolkit (*31*). After splitting and trimming of the 6 bp barcode, FASTQ files for individual reads were aligned to the reference genome (*E. coli* K-12 MG1655 or the C321.ΔA predicted genome sequence) using Bowtie2 version 2.0.0-beta5 (*32*) with local alignment and soft-clipping enabled. PCR duplicates were removed using the Picard toolkit <http://picard.sourceforge.net/> and reads were realigned around short indels using the Genome Analysis Toolkit (*33*).

SNP calling using Freebayes: SNPs were called using the Freebayes package (arXiv:1207.3907v2 [q-bio.GN]). SNP calls were made using a *--ploidy* flag value of 2, in order to catch SNPs that occur in duplicated regions. These SNPs show up as heterozygous calls in the output. The minimum alternate fraction for such calls was set at 0.4. The p-value cutoff was set at 0.001. SNPs from all genomes were called simultaneously, using the *--no-ewens-priors* and *--no-marginals* flags. The *--variant-input flag was used to provide Freebayes with the recoded*

SNP (UAG-to-UAA) positions as putative variants to call regardless of evidence. Reads supporting SNPs were required to have a minimum mapping quality of 10 and a minimum base quality of 30. Mapping quality was not otherwise used to assess SNP likelihoods (*--use-mapping-quality* was disabled). We ran Freebayes as described above to generate a single VCF file containing all variants for all samples. This VCF file was then further analyzed and filtered before as described below, before generating the summarizing diagram Fig. S3.

SNP Effect using snpEFF: SnpEff 2.0.5d (*34*) was used to annotate variants and to predict effects for called SNPs. First, the reference genome's annotated GenBank Record (GI:48994873) was used to create a genome database, and the VCF records were annotated for coding effects only.

Final SNP filtering: In addition to the Freebayes SNP identification criteria, we used additional metrics to filter out SNPs that could not be called with high confidence. This additional filtering helped to reduce the complexity of the relationship of variants across all sequenced genomes in order to plot Fig. S3. Note that this filtering resulted in some low-evidence variants being temporarily ignored in the aggregate analysis. However, these were carefully triaged and identified in the process of generating the sequence annotation file for the final C321.ΔA strain.

i. All 'heterozygous' calls were filtered out, as these represent SNPs whose reads map to multiple locations in the genome.
ii. SNPs that were present in fewer than three samples and could not be called either present or absent in >20 strains due to poor coverage or read mapping quality were filtered out.
iii. SNPs were filtered out if they were covered by ≤ 20 reads with good mapping quality across all genomes.
iv. SNPs that could be called absent or present in fewer than three genomes were removed.

Structural variants using Pindel and Breakdancer: Pindel (*35*) and Breakdancer (*36*) were both used to find potential structural variants in the genomes. First, Picard <http://picard.sourceforge.net/> was used to gather insert size metrics per genome. This information, along with the aligned BAM data, was run through Pindel. The Pindel output was converted to VCF using the *pindel2vcf* tool. We required at least 20 reads to support a breakpoint or junction. The *breakdancer_max* program in Breakdancer was also used to find structural variants. For Breakdancer, at least 8 read pairs were required to support a called structural event.

We manually corroborated structural variant calls from Pindel and Breakdancer through visual examination of read alignments. Since we observed a high-rate of false-positive and false-negative calls with these toolswe did not include them in our final strain analysis in the main text. Still, the Pindel and Breakdancer data were useful in troubleshooting cassette insertions and intentional gene knockouts and replacements.

Future work to combine evidence from these and additional tools might lead to a more robust, comprehensive, and high throughput method to validate structural variants using only short-read sequencing data.

Breakdancer predicted 49 unique events, and 187 total events across 69 strains. Because Breakdancer cannot call across multiple strains simultaneously and only gives approximate event locations based on read-pair distances, events that occurred in multiple samples were identified by using similar event start and end locations. Breakdancer predicted a total of 21 unique deletions, 5 unique inversions, and 23 unique translocations.

Pindel used split read data to predict both uncharacterized breakpoints and whole structural events. 258 unique uncharacterized breakpoints were found; 230 of these occur in only a single sample. Pindel also predicted 79 unique structural events. 9 were large deletions, 59 were insertions of unknown size, and 11 were inversions.

Coverage analysis: Coverage for each genome was analyzed using the bedtools (*37*) programs *makewindows* and *multicov.* The genome was split into 50 bp windows and BAM coverage was assessed for each window. A custom python script was used to take this information and find contiguous windows of low and high coverage, indicative of gene amplifications and deletions. These results are included as supplementary Table S31.

Confirming cassette insertion sites: Known insertion sites of CAGE antibiotic resistance markers were confirmed by selecting the reads that were soft clipped and/or not aligned to the MG1655, and aligning them to the known cassette sequences using Bowtie. Cassette insertion locations were inferred using the alignment locations of paired reads in which one read mapped to a cassette and the other mapped to a location on the genome.

Visually confirming SNPs and structural variants: The *tview* tool in the *Samtools* package (*38*) was used to visually inspect individual UAG SNPs and to assess the veracity of low-confidence SNP and structural variant calls.

Generating genome figures: Fig. S3 was created using custom software written in R and Processing.

Fitness analysis

To assess fitness, strains were grown in flat-bottom 96-well plates (150 μL LB^L, 34 °C, 300 rpm). Kinetic growth (OD_{600}) was monitored on a Biotek H4 plate reader at 5 minute intervals. Doubling times were calculated by $t_{double} = c*\ln(2)/m$, where c = 5 minutes per time point and m is the maximum slope of $\ln(OD_{600})$. Since some strains achieved lower maximum cell densities, slope was calculated based on the linear regression of $\ln(OD_{600})$ through 5 contiguous time points (20 minutes) rather than between two pre-determined OD_{600} values. To monitor fitness changes in the CAGE lineage, growth curves were measured in triplicate, and their average was reported in Fig. 2 and Table S1. To determine the effect of RF1 removal and NSAA incorporation on the panel of recoded strains (Table 1), growth curves were measured in triplicate (Fig. 3A, Fig. S8). Statistics were based on a Kruskal-Wallis one-way ANOVA followed by Dunn's multiple comparison test, where $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$.

To assess re-growth phenotypes from long-term NSAA expression, overnight cultures were first grown in LB^L supplemented with chloramphenicol to maintain the pEVOL plasmids. These

cultures were passaged into LB^L containing chloramphenicol, arabinose (to induce pEVOL), and either pAcF or pAzF depending on whether pEVOL-pAcF or pEVOL-pCNF was used. Growth with shaking at 34°C was monitored using a Biotek H1 or a Biotek Eon plate reader with OD_{600} readings every 10 minutes (pAcF) or 5 minutes (pAzF). After 16 hours of growth, the expression cultures were passaged into identical expression conditions and the growth curves were monitored with the same protocols.

NSAA incorporation assays

Plasmids and strains for NSAA incorporation: p-acetyl-L-phenylalanine (pAcF) incorporation was achieved using pEVOL-pAcF (*9*) which contains two copies of pAcF-RS and one copy of $tRNA_{CUA}^{opt}$. The pEVOL-pAcF plasmid was maintained using chloramphenicol resistance. One copy of pAcF-RS and $tRNA_{CUA}^{opt}$ were constitutively expressed, and the second copy of pAcF-RS was under araBAD-inducible control (0.2% L-arabinose).

O-phospho-L-serine (Sep) incorporation was achieved by expression of $tRNA^{Sep}$ from pSepT and both EFSep (EF-Tu variant capable of incorporating Sep) and SepRS from pKD-SepRS-EFSep (*21*). To prevent enzymatic dephosphorylation of Sep *in vivo*, the gene encoding phosphoserine phosphatase (*serB*), which catalyzes the last step in serine biosynthesis, was inactivated. Specifically, Glu93 (GAA) was mutated to a premature UAA stop codon *via* MAGE. The pKD-SepRS-EFSep plasmid was maintained using kanamycin resistance and both SepRS and EFSep were induced using IPTG. The pSepT plasmid was maintained using tetracycline resistance, and $tRNA^{Sep}$ was constitutively expressed.

Effect of RF1 deletion, aaRS expression, and NSAA incorporation on fitness: Stationary phase pre-cultures were obtained by overnight growth with shaking at 34 °C in 150 µl LB^L supplemented with chloramphenicol for plasmid maintenance. Stationary phase cultures were diluted 100-fold into 150 µl LB^L containing chloramphenicol and 0.2% L-arabinose and/or 1 mM pAcF where indicated. Growth was monitored on a Biotek Synergy H1 plate reader. OD_{600} was recorded at 10-minute intervals for 16 hours at 34 °C with continuous shaking. All data were measured in triplicate. Doubling time was determined for each replicate as described above, and replicates were averaged for Fig. 3A.

GFP variant synthesis: GFP variants (Table S33) were synthesized as gBlocks by IDT and modified with an N-terminal 6His tag *via* PCR. His-tagged GFP variants were isothermally assembled (*39*) into the pZE21 plasmid backbone (*40*) to yield the array of GFP reporter plasmids used in this study. Reporter plasmids were maintained using kanamycin resistance and induced using 30 ng/mL anhydrotetracycline (aTc).

UAG suppression and GFP Fluorescence: Stationary phase pre-cultures were obtained by overnight growth with shaking at 34 °C in 150 µl LB^L supplemented with appropriate antibiotics for plasmid maintenance. Stationary phase cultures were diluted 100-fold into 150 µl fresh LB^L containing the same antibiotics as the overnight pre-culture. These cultures were grown to mid-log phase and diluted 100-fold into 150 µl fresh LB^L containing the same antibiotics plus 30 ng/ml aTc, 0.2% L-arabinose, and/or 1 mM pAcF (where indicated). Protein expression proceeded for 16 hours at 34 °C with continuous shaking. Following 16 hours of expression, cultures were transferred to V-bottomed plates, pelleted, and washed once in 150 µL of PBS (pH 7.4). Washed pellets were resuspended in 150 µL of PBS (pH 7.4) and transferred to a black-walled, clear-bottom plate to measure GFP fluorescence for each strain. Both OD_{600} and GFP fluorescence (Ex: 485 nm, Em: 528 nm) were measured on a Biotek Synergy H1 plate reader. Fluorescence and OD_{600} measurements were corrected by subtracting background fluorescence and OD_{600} (determined using PBS blanks). Relative fluorescence (in rfu) was calculated by the ratio fluorescence/OD_{600}. Reported values represent an average of four replicates. After

measurements were complete, the cells were pelleted, the supernatant was aspirated, and the pellets were frozen at -80 °C for subsequent protein purification and Western blot analysis.

Protein extraction and Western blots: Cell pellets were obtained as described above. Cells were lysed using a lysis cocktail containing 150 mM NaCl, 50 mM Tris-HCl, 0.5x BugBuster reagent, 5% glycerol, 50 mM Na_3VO_4, 50 mM NaF, protease inhibitors (Roche), and 1 mM DTT. The resulting lysates were spun at 4 °C for 15 minutes at 3200 x g only in cases where soluble and insoluble fractions were separately analyzed. Protein lysate concentrations were determined using the BioRad-DC colormetric protein assay. Lysates were normalized by optical density at 600 nm, resolved by SDS-PAGE, and electro-blotted onto PVDF membranes (Millipore, # ISEQ00010). Western blot analysis was performed with mouse monoclonal antibody directed against GFP (Invitrogen, # 332600), and membranes were imaged with an HRP secondary antibody (Jackson Immunoresearch, JAC-715035150) *via* chemiluminescence on a ChemiDoc system (BioRad).

Mass spectrometry

Materials: Urea, Tris-HCl, $CaCl_2$, iodoacetamide (IAA), Pyrrolidine, DL-lactic acid, HPLC grade water and acetonitrile (ACN) were from Sigma-Aldrich (St. Louis, MO). Chloroform and dithiothretitol (DTT) were from American Bioanalytical (Natick, MA). Methanol, trifluoroacetic acid (TFA), ammonium hydroxide and formic acid (FA) were obtained from Burdick and Jackson (Morristown, NH). Sequencing grade modified trypsin was from Promega (Madison,WI). Anionic acid cleavable surfactant II (ALS) was from Protea (Morgantown, WV). UltraMicroSpin^{TM} columns, both the C_{18} and the DEAE PolyWAX variety were from The Nest Group, Inc. (Southborough, MA). Titaniumdioxide (TiO_2) with a particle size of 5 µm was obtained from GL Sciences Inc. (Torrance, CA).

Cell culture and lysis: Strains were routinely grown in LB^L media with the following concentration of antibiotics when appropriate: tetracycline (12 µg/mL), kanamycin (50 µg/mL), chloramphenicol (12 µg/mL), and zeocin (25 µg/mL). Bacterial cell cultures were grown at 30°C while shaking at 230 rpm until late log phase, quenched on ice and pelleted at 10,000 x g (10 min). The media was discarded and the cell pellets were frozen at -80°C to assist with subsequent protein extraction. Frozen cell pellets were thawed on ice and lysed in lysis buffer consisting of BugBuster reagent, 50 mM Tris-HCl (pH 7.4, 23°C), 500 mM NaCl, 0.5 mM EGTA, 0.5 mM EDTA, 14.3 mM 2-mercaptoethanol, 10 % glycerol, 50 mM NaF, and 1 mM Na_3O_4V, Phosphatase inhibitor cocktail 3 and complete protease inhibitor cocktail (Sigma Aldrich) were added as recommended by the corresponding manufacturer. Cell suspensions were incubated on ice for 30 min and the supernatant was removed after ultracentrifugation. The remaining pellet was re-extracted and resulting fractions were combined.

Protein lysates: Protein was precipitated with the methanol/chloroform method as previously described (*41*). One third of the resulting protein pellet was dissolved in 1.5 ml freshly prepared 8 M Urea/0.4 M Tris-HCl buffer (pH= 8.0, 23 °C). 5 mg protein was reduced and alkylated with IAA and digested overnight at 37°C using sequencing grade trypsin. The protein digest was desalted using C_{18} Sep-Pak (Waters) and the purified peptides were lyophilized and stored at -80°C.

Digestion of intact E. coli for shotgun proteomics: Cells were grown overnight to stationary phase, quenched on ice, and 2 ml culture was used for protein extraction and mass spectrometry. Cells were pelleted for 2 min at 2000 x g and the resulting pellet was washed twice with 1 ml ice cold Tris-HCl buffer pH=7.4, 23°C. The cells were then re-suspended in 100 µl Tris-HCl buffer pH=7.4, 23°C, split into 4 equal aliquots of 25 ul and the cell pellet was frozen at -80 °C. Frozen pellets were lysed with 40 µl lysis buffer consisting of 10 mM Tris-HCl buffer pH = 8.6 (23°C) supplemented with 10 mM DTT, 1 mM EDTA and 0.5 % ALS. Cells were lysed by vortex for 30 s and disulfide bonds were reduced by incubating the reaction for 35 min. at 55 °C in a heating block. The reaction was briefly quenched on ice and 16 µl of a 60 mM IAA solution was added. Alkylation of cysteines proceeded for 30 min in the dark. Excess IAA was quenched with 14 µl of a 25 mM DTT solution and the sample was then diluted with 330 µl of 183 mM Tris-HCl buffer pH=8.0 (23 °C) supplemented with 2 mM $CaCl_2$. Proteins were digested overnight using 12 µg sequencing grade trypsin for each protein aliquot, and the reaction was then quenched with 64 µl of a 20 % TFA solution, resulting in a sample pH<3. Remaining ALS reagent was cleaved for 15 min at room temperature. An aliquot of the sample consisting of ~30 µg protein (as determined by UV_{280} on a nanodrop) was desalted by reverse phase clean-up using C_{18} UltraMicroSpin columns. The desalted peptides were dried at room temperature in a rotary vacuum centrifuge and reconstituted in 30 µl 70 % formic acid 0.1 % TFA (3:8 v/v) for peptide quantitation by UV_{280}. The sample was diluted to a final concentration of 0.6 µg/µl and 4 µl (2.4 µg) were injected for LC-MS/MS analysis of the unfractionated digest using a 200 min method.

Phosphopeptide enrichment: Offline phosphopeptide enrichment was carried out with Titanium dioxide (TiO_2) using a bulk enrichment strategy adapted from Kettenbach *(42)*. Briefly, between 0.4 and 1 mg of desalted peptide digest was transferred into a 1.5 ml PCR tube and dissolved at a concentration of 1mg/ml in "binding solution" consisting of 2 M lactic acid in 50 % ACN. Activated TiO_2 was prepared as a concentrated slurry in binding solution and added to the peptide solution to obtain a TiO_2 to peptide ratio of 4:1 by mass. The mixture was incubated for 2 h at room temperature on an Orbit M60 laboratory shaker operated at 140 rpm. The suspension was centrifuged for 20 s at 600 x g and the supernatant was removed. The TiO_2 beads were washed twice with 50 µl of the binding solution and then 3 times with 100 µl 50 % ACN, 0.1 % TFA. Stepwise elution of phosphopeptides from the beads was carried out using 20 µl of 0.2 M sodium phosphate buffer pH=7.8 followed by 20 µl 5 % ammonium hydroxide and 20 µl 5 % pyrrolidine solution. The pH of the combined extracts was adjusted with 30 µl of ice cold 20 % TFA resulting in a sample pH <3.0. Peptides were desalted on C_{18} UltraMicroSpin columns as described above and the peptide concentration was estimated by UV_{280}.

Offline fractionation of tryptic digests: Offline electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) *(43)* was performed on disposable DEAE PolyWAX UltraMicroSpin columns. Columns were activated as recommended by the manufacturer and then conditioned with 3 x 200 µl washes with 90 % ACN, 0.1 % acetic acid (buffer A). For this purpose, the columns were centrifuged for at 200 x g for 1 min at 4°C. The column was then loaded with 50 µg of a desalted peptide digest prepared in 25 µl buffer A, and the flow-through was collected. Stepwise elution of the peptides was carried out using brief centrifugation steps carried out for 30 s at 200 x g with 50 µl eluent unless noted otherwise. The elution steps consisted of the following volumetric mixtures of buffer A and buffer B (0.1 % formic acid in 30 % ACN): (1) 100:0 (2) 96:4 (3) 90:10 (4) 80:20 (5) 60:40 (6) 100 µl of 20:80 (7) 100 µl of 0:100. Additional

elution steps consisted of: (8) 1 M triethylamine buffer adjusted with formic acid to pH=2.0. (9) 0.2 % ammonia (10) 0.2 % ammonia and finally (11) 100 µl 70 % formic acid. The collected fractions were dried in a vacuum centrifuge and reconstituted in 15 µl solvent consisting of 3:8 by volume of 70 % formic acid and 0.1 % TFA. Fractions were analyzed by LC-MS/MS using a 400 min gradient.

Liquid chromatography and mass spectrometry: Capillary LC-MS was performed on an Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) connected to a nanoAcquity UPLC (Waters, Milford, MA). Liquid chromatography was performed at 35 °C with a vented split setup consisting of a commercially available 180 µm x 20 mm C_{18} nanoAcquity UPLC trap column and a BEH130C18 Waters symmetry 75 µm ID x 250 mm capillary column packed with 5 and 1.7 µm particles respectively. Mobile phase A was 0.1 % formic acid (FA) and mobile phase B was 0.1 % FA in acetonitrile. The injection volume was 4-5 µl depending on the sample concentration. Up to 2.4 µg peptides were injected for each analysis. Peptides were trapped for 3 min in 1 % B with and a flow rate of 5 µl/min. Gradient elution was performed with 90, 200 and 400 min methods with a flow rate of 300 nl/min. Two blank injections were performed between samples to limit potential carryover between the runs. The gradient for the 90 min method was 1-12 % B over 2 min, 12-25 % B over 43 min, 25-50 % B over 20 min, followed by 6 min at 95 % B and column re-equilibration in 1 % B. The gradient for the 200 min was 1-10 % B over 2 min, 10-25 % B over 150 min, and 25-50 % B over 20 min, followed by 7 min at 95 % B and re-column equilibration at 1 % B. The gradient for the 400 min was 1 min in 1 % B, 1-7 % B over 2 min, 7-20 % B over 298 min, and 20-50 % B over 60 min, followed by a 1 min flow ramp to 95 % B. The column was flushed for 9 min using 95 % B and then re-equilibrated for 27 min at 1 % B prior to the next injection. Mass spectrometry was performed with a spray voltage of 1.8 kV and a capillary temperature of 270 °C. A top 10 Higher Collisional Energy Dissociation (HCD) method with one precursor survey scan (300-1750 Da) and up to 10 tandem MS spectra performed with an isolation window of 2 Da and a normalized collision energy of 40 eV. The resolving power (at m/z = 400) of the Orbitrap was 30,000 for the precursor and 7500 for the fragment ion spectra, respectively. Continuous lock mass calibration was enabled using the polycyclodimethylsiloxane peak (m/z = 445.120025) as described *(44)*. Dynamic exclusion criteria were set to fragment precursor ions exceeding 3000 counts with a charge state >1 twice within a 30 s period before excluding them from subsequent analysis for a period of 60 s. The exclusion list size was 500 and early expiration was disabled.

Proteomics data processing: Raw files from the Orbitrap were processed with Mascot Distiller and searched in-house with MASCOT (v. 2.4.0) against the EcoCyc *(45)* protein database release 16.0 for *E. coli* K-12 substr. MG1655 with a custom database and search strategy designed to identify amber suppression (Aerni et al. manuscript in preparation). Forward and decoy database searches were performed with full trypsin specificity allowing up to 3 missed cleavages and using a mass tolerance of ±30 ppm for the precursor and ±0.1 Da for fragment ions, respectively. Cysteines were considered to be completely alkylated with IAA unless samples were processed by a gel-based workflow. In that case Propionamide (C) was considered as a variable modification. Additional variable modifications for all searches were oxidation (M) and deamidation (NQ) for samples processed with urea Carbamyl (K, R, N-term). In order to detect pAcF containing peptides, a variable custom modification for Y was introduced with the composition C_2H_2 and monoisotopic mass of 26.015650 Da. Typical FDR were <1 % for

peptides above identity threshold and <2% considering all peptides above identity or homology threshold respectively. The MASCOT search results were deposited in the Yale Protein Expression Database (YPED) *(46)*. The following filter rules were specified in YPED for reporting of protein identifications: (i) At least 2 bold peptides and peptide scores ≥20 or (ii) 1 bold red peptide with a peptide score ≥20 with at least one additional bold red peptide with a score between 15 and 20.

Bacteriophage assays

For all phage experiments, growth was carried out in LB^L at 30 °C. Liquid cultures were aerated with shaking at 300 rpm. Before each experiment, a fresh phage lysate was prepared. To do this, *Escherichia coli* MG1655 was grown to mid-log phase in 3 mL of LB^L, then ~2 uL of T7 bacteriophage (ATCC strain BAA-1025-B2) or T4 bacteriophage (ATCC strain 11303-B4) was added directly from a glycerol stock into the bacterial culture. Lysis proceeded until it was complete (lysate appears clear after ~4 hours). The entire lysate was centrifuged to remove cell debris (10,000 rcf, 10 minutes), and 3 mL of lysate was transferred to a glass vial supplemented with 150 mg NaCl for phage preservation. Lysates were prepared fresh, titered, and stored at 4 °C for the duration of each experiment. One lysate was used for all replicates of a given experiment.

Phage titering: Phage lysate was titered by serial dilution into LB^L (10-fold dilution series). Before plating on LB^L agar, 10 µL of the diluted phage lysate was mixed with 300 µL of mid-log *E. coli* MG1655 culture and 3 mL of molten top agar. Plaques matured for ~4 hours at 30 °C. Titers (pfu/mL) were calculated based on the lysate dilutions that produced 20-200 pfu.

Plaque area: For plaque area assays, bacterial cultures were grown to mid-log phase in 3 mL LB^L. To accommodate different doubling times, faster-growing cultures were continually diluted until all strains reached OD_{600} ~0.5. Immediately prior to infection, OD_{600} was normalized to 0.50 for all cultures. Approximately 30 pfu of T7 bacteriophage were mixed with 300 µL of OD_{600} = 0.50 culture and 3 mL of molten top agar, and then immediately plated on LB^L agar. Plaques were allowed to mature at 30 °C for 7 hours, then the plates were imaged on a Bio-Rad Gel Doc system, and plaque areas were measured using ImageJ (*47*). Statistics were based on a Kruskal-Wallis one-way ANOVA followed by Dunn's multiple comparison test, where *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

T7 Fitness: Fitness was assessed in triplicate at low MOI based on protocols by Heineman *et al.* (*22*). Briefly, bacterial glycerol stocks were inoculated directly into 3 mL LB^L and serially diluted in LB^L. Serial dilutions were grown overnight (30 °C, 300 rpm), so that one of the dilutions would be at mid-log growth phase in the morning. Prior to infection, a second dilution series was performed so that host strains would be at optimal growth phase over the course of the serial infection. Starting cultures were normalized to OD_{600} = 0.50 by adding LB^L immediately before infecting the cultures (MOI = 0.015) at t = 0. Infected culture was diluted 1/10 into 3 mL of uninfected mid-log phase culture at 30 minute intervals. Aliquots of the infection were taken at t = 4, 10, 60, and 120 minutes. At t = 4, the aliquot was treated with chloroform to quantitate non-adsorbed phage particles. For all other time points (t = 10, 60, and 120), aliquots were immediately mixed with 300 µL of mid-log *E. coli* MG1655 and 3 mL molten top agar and then

spread on LB^L agar. Plaques were counted after maturing for ~4 hours at 30 °C, and then pfu/mL was calculated for each time point, correcting for dilutions. Adsorption efficiency was consistently >95% as determined by $(N_{t=4} - N_{t=10}) / N_{t=10}$, and fitness was determined by $[\log_2(N_{t=120}/N_{t=60})]/(\Delta t/(60 \text{ min/hr}))$, where N is the number of phages at time t minutes and Δt = 60 min.

Kinetic lysis time: Mean lysis time was determined with 12 replicates based on protocols from Heineman *et al.* (*22*), except that OD_{600} was monitored instead of OD_{540}. Mid-log phase cells (as in the fitness assay) were infected at MOI = 5, then 150 µL aliquots of infected culture were distributed into a 96-well flat bottomed plate and sealed with Breathe-Easy™ sealing membrane. Lysis was monitored at 30 °C with shaking at 300 rpm on a Biotek H4 plate reader with OD_{600} measurements taken every 5 minutes. Each lysis curve was fit to a cumulative normal distribution using the normcdf function in MATLAB. Mean lysis time, mean lysis OD_{600}, and mean lysis slope were calculated using this cumulative normal distribution function.

B. Time and cost

In order to demonstrate the efficiency and cost-effectiveness of our recoding strategy, we explicitly present the total full time equivalents (10.75 FTE years) and DNA costs ($20,333) required to complete this project. Because much of our research time was spent developing and optimizing these genome engineering tools as described below, we estimate the actual time spent constructing a fully recoded genome (5.5 FTE years), and the minimum amount of time that it would take to repeat its construction with current knowledge (0.5 FTE years) (Tables S10 and S11). By contrast, the design, synthesis, and assembly of the 1.08–mega–base pair *Mycoplasma mycoides* JCVI-syn1.0 genome required $40 million and more than 200 FTE years (*48*). While future *de novo* genome synthesis projects will likely improve on these figures by incorporating chip-based DNA synthesis (*49*), our strategy nevertheless demonstrates considerable advantages in the cost and efficiency of making hundreds of genome changes.

Table S17. Time required to reassign UAG

Phase	Technology development	Actual strain construction	Time to repeat	Time with CoS-MAGE[a]
MAGE	3.75	1.50	0.15	0.24
CAGE	7.00	4.00	0.35	0.12
Total	**10.75**	**5.50**	**0.50**	**0.36**

[a]Suggested improvements: make 40 changes per strain using improved CoS-MAGE strains (*50, 51*)

Table S18. DNA cost for reassigning the UAG codon

Oligo type	MAGE oligos	mascPCR primers	Cassette amplification primers	Cassette screening primers	Deletion oligos	Total
Description	320 x 90-mer oligos with 4 PTO bonds	978 oligos (~23 bp)	190 x 72-mer oligos	144 x 25-mer oligos	25 x 90-mer oligos	-
Yield	100 nmole DNA plate	25 nmole DNA plate	25 nmole DNA plate	25 nmole DNA plate	100 nmole DNA plate	-
Price per base*	$0.28 per base, $3.50 per PTO bond	$0.18 per base	$0.18 per base	$0.18 per base	$0.28 per base	-
Total price	$12,544.00	$4,048.92	$2,462.40	$648.00	$630.00	$20,333.32

*IDT standard price

Since we developed MAGE and CAGE at the same time as we were using them to reassign UAG, a considerable portion of our effort was devoted to technology optimization and changing strategies. For instance, since *tolC* negative selection yields scarless conjugal junctions, desired conjugants can be prepared for subsequent conjugations in one step by inserting *kanR*-oriT or *tolC* directly into one of the existing positive markers (*11*). Therefore, 6 modular cassettes targeting *kanR*-oriT or *tolC* to replace *specR* (spectinomycin), *zeoR* (zeocin), or *gentR* (gentamycin) are adequate for all conjugations beginning with the second round. Our initial designs did not take this into account, so we first had to remove one or both positive markers via a two step replacement and deletion procedure using *tolC* or *galK*. However, now that we better understand the homology requirements for precisely assembling genome segments of various sizes (Table S20), selectable markers can be placed to permit one-step turnaround between conjugations. Therefore, we report both the FTE time required to complete the construction of C321 and the estimated FTE time required to repeat the project with current knowledge (Table S17).

Table S20. Positions of markers for CAGE and window sizes for conjugal junctions

Conjugation	Donor oriT position	Recipient PN marker position	lo Positive marker position	hi Positive marker position	oriT/tolC junction[a]	Positive marker junction
Conj1	4019968	none	3921005	4417928	undefined	4142298
Conj2	4497524	none	4417928	4612628	undefined	4444521
Conj3	189613	182395	4612628	374608	7218	4238020
Conj4	480320	474528	36400	629000	5792	4046621
Conj5	781100	788054	608541	903110	6954	4344652
Conj6	1145180	1124600	892756	1255700	20580	4276277
Conj7	1416412	1415470	1255700	1542300	942	4352621
Conj8	MAGE	MAGE	MAGE	MAGE	MAGE	MAGE
Conj11	2438300	2428900	2223738	2627100	9400	4235859
Conj12	2784761	2783150	2627100	2840467	1611	4425854
Conj13	2967175	2968028	2840467	3014000	853	4465688
Conj14	3176034	3184259	3010540	3334920	8225	4314841
Conj15	3544352	none	3331657	4245059	undefined	3725819
Conj16	3816822	none	3735445	4245059	undefined	4129607
Conj17	4417928	4417928	3921005	4612628	0	3947598
Conj18	374608	36400	4612628	629000	338208	3983628
Conj19	892756	903110	608541	1255700	10354	3992062
Conj20	1529620	1542300	1255700	1702450	12680	4192471
Conj21	MAGE	MAGE	MAGE	MAGE	MAGE	MAGE
Conj22	2627100	none	2223738	2840467	undefined	4022492
Conj23	3014000	3010540	2840467	3334920	3460	4144768
Conj24	3734278	none	3332800	3921005	undefined	4051016
Conj25	4610360	4612400	3921005	629000	2040	3292005
Conj26	1255700	none	608541	1702450	undefined	3545312

Conj27	2223738	2209114	1710450	2840467	14624	3509204
Conj28	3332800	3346270	2840467	3921005	13470	3558683
Conj29	608541	791470	1702450	2627225	182929	924775
Conj30_Cn2	2848625	2840467	1710450	3921005	8158	2428666
Conj30_Cn7	2840467	2209114	1710450	3921005	631353	2428666
Conj30_5	1719000	1663210	608541	3921005	55790	1326757
Conj31	3864420	3921005	1255700	1719000	56585	463300

[a]Undefined means that there was no selection for the desired crossover position during conjugation

Minimal time required to repeat the construction of C321 with current knowledge

MAGE: 40 days

- 2 days of continuous cycling for 18 cycles
- 16 days to screen 32 MAGE populations (screen 2 populations per day)
- 1 day for 7 additional cycles
- 16 days to screen MAGE populations (screen 2 populations per day)
- 5 days to introduce the remaining UAG alleles and screen for desired clones

CAGE: 90 days

- 1 day to prepare selectable marker cassettes
- 1 day to recombine *specR*, *gentR*, or *zeoR* marker into rEc strains
- 1 day to screen for desired recombinants
- 1 day to recombine marker *tolC* or *kanR*-oriT into recombinants
- 1 day to screen for desired double recombinants
- 85 days for 6 conjugations (minimum of 5 days per conjugation, maximum of 2 conjugations per day)
    - Phase 1: 16 conjugations = 40 days
    - Phase 2: 8 conjugations = 20 days
    - Phase 3: 4 conjugations = 10 days
    - Phase 4: 2 conjugations = 5 days
    - Phase 5: 1 conjugation = 5 days
    - Phase 6: 1 conjugation = 5 days

C. Construction of a recoded genome

Starting from EcNR2 (*Escherichia coli* MG1655 Δ*mutS*::*cat* Δ(*ybhB-bioAB*)::[λcI857 N(*cro-ea59*)::*tetR-bla*]), we removed 305/321 UAG codons across 32 "rEc" strains. Each strain had 10 adjacent UAG codons that we converted to UAA using MAGE. None of these strains exhibited impaired fitness. We then used CAGE to hierarchically assemble the recoded segments ("Conj" strains) into a fully recoded strain (summarized in Fig. 2). We identified and overcame several barriers during genome construction. Below, we describe all deviations from our initial design, which was to (1) create 32 strains each with 10 UAG codons replaced by UAA, (2) hierarchically combine adjacent recoded segments into a strain completely lacking UAG, (3) remove release factor 1 (RF1) so that UAG would not cause translational termination. UAG IDs are based on Table S16.

MAGE phase:

UAGs that were not converted (false positives from MASC-PCR analysis): (Table S16)
rEc4 retained UAGs 4.9 and 4.10.
rEc5 retained UAGs 5.1 and 5.2.
rEc12 retained UAG 12.9.
rEc14 retained UAG 14.5.
rEc15 retained UAG 15.8.
rEc19 retained UAG 19.7.
rEc30 retained UAG 30.3.

UAGs that were converted in addition to the targeted set (Probably from MAGE oligo mix-ups): (Table S16)
rEc29 had UAG→UAA 16.1-16.4, 30.5
rEc30 had UAG→UAA 6.7
rEc31 had UAG→UAA 6.7

CAGE phase:

CAGE design for Conj1, Conj2, Conj3, Conj31, and Conj32: We were still optimizing the conjugation selection criteria at the beginning of the CAGE phase. For the first few conjugations, we used no selections or positive selections at conjugal junctions. As the CAGE phase proceeded, we adopted *tolC* negative selection at the conjugal junction between recoded genome segments to permit scarless genome assembly.

Conj8 MAGE construction: Instead of conjugating rEc15 + rEc16 to produce Conj8, we performed additional MAGE cycling in rEc15 to convert 16.1-16.4. This strain was renamed Conj8, and rEc16 was not used in the final recoded genome assembly.

Conj11 IS insertion into tolC: IS5 was inserted into *tolC* rather than the desired *tolC* deletion. This undesired feature was automatically lost during Conjugation 22.

Conj21 and Conj23 dysfunctional *tolC*: Robust negative selection is important for creating scarless conjugal junctions while ensuring that all donor alleles are transferred during CAGE (*11*). We previously reported that two of our 1/8 recoded strains (Conj21 and Conj23) were found to simultaneously survive positive selection (SDS resistance) and negative selection (Colicin E1 resistance). We were able to correct this phenotype in Conj23 by removing the dysfunctional *tolC* cassette and introducing a functional *tolC* elsewhere in the genome; however, the dysfunctional allele appeared to map elsewhere in Conj21 (*11*). Additionally, the Conj10 parental strain used to create Conj21 appeared to also have the dysfunctional *tolC* phenotype. Since we were unable to readily identify the causative allele via whole genome sequencing (obvious candidate genes such as *tolQ*, *tolR*, *tolA*, and *butB* were not mutated), we re-made Conj21 using CoS-MAGE (*14*). This process took 8 cycles of CoS-MAGE and MASC-PCR screening (25 calendar days) to convert 30 UAG codons to UAA. We have found that using PCR to confirm the loss of *tolC* during conjugation is generally adequate to ensure robust isolation of a desired genotype when it is present at a frequency of greater than 1E-5 in the pre-selection population. Therefore, it is advantageous to perform the post-conjugation positive selections first to remove undesired genotypes from the population prior to Colicin E1 selection. Additionally, we are currently working on identifying dysfunctional *tolC* alleles with the goal of mitigating escape mechanisms and thereby increasing the selective power of the *tolC* negative selection.

Potential recombination hotspot caused UAGs to be retained in Conj6, Conj19, and Conj26: Although rare, several UAG codons were unexpectedly retained (Table S16) during CAGE despite proper *tolC*/*kan*^{R}-oriT conjugal junction placement. The Conj6 donor failed to transfer UAGs 10.6 – 10.10 during Conjugation 19. In turn, the Conj19 donor failed to transfer UAGs 9.4, 9.5, 9.10, 10.5, 11.3, and 11.8 during Conjugation 26 (S[UAGs converted in all strains]). This region may be a recombination hotspot that promotes several crossovers. We used MAGE to convert these undesired UAGs.

Conj25 *tolC* positive selection: We introduced *tolC* into the Conj18 donor instead of the Conj17 recipient for Conjugation 25. Therefore, we performed SDS selection rather than ColE1 selection. After isolating a desired Conj25 clone, we removed the *tolC* via λ Red before replacing selectable markers and proceeding to the next conjugation.

Conj20, Conj26, and Conj29 putative rearrangement: We found that *tolC* repeatedly recombined into an unknown location when we attempted to use it to delete *spec*^{R} from Conj20. Therefore, we performed Conjugation 26 without *tolC* negative selection. Unfortunately, the same *tolC* mistargeting was observed in strain Conj26, indicating that the genome feature causing the mistargeting had been inherited. Therefore, we identified the position of the undesired *tolC* insertion so that we could remove it. To this end, we first tested several different selectable cassettes and found that the *tolC* cassette's promoter and terminator sequences were both necessary and sufficient for the Conj20 mistargeted *tolC* insertion (Fig. S11A).

Next, we used inverse PCR and Sanger sequencing to locate the exact position of the *tolC* mistargeting (Fig. S11B). Briefly, we purified genomic DNA from a Δ*tolC*:*kan*^{R} recombinant, sheared it to ~2 kb fragments on a Covaris AFA Ultrasonication machine, end repaired the gDNA fragments (NEB end repair kit), and ligated standard Illumina adapters (T4 DNA Ligase). We then used 3 cycles of nested PCR in which one primer annealed to the Illumina adapter and

the other 3 primers annealed facing outwards from the *kan*^{R} gene to amplify each junction between the *kan*^{R} gene and the surrounding genomic sequence. We gel purified the portion of the smear corresponding to ~1 kb PCR products, re-amplified with the third nested primer, purified the product (Qiagen PCR purification kit), and then directly Sanger sequenced without subcloning to identify the genome sequence flanking *tolC*. Sequencing indicated that the *kan*^{R} N-terminus was inserted just downstream of nt 3,176,063 (endogenous position of *tolC*), and that the *kan*^{R} C-terminus was inserted just upstream of nt 3,421,404. These loci are 245,341 kb apart in the *E. coli* MG1655 genome. Although we were unable to identify the structural variant in Conj20 via whole genome sequencing, we confirmed the putative rearrangement via colony PCR using primers that hybridize ~150 bp on either side of the putative *kan*^{R} insertion site, and we observed the expected 1.5 kb amplicon (1.2 kb *kan*^{R} + 300 bp of flanking genome sequence, verified via Sanger sequencing). The same PCR in Conj20 (without Δ*tolC*:*kan*^{R} inserted) did not produce an amplicon, and PCR amplification of the endogenous *tolC* locus of both strains produced the expected amplicon for a *tolC* deletion. Taken together, this indicates that the region near the endogenous *tolC* was duplicated and inserted near nt 3,421,404, that a large sequence (too large to be detected by PCR) is deleted by Δ*tolC*:*kan*^{R}, and that the endogenous *tolC* region was not impacted by the mistargeting.

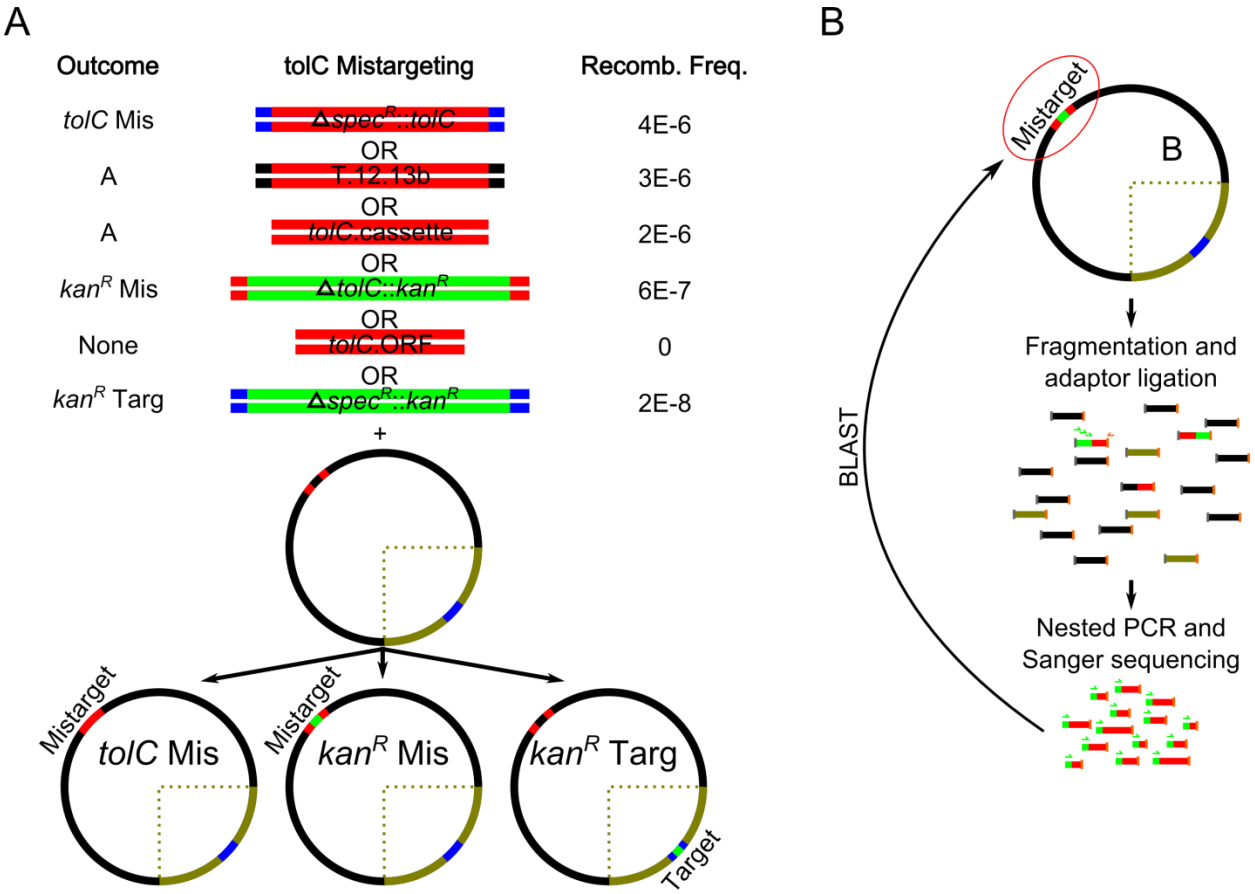


Fig. S11. Putative Conj20 rearrangement causing *tolC* mistargeting. (A) Several different *tolC* cassettes repeatedly recombined into an unknown locus (*tolC* Mis), a *kan*^{R} cassette having homology to the *tolC* cassette's promoter and terminator sequences efficiently recombined into an unknown locus (*kan*^{R} Mis), a *kan*^{R} cassette having homology to *spec*^{R} efficiently recombined

into the expected locus (*Kan*^{R} Targ), and the *tolC* ORF lacking a promoter and terminator was not recombinogenic in Conj20. Therefore, the *tolC* cassette's promoter and terminator were necessary and sufficient to mediate *tolC* mistargeting in Conj20. (B) The position of mistargeting was identified by purifying the genome of C20.DT:*kan*^{R}, fragmenting to ~2 kb pieces on a Covaris AFA Ultrasonication machine, repairing DNA ends with a NEB End Repair kit, ligating Illumina adaptors, and performing 3 rounds of nested inverse PCR. The amplicons were gel purified, re-amplified, Sanger sequenced, and BLASTed against the *E. coli* MG1655 genome (taxid:511145). The N-terminal insertion site was nt 3,176,063 (endogenous *tolC* position) and the C-terminal insertion site was nt 3,421,404 (245,341 bp away).

Since the putative rearrangement in Conj20 and Conj26 (region including nt 3,176,063 – nt 3,421,404) was distant from the recoded region (nt 633,969 – nt 1,663,144), we easily prevented its transfer during Conjugation 29 by placing the Conj25 recipient's positive selectable marker at SIR.22.23c (nt 2,627,225) instead of SIR.32.1 (nt 3,921,005). This marker placement permitted a *tolC*/*kan*^{R}-oriT junction between nt 608,541 – nt 629,000 (20,459 bp) and a *gent*^{R}/*zeo*^{R} junction between nt 1,702,450 and nt 2,627,225 (924,775 bp) (Fig. S12).

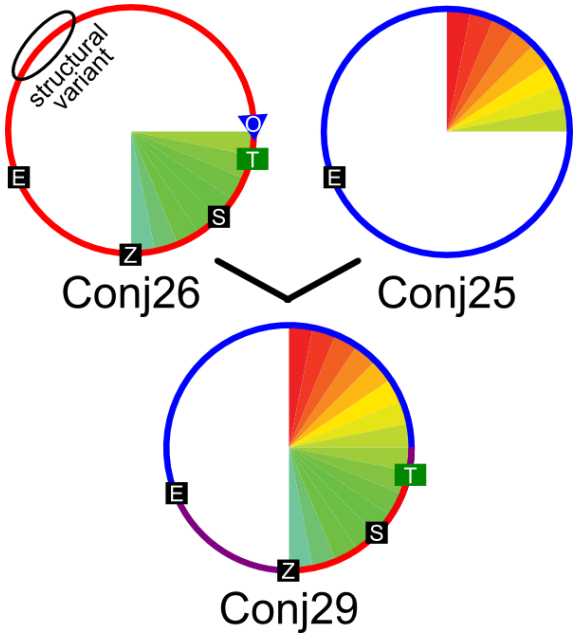


Fig. S12. Strategic marker placement allowed removal of the undesired structural variant from Conj26. Rather than placing *gent*^{R} at the boundary of the Conj25 recoded region, it was placed further away to select against inheritance of the Conj26 structural variant. Red lines represent Conj26 donor genome sequence, blue lines represent Conj25 recipient genome sequence, and purple lines indicate conjugal junction regions.

Inadequate homology for conjugal junction in Conj28 and Conj30: There is an average of 14.3 kb spanning adjacent UAG codons in *E. coli* MG1655, but many of these regions are inadequate for transferring large genome segments, since conjugal transfer frequency decreases exponentially with increasing distance (*52*). Our first attempts at using small homology regions to transfer large genome segments either led to failed selections (Conj28) or produced low complexity populations consisting of few recombinants (Conj30). By increasing the distance

between *kan*^{R}-oriT and *tolC* (Table S20), complete transfer of the recoded segment was achieved, but marker placement sometimes allowed recoded alleles near conjugal junctions to be lost (Fig. 2, Table S16).

Our initial attempts at Conjugation 28 failed because 2120 bp of homology between the donor's *kan*^{R}-oriT and the recipient's *tolC* were inadequate to transfer all 573,882 bp of recoded donor DNA. Instead, the putative Conj28 candidates all retained *tolC* and 25 or more UAG codons proximal to *kan*^{R}-oriT. Therefore, our selections yielded the dysfunctional *tolC* phenotype that was described above for Conj21 and Conj23. However, when we moved *tolC* so that it was 13,470 bp away from *kan*^{R}-oriT and repeated Conjugation 28, we easily selected desired clones.

In another case, the inefficient 1/4 genome transfer during Conjugation 30 yielded a low complexity population retaining 30 undesired UAGs (segments 18-20) in the middle of the donor region (Fig. S13). Such double crossovers may be caused by two separate conjugations (*52*), or may be formed when the excised recipient genome is partially degraded and recombined back into the donor segment that originally displaced it (*53*). Although the selections did not fail, recombination occurred rarely in the desired 8,158 bp *tolC*/*kan*^{R}-oriT conjugal junction, yielding a single isogenic population (46 out of 46 screened clones) retaining the same 30 UAGs from segments18-20. Rather than repeating the conjugation with the original conjugants, we chose a clone from the first conjugation to carry forward as the recipient in a second conjugation. We moved the selectable markers in Conj27 and the new recipient so that there would be 631,353 bp between *tolC* and *kan*^{R}-oriT, and then repeated the conjugation. This time, all remaining alleles were properly transferred (Fig. S13).

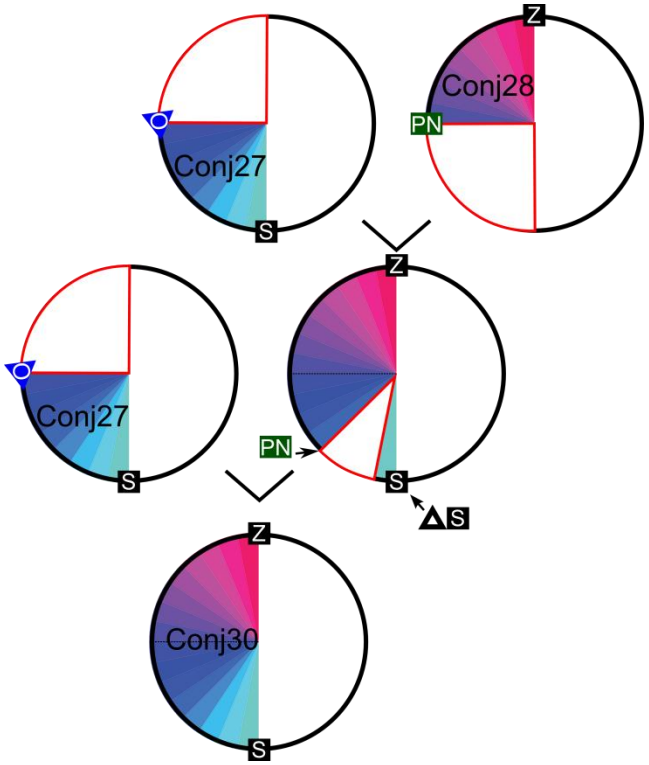


Fig. S13. Strain Conj30 was prepared by two serial conjugations. The first Conjugation 30 was performed using Conj27 and Conj28 (with 8,158 bp of homology between *tolC* and *kan*^{R}-oriT).

After selecting for *Spec*^{R}, *Zeo*^{R}, and ColE1^R, 46 out of 46 clones retained ~30 UAG codons in sets 18-20. After removing *spec*^{R} (replaced with *tolC* and then deleted *tolC*) and inserting a new *tolC* near the remaining UAG alleles in the conjugal progeny (providing 631,353 bp between *tolC* and *kan*^{R}-oriT for proper recombination), we performed a second conjugation to transfer the remaining alleles to produce Conj30.

Redundant recoding for Conjugation 31: Based on the above results, the 16.2 kb (*kan*^{R}-oriT/*tolC*) and 61.5 kb (*gent*^{R}/*spec*^{R}) conjugal junctions originally planned for Conjugation 31 were unlikely to accommodate transfer of 1/2 of the genome. Therefore, prior to attempting Conjugation 31, we transferred 1029 kb of recoded genome from Conj26 into Conj30 (C30.5, Fig. 2) so that this region would be redundantly recoded in both parental strains for Conj31. Additionally, to decrease the chance of a failed *tolC* selection, we inserted *tolC* into the donor strain so that we could positively select on SDS. Thus, Conjugation 31 was successfully performed using a 56.6 kb oriT/*tolC* junction and a 463 kb *gentR*/*specR* junction (Fig. 2, Fig. S14).

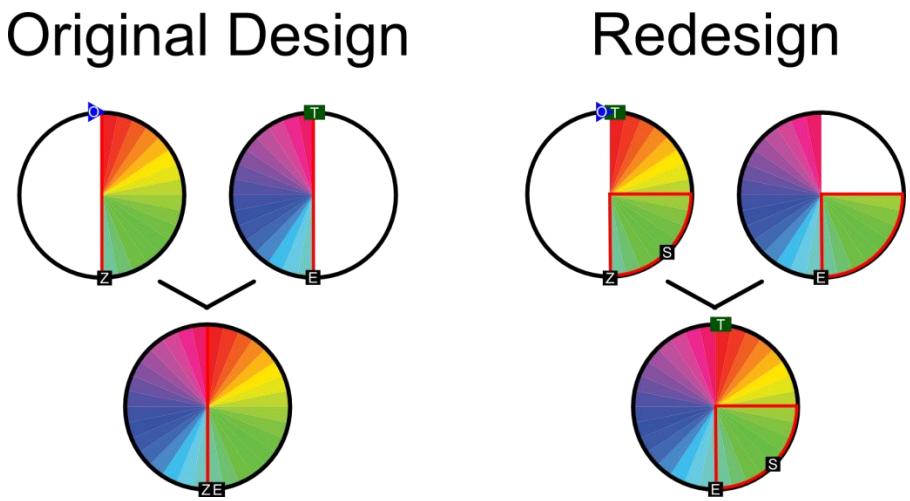


Fig. S14. Redundant recoding for Conjugation 31. Conj29 and Conj30 only provide 16.2 kb and 61.5 kb of homology for their *kan*^{R}-oriT/*tolC* and *gent*^{R}/*spec*^{R} junctions, respectively. Therefore, we moved the *kan*^{R}-oriT/*tolC* junction and created Conj30.5, which has the third quadrant of the genome redundantly recoded. This provides a 56.6 kb oriT/*tolC* junction and a 463 kb *gentR*/*specR* junction. Additionally, we used *tolC* in the donor genome to permit SDS selection, which has a lower escape rate than ColE1 selection. Colored wedges represent recoded segments containing 10 UAG→UAA conversions, O = *kan*^{R}-oriT, T = *tolC*, E = *gent*^{R}, S = *spec*^{R}.

Removing remaining UAG codons: After the final conjugation, 3 selectable markers (*tolC*, *gent*^{R}, and *spec*^{R}) and 11 UAG codons (Table S21) from the original design of 314 UAGs were retained. We used *tolC* to delete these undesired selectable markers and MAGE to convert the UAG codons to UAA.

Table S21. UAG codons that were retained in Conj31 after CAGE

Gene	UAG Pos	UAG ID	Trans Dir	Replichore	Why UAG
b4273	4497523	3.10	+	1	Lost during Conjugation 2
ybaA	476249	8.1	+	1	Lost during Conjugation 4
sucB	761962	9.10	+	1	Lost during Conjugation 26
ybiR	853988	10.6	+	1	Lost during Conjugation 19
yceF	1145234	11.10	-	1	Lost during Conjugation 6
ydfP	1637054	15.9	-	2	Lost during Conjugation 20
rzpQ	1647065	15.10	+	2	Lost during Conjugation 20
yegW	2180057	20.8	-	2	Reverted by yegV oligo
ascB	2840436	24.10	+	2	Reverted by Z.24.25 recombination
hycI	2840595	25.1	-	2	Lost during Conjugation 30
atpE	3918973	32.10	-	2	Lost during Conjugation 16

Upon closer inspection, we observed that *yegV* and *yegW* had overlapping, convergent open reading frames so that MAGE oligos individually converting the UAG of one gene would revert the UAG of the other gene. Therefore, we designed a MAGE oligo that would simultaneously convert the UAGs of both *yegV* and *yegW* (Fig. S15). Such design clashes will become more common as genome designs incorporate more mutations in closer proximity.

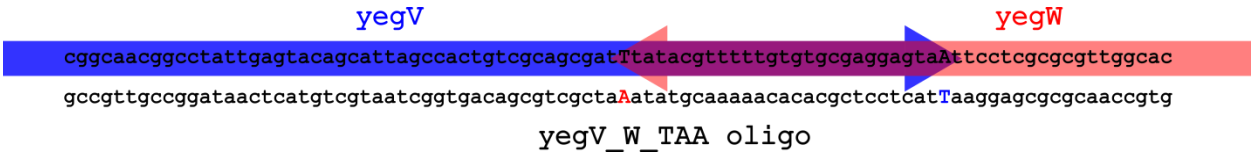


Fig. S15. MAGE oligo simultaneously converting UAGs of convergently overlapping yegV and yegW genes. The top sequence is the desired genomic sequence (shown 5' → 3'). The bottom sequence is the MAGE oligo that simultaneously converts the UAG codons in *yegV* and *yegW* (shown 3' → 5').

Removing new UAG codons: Genome annotations and interpretations are incomplete and are continually being updated based on empirical results. We initially designed the MAGE oligos based on 314 predicted UAGs (NCBI, NC_000913, Feb. 07, 2006). However, we subsequently identified 8 additional UAGs from the Apr. 24, 2007 NCBI update. Further analysis of the ecocyc.org (*45*) database (Mar. 19, 2012) identified 3 more UAGs (Table S22). Ecocyc also flagged 4 previously identified putative UAGs as part of phantom genes (sequences previously annotated, but that are not genes, Table S23). We efficiently converted the remaining 11 UAGs via MAGE. However, the fact that we needed to update our design highlights a central problem with using incomplete data to design genomes. Such trivial design changes distributed throughout the genome would require significant effort to implement via whole genome synthesis.

Table S22. UAG codons that were not targeted in the original design

Gene	UAG Pos	Trans Dir	Replichore	Identified
yafF	239378	+	1	NC_000913 (NCBI) 02/07/2006 update
yliI	879080	+	1	NC_000913 (NCBI) 02/07/2006 update
ymdF	1067477	+	1	NC_000913 (NCBI) 04/24/2007 update
yheV	3476614	-	2	NC_000913 (NCBI) 04/24/2007 update
yjbS	4266832	-	1	NC_000913 (NCBI) 04/24/2007 update
yjdO	4351104	+	1	NC_000913 (NCBI) 04/24/2007 update
insB	4517037	+	1	NC_000913 (NCBI) 04/24/2007 update
ytjA	4610312	+	1	NC_000913 (NCBI) 04/24/2007 update
mntS	852092	-	1	Ecocyc.org flat file 03/19/2012
yahH	339313	+	1	Ecocyc.org flat file 03/19/2012
ykgN	279248	-	1	Ecocyc.org flat file 03/19/2012

Table S23. UAG codons found in genes re-annotated as phantom

Gene	UAG Pos	UAG ID	Trans Dir	Replichore
b4250	4481621	3.8	+	1
b1354	1426575	14.2	+	1
b1367	1433519	14.4	+	1
b2191	2296256	21.5	+	2

Cleanly removing RF1 without impairing fitness: The complete deletion of *prfA* also removes the ribosomal binding site (RBS) from the overlapping essential gene, *prmC*. Therefore, we tested three *prfA* deletion cassettes (Δ*prfA*::*spec*^{R}, Δ*prfA*::*tolC*, and a clean deletion) to remove the ability of UAG to terminate translation. While *spec*^{R} contains an appropriately placed RBS, the C-terminus of *tolC* is C/T rich, so we added a synthetic RBS to ensure adequate *prmC* expression. Finally, we cleanly deleted Δ*prfA:tolC* while retaining the synthetic RBS for *prmC*. All three designs produced viable Δ*prfA* strains without significantly impairing fitness (Fig. 3).

>Δ*prfA*::*spec*^{R}

ctggagtaacagtacatcattttctttttttacagggtgcatttacgcctatgcgctcacgcaactggtccagaaccttgaccgaacgcagcggt
ggtaacggcgcagtggcggttttcatggcttgttatgactgttttttttggggtacagtctatgcctcgggcatccaagcagcaagcgcgttacg
ccgtgggtcgatgtttgatgttatggagcagcaacgatgttacgcagcagggcagtcgccctaaaacaaagttaaacatcatgagggaagc
ggtgatcgccgaagtatcgactcaactatcagaggtagttggcgtcatcgagcgccatctcgaaccgacgttgctggccgtacatttgtacg
gctccgcagtggatggcggcctgaagccacacagtgatattgatttgctggttacggtgaccgtaaggcttgatgaaacaacgcggcgagc
tttgatcaacgacctttttggaaacttcggcttcccctggagagagcgagattctccgcgctgtagaagtcaccattgttgtgcacgacgacatc
attccgtggcgttatccagctaagcgcgaactgcaatttggagaatggcagcgcaatgacattcttgcaggtatcttcgagccagccacgatc
gacattgatctggctatcttgctgacaaaagcaagagaacatagcgttgccttggtaggtccagcggcggaggaactctttgatccggttcct
gaacaggatctatttgaggcgctaaatgaaaccttaacgctatggaactcgccgcccgactgggctggcgatgagcgaaatgtagtgcttac
gttgtcccgcatttggtacagcgcagtaaccggcaaaatcgcgccgaaggatgtcgctgccgactgggcaatggagcgcctgccggccc
agtatcagcccgtcatacttgaagctagacaggcttatcttggacaagaagaagatcgcttggcctcgcgcgcagatcagttggaagaattt
gtccactacgtgaaaggcgagatcaccaaggtagtcggcaaataatggaatatcaacactggttacgtgaagcaataagccaacttcaggc
gagc

>Δ*prfA*::*tolC*
ctggagtaacagtacatcattttctttttttacagggtgcatttacgcctatgaagaaattgctccccattcttatcggcctgagcctttctgggttc
agttcgttgagccaggccgagaacctgatgcaagtttatcagcaagcacgccttagtaacccggaattgcgtaagtctgccgccgatcgtga
tgctgcctttgaaaaaattaatgaagcgcgcagtccattactgccacagctaggtttaggtgcagattacacctatagcaacggctaccgcga
cgcgaacggcatcaactctaacgcgaccagtgcgtccttgcagttaactcaatccatttttgatatgtcgaaatggcgtgcgttaacgctgcag
gaaaaagcagcagggattcaggacgtcacgtatcagaccgatcagcaaaccttgatcctcaacaccgcgaccgcttatttcaacgtgttgaa
tgctattgacgttctttcctatacacaggcacaaaaagaagcgatctaccgtcaattagatcaaaccacccaacgttttaacgtgggcctggta
gcgatcaccgacgtgcagaacgcccgcgcacagtacgataccgtgctggcgaacgaagtgaccgcacgtaataaccttgataacgcggt
agagcagctgcgccagatcaccggtaactactatccggaactggctgcgctgaatgtcgaaaactttaaaaccgacaaaccacagccggtt
aacgcgctgctgaaagaagccgaaaaacgcaacctgtcgctgttacaggcacgcttgagccaggacctggcgcgcgagcaaattcgcca
ggcgcaggatggtcacttaccgactctggatttaacggcttctaccgggatttctgacacctcttatagcggttcgaaaacccgtggtgccgct
ggtacccagtatgacgatagcaatatgggccagaacaaagttggcctgagcttctcgctgccgatttatcagggcggaatggttaactcgca
ggtgaaacaggcacagtacaactttgtcggtgccagcgagcaactggaaagtgcccatcgtagcgtcgtgcagaccgtgcgttcctccttc
aacaacattaatgcatctatcagtagcattaacgcctacaaacaagccgtagtttccgctcaaagctcattagacgcgatggaagcgggctac
tcggtcggtacgcgtaccattgttgatgtgttggatgcgaccaccacgttgtacaacgccaagcaagagctggcgaatgcgcgttataacta
cctgattaatcagctgaatattaagtcagctctgggtacgttgaacgagcaggatctgctggcactgaacaatgcgctgagcaaaccggtttc
cactaatccggaaaacgttgcaccgcaaacgccggaacagaatgctattgctgatggttatgcgcctgatagcccggcaccagtcgttcag
caaacatccgcacgcactaccaccagtaacggtcataaccctttccgtaactaagaggaataatggaatatcaacactggttacgtgaagca
ataagccaac

>Clean deletion
gggctggagtaacagtacatcattttctttttttacagggtggaggaggaataatggaatatcaacactggttacgtgaagcaataagcc

D. GRO nomenclature and applications

Although for clarity we have assigned informal names to describe our key recoded strains, we have also developed the following GRO nomenclature: C(F/E,M,A)_I, where C is the number of codons instances changed, F/E is the number of codons completely removed from the full genome (F), or all essential genes (E), M is the number of previously essential codon functions manipulated (*e.g.* release factors, tRNAs, aminoacyl-tRNA synthetases), A is the number of codons reassigned to a new amino acid (A_{wt} is wild type function and A_o is without any assigned function), and I is a descriptive index to differentiate strain variants. For example, C7(E1,M1,A_1)_Δ*prfA*::*spec*^{R}.Δ*mutS*::*tolC* has UAG changed to UAA in all 7 essential genes, has RF1 replaced by *spec*^{R}, incorporates one NSAA at UAG codons, and has *mutS* replaced by *tolC*. Similarly, C321(F1,M1,A_o)_Δ*prfA*.Δ*mutS*::*zeo*^{R}.Δ*tolC* has all 321 known UAG codons changed to UAA, has RF1 cleanly deleted, stalls translation at UAG codons, has *mutS* replaced by *zeo*^{R}, and has *tolC* deleted.

Table S37. Recoded strains and their genotypes

Strain[a]	GRO nomenclature	Essential codons changed[b]	Total codons changed[c]	Previously essential codon functions manipulated[d]	Expected (obs.) UAG translation function[e]
EcNR2	-	0/7	0/321	None	Stop
C0.B*	C0(M1,A_{wt})_Δ*mutS*::*zeo*^{R}.*prfB*	0/7	0/321	*prfB*‡	Stop
C0.B*.ΔA::S	C0(M2,A_o)_B*.Δ*prfA*::*spec*^{R}.Δ*mutS*::*zeo*^{R}.*prfB*	0/7	0/321	*prfB*‡, Δ*prfA*::*spec*^{R}	None (stop*)
C7	C7(E1,A_{wt})_Δ*mutS*::*tolC*	7/7	7/321	None	Stop
C7.ΔA::S	C7(E1,M1,A_o)_Δ*prfA*::*spec*^{R}.Δ*mutS*::*tolC*	7/7	7/321	Δ*prfA*::*spec*^{R}	None (sup)
C13	C13(E1,A_{wt})_Δ*mutS*::*tolC*	7/7	13/321	None	Stop
C13.ΔA::S	C13(E1,M1,A_o)_Δ*prfA*::*spec*^{R}.Δ*mutS*::*tolC*	7/7	13/321	Δ*prfA*::*spec*^{R}	None (sup)
C321	C321(F1,A_{wt})_Δ*mutS*::*zeo*^{R}.Δ*tolC*	7/7	321/321	None	Stop
C321.ΔA::S	C321(F1,M1,A_o)_Δ*prfA*::*spec*^{R}.Δ*mutS*::*zeo*^{R}.Δ*tolC*	7/7	321/321	Δ*prfA*::*spec*^{R}	None (nc)
C321.ΔA::T	C321(F1,M1,A_o)_Δ*prfA*::*tolC*.Δ*mutS*::*zeo*^{R}	7/7	321/321	Δ*prfA*::*tolC*	None (nc)
C321.ΔA	C321(F1,M1,A_o)_Δ*prfA*.Δ*mutS*::*zeo*^{R}.Δ*tolC*	7/7	321/321	Δ*prfA*	None (nc)

[a]All strains are based on EcNR2 (*Escherichia coli* MG1655 Δ*mutS*::*cat* Δ(*ybhB-bioAB*)::[λcI857 N(*cro-ea59*)::*tetR-bla*]) which is mismatch repair deficient (Δ*mutS*) to achieve high frequency allelic replacement; C0 and C321 strains are Δ*mutS*::*zeo*^{R}; C7 and C13 strains are Δ*mutS*::*tolC*; C7, C13, and C321 strains have the endogenous *tolC* deleted, making it available for use as a selectable marker. Spectinomycin resistance (S) or tolC (T) were used to delete *prfA* (A). Bacterial genetic nomenclature describing these strains includes :: (insertion) and Δ (deletion).

[b]Out of a total of 7

[c]Out of a total of 321

[d]*prfA* encodes RF1, terminating UAG and UAA; *prfB* encodes RF2, terminating UGA and UAA; *prfB*‡ = RF2 variant (T246A, A293E, and removed frameshift) exhibiting enhanced UAA termination (*16*) and weak UAG termination (*17*).

[e]Observed translation function: Stop = expected UAG termination; stop* = weak UAG termination from RF2 variant; sup = strong selection for UAG suppressor mutations; nc = near-cognate suppression in the absence of all other UAG translation function.

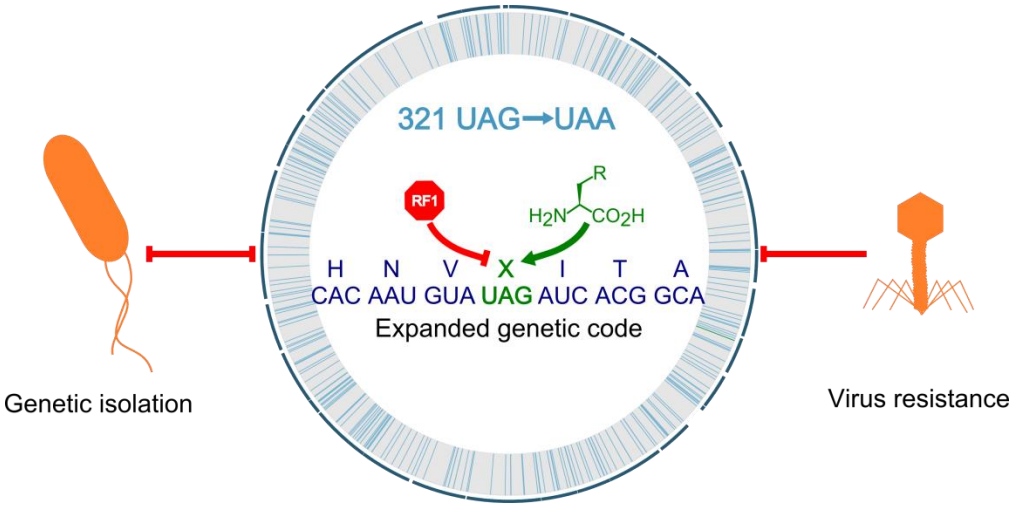


Fig. S1. Properties of genomically recoded organisms (GROs). We have removed all 321 UAG codons (blue radial lines) and release factor 1 (RF1; terminates translation at UAG) from *E. coli* MG1655. Our recoded strain provides a dedicated UAG codon for plug-and-play translation of nonstandard amino acids (NSAAs). This enables efficient expression of GFP variants containing several UAG codons, provides increased resistance to bacteriophage T7 infection, and establishes a basis for the genetic isolation of GROs.

E. Partial recoding strategies for reassigning UAG codon function

Three hypotheses have attempted to explain why RF1-mediated UAG termination is essential: (i) inadequate RF2-mediated UAA termination (*16, 54*), (ii) essential gene (Table S5) loss of function due to UAG read-through (*15*), and/or (iii) translational stalling in the absence of UAG function (*15*). The UAG codon appears to tolerate sense suppression at the majority of UAG codons (*15, 16, 54*). As reported by Mukai et al. (*15*) and illustrated in Fig. S16, this appears to be an evolutionary feature, given that UAA and UGA stop codons are overrepresented at short distances triplets downstream of UAG codons. We analyzed GO terms using <http://biit.cs.ut.ee/gprofiler/index.cgi>, but we observed no enrichment for any specific component, process, or function.

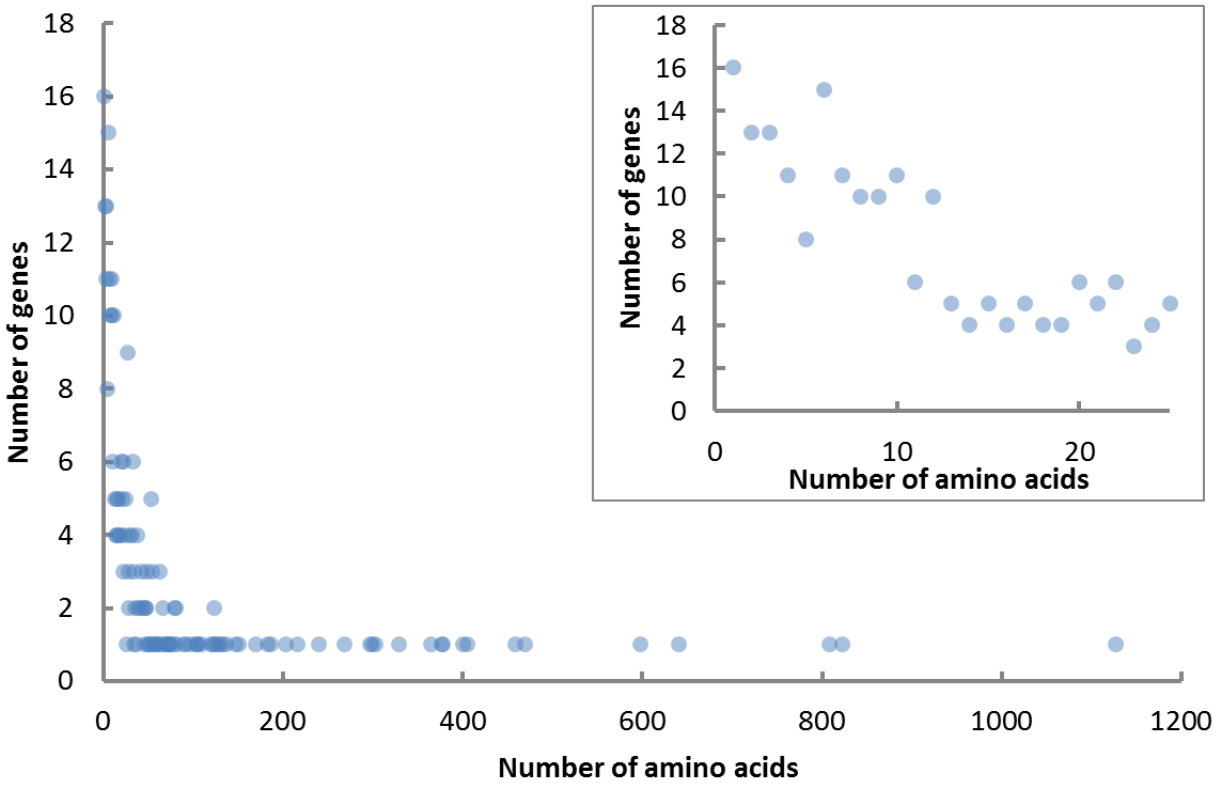


Fig. S16. Distribution of the number of amino acids added to the C-terminus of genes as a result of UAG read-through. The inset is zoomed in on the first 20 triplets following the UAG codon.

Table S5. Essential and important genes terminating with UAG.

Gene	Strand	Gene size (bp)	MG1655 UAG coordinate	Essential[a]	Function[b]	Deletion phenotype[c]
murF	+	1358	96008	Yes	Peptidoglycan biosynthesis	Essential
lolA	+	611	937206	Yes	Periplasmic lipoprotein chaperone	Essential
lpxK	+	986	968575	Yes	LPS biosynthesis	Essential
hemA	+	1256	1264193	Yes	Porphyrin biosynthesis	Essential
hda	-	746	2616097	Yes	Replication initiation regulation	Barely affected (false negative)
mreC	-	1103	3396897	Yes	Peptidoglycan biosynthesis and chromosome segregation	Essential
coaD	+	479	3808327	Yes	Coenzyme A biosynthesis	Essential
yafF	+	188	239378	No	Conserved protein, pseudogene	Essential
pgpA	+	518	436331	No	Phospholipid processing	Moderate fitness decrease
sucB	+	1217	761962	No	Energy regeneration	Major fitness decrease
fabH	+	953	1148935	No	Fatty acid biosynthesis	Major fitness decrease
fliN	+	413	2019525	No	Component of flagellar motor's switch complex	Moderate fitness decrease
atpE	-	239	3918973	No	Energy regeneration	Major fitness decrease

[a] Essentiality was from the PEC database <http://www.shigen.nig.ac.jp/ecoli/pec/index.jsp> (*55*). Genes in white are essential genes with their UAG replaced in C7.ΔA::S. Genes in gray are additional genes with their UAG replaced in C13.ΔA::S.

[b] Gene functions were referenced from <http://www.ecocyc.org> (*45*).

[c] The deletion phenotype was based on results from the Keio collection (*56*).

F. Analysis of MAGE and CAGE

Doublings

Our recoded strain construction was performed in an EcNR2 background (*Escherichia coli* MG1655 Δ*mutS*::*cat* Δ(*ybhB-bioAB*)::[λcI857 N(*cro-ea59*)::*tetR-bla*]), which is defective for mismatch repair. While this background permits efficient allele replacement, it also increases the transition mutation rate ~100 fold. Therefore, continued culturing introduces additional diversity due to spontaneous mutagenesis, which can provide beneficial mutations that compensate for unforeseen genome design flaws. Additionally, these mutations can introduce deleterious mutations that introduce auxotrophies and slow growth, especially when diverse populations are forced through monoclonal bottlenecks (*57*). Although, we have confirmed that the C321.ΔA strains are not auxotrophic, off-target mutagenesis probably underlies their reduced fitness. Therefore, we have calculated the approximate number of doublings for each genome manipulation used in the construction of construction C321.ΔA. Using this information, we estimate the maximum number of doublings during strain construction, the maximum number of doublings that would be expected if we repeated our strain construction, and the maximum number of doublings that would be expected if we improved our strategy by using CoS-MAGE (*14*) to replace 40 UAGs per strain before commencing CAGE. After an estimated 7340 doublings, the 305 off-target mutations detected in C321.ΔA suggests net mutation rate of 9E-9 mutations/bp/doubling, which is consistent with a *mutS*$^-$ phenotype (*58*).

MAGE

Step	**Divisions per**	**Number**	**Cell divisions**
MAGE cycles	6	25	150
o/n growths	15	6	90
Re-dilution	5	6	30
Colony/plating	30	2	60
Outgrowth	12	3	36
Dilute, re-grow, freeze	6	1	6
MAGE total			**372**

Selectable marker dsDNA recombinations

Step	**Cell divisions per repetition**
o/n growths	10
Outgrowth, mid-log	10
Induce @ 42, 15 min	0
Electroporation	0
Recover 1 hour	1
Colony/plating	30
colony outgrowth, mid-log	10
Dilute, re-grow, freeze	6
Divisions/Recombination	67
Total	134

Oligo-mediated tolC deletion

Step	Cell divisions per repetition
o/n growths	10
Outgrowth, mid-log	10
Induce @ 42, 15 min	0
Electroporation	0
Recover to stationary	10
Dilute 1/100, outgrowth, mid-log	6
Dilute 1/100, colE1 selection	16[a]
Colony/plating	30
colony outgrowth, mid-log	10
Dilute, re-grow, freeze	6
Divisions/Recombination	98
Total	196

[a]Assumes 1E-3 frequency of tolC deletion

Final MAGE (Conj31->C321.ΔA

Step	Divisions per	Number	Cell divisions
MAGE cycles	6	39	234
o/n growths	15	4	60
Re-dilution	5	14	70
Colony/plating	30	7	210
Outgrowth	12	7	84
Dilute, re-grow, freeze	6	2	12
MAGE total			**670**

CoS-MAGE (off/on cycle):

Step	Cell divisions per repetition
o/n growths	10
Dilute 1/100, outgrowth, mid-log	6
Induce @ 42, 15 min	0
Electroporation (inactivate tolC)	0
Recover for 7 hours, stationary	12
Dilute 1/100, re-growth	6
Dilute 1/100, colE1 selection	14[a]
Colony/plating	30
Outgrowth, mig-log	10
Induce @ 42, 15 min	0
Electroporation (revert tolC)	0
Recover for 3 hours, mid-log	10
Colony/plating	30
Outgrowth	12
Dilute, re-grow, freeze	6

Total	146

[a]Assumes 1% frequency of desired *tolC* genotype

Table S2. Total estimated number of doublings required to reassign UAG

	Actual[a]		**Re-do[b]**		**CoS-MAGE[c]**	
Manipulation	Number	Doublings	Number	Doublings	Number	Doublings
MAGE	n/a	372	n/a	372	n/a	0
CoS-MAGE	0	0	0	0	3	438
dsDNA Recombinations	19	2546	9	1206	8	1072
Conjugations	7	1792	6	1536	3	768
tolC deletions	10	1960	2	392	3	588
Post-assembly MAGE	n/a	670	n/a	0	n/a	0
Total		7340		3506		2866

[a]Estimated maximum number of actual doublings
[b]Estimated maximum number of doublings to repeat C321.ΔA
[c]Estimated maximum number of doublings using CoS-MAGE to convert 40 UAG codons per strain prior to CAGE

G. Analysis of recoded lineage

Cell morphology in the presence or absence of RF1

Given the extreme degree of genome manipulation necessary to remove all native UAG codons, we wanted to confirm that the cell morphology was not changed (*e.g.* cell elongation or a filamentous phenotype, which might indicate stress response or problems with cell division (*59*). We imaged MG1655, EcNR2, C321, and C321.ΔA::S on bright field using a Zeiss Axio Observer Z1 with a 100X oil immersion objective supplemented with a 1.6X internal lens. Cell morphology was consistent across all strains. The slightly shorter cell lengths for C321 and C321.ΔA::S may be because these strains grow more slowly than MG1655 and EcNR2.

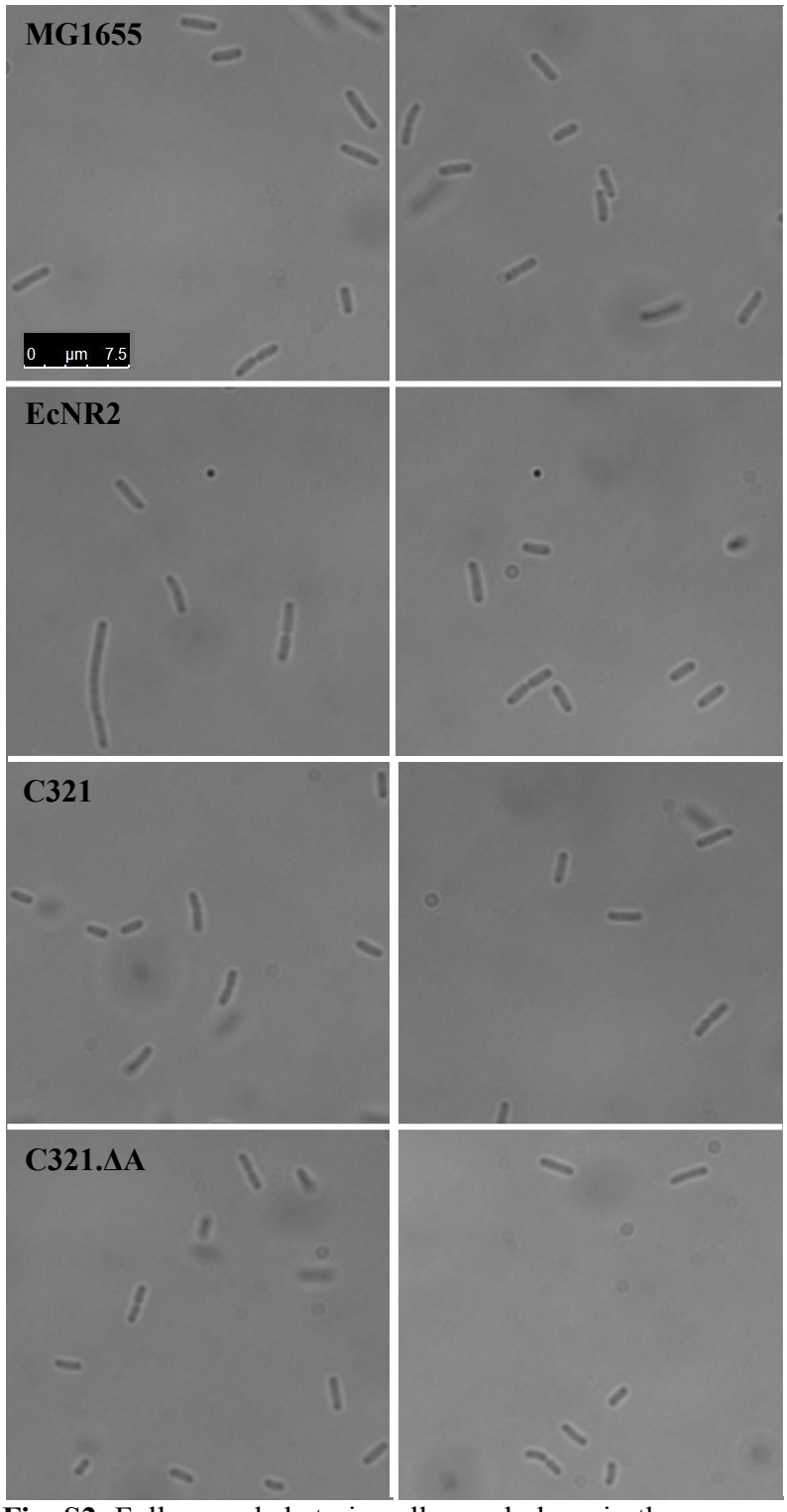


Fig. S2. Fully recoded strain cell morphology in the presence or absence of RF1. Recoding and RF1 removal does not cause cell aggregation or a filamentous phenotype, which are indictors of cell stress.

Doubling times for each strain in recoded lineage

Doubling times were determined for each strain in the C321.ΔA lineage, represented with a heat map in Fig. 2, tabulated in Table S1.

Table S1. Doubling times and Max OD_{600} of recoded genome lineage

Strain	Doubling time (min.)	Doubling time standard deviation	Max OD_{600}	Max OD_{600} standard deviation
MG1655	47	1	1.09	0.01
EcNR2	47	1	1.04	0.03
rEc1	51	2	0.94	0.01
rEc2	49	1	1.02	0.03
rEc3	49	2	1.09	0.02
rEc4	48	1	1.03	0.01
rEc5	49	1	0.90	0.03
rEc6	50	1	0.92	0.02
rEc7	48	1	1.06	0.02
rEc8	49	1	1.00	0.02
rEc9	50	1	1.01	0.01
rEc10	49	1	1.02	0.02
rEc11	47	2	1.02	0.01
rEc12	51	1	1.03	0.02
rEc13	52	2	1.07	0.02
rEc14	49	3	1.05	0.00
rEc21	46	2	1.08	0.01
rEc22	49	2	1.05	0.01
rEc23	48	1	1.05	0.02
rEc24	48	1	0.99	0.01
rEc25	45	2	1.04	0.02
rEc26	48	2	1.10	0.01
rEc27	50	3	1.03	0.01
rEc28	49	1	1.00	0.01
rEc29	44	1	1.01	0.01
rEc30	53	3	1.01	0.02
rEc31	48	1	1.13	0.01
rEc32	49	1	1.06	0.02
Conj1	54	3	1.03	0.04
Conj2	52	2	1.09	0.04
Conj3	71	0	0.59	0.08
Conj4	46	1	1.19	0.02
Conj5	54	1	1.10	0.05

Conj6	57	2	1.07	0.04
Conj7	52	4	1.01	0.03
Conj8	47	2	1.05	0.01
Conj11	86	19	0.73	0.16
Conj12	49	1	1.13	0.02
Conj13	46	2	1.10	0.03
Conj14	47	2	1.14	0.01
Conj15	90	31	0.78	0.32
Conj16	49	1	1.05	0.13
Conj17	50	1	1.02	0.03
Conj18	56	3	1.02	0.01
Conj19	54	1	1.03	0.04
Conj20	50	2	1.01	0.01
Conj21[a]	54	4	1.16	0.05
Conj22	55	0	1.06	0.01
Conj23	74	5	1.05	0.02
Conj24	75	5	1.06	0.06
Conj25	56	3	0.97	0.03
Conj26	52	3	1.00	0.02
Conj27	55	1	1.11	0.02
Conj28	66	3	1.01	0.01
Conj29	63	0	0.96	0.03
Conj30	68	4	0.99	0.04
Conj30.5	90	8	0.62	0.13
C321.ΔA[a]	75	1	0.95	0.01

[a] Conj21 and C321.ΔA growth curves were performed separately from the others

Whole-genome sequencing

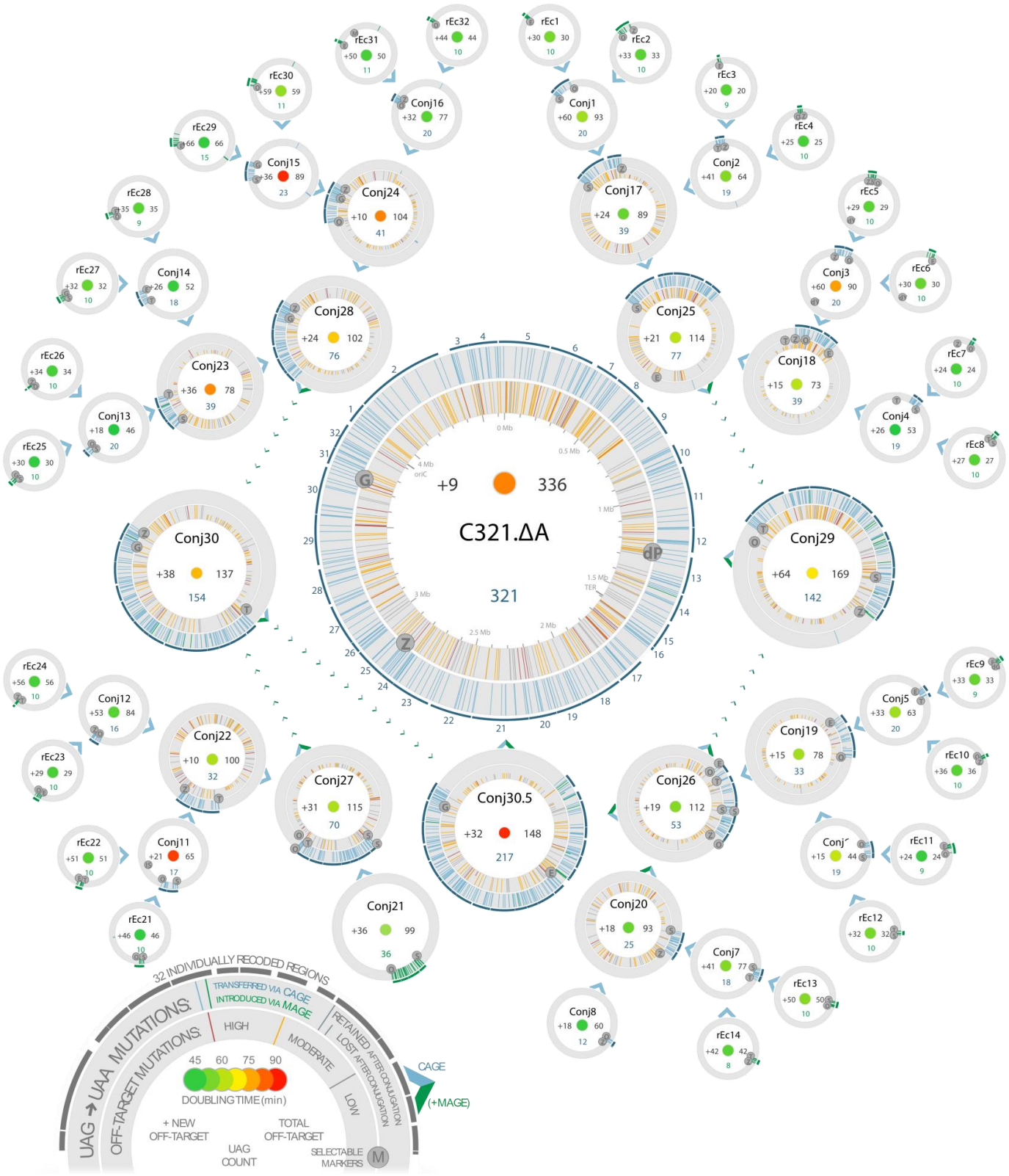


Fig. S3. Construction and analysis of C321.ΔA. The genome was conceptually divided into 32 segments, each containing 10 UAG codons. MAGE (*13*) was used to convert all 10 UAG codons to the synonymous UAA codon in each segment across 32 parallel strains, and CAGE (*11*) was used to hierarchically assemble recoded genome segments into a fully recoded chromosome. Blue arrows point from each strain to its conjugal progeny; blue and green arrows indicate when MAGE was used to convert remaining UAG codons. Strain names (top), total UAGs removed (bottom, Table S3), new off-target mutations (left), total off-target mutations (right, Table S2), and doubling times (green to yellow to red gradient indicates increasing doubling times; Table S1) are reported at the center of each genome. Radial lines in each genome indicate the positions of mutations. The outer circle shows all UAG codons that have been replaced with UAA (green indicates UAG→UAA introduced *via* MAGE and blue indicates UAG→UAA transferred *via* CAGE). The inner circle indicates all off-target mutations acquired during recoded genome construction (color indicates mutation severity according to snpEFF (*34*): gray = low, orange = medium, and red = high). Full lines are mutations that were transferred by CAGE, and half lines are mutations that were lost during conjugation. Approximate positions of conjugal crossovers can be inferred based on which mutations were transferred. A complete list of mutations can be found in Table S4. Gray circles indicate positions of selectable markers immediately before conjugation (O = kan^R-oriT, T = *tolC*, G = *galK*, M = *malK*, S = $spec^R$, E = $gent^R$, Z = zeo^R, dP = Δ*prfA*, IS = *tolC*::IS5). In cases where marker symbols overlapped, they were repositioned for clarity. Strains rEC15 through rEC20 are not included because Conj21 was constructed entirely via CoS-MAGE.

Overview of genome sequencing: Genome sequencing confirmed that all 321 known UAGs have been removed from its genome and that 355 additional mutations were acquired during strain construction (1E-8 mutations/bp/doubling over ~7340 doublings; Fig. S3, Table S2). Only 51 of these unintended mutations were predicted to be highly disruptive by snpEFF (Table S3) (*34*), providing a tractable number of alleles that could be reverted *via* MAGE to potentially improve fitness. Only one *bona fide* IS element transposition event (IS5 in Conj11) and one putative rearrangement (Conj20) were observed, suggesting that structural variants are rare. We also sequenced and characterized the complete CAGE lineage, and observed that the intermediate strains exhibited varying fitness (Fig. 2), as expected for mutator (*i.e.*, Δ*mutS*) strains forced through monoclonal bottlenecks (*57*). Notably, the fitness defects in Conj3, Conj11, Conj15, Conj23, and Conj24 were mitigated in their conjugal progeny even though the UAG→UAA mutations from these strains were inherited (Fig. S3 and Table S4). This suggests that off-target mutations likely caused the observed fitness defects, and that CAGE can eliminate deleterious mutations by preferentially selecting healthy alleles from one parent. Sequencing indicated that MAGE cycling in the rEc strains resulted in an average of 37.4 unintended mutations per strain after ~372 doublings (2E-8 mutations/bp/doubling). Across the entire lineage, we observed only 39 putative MAGE oligonucleotide synthesis errors and 6 putative oligonucleotide mistargeting events resulting in mutations at homologous sequences elsewhere in the genome, rather than the desired target. Therefore, MAGE oligonucleotides do not appear to be a major cause of mutagenesis. Of the remaining 2,225 off-target mutations in the lineage (Table S3), 92% were transitions (A•T→G•C and G•C→A•T) (*58*), suggesting that MutS inactivation underlies most of the unintended mutagenesis (*58*).

Off-target mutations: There are many ways that unintended mutations occur. Mismatch repair deficiency probably accounted for the majority of the 2270 off-target mutations across all 69 strains that were sequenced. Additionally, MAGE oligos can introduce off-target mutations *via* recombination. Oligos that contain chemical synthesis errors can introduce off-target mutations near their desired UAG→UAA mutation, and oligos can mistarget to homologous sequences elsewhere in the genome.

Summary of SNPs: The number of mutations introduced into each strain of the C321.ΔA lineage is summarized in Table S3, including the breakdown of SNP severity according to snpEFF (Table S24) (*34*). All mutations and their predicted severity are tabulated in Table S4. This information could be used to identify off-target mutations that were responsible for the transiently reduced fitness of Conj3, Conj11, Conj15, Conj23, and Conj24, but that were not propagated inherited *via* CAGE. Furthermore, by comparing the severity and location of off-target mutations in C321.ΔA, candidate alleles could be identified for reversion in an attempt to ameliorate its reduced fitness.

Table S3 is attached separately, and contains a summary of SNP types per strain (UAG→UAA mutations, SNPs originating from off-target mutagenesis, SNPs due to oligo-synthesis errors and MAGE oligo mistargeting) and the number of SNPs transferred by each strain during CAGE. This table also summarizes the number of SNPs in each strain according to snpEFF severity (*34*). The categories are as follows:

- SAMPLE = Name of the sample.
- STRAIN_NUM = Identification number for this strain.
- NEW_OT_OLIGO = Number of new off-target SNPs in this strain that fall in regions targetted by MAGE oligos.
- NEW_OT = Total number of new off-target SNPs in this strain.
- NEW_OT_MT = Number of new off-target SNPs in this strain that fall into regions with significant homology to MAGE oligos (indicative of MAGE mistargetting).
- NEW_OT_TS = Number of new off-target SNPs in this strain that are transitions.
- NEW_OT_NOT_OLIGO_TS = Number of new off-target SNPs in this strain that are transitions and not in regions targetted by MAGE oligos.
- NEW_OT_NOT_XFER = Number of new off-target SNPs in this strain that are transferred to the child strain via CAGE.
- TOTAL_OT = Total number of off-target mutations in this strain. TOTAL_MT = Total number of mutations in this strain that fall into regions with significant homology to MAGE oligos (indicative of MAGE mistargetting).
- NEW_AMBER = Number of new UAG to UAA SNPs in this strain.
- TOTAL_AMBER = Total number of UAG to UAA SNPs in this strain.
- EFF_NONE* = Number of SNPs in this strain with no known effect on genic regions.
- EFF_LO* = Number of SNPs in this strain with an effect characterized by snpEFF as "low".
- EFF_MED* = Number of SNPs in this strain with an effect characterized by snpEFF as "moderate".
- EFF_HI* = Number of SNPs in this strain with an effect characterized by snpEFF as "high".

* In cases where SNPs have multiple effects, the highest is reported.

Table S24. Summary of snpEFF types

High	START_LOST FRAME_SHIFT STOP_GAINED STOP_LOST
Moderate	NON_SYNONYMOUS_CODING CODON_CHANGE CODON_INSERTION CODON_CHANGE_PLUS_CODON_INSERTION CODON_DELETION CODON_CHANGE_PLUS_CODON_DELETION
Low	SYNONYMOUS_START NON_SYNONYMOUS_START START_GAINED SYNONYMOUS_CODING SYNONYMOUS_STOP

Table S4 is attached separately, and contains an exhaustive list of all called SNPs per strain, including those that passed the initial Freebayes filtering but not the more stringent downstream filters. The categories are as follows:

- SAMPLE = Name of the strain.
- POS = Chromosome name and position.
- seqnames = Chromosome name.
- start = SNP start position.
- end = SNP end position.
- width = Width of event in bases.
- REF = Reference allele.
- ALT = Alternate allele(s).
- QUAL = SNP quality metric.
- NS = Number of samples in which the SNP was called.
- DP = Total depth across all samples.
- AC = Total number of alternate alleles in called genotypes.
- AF = Estimated allele frequency in the range (0,1].
- RO = Reference allele observations.
- AO = Alternate allele observations.
- AB = Allele balance ratio.
- RUN = Run length (the number of consecutive repeats of the alternate allele in the reference genome).
- DPRA = Alternate allele depth ratio (ratio between ALT SNP calls and WT SNP calls for a given allele and strain)

- TYPE = The type of allele (snp, mnp, ins, del, or complex).
- LEN = Allele length.
- MQM = Mean mapping quality of observed alternate alleles.
- MQMR = Mean mapping quality of observed reference alleles.
- PAIRED = Proportion of observed alternate alleles which are supported by properly paired read fragments.
- PAIREDR = Proportion of observed reference alleles which are supported by properly paired read fragments.
- EFF = Effect string from snpEFF.
- EFF_TYPE = Effect types.
- EFF_SEV = Effect severities.
- EFF_FUNC = Effect functional class.
- EFF_CODON = Effect codon data, if SNP changes a codon.
- EFF_AA = Effect amino acid data, if SNP changes an amino acid.
- EFF_GENE = Gene(s) which this SNP affects.
- EFF_SEV_HIGHEST = The highest severity of all effects for this SNP.
- S_GT = Sample genotype.
- S_GQ = Genotype quality, the Phred-scaled probability of the called genotype.
- S_DP = Sample read depth.
- S_RO = Sample read observations.
- S_QR = Sum of quality of the alternate observations.
- S_QA = Sum of quality of the reference observations.
- S_AO = Alternate allele observation count.
- GT.A = If heterozygous, WT/ALT status 1.
- GT.B = If heterozygous, WT/ALT status 2.
- HET = Is this SNP called as 'heterozygous' (see supplemental methods).
- NC = Is this SNP not called for this genome.
- CALL = Call status (0 for WT, 1+ for ALT).
- VAR = Is this SNP not WT.
- DISPLAY_NAME = Display name of the sample.
- PARENT = Parent strains for this strain.
- CHILD = Child strains for this strain.
- STRAIN = Strain name.
- STRAIN_TYPE = Strain type.
- STRAIN_ID = Strain ID.
- IN_OLIGO = Is this SNP in a region targetted by a MAGE oligo
- AMBER = Is this an UAG to UAA SNP?
- AMBER_COUNT = Number of UAG to UAA mutations made in this SNP.
- IN_CHILD = Number of child strains that received this SNP from this strain.
- IN_PARENT = Number of parent strains that passed this SNP to this strain.
- NO_CALL = Was this SNP not called for this strain?
- NC_COUNT = Number of strains in which this SNP was not called WT/ALT.
- C_COUNT = Number of strains in which this SNP was called WT/ALT.
- NC_PCT = Percentage of strains in which this SNP could not be called.

- INSUFF_CALLS = Flag for whether or not this SNP was called in too few samples.
- AO_TOTAL = Total number of alternate observations across all alternate alleles.
- INSUFF_READS = Flag for whether or not this SNP had too few good quality mapped reads across all samples.
- INSUFF_SAMPLES = Flag for whether or not this SNP was called in too few samples.
- BAD = Flagged if this SNP had either insufficient calls, reads, or called samples.
- ANCESTRAL = Does this SNP occur in MG1655 or EcNR2?
- FILTER = Does this SNP match all the criteria described in the supplemental SNP filtering methods?
- DISPLAY = Should this SNP be displayed in Fig. 2? (FILTER + !ANCESTRAL)
- TS = Is this SNP a transition mutation (A→G, G→A, C→T, or T→C)?

Chemical synthesis errors: We detected 39 off-target mutation events in regions targeted by MAGE oligos in the strains that underwent extensive MAGE cycling (rEc strains and C321). Of these, 16 were mismatches, 23 were deletions, and 0 were insertions. A subset of these mutations may be caused by spontaneous mutagenesis (Δ*mutS*).

MAGE oligo mistargeting: We used blastn (default parameters, http://blast.ncbi.nlm.nih.gov/) to identify 31 MAGE oligos in regions of the genome that shared homology with the intended oligo site (Table S25).

Table S25. Summary of blastn results for potential MAGE oligo mistargeting regions

Oligo ID	Avg. align length	Avg. nt identity	Number of alignments[a]	Total align length
ascB	28.10112	100.0	89	2501
aslB	91.00000	95.6	1	91
b0299	76.80000	98.4	5	384
b0361	58.85714	97.7	7	412
b1228	91.00000	92.3	1	91
b1402	56.25000	98.2	8	450
b1578	56.25000	98.2	8	450
b1996	57.25000	98.3	8	458
b2860	57.50000	98.3	8	460
b3045	59.00000	97.2	8	472
b4273	57.14286	98.3	7	400
b4283	54.66667	96.6	3	164
eaeH	65.00000	98.5	5	325
hda	28.00000	100.0	1	28
hokE	61.00000	95.9	2	122
insB	88.00000	89.8	7	616
rcsC	35.63333	95.9	60	2138
rhsA	77.00000	94.8	2	154
tfaE	90.00000	95.6	1	90
tfaS	90.00000	94.4	1	90
tra5_1	78.00000	98.0	5	390

tra5_2	78.16667	98.3	6	469
tra5_3	76.83333	98.3	6	461
tra5_4	77.66667	97.9	6	466
yafF	35.33333	100.0	3	106
yafL	34.84483	96.5	58	2021
yahH	44.83333	91.1	12	538
ygeP	45.75000	99.2	8	366
yghQ	48.00000	98.5	11	528
yjjV	37.32979	96.4	94	3509
yrhA	76.25000	97.5	4	305

[a] Number of times each oligo aligns at genomic locations other than the desired target location

There were 61 total unique mutations in the regions identified by BLAST. Of the 44 that passed filter, 4 were already present in EcNR2, 16 were on-target UAG→UAA mutations, and 28 were potentially caused by oligo mistargeting. Because some mutations were found in multiple strains, we detected 32 total off-target mutations that shared homology with at least one MAGE oligo. To verify putative mistargeting events, we identified all oligos that satisfied the following requirements: (i) the oligo had been MAGE cycled in the mutated strain in question and (ii) the oligo was homologous to the region in which the mutation occurred. According to these criteria, there were only 6 likely mistargeting events (Table S26).

- There were 5 *bona fide* mistargeting events—putative mistargeting resulted in mutations that matched the oligo sequence.
- There was 1 putative mistargeting event—putative mistargeting resulted in mutation that may have been caused by a chemical synthesis error in the MAGE oligo.
- There were 26 putative false positives:
    - There were 7 putative synthesis errors from proper MAGE oligo targeting that were identified as off-target homologies for other oligos (some oligos that target repetitive elements share similar sequences to each other).
    - There were 9 putative spontaneous mutations (mutations in mistargeting homology regions for MAGE oligos that were not used in the mutated strain).
    - There were 10 heterozygous mutations toward the b1228 oligo sequence in strains that had never been exposed to this oligo (probably an artifact of binary heterozygous SNP calling).

Off-target structural variants: With the possible exception of the Conj 20 and Conj 26 rearrangement described above, we found few instances of structural variants that could be caused by CAGE. This analysis is based on Pindel (*35*) and Breakdancer (*36*) output, which primarily identified the known marker insertion sites. Table S27 and Table S28 report all uncharacterized Pindel breakpoint events and all complete structural events, respectively. All reported events have at least 20 split reads supporting them. Additionally, Table S29 reports all high quality Breakdancer events that are supported by a minimum of 8 reads and have a quality score of at least 20. False positives and false negatives were observed in output from both Pindel and Breakdancer. Therefore, as described in the methods section each structural variant must be confirmed by hand using samtools tview <http://samtools.sourceforge.net/tview.shtml> (*38*).

CAGE removes deleterious alleles: We observed several cases in which CAGE improved fitness in conjugal progeny by allowing preferential inheritance of healthy alleles from one parental strain. This effect is most pronounced during the early stages of CAGE in which the recoded segment is small, and the conjugal junctions are less constrained. However, it diminishes with increasing recoded region sizes, since random mutations become less likely to be removed by chance, and the population of desired genotypes becomes smaller (Table S16) (*53*).

Generating C321.ΔA sequence annotation file (genbank format): We generated an annotated sequence file in Genbank format for C321.ΔA using custom software. This process required us to scrutinize the above SNP and structural variant analysis at a deeper level and resulted in accepting an additional 19 SNPs and 2 deletions that had been previously identified by Freebayes or Pindel or Breakdancer, but which had been triaged based on heuristics intended to remove false positives.

The software takes as input:

- MG1655 reference Genbank with accession number NC_000913 from NCBI
- List of UAG positions in MG1655 (Table S34).
- List of manual fixes which include cassette insertions and deletions (*e.g.* delete *prfA*, insert lambda prophage), as well as the 2 structural variations and 19 SNPs that were hand-validated as described above (Table S35)
- List of remaining off-target variants as called by Freebayes (Table S36)

Our software applies these changes and outputs an annotated file in Genbank format. We then realigned the C321.ΔA fastQ sequencing reads to this genbank file, and re-ran the variant-calling pipeline to identify any discrepancies. By repeating this process iteratively, we were able to identify variants that were previously filtered out due to insufficient evidence based on the MG1655 reference sequence.

Finally, we wrote another custom script to convert our Genbank file into the .sqn submission format required by NCBI. This was done by generating a five-column table format representing the feature annotations which is then fed into the NCBI script tbl2asn. This script performs an additional layer of validation on the annotated sequence according to well-established biological rules, and generates the submission file to be sent to NCBI. The sequence and annotation were submitted to NCBI for release at time of publication.

Current technologies are inadequate:

Modern next-generation sequencing (*e.g.* Illumina HiSeq) now allows for dozens of bacterial strains to be sequenced simultaneously and in a matter of days. Despite the increasing ease of generating raw sequencing data for bacterial genomes, there are a lack of purpose-built tools to deal with this data.

Our current pipeline combines almost a dozen modular tools, many of which are designed for human genome assembly and human population genetics. We know of no existing tools that integrate multi-step genome-scale design, short-read assembly, and SNP and structural variant

detection. The development of such tools would allow for rapid iteration, testing, and troubleshooting of engineered genomes.

Additionally, while the small size of bacterial genomes makes short-read sequencing assembly relatively simple, many genomic variants remain beyond the reach of short read sequencing alone because they occur in duplicated regions (*e.g.* tRNAs, IS elements, highly paralogous genes, etc.). In many cases, short reads align to all copies of such regions with equal likelihood, making it difficult to call SNPs and structural variants in these regions. The creation of genomes with removed or diversified paralogous sequences could be combined with longer sequencing read lengths to produce correct, short-read genome sequences *via* resequencing.

H. Mass spectrometry

We hypothesized that NSAA incorporation was occurring at native UAG positions of unrecoded genomes and we thus aimed to investigate this by directly measuring this effect in the native proteome. This has not been achieved for multiple native genes and previous work relied on tagging methods (altered genes) or plasmid-based single ORFs. We chose an in-depth proteomics approach to provide an unbiased view of the native proteome. This approach comes with a few expected technological limitations of mass spectrometry. Currently, no single proteomics method, or combination of methods, allows for 100% sequence coverage of all proteins. Our shotgun methods, which are slightly better than recent reports (*60*), have an inherent bias towards the detection of higher abundance proteins. We detected over 1,000 proteins (~1/4th of the proteome) and only 40 to 60 of the proteins detected were UAG containing ORFs. The major reason we do not observe more NSAA peptides is that the majority of UAG ORFs are lower in abundance and not in the top 1,000 proteins in the cell. We therefore applied a more robust method described in the SOM that nearly doubled the number of detectable proteins and more than tripled the number of UAG ORFs detected. However, limitations such as depth of peptide covered per ORF, observable peptides with mass spectrometry compatibility properties (such as peptide length, ionization properties, and ideal trypsin cleavage sites), and non-UAG dependent termination sequences are factors that reduce the number of NSAA peptides observed. We also expect that UAG read through and NSAA incorporation would destabilize proteins and reduce their expression below detectable levels. Based on these limitations, we think our list of natural UAG suppression, which is obtained from the most technologically advanced MS methods, underrepresents the total number of natural UAG suppression events. Nevertheless, we observed a highly reproducible sampling of multiple native ORFs that tolerated two distinct types of NSAA insertions. Importantly, these events were erased from the proteome by recoding, a property we confirmed by direct observation of the proteome (Fig. 3D and Fig. S7). We think the native, off target NSAA insertions are relevant at any level and we confirmed that NSAA insertions occur at genes essential for viability and fitness (*e.g. mreC* and *sucB*; Fig. 3C, Table S8, and Table S11).

Table S6. Summary of survey proteomic analysis of strains incorporating pAcF

Strain	OTS	NSAA	Protein ID's[a] #	UAG ORF's[b] #	UAG peptides #	FDR[c] %	FDR[d] %
C0.B*.ΔA::S	none	none	1101	49	0	1.00	1.34
C0.B*.ΔA::S	none	pAcF	1149	53	0	0.86	1.19
C0.B*.ΔA::S	pEVOL-pAcF	none	1130	55	0	0.84	1.19
C0.B*.ΔA::S	pEVOL-pAcF	pAcF	1131	40	3	1.02	1.29
C314.ΔA::S	none	none	1139	60	0	0.85	1.22
C314.ΔA::S	none	pAcF	1138	64	0	0.81	1.22
C314.ΔA::S	pEVOL-pAcF	none	1042	62	0	0.97	1.31
C314.ΔA::S	pEVOL-pAcF	pAcF	1006	55	0	0.96	1.34

[a] Protein ID statistics from Yale Protein Expression Database (YPED)
[b] Identified by searching UAG only DB, retrieved from MASCOT, 5 % False Discovery Rate (FDR)
[c] Peptide matches above identity threshold (YPED)
[d] Peptide matches above homology or identity threshold

Table S7. Summary of in-depth proteomics of strains incorporating pAcF

Strain	OTS	NSAA	Protein ID's[a] #	UAG ORF's[b] #	UAG peptides #	FDR[c] %	FDR[d] %
C0.B*.ΔA::S	pEVOL-pAcF	pAcF	1814	137	9[e]	0.87	2.45
C314.ΔA::S	pEVOL-pAcF	pAcF	1803	163	0	1.05	2.58

[a] Protein ID statistics from YPED
[b] Identified by searching UAG only DB, pulled from MASCOT
[c] Peptide matches above identity threshold (YPED)
[d] Peptide matches above homology or identity threshold
[e] 11 suppressed UAG codons (two UAG codons each in SucB and YbjK peptides)

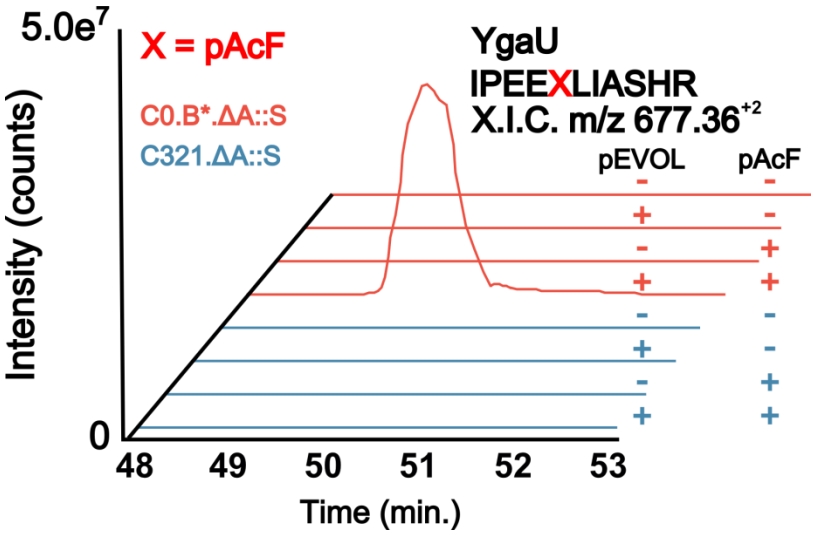


Fig. S7. Extracted ion chromatograms are shown for pAcF incorporation into the YgaU peptide. Peptides containing pAcF were only observed in C0.B*.ΔA::S, and not in C321.ΔA::S, when pEVOL-pAcF was induced and pAcF was supplemented.

Table S8. Summary of identified pAcF containing peptides

Protein	Peptide sequence[a]	Experimental MW	Calculated MW	Delta mass ppm	MASCOT Ion score
FrmR	XLNLLPY	920.5022	920.5007	1.6	16.47
SucB	LLLDVXXFK	1224.6668	1224.6794	10.3	26.44
YbjK	VAGXXISFR	1126.5826	1126.5811	1.3	55.27
MarA	FLHPLNHYNSXLK	1671.8607[b]	1670.8569	600.8[b]	33.03
SpeG	TPGQTLLKPTAQXH	1579.8371	1579.8358	0.8	67.53
YgaU	IPEEXLIASHR	1352.7096	1352.7088	0.6	88.36
LuxS	LQELHIXSVNYLHN	1767.8927	1767.8944	1.0	62.2
LldD	GNAAXSFAPPHPNPLPQGEGTVR	2402.1769	2402.1767	0.1	54.39
IlvA	LMXPLFLR	1077.6050	1077.6045	0.5	30.95

[a] X = pAcF
[b] ^{13}C isotope

Table S9. Summary of all identified proteins with pAcF incorporation at UAG codon(s)

Protein	Description[a]	**C0**.ΔA::S + **pEVOL** + **pAcF**	C314.ΔA::S + **pEVOL** + **pAcF**
FrmR	Regulator protein that represses frmRAB operon	+	-
SucB	Dihydrolipoyltranssuccinase	+	-
YbjK	Predicted DNA-binding transcriptional regulator	+	-
MarA	DNA-binding transcriptional dual activator of multiple antibiotic resistance	+	-
SpeG	Spermidine N1-acetyltransferase	+	-
YgaU	Predicted protein	+	-
LuxS	S-ribosylhomocysteine lyase	+	-
LldD	L-lactate dehydrogenase, FMN-linked	+	-
IlvA	Threonine deaminase	+	-

[a] Gene functions were referenced from <http://www.ecocyc.org> (*45*).

Supplemental Information: Phosphoserine

Table S10a. Summary from the proteomic analysis of the TiO_2 enriched fraction of strains containing Sep-TECH

Strain[a]	OTS	NSAA	Protein ID's[b] #	UAG ORF's[c] #	UAG peptides #	FDR[d] %	FDR[e] %
EcNR2.ΔserB	SepRS/tRNA^{Sep}	Sep	313	17	0	1.15	3.26
EcNR2.ΔserB	SepRS/tRNA^{Sep}	Sep	292	21	0	0.55	1.76
C0.B*.ΔA::S.ΔserB	SepRS/tRNA^{Sep}	Sep	325	23	6	0.64	1.93
C0.B*.ΔA::S.ΔserB	SepRS/tRNA^{Sep}	Sep	249	21	5	0.82	2.42
C13.ΔA::S.ΔserB	SepRS/tRNA^{Sep}	Sep	188	20	4	1.63	3.05
C13.ΔA::S.ΔserB	SepRS/tRNA^{Sep}	Sep	314	16	5	0.92	1.88
C321.ΔA::S.ΔserB	SepRS/tRNA^{Sep}	Sep	227	12	1[f]	1.25	2.45
C321.ΔA::S.ΔserB	SepRS/tRNA^{Sep}	Sep	335	20	1[f]	0.90	2.65

[a] All strains harbored pKD-SepRS-EFsep and pSepT (Sep OTS) and were supplemented with Sep
[b] Protein ID statistics from Yale Protein Expression Database (YPED), results from biological replicates are listed separately
[c] Identified by searching UAG only DB, retrieved from MASCOT, 5 % False Discovery Rate (FDR)
[d] Peptide matches above identity threshold (YPED)
[e] Peptide matches above homology or identity threshold
[f] Carryover levels observed (source of carryover from prior MS run: C13.ΔA::S.ΔserB

We observed only a single NSAA peptide (resulting from native UAG suppression) in two samples from C321.ΔA::S.ΔserB. We loaded 4µg of peptides for these LC-MS runs and followed each run with 2 different types of blank runs designed to clean the LC column. However, we still observed a small amount of carryover, after the two blanks that introduced a small amount of a phosphoserine peptide into the C321.ΔA::S.ΔserB sample from the previous C13.ΔA::S.ΔserB run. We re-ran the set of 4 samples at 1ug loads with the same blank runs and saw no detectable carryover (i.e. this eliminated the detection of the single carryover phosphopeptide from the C321.ΔA::S.ΔserB sample).

Table S10b. Summary from the proteomic analysis of the TiO_2 enriched fraction of strains containing Sep-TECH

Strain[a]	OTS	NSAA	Protein ID's[b] #	UAG ORF's[c] #	TAG peptides #	FDR[d] %	FDR[e] %
EcNR2.ΔserB	SepRS/tRNA^{Sep}	Sep	249	7	0	0.82	2.42
C0.B*.ΔA::S.ΔserB	SepRS/tRNA^{Sep}	Sep	202	9	3	0.29	2.43
C13.ΔA::S.ΔserB	SepRS/tRNA^{Sep}	Sep	188	12	2	1.63	3.05
C321.ΔA::S.ΔserB	SepRS/tRNA^{Sep}	Sep	198	6	0	0.88	1.96

[a] All strains harbored pKD-SepRS-EFsep and pSepT (Sep OTS) and were supplemented with Sep
[b] Protein ID statistics from Yale Protein Expression Database (YPED), results from biological replicates are listed separately
[c] Identified by searching UAG only DB, retrieved from MASCOT, 5 % False Discovery Rate (FDR)
[d] Peptide matches above identity threshold (YPED)
[e] Peptide matches above homology or identity threshold

Table S11. Summary of Sep-containing peptides identified by proteomics from two biological replicates each

Protein	Peptide sequence[a]	Experimental MW	Calculated MW	Delta mass ppm	MASCOT Ion score
LuxS	LQELHIXSVNYLHN	1745.8160	1745.8138	1.3	45.52
SpeG	TPGQTLLKPTAQXH	1557.7579	1557.7552	1.7	76.26
RlpA	LQTEAQLQSFITTAQXR	2000.9606	2000.9568	1.9	58.17
MreC	APGGQXWR	937.3808	937.3807	0.1	39.2
Nei	FGAXVEINR	1071.4735	1071.4750	1.4	53.06
LldD	GNAASXFAPPHPNPLPQGEGTVR	2380.1013	2380.0961	2.2	43.21
YhbW	EELLGXCVLTR	1355.6204	1355.6156	3.5	39.57
LpxK	LLTQLTLLASGNXLR	1678.9069	1678.9019	3.0	35.78

[a] X = Sep

Table S12. Summary of all identified proteins with Sep incorporation at an amber stop codon

Protein	Description[a]	EcNR2.ΔserB + OTS[b]	C0.B*.ΔA::S.ΔserB + OTS[b]	C13.ΔA::S.ΔserB + OTS[b]	C321.ΔA::S.ΔserB + OTS[b]
LuxS	S-ribosylhomocysteine lyase	-	+	+	+[c]
SpeG	Spermidine N1-acetyltransferase	-	+	+	-
RlpA	Septal ring protein, suppressor of prc, minor lipoprotein	-	+	+	-
MreC	Cell wall structural complex MreBCD transmembrane component MreC	-	+	-	-
Nei	Endonuclease VIII/ 5-formyluracil/5-hydroxymethyluracil DNA glycosylase	-	+	+	-
LldD	L-lactate dehydrogenase, FMN-linked	-	+	+	-
YhbW	Predicted enzyme	-	+	-	-
LpxK	Lipid A 4'kinase	-	+	-	-

[a] Gene functions were referenced from <http://www.ecocyc.org> (*45*).
[b] OTS = pKD-SepRS-EFsep and pSepT
[c] Contaminant levels observed (source of contamination from prior MS run: C13.ΔA::S.ΔserB)

I. NSAA incorporation

One of the main goals of reassigning the genetic code is to provide a dedicated channel for plug-and-play incorporation of NSAAs. To this end, we have created a robust chassis completely lacking UAG function, which is capable of accepting orthogonal aaRS/tRNA pairs. We have shown that the only known strategy to completely abolish UAG function is to remove all instances of UAG from the genome and then delete RF1. We have verified previous reports (*17, 54*) that the RF2 variant (frameshift removed, T246A, A293E) can permit RF1 deletion, but also weakly terminates at UAG codons (Fig. 3B). Additionally, NSAA incorporation in these strains is highly toxic ((*54*) and Fig. 3) probably because it outcompetes termination in some essential genes. This effect is particularly apparent upon outgrowth from overnight expression of pAcF and pAzF (Fig. S5). In contrast, removing essential UAGs permits the efficient incorporation of NSAAs, but plug-and-play UAG reassignment is difficult because UAG function cannot be abolished in these strains (new UAG function must be introduced prior to RF1 deletion (*15, 17*)). Although we were able to delete RF1 without introducing a suppressor in C7.ΔA::S and C13.ΔA::S, both strains rapidly selected for efficient natural suppression. C321.ΔA::S, C321.ΔA::T, and C321.ΔA were not affected by NSAA expression. All growth curves used for this analysis are in Fig. S17.

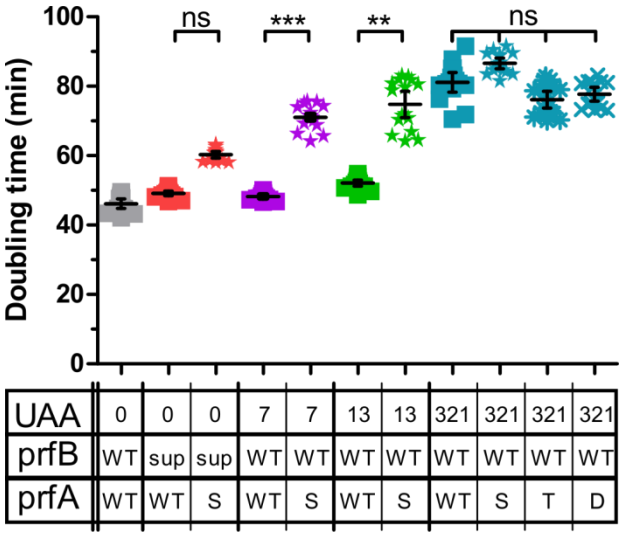


Fig. S4. Doubling times in recoded strains +/- RF1. The number of UAG→UAA conversions are indicated by UAA. RF1 status is denoted as wt *prfA* (WT), Δ*prfA*::*spec*^{R} (S), Δ*prfA*::*tolC* (T), or Δ*prfA* (Δ). RF2 sup indicates a variant (frameshift removed, T246A, A293E) capable of suppressing lethality of RF1 deletion. While C321 has a slower growth rate than the other RF1 strains (probably due to off-target mutagenesis; see discussion in main text), RF1 deletion does not affect fitness. All other strains (C0.prfB*, C7, and C13) exhibited reduced fitness upon RF1 deletion. The gray symbols in the first column correspond to MG1655 (wild type) doubling time.

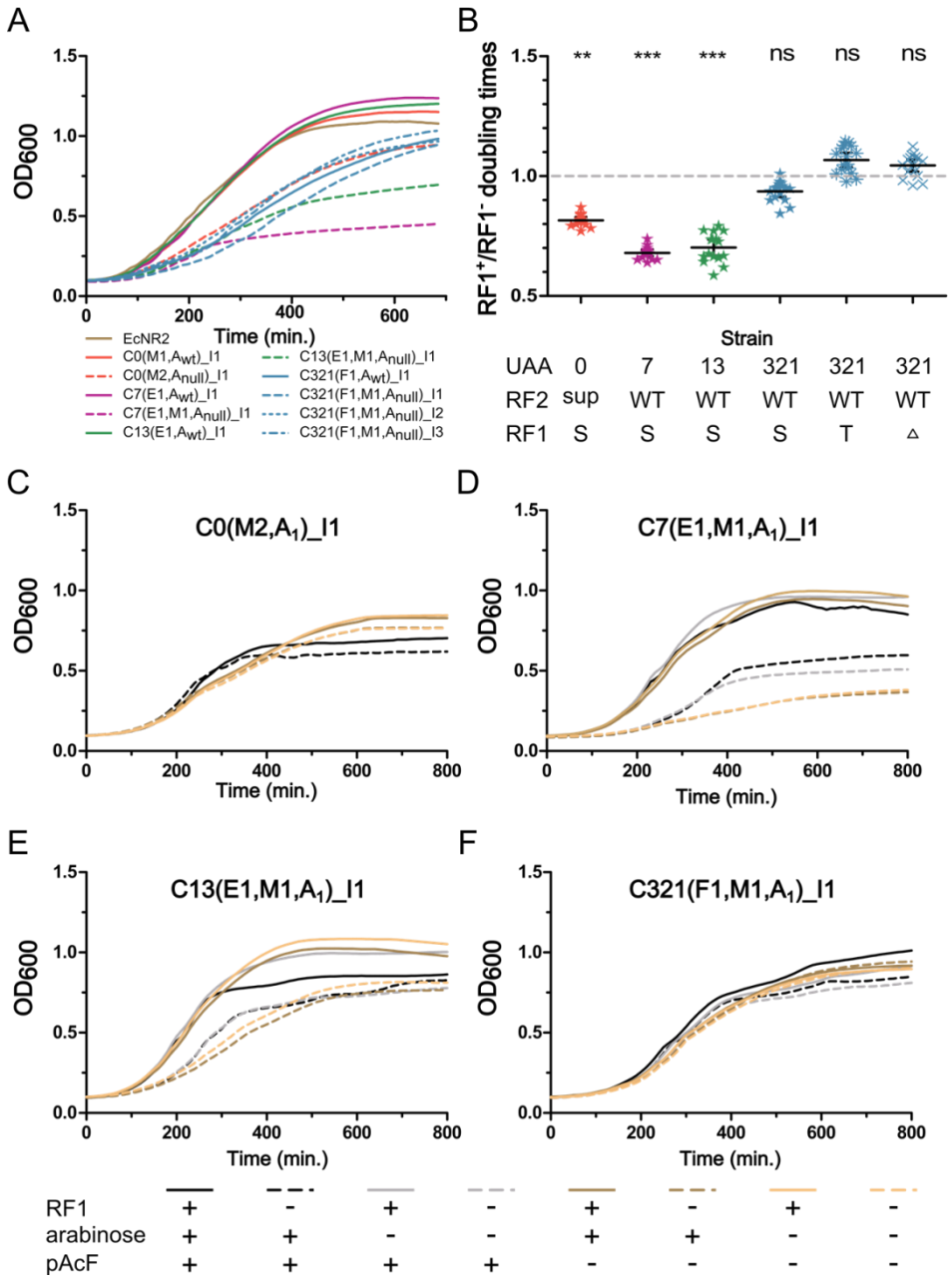


Fig. S17. Native UAGs cause detrimental pleiotropic effects after codon reassignment. RF1 status is denoted as wt *prfA* (WT), Δ*prfA*::*spec*^{R} (S), Δ*prfA*::*tolC* (T), or Δ*prfA* (Δ). RF2 sup indicates a variant (frameshift removed, T246A, A293E) capable of suppressing lethality of RF1 deletion. **(A)** Averaged kinetic growth curves of RF1$^+$ (solid lines) and RF1$^-$ (dashed lines) strains with no UAG suppression. **(B)** Ratios of doubling times for RF1$^+$/RF1$^-$ strains with no aaRS supplemented to reassign UAG (n = 16). Statistical significance was determined using the Kruskal-Wallis test ($p < 0.0001$) followed by Dunn's multiple comparison test to compare each ratio to unity (* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$). RF1 deletion increased doubling time and decreased maximum cell density for RF2 variants and partially recoded strains, but not for fully recoded strains. **(C-F)** Average kinetic growth curves of RF1$^+$ (solid lines) and RF1$^-$ (dashed lines) strains with pEVOL-pAcF expression and pAcF supplementation. The sense suppression of UAG impairs fitness in recoded RF2 variants (natural amino acids are

incorporated and impair fitness in the presence of pEVOL-pAcF even when pAcF is not supplemented) **(C)**, improves fitness in partially recoded strains **(D)** and **(E)**, and does not affect fitness in fully recoded strains **(F)**.

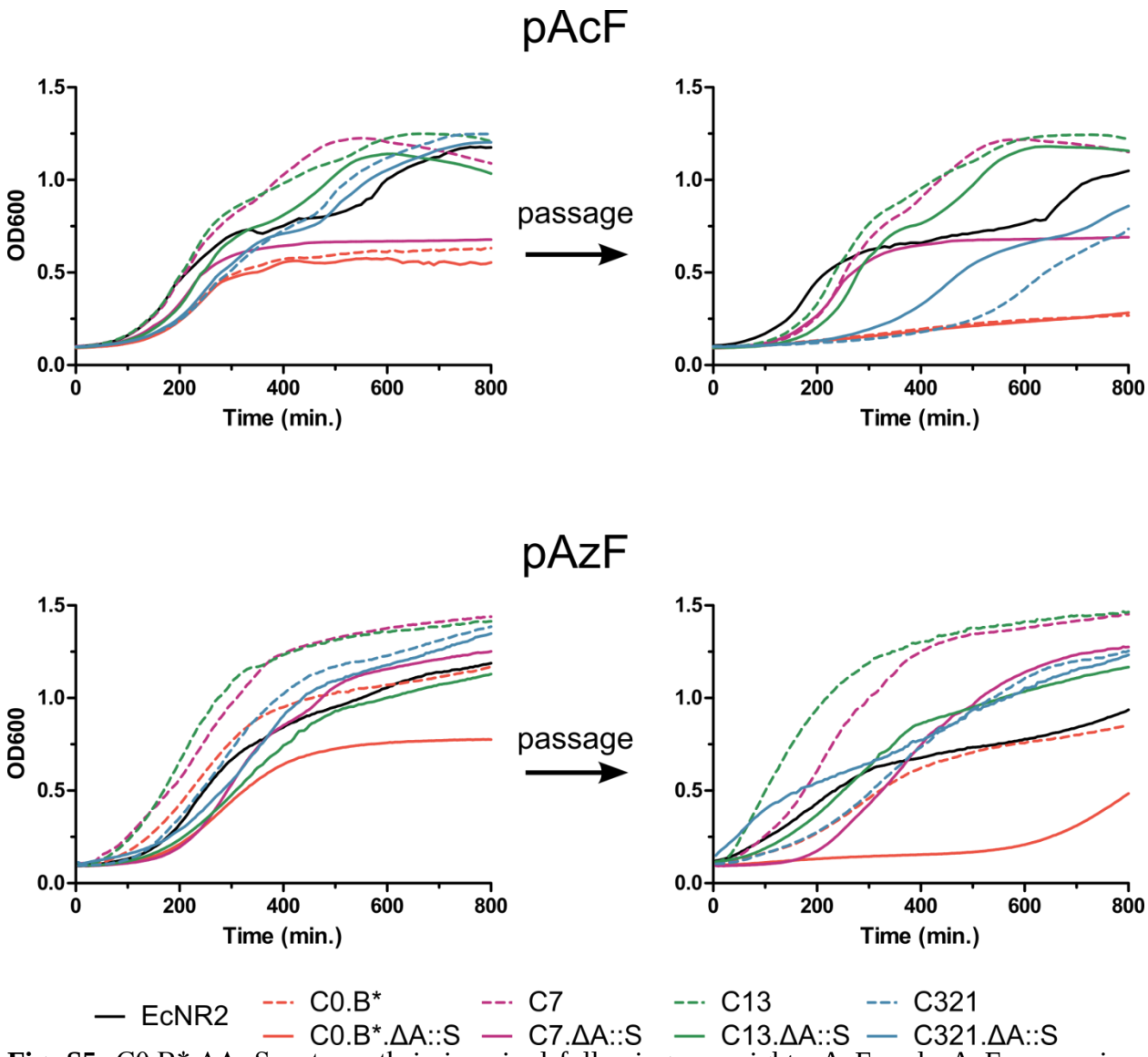


Fig. S5. C0.B*.ΔA::S outgrowth is impaired following overnight pAcF and pAzF expression. Overnight cultures were grown in LB^L supplemented with chloramphenicol (pEVOL maintenance), arabinose (induces the aaRS), and NSAA. After 16 hours of growth, these cultures were passaged into identical media. Growth at 34°C was monitored *via* OD_{600} readings at 10-minute (pAcF) or 5-minute (pAzF) intervals using a biotek H1 plate reader.

GFP expression with UAG reassigned to p-acetylphenylalanine (pAcF)

For each recoded strain, three GFP reporters (0UAG, 1UAG, and 3UAG) were expressed in the presence and absence of pAcF, pAzF, and NapA. Fig. S6 reports the raw fluorescence for each

strain, amino acid, and reporter gene. Therefore, fluorescence readings take into account both expression levels and cell density, which are both relevant with respect to protein overexpression. Regardless of whether this is caused by UAG recoding or off-target mutations that non-specifically increase protein production, C321.ΔA::S consistently produces the highest fluorescence on par with the wt GFP controls after 17 hours of pAcF, pAzF, or NapA expression (Fig. S6). C0.B*.ΔA::S exhibited low fluorescence, while C7.ΔA::S and C13.ΔA::S appeared to read through UAG using canonical amino acids. C321.ΔA::S produced high levels of fluorescence, but only when the relevant NSAA was supplemented. Finally, we note that the 3UAG GFP variant produced higher fluorescence than expected in EcNR2. We verified the EcNR2 genotype, confirmed that the correct plasmid was present, and repeated the transformation of fresh pZE21G-3UAG into fresh EcNR2, but the 3UAG expression was consistently higher than the 1UAG expression in this strain for unknown reasons.

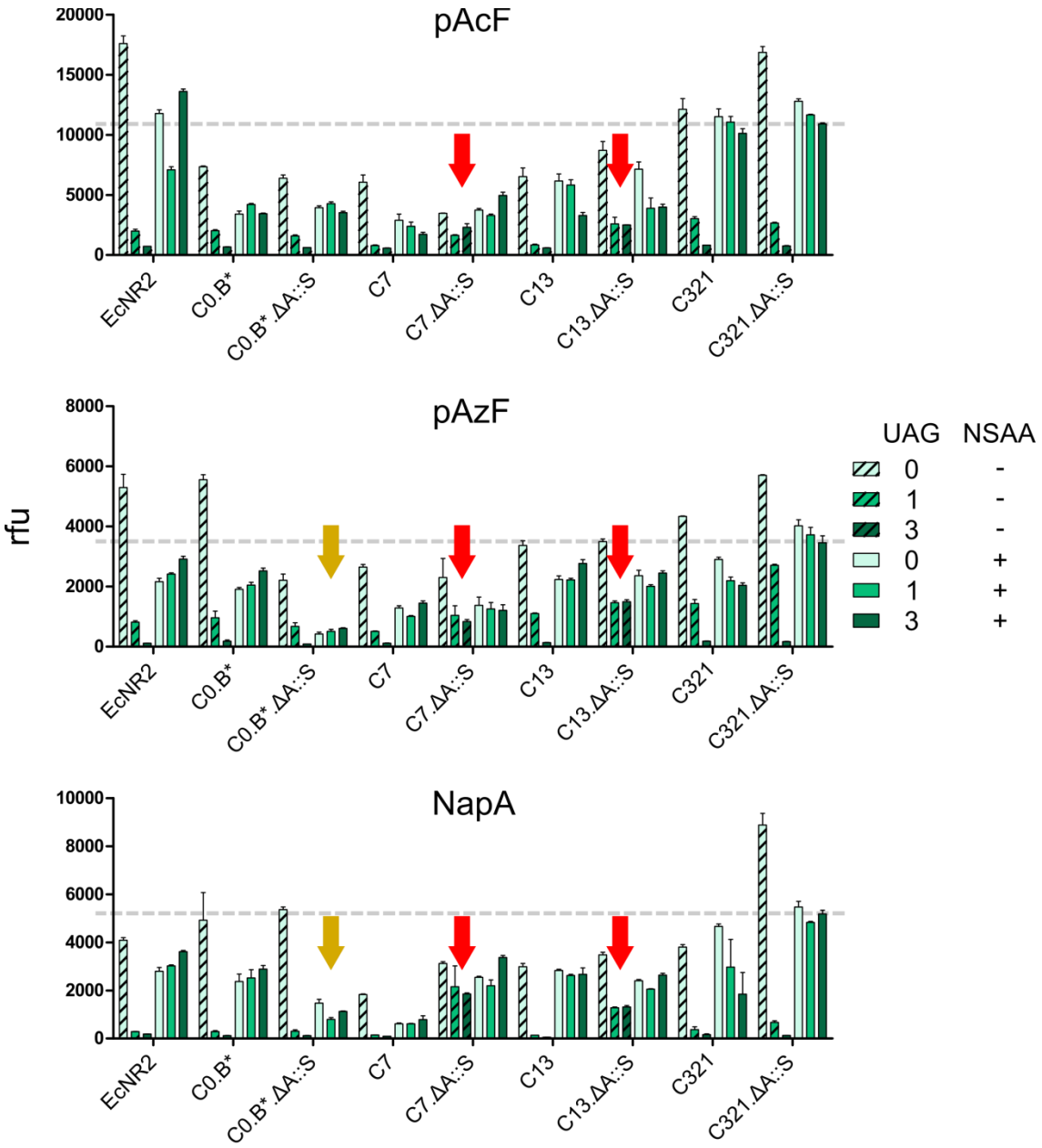


Fig. S6. Complete removal of all native UAGs permits robust NSAA incorporation. Regardless of whether this is caused by UAG recoding or off-target mutations that non-specifically increase protein production, C321.ΔA::S consistently produces the highest fluorescence after 17 hours of pAcF, pAzF, or NapA expression (see gray dashed horizontal lines as a benchmark). We report raw fluorescence without taking OD600 into account, which may contribute to the reduced fluorescence of the partially recodeded strains. We expressed GFP variants containing 0, 1, or 3 UAG codons in our panel of recoded strains (Table 1) with UAG reassigned to pAcF (top panel; using pEVOL-pAcF (*9*)), pAzF (middle panel; using pEVOL-pCNF), and NapA (bottom panel; using pEVOL-pAcF). As evidenced by strong fluorescence for all reporters in the RF1+ strains, the pEVOL expression system is extremely active and strongly outcompetes RF1 in genes containing up to 3 UAG codons. Notably, C0.B*.ΔA::S yielded less fluorescence than its C0.B* precursor (yellow arrows for pAzF and NapA), probably due to toxicity from UAG read-through

in essential genes. In contrast, C7.ΔA::S and C13.ΔA::S produced consistent levels of fluorescence in the 1 UAG and 3 UAG GFP reporters even when NSAAs were not supplemented in the media (red arrows), suggesting that these strains have acquired spontaneous UAG suppressors. Unlike the partially recoded strains, C321.ΔA::S yielded robust fluorescence without acquiring a mutational UAG suppressor. Although near-cognate suppression (*18*) may have resulted in residual expression of 1 UAG GFP, the expression was extremely low for 3 UAG GFP.

Spontaneous UAG suppressors in C7.ΔA::S and C13.ΔA::S

GFP fluorescence (Fig. S6, red arrows) and Western blots (Fig. 3B) indicated that C7.ΔA::S and C13.ΔA::S had spontaneously acquired efficient natural UAG suppressors. Therefore, we investigated this putative natural suppression in C13.ΔA::S *via* LC-MS/MS. To this end, we expressed an E17* GFP variant in C13.ΔA::S and used LC-MS/MS to identify the amino acid(s) incorporated in response to UAG. This analysis found efficient suppression with Lys, Gln, and Tyr (Table S13).

Cells were cultured and lysed as described in the methods section. Cell free extracts were obtained by ultracentrifugation and clarified lysates were applied to Ni-NTA metal affinity resin and purified according to the manufacturer's instructions. Wash buffer contained 50 mM Tris pH 7.5, 500 mM NaCl, 0.5 mM EDTA, 0.5 mM EGTA, 10 mM beta-mercaptoethanol, 50 mM NaF, 1 mM Na_3VO_4 and 5 mM imidazole. Proteins were eluted with buffer containing 500 mM imidazole. Purified protein fractions were subjected to SDS-PAGE electrophoresis, and the gel was stained with Coomassie blue. Protein bands corresponding to the molecular weight of GFP (28.5 kDa) were subjected to in-gel digestion using trypsin as previously described (*61*), and peptides were quantified by UV_{280}. LC-MS was carried out using a 90 min gradient with 100 ng of the digest for each analysis as described above.

Table S13. LC-MS/MS of C13.ΔA::S after appearance of natural suppression

Protein	Peptide sequence[a,b]	Exp. MW Da	Calc. MW Da	Δ m ppm	MASCOT Ion score[c]
GFP E17*	**SKGEELFTGVVPILVK**	1714.9869	1714.9869	0.00	38.81
GFP E17*	**GEELFTGVVPILVK**	1499.8608	1499.8599	0.60	29.9
GFP E17*	**SKGEELFTGVVPILVQLDGDVNGHK**	2651.3786	2651.3807	-0.75	84.97
GFP E17*	**SKGEELFTGVVPILVQLDGDVNGHK**	2650.3889	2650.3967	-2.94	68.35
GFP E17*	**GEELFTGVVPILVQLDGDVNGHK**	2435.2598	2435.2697	-4.02	43.59
GFP E17*	**MSKGEELFTGVVPILVQLDGDVNGHK**	2781.4338	2781.4371	-1.19	34.41
GFP E17*	**SKGEELFTGVVPILVYLDGDVNGHK**	2685.3957	2685.4014	-2.12	29.64

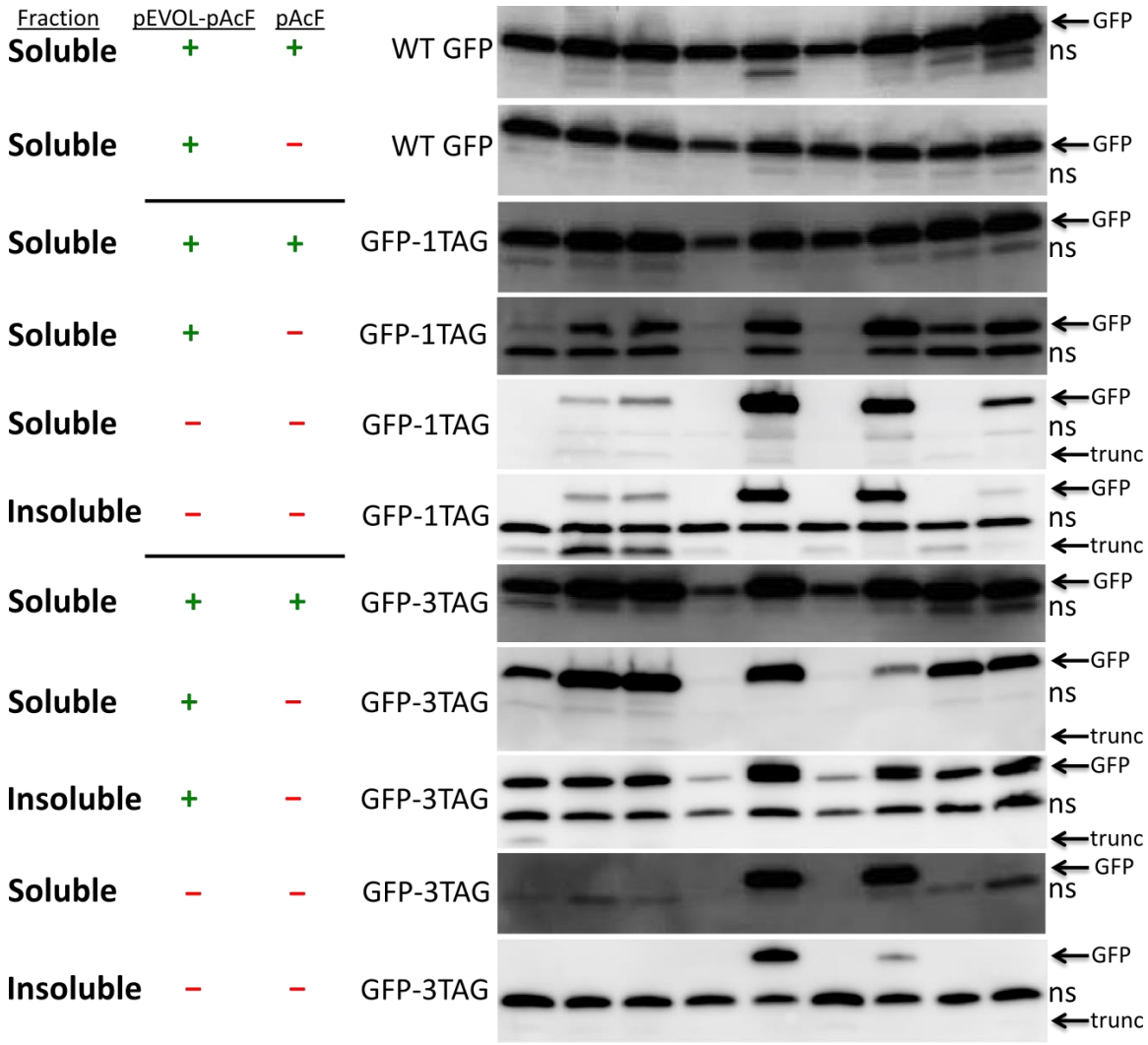
[a]Underlined residues are deamidated.
[b]Lysine (K) insertion adds a unique trypsin cleavage site and produces two unique peptides.
[c]All reported peptides have MASCOT scores above identity.

Western blots: soluble and insoluble fractions

All Western blots not included in Fig. 3B are included below. Because the anti-GFP antibody binds to an epitope between Y45 and Y151, only the 1UAG GFP variant produced truncation products that could be probed. We tested an anti-His antibody that would recognize the N-terminal 6His tag, but the affinity was too low for robust visualization. The soluble fraction primarily contains full-length GFP, while the insoluble fraction primarily contains the truncation products. Our strain is based on MG1655, which, unlike BL21, does not have important proteases (*lon* and *ompT*) inactivated. Therefore, it is possible that the insoluble truncation products are being degraded and underrepresenting the total amount of UAG-mediated termination.

The supernatant Western blots show that C7.ΔA::S and C13.ΔA::S acquired natural suppressors of UAG, that pEVOL-pAcF is capable of incorporating natural amino acids when pAcF is not supplemented, and that near-cognate UAG suppression (UAG recognition by an anticodon that is not CUA) does not cause strong UAG read-through (Fig. S18).

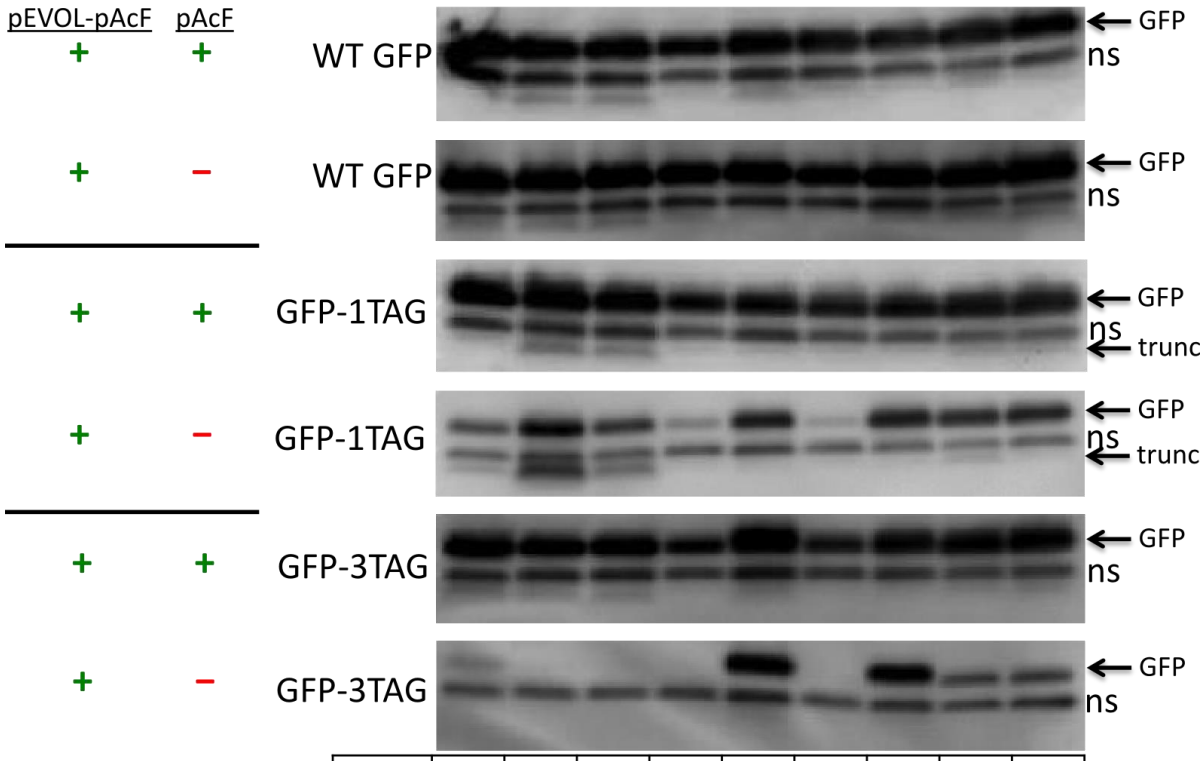
The crude lysate Western blots were performed in an attempt to show the soluble full length GFP and the insoluble truncation products on the same Western blot. Unfortunately, the supernatant overwhelms the insoluble fraction, making it difficult to simultaneous visualization of full-length GFP (soluble) and truncated peptides (insoluble) (Fig. S19).



UAA	0	0	0	7	7	13	13	321	321
RF2	WT	sup	sup	WT	WT	WT	WT	WT	WT
RF1	WT	WT	S	WT	S	WT	S	WT	S

Fig. S18. Western blots of GFP variants in the soluble/insoluble fractions. GFP variants containing 0, 1, 2, or 3 UAG codons (Table S33) were expressed in recoded strains with UAG reassigned to pAcF (strains harbored pEVOL-pAcF (*9*)). Strain genotypes are indicated as follows: RF1 status is denoted as wt *prfA* (WT), Δ*prfA*::*spec*^{R} (S), Δ*prfA*::*tolC* (T), or Δ*prfA* (Δ). RF2 sup indicates a variant (frameshift removed, T246A, A293E) capable of compensating for RF1 deletion. Western blots of the soluble fraction were probed with an anti-GFP antibody that recognizes an N-terminal epitope. The "ns" signifies a non-specific band. Truncation products ("trunc") were present primarily in the insoluble fractions. Truncation products are most visible for the 1UAG variant because our anti-GFP antibody recognizes an epitope that is not translated in the truncated portion of the 3UAG variant (see Table S33 for UAG positions). Still, the 3UAG pellet fractions show faint bands corresponding to the expected size for the 1UAG variant, probably due to read-through at the first UAG. C7.ΔA::S and C13.ΔA::S efficiently produced all variants of GFP regardless of UAG number and pAcF supplementation, suggesting that these

strains have acquired natural suppressors of UAG. Additionally, full-length 1UAG GFP was visible in all strains lacking RF1 when pEVOL-pAcF was expressed even when pAcF was not supplemented, showing that pEVOL-pAcF is also capable of weakly incorporating natural amino acids. When pEVOL-pAcF was not induced (only expression of constitutive gene copy), a small amount of UAG suppression was observed in C0.B*, C0.B*.ΔA::S, and C321.ΔA::S. This suppression may be caused by weaker expression of the constitutive pAcF-RS copy or by near-cognate suppression (*18*). However, no full-length 3UAG was observed in the absence of pEVOL-pAcF induction and pAcF supplementation, indicating that UAG read-through is weak unless UAG is explicitly reassigned to new function.



UAA	0	0	0	7	7	13	13	321	321
RF2	WT	sup	sup	WT	WT	WT	WT	WT	WT
RF1	WT	WT	S	WT	S	WT	S	WT	S

Fig. S19. Western blots of GFP variants in a crude lysate. GFP variants containing 0, 1, 2, or 3 UAG codons (Table S33) were expressed in recoded strains with UAG reassigned to pAcF (strains harbored pEVOL-pAcF (*9*)). Strain genotypes are indicated as follows: RF1 status is denoted as wt *prfA* (WT), Δ*prfA*::*spec*^{R} (S), Δ*prfA*::*tolC* (T), or Δ*prfA* (Δ). RF2 sup indicates a variant (frameshift removed, T246A, A293E) capable of compensating for RF1 deletion. Western blots of crude lysates were probed with an anti-GFP antibody that recognizes an N-terminal epitope. The "ns" signifies a non-specific band. Truncation products ("trunc") were present in the insoluble fraction, but were faint in the Western blots of crude lysates, perhaps due to proteolysis.

J. Increased T7 resistance

Although T4 bacteriophage did not appear to be affected, T7 bacteriophage exhibited reduced fitness in strains lacking UAG function. Further experimentation is required to fully explain this difference in behavior, but previous work may offer some clues. We have considered which genes might be affected by UAG reassignment for each bacteriophage.

T4: 3 of 19 genes terminating with UAG are essential (Table S30a) (*62*).

- Gene 60 (DNA topoisomerase): Gene 60 mRNA contains a short region that must be skipped by translational bypassing in order to produce full length DNA topoisomerase (*63*). A UAG codon plays a role in bypassing efficiency, and UAG stalling may even aid in the translational bypassing.
- Gene 41 (DNA primase/helicase): The C-terminus of gene 41 helicase is involved in Gp59 binding, which is necessary for recombination-dependent replication and for double-strand break repair (*64*). UAG stalling did not significantly impair T4 plaque formation, suggesting that there may have been adequate levels of ribosome rescue by arfA (*65*) and/or yeaJ (*66*) to support normal replication under the conditions tested.
- Gene 15 (Proximal tail sheath stabilizer): Gp15 plays a crucial role in stabilizing the contractile sheath, and forms hexamers that make important contacts with Gp3 and Gp18 (*67*). Hexamer formation occurred even with a C-terminal truncation variant. UAG stalling did not significantly impair T4 plaque formation, suggesting that there may have been adequate levels of ribosome rescue by arfA (*65*) and/or yeaJ (*66*) to support normal tail sheath formation under the conditions tested.

T7: 1 of 6 genes terminating with UAG is essential (Table S30b) (*68*).

- Gene 6 (gp6, T7 exonuclease): Gp6 amber mutants are lysis delayed, suggesting that the C-terminus of gp6 may be important for function (*69*). Therefore, ribosome stalling, tmRNA-mediated degradation, and/or C-terminal extension could decrease gp6 activity in the absence of RF1. This in turn could cause a shortage of nucleotides for phage replication and/or inhibit RNA primer removal, recombination, and concatemer processing during T7 replication (*68*).

Table S30a. UAG terminating genes in bacteriophage T4 (excerpted from (*62*))

Gene	Essential	Function
60	Yes	DNA topoisomerase subunit
modA.3	No	Hypothetical protein
41	Yes	Replicative and recombination DNA primase/helicase
mobB	No	Putative site-specific intron-like DNA endonuclease
a-gt.2	No	Hypothetical protein
55.8	No	Conserved hypothetical predicted membrane-associated protein
I-TevII	No	Endonuclease for nrdD-intron homing
nrdC.5	No	Conserved hypothetical protein
nrdC.9	No	Conserved hypothetical protein
tk	No	Thymidine kinase
vs.5	No	Conserved hypothetical protein

e.2	No	Conserved hypothetical predicted membrane-associated protein
5.4	No	Conserved hypothetical protein
15	Yes	Proximal tail sheath stabilizer, connector to gp3 and/or gp19
segD	No	Probable site-specific intron-like DNA endonuclease
uvsY.-2	No	Hypothetical protein
alt.-2	No	Hypothetical protein
I-TevIII	No	Defective intron homing endonuclease
frd.2	No	Conserved hypothetical protein

Table S30b. UAG terminating genes in bacteriophage T7 (excerpted from (*68*))

Gene	Essential	Function
0.6B	No	Unknown function
3.8	No	Homing endonuclease
5.3	No	Homing endonuclease
6	Yes	5'->3' dsDNA exonuclease activity, RNase H
18.5	No	Holin (lambda Rz analog)
19.5	No	Holin (suppresses gp17.5 mutants)

Plaque area

RF1$^-$ strains yielded smaller plaques, indicating increased T7 resistance (Fig. 4). The raw images of plaques on each recoded host are shown in Fig. S8. We included MG1655 (fastest growth) and C30.5 (slowest growth) as benchmarks to demonstrate that plaque area is not affected by strain doubling time.

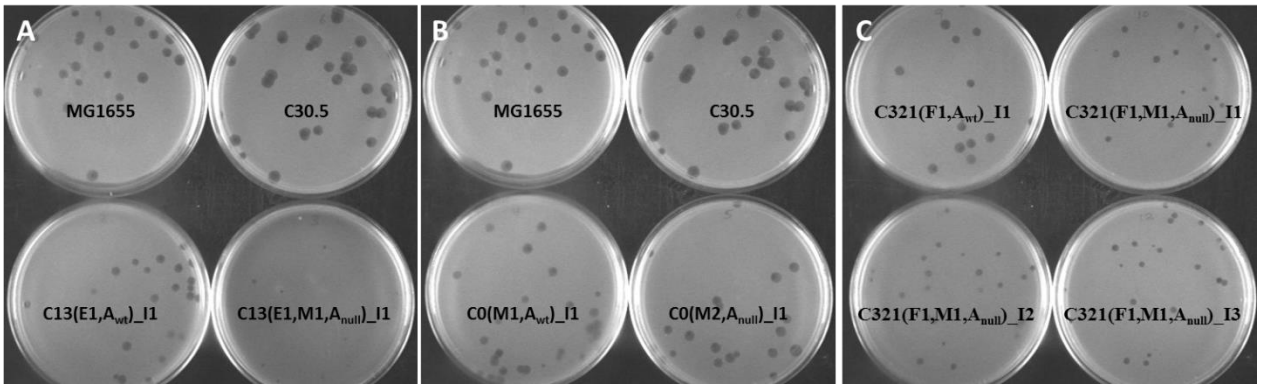


Fig. S8. Bacteriophage T7 plaques on recoded host strains. With the exception of C0.B*.ΔA::S, all RF1$^-$ strains yielded smaller plaques than their RF1$^+$ counterparts. C13.ΔA::S yielded the smallest plaques, perhaps because translational stalling at native UAG codons may sequester ribosomes and reduce translation or because mutational suppression introduces C-terminal extensions that impair important phage proteins.

Plaque areas were significantly different ($p < 0.0001$) based on RF1 status according to a Kruskal-Wallis one-way ANOVA followed by Dunn's multiple comparison test, where *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$ (Fig. S20). The complete results of the multiple comparison test are tabulated in Table S14.

Table S14. Pairwise statistical comparison of plaque areas.

	C13	C13.ΔA::S	C0.B*	C0.B*.ΔA::S	C321	C321.ΔA::S	C321.ΔA::T	C321.ΔA
MG1655	ns	***	ns	ns	ns	***	***	***
C13		*******	ns	ns	ns	*	*	*
C13.ΔA::S			***	***	***	ns	ns	ns
C0.B*				**ns**	ns	**	**	**
C0.B*.ΔA::S					ns	***	***	***
C321						*******	*******	*******
C321.ΔA::S							ns	ns
C321.ΔA::T								ns
C321.ΔA								

Statistical significances for pair wise plaque area comparisons were calculated using a Kruskal-Wallis one-way ANOVA ($p < 0.0001$) followed by Dunn's multiple comparison test. On the star system, * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$. Strains with UAG removed from all essential genes are highlighted in green, strains with a compensatory RF2 variant are highlighted in magenta, and strains with UAG removed from all genes are highlighted in blue. C0.B*.ΔA::S was the only strain that did not show a statistically significant decreased plaque area after RF1 inactivation.

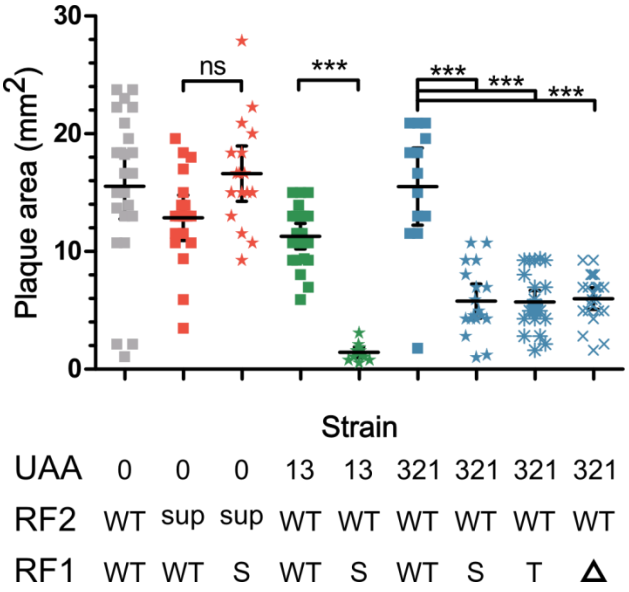


Fig. S20. Bacteriophage T7 infection is attenuated in GROs lacking RF1. RF1 (*prfA*) status is denoted by symbol shape: ■ is wt *prfA* (WT), ★ is Δ*prfA*::*spec*^{R} (S), ❋ is Δ*prfA*::*tolC* (T), and ✕ is a clean deletion of *prfA* (Δ). RF2 "sup" indicates a variant (frameshift removed, T246A, A293E) capable of suppressing lethality of RF1 deletion. **(A)** Plaque area (mm^2) distributions for strains with or without RF1. Plaque areas were calculated using ImageJ, and means +/- 95% confidence intervals are presented with the raw plaque area data ($n > 12$ for each strain). In the absence of RF1, all strains except for C0.B*.ΔA::S yielded significantly smaller plaques, indicating that the RF2 variant can terminate UAG adequately to maintain T7 fitness. Statistics are based on a Kruskal-Wallis one-way ANOVA followed by Dunn's multiple comparison test (*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$). A statistical summary can be found in Table S8.

Kinetic lysis

To confirm the plaque area observations, we performed kinetic lysis curves with T7 infected at a multiplicity of infection (MOI) of 5. This ensured that all host cells were rapidly and synchronously infected by phage particles. We monitored lysis on a Biotek H4 plate reader with OD_{600} measurements taken every 5 minutes. The mean lysis curves were plotted using average OD_{600} values for each time point (n = 12), and a two-way ANOVA showed that the lysis curves were significantly different ($p < 0.0001$). Each lysis curve was fit to a cumulative normal distribution from which mean lysis parameters were calculated using the normcdf function in MATLAB (Fig. S9).

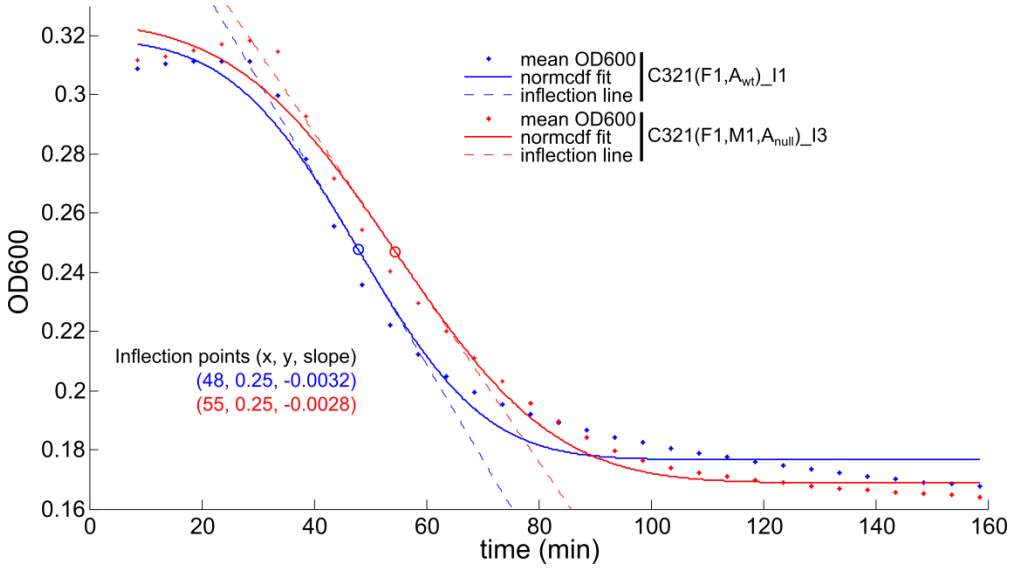


Fig. S9. T7 kinetic lysis curves (MOI = 5). Mean lysis time (+/- standard error of the mean) was 47.9 (+/- 0.1) minutes for C321 and 54.5 (+/- 0.2) minutes for C321.ΔA::S, indicating that lysis is delayed in the absence of RF1 (n = 12, $p < 0.0001$, unpaired t test with Welch's correction). Mean lysis OD_{600} was 0.25 (with negligible standard error of the mean) for both strains, showing that both hosts were infected under identical conditions and could be completely lysed by T7.

One-step growth curves

To determine burst size and eclipse time, one step growth curves were performed as previously described (*70*). Briefly, mid-log phase cultures were infected at MOI = 0.1. At 3 minutes post infection, 30 µL of infected culture was diluted 500-fold into 15 mL LB^L to minimize further phage adsorption. Two aliquots were taken at t = 6 minutes—one aliquot was titered directly and the other was treated with chloroform before titering. Adsorption efficiency was determined by $(pfu_{noCHCL3} - pfu_{CHCl3})/ pfu_{noCHCL3}$. Additional aliquots of the infection were taken at the following time points and were immediately treated with chloroform to release intracellular

phage particles and to halt infection: 6, 18, 21, 24, 27, 29, 31, 33, 35, 37, 39, 41, 45, and 50 minutes. These samples were then titered to monitor intracellular phage assembly during a single phage life cycle. Six replicates were performed, and each one-step growth curve was analyzed separately before their parameters were averaged. We estimated one-step growth parameters by using the scipy.optimize.curve_fit function to fit pfu/mL to

$$\phi = \begin{cases} 0, & 0 < t < a \\ r(t-a), & a < t < \frac{B}{r} + a \\ B, & t > \frac{B}{r} + a \end{cases}$$

where ϕ is the number of phage progeny as a function of time (t), a is eclipse time, r is rise rate, and B is burst size (*70*).

Adsorption efficiency ranged from ~20% – 60%, which is considerably lower than the >95% that we observed during the T7 fitness assay (Fig. 4C). This discrepancy is probably because we performed this assay using phage lysate that had been stored at 4 °C for several weeks. Although we re-titered before each replicate, infection became less efficient as the phage lysate aged. Although this increased variance, T7 infection consistently proceeded more efficiently in C321 than in C321.ΔA (Fig. S21).

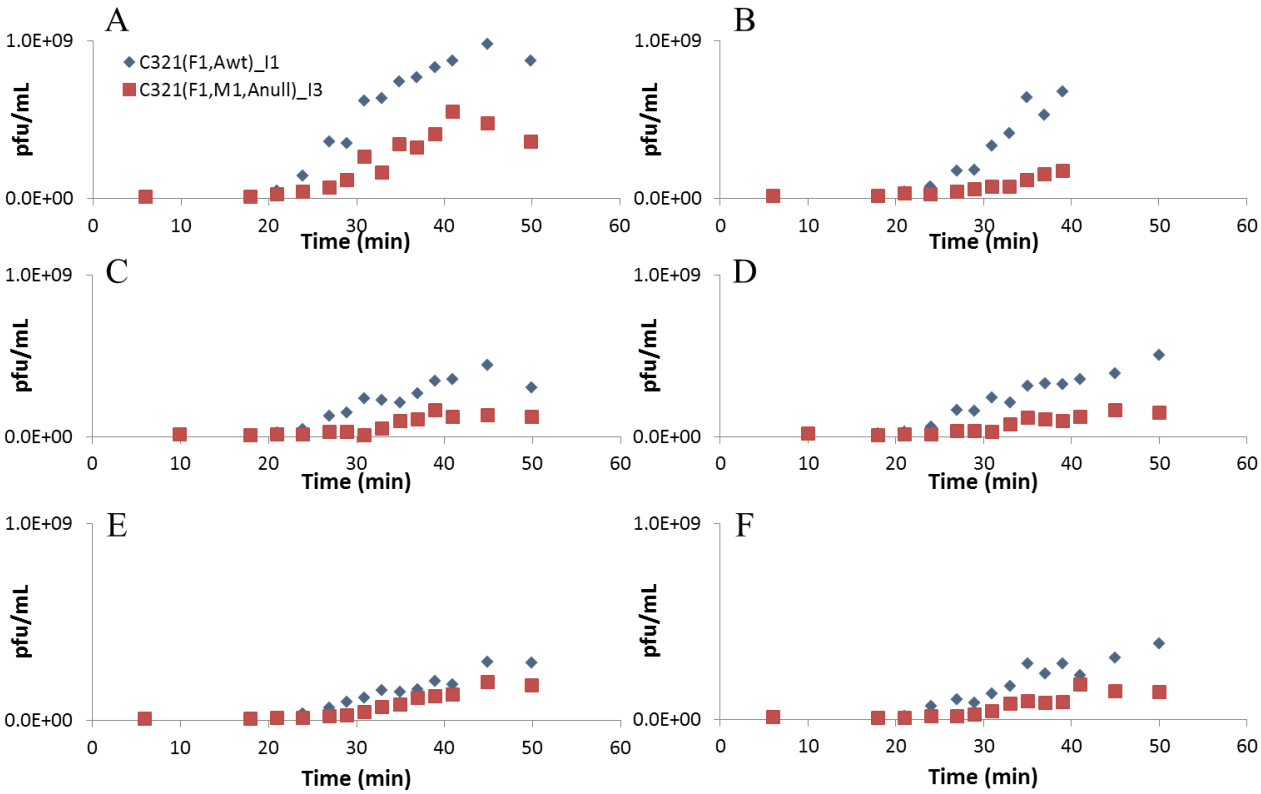


Fig. S21. One step growth curves were performed using hosts C321 and C321.ΔA in six replicates: A = replicate 1, B = replicate 2, C = replicate 3, D = replicate 4, E = replicate 5, and F = replicate 6. Although the T7 lysate was properly stored at 4 °C in LB^{L} supplemented with 900 mM sodium chloride, we found that longterm storage decreased adsorption efficiency (and apparent burst size) in both hosts. Therefore, replicates 3 and 4 were taken on the same day and replicates 5 and 6 were taken two days later to minimize variance. Even despite the variance, C321 consistently yielded larger burst sizes than C321.ΔA in all replicates.

Because the first two replicates yielded higher phage titers than the others, we combined replicates 3-6, which were performed over the course of four days (Fig. S10). RF1 removal caused a 30% longer eclipse time ($p = 0.01$), a 60% smaller burst size ($p = 0.02$), and a 35% slower rise rate ($p = 0.04$) (Fig. S22, Table S15). Percentage changes were calculated by $(RF1^{-}_{param} - RF1^{+}_{param}) / RF1^{+}_{param}$.

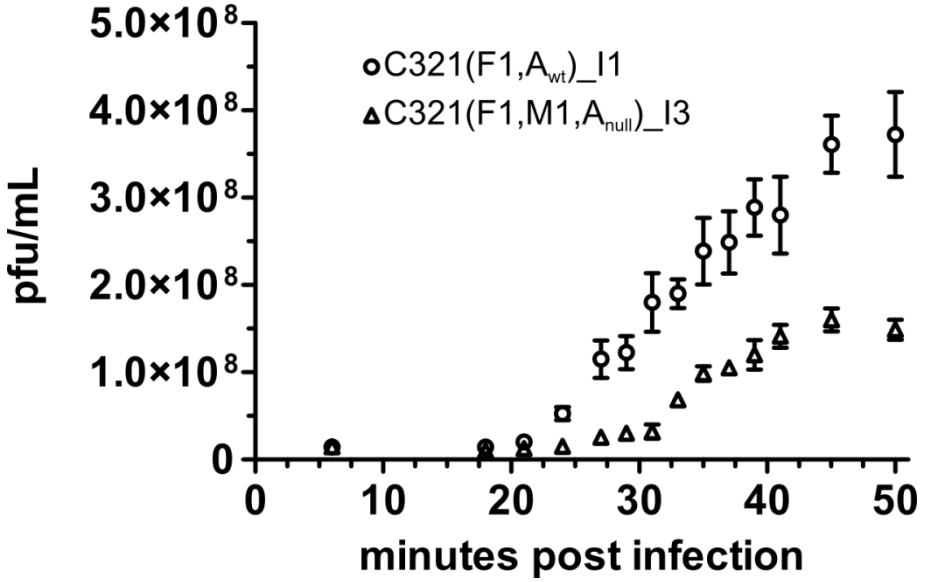


Fig. S10. One step growth curve averaged across replicates 3-6 (Fig. S21C-F). Mean pfu/mL +/- SEM are plotted for each time point. The one step growth curves for each host were significantly different (p = 0.002), as determined by a two way repeated measures ANOVA.

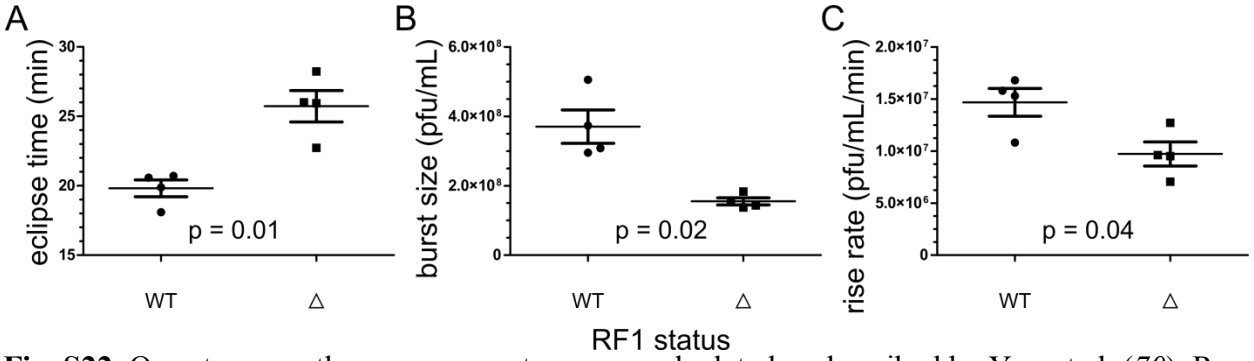


Fig. S22. One step growth curve parameters were calculated as described by You et al. (*70*). Raw data points are plotted with mean +/- SEM. The p values were calculated using an unpaired t test with Welch's correction. Compared to C321 (WT), the C321.ΔA (Δ) supports T7 infection with (A) a 30% (+/- 2%) longer eclipse time (p = 0.01), (B) a 59% (+/- 9%) smaller burst size (p = 0.02), and (C) a 35% (+/- 5%) slower rise rate (p = 0.04).

Table S15. One-step growth parameters: eclipse time, rise rate, and burst size

Metric[a]	**C321**	**C321.ΔA**	**% change**[b]	**p value**[c]
Eclipse time (min)	19.8 (+/- 0.6)	25.7 (+/- 1.1)	30% longer	0.01
Burst size (pfu/mL)	3.7E8 (+/- 4.8E7)	1.5E8 (+/- 1.0E7)	59% smaller	0.02
Rise rate (pfu/mL/min)	1.5E7 (+/- 1.3E6)	9.7E6 (+/- 1.2E6)	35% slower	0.04

[a] Data is based on 4 replicates (Replicates 3-6, Fig. S21C-F)
[b] % change in C321.ΔA with respect to C321; $(RF1^{-}_{param} - RF1^{+}_{param})/ RF1^{+}_{param}$
[c] p values were calculated using an unpaired t test with Welch's correction

K. Selectable markers used in this study

>*mutS*::*cat* (1017 bp)
gctgatccactagttctagagcggccgctctaaaaaatgccctcttgggttatcaagagggtcattatatttcg
cggaattcatgctatcgacgtcgatatctggcgaaaatgagacgttgatcggcacgtaagaggttccaactttc
accataatgaaataagatcactaccgggcgtatttttgagttatcgagattttcaggagctaaggaagctaaa
atggagaaaaaaatcactggatataccaccgttgatatatcccaatggcatcgtaaagaacattttgaggcatt
tcagtcagttgctcaatgtacctataaccagaccgttcagctggatattacggccttttaaagaccgtaaaga
aaaataagcacaagttttatccggcctttattcacattcttgcccgcctgatgaatgctcatccggaatttcgt
atggcaatgaaagacggtgagctggtgatatgggatagtgttcacccttgttacaccgttttccatgagcaaac
tgaaacgttttcatcgctctggagtgaataccacgacgatttccggcagtttctacacatatattcgcaagatg
tggcgtgttacggtgaaaacctggcctatttccctaaagggtttattgagaatatgtttttcgtctcagccaat
ccctgggtgagtttcaccagttttgatttaaacgtggccaatatggacaacttcttcgcccccgttttcaccat
gggcaaatattatacgcaaggcgacaaggtgctgatgccgctggcgattcaggttcatcatgccgtctgtgatg
gcttccatgtcggcagaatgcttaatgaattacaacagtactgcgatgagtggcagggcggggcgtaatttgat
atcgagctcgtcagcaggcgcgcctgtaatcacactggctcaccttcgggtgggcctttctgcgtttaaaaaaa
acgggccggcgcgaacgccggcccgcggccgccacccagcttttgttccctttag

>*kan*^{R}-oriT (1949 bp); oriT is from RK2 (*28*)
cctgtgacggaagatcacttcgcagaataaataaatcctggtgtccctgttgataccgggaagccctgggccaa
cttttggcgaaaatgagacgttgatcggcacgtaagaggttccaactttcaccataatgaaataagatcactac
cgggcgtatttttgagttgtcgagattttcaggagctaaggaagctaaaatgagccatattcaacgggaaacg
tcgaggccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggca
atcaggtgcgacaatctatcgcttgtatgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggta
gcgttgccaatgatgttacagatgagatggtcagactaaactggctgacggaatttatgcctcttccgaccatc
aagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccccggaaaaacagcattccaggt
attagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcga
ttcctgtttgtaattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatcacgaatgaataac
ggtttggttgatgcgagtgattttgatgacgagcgtaatggctggcctgttgaacaagtctggaaagaaatgca
taaacttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataaccttatttttgacg
aggggaaattaataggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatccta
tggaactgcctcggtgagttttctccttcattacagaaacggctttttcaaaaatatggtattgataatcctga
tatgaataaattgcagtttcatttgatgctcgatgagtttttctaatttttttaaggcagttattggtgcccctt
aaacgcctggttgctacgcctgaataagtgataataagcggatgaatggcagaaattcgaaagcaaattcgacc
cggtcgtcggttcagggcagggtcgttaaatagccgcttatgtctattgctggttggcgctcggtcttgccttg
ctcgtcggtgatgtacttcaccagctccgcgaagtcgctcttcttgatggagcgcatggggacgtgcttggcaa
tcacgcgcacccccggccgttttagcggctaaaaaagtcatggctctgccctcgggcggaccacgcccatcat
gaccttgccaagctcgtcctgcttctcttcgatcttcgccagcagggcgaggatcgtggcatcaccgaaccgcg
ccgtgcgcgggtcgtcggtgagccagagtttcagcaggccgcccaggcggcccaggtcgccattgatgcgggcc
agctcgcggacgtgctcatagtccacgacgcccgtgattttgtagccctggccgacggccagcaggtaggccga
caggctcatgccggccgccgccgccttttcctcaatcgctcttcgttcgtctggaaggcagtacaccttgatag
gtgggctgcccttcctggttggcttggtttcatcagccatccgcttgccctcatctgttacgccggcggtagcc
ggccagcctcgcagagcaggattcccgttgagcaccgccaggtgcgaataagggacagtgaagaaggaacaccc
gctcgcgggtgggcctacttcacctatcctgcccggctgacgccgttggatacaccaaggaaagtctacacgaa
ccctttggcaaaatcctgtatatcgtgcgaaaaaggatggatataccgaaaaaatcgctataatgaccccgaag
cagggttatgcagcggaaaagcgct

>*gent*^{R} (831 bp)
acgcacaccgtggaaacggatgaaggcacgaacccagttgacataagcctgttcggttcgtaaactgtaatgca
agtagcgtatgcgctcacgcaactggtccagaaccttgaccgaacgcagcggtggtaacggcgcagtggcggtt
ttcatggcttgttatgactgtttttttgtacagtctatgcctcgggcatccaagcagcaagcgcgttacgccgt
gggtcgatgtttgatgttatggagcagcaacgatgttacgcagcagcaacgatgttacgcagcagggcagtcgc
cctaaaacaaagttaggtggctcaagtatgggcatcattcgcacatgtaggctcggccctgaccaagtcaaatc
catgcgggctgctcttgatcttttcggtcgtgagttcggagacgtagccacctactcccaacatcagccggact

ccgattacctcgggaacttgctccgtagtaagacattcatcgcgcttgctgccttcgaccaagaagcggttgtt
ggcgctctcgcggcttacgttctgcccaggtttgagcagccgcgtagtgagatctatatctatgatctcgcagt
ctccggcgagcaccggaggcagggcattgccaccgcgctcatcaatctcctcaagcatgaggccaacgcgcttg
gtgcttatgtgatctacgtgcaagcagattacggtgacgatcccgcagtggctctctatacaaagttgggcata
cgggaagaagtgatgcactttgatatcgacccaagtaccgccacctaacaattcgttcaagccgagatcggctt
cccgg

>*zeo*^{R} (761bp)
ggtgttgacaattaatcatcggcatagtatatcggcatagtataatacgacaaggtgaggaactaaaccatggc
caagttgaccagtgccgttccggtgctcaccgcgcgcgacgtcgccggagcggtcgagttctggaccgaccggc
tcgggttctcccgggacttcgtggaggacgacttcgccggtgtggtccgggacgacgtgaccctgttcatcagc
gcggtccaggaccaggtggtgccggacaacaccctggcctgggtgtgggtgcgcggcctggacgagctgtacgc
cgagtggtcggaggtcgtgtccacgaacttccgggacgcctccgggccggccatgaccgagatcggcgagcagc
cgtgggggcgggagttcgccctgcgcgacccggccggcaactgcgtgcacttcgtggccgaggagcaggactga
cacgtccgacggcggcccacgggtcccaggcctcggagatccgtccccctttccctttgtcgatatcatgtaat
tagttatgtcacgcttacattcacgccctccccccacatccgctctaaccgaaaaggaaggagttagacaacct
gaagtctaggtccctatttatttttttatagttatgttagtattaagaacgttatttatatttcaaatttttct
tttttttctgtacagacgcgtgtacgcatgtaacattatactgaaaaccttgcttgagaaggttttgggacgct
cgaaggctttaatttgcaagct

>*spec*^{R} (1201 bp)
cagccaggacagaaatgcctcgacttcgctgctgcccaaggttgccgggtgacgcacaccgtggaaacggatga
aggcacgaacccagtggacataagcctgttcggttcgtaagctgtaatgcaagtagcgtatgcgctcacgcaac
tggtccagaaccttgaccgaacgcagcggtggtaacggcgcagtggcggttttcatggcttgttatgactgttt
ttttggggtacagtctatgcctcgggcatccaagcagcaagcgcgttacgccgtgggtcgatgtttgatgttat
ggagcagcaacgatgttacgcagcagggcagtcgccctaaaacaaagttaaacatcatgagggaagcggtgatc
gccgaagtatcgactcaactatcagaggtagttggcgtcatcgagcgccatctcgaaccgacgttgctggccgt
acatttgtacggctccgcagtggatggcggcctgaagccacacagtgatattgatttgctggttacggtgaccg
taaggcttgatgaaacaacgcggcgagctttgatcaacgacctttttggaaacttcggcttcccctggagagagc
gagattctccgcgctgtagaagtcaccattgttgtgcacgacgacatcattccgtggcgttatccagctaagcg
cgaactgcaatttggagaatggcagcgcaatgacattcttgcaggtatcttcgagccagccacgatcgacattg
atctggctatcttgctgacaaaagcaagagaacatagcgttgccttggtaggtccagcggcggaggaactcttt
gatccggttcctgaacaggatctatttgaggcgctaaatgaaaccttaacgctatggaactcgccgcccgactg
ggctggcgatgagcgaaatgtagtgcttacgttgtcccgcatttggtacagcgcagtaaccggcaaaatcgcgc
cgaaggatgtcgctgccgactgggcaatggagcgcctgccggcccagtatcagcccgtcatacttgaagctaga
caggcttatcttggacaagaagaagatcgcttggcctcgcgcgcagatcagttggaagaatttgtccactacgt
gaaaggcgagatcaccaaggtagtcggcaaataaagctttactgagctaataacaggactgctggtaatcgcag
gccttttttatttctgca

>*tolC* (1764 bp)
ttgaggcacattaacgccctatggcacgtaacgccaacctttgcggtagcggcttctgctagaatccgcaata
attttacagtttgatcgcgctaaatactgcttcaccacaaggaatgcaaatgaagaaattgctccccattctta
tcggcctgagcctttctgggttcagttcgttgagccaggccgagaacctgatgcaagtttatcagcaagcacgc
cttagtaacccggaattgcgtaagtctgccgccgatcgtgatgctgcctttgaaaaaattaatgaagcgcgcag
tccattactgccacagctaggtttaggtgcagattacacctatagcaacggctaccgcgacgcgaacggcatca
actctaacgcgaccagtgcgtccttgcagttaactcaatccatttttgatatgtcgaaatggcgtgcgttaacg
ctgcaggaaaaagcagcagggattcaggacgtcacgtatcagaccgatcagcaaaccttgatcctcaacaccgc
gaccgcttatttcaacgtgttgaatgctattgacgttctttcctatacacaggcacaaaaagaagcgatctacc
gtcaattagatcaaaccacccaacgttttaacgtgggcctggtagcgatcaccgacgtgcagaacgcccgcgca
cagtacgataccgtgctggcgaacgaagtgaccgcacgtaataaccttgataacgcggtagagcagctgcgcca
gatcaccggtaactactatccggaactggctgcgctgaatgtcgaaaactttaaaaccgacaaaccacagccgg
ttaacgcgctgctgaaagaagccgaaaaacgcaacctgtcgctgttacaggcacgcttgagccaggacctggcg
cgcgagcaaattcgccaggcgcaggatggtcacttaccgactctggatttaacggcttctaccgggatttctga
cacctcttatagcggttcgaaaacccgtggtgccgctggtacccagtatgacgatagcaatatgggccagaaca

aagttggcctgagcttctcgctgccgatttatcagggcggaatggttaactcgcaggtgaaacaggcacagtac
aactttgtcggtgccagcgagcaactggaaagtgcccatcgtagcgtcgtgcagaccgtgcgttcctccttcaa
caacattaatgcatctatcagtagcattaacgcctacaaacaagccgtagtttccgctcaaagctcattagacg
cgatggaagcgggctactcggtcggtacgcgtaccattgttgatgtgttggatgcgaccaccacgttgtacaac
gccaagcaagagctggcgaatgcgcgttataactacctgattaatcagctgaatattaagtcagctctgggtac
gttgaacgagcaggatctgctggcactgaacaatgcgctgagcaaaccggtttccactaatccggaaaacgttg
caccgcaaacgccggaacagaatgctattgctgatggttatgcgcctgatagcccggcaccagtcgttcagcaa
acatccgcacgcactaccaccagtaacggtcataacccttccgtaactgaaagcttgatatcgaattcctgca
gcccgggggatcccatggtacgcgtgctagaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctt
tcgttttatctgttgtttgtcggtgaacgctctcctgagtaggacaaatccgccgccctaga

>*galK* (1270 bp)

cctgttgacaattaatcatcggcatagtatatcggcatagtataatacgacaaggtgaggaactaaacccagga
ggcagatcatgagtctgaaagaaaaaacacaatctctgtttgccaacgcatttggctaccctgccactcacacc
attcaggcgcctggccgcgtgaatttgattggtgaacacaccgactacaacgacggtttcgttctgccctgcgc
gattgattatcaaaccgtgatcagttgtgcaccacgcgatgaccgtaaagttcgcgtgatggcagccgattatg
aaaatcagctcgacgagttttccctcgatgcgcccattgtcgcacatgaaaactatcaatgggctaactacgtt
cgtggcgtggtgaaacatctgcaactgcgtaacaacagcttcggcggcgtggacatggtgatcagcggcaatgt
gccgcagggtgccgggttaagttcttccgcttcactggaagtcgcggtcggaaccgtattgcagcagctttatc
atctgccgctggacggcgcacaaatcgcgcttaacggtcaggaagcagaaaaccagtttgtaggctgtaactgc
gggatcatggatcagctaatttccgcgctcggcaagaaagatcatgccttgctgatcgattgccgctcactggg
gaccaaagcagtttccatgcccaaaggtgtggctgtcgtcatcatcaacagtaacttcaaacgtaccctggttg
gcagcgaatacaacacccgtcgtgaacagtgcgaaaccggtgcgcgtttcttccagcagccagccctgcgtgat
gtcaccattgaagagttcaacgctgttgcgcatgaactggacccgatcgtggcaaaacgcgtgcgtcatatact
gactgaaaacgcccgcaccgttgaagctgccagcgcgctggagcaaggcgacctgaaacgtatgggcgagttga
tggcggagtctcatgcctctatgcgcgatgatttcgaaatcaccgtgccgcaaattgacactctggtagaaatc
gtcaaagctgtgattggcgacaaaggtggcgtacgcatgaccggcggcggatttggcggctgtatcgtcgcgct
gatcccggaagagctggtgcctgccgtacagcaagctgtcgctgaacaatatgaagcaaaaacaggtattaaag
agacttttacgtttgtaaaccatcacaaggagcaggacagtgctgaaaaaaaaaaccccgcccctgacagggc
ggggttttttt

>*malK* (1116 bp)

atggcgagcgtacagctgcaaaatgtaacgaaagcctggggcgaggtcgtggtatcgaaagatatcaatctcga
tatccatgaaggtgaattcgtggtgtttgtcggaccgtctggctgcggtaaatcgactttactgcgcatgattg
ccgggcttgagacgatcaccagcggcgacctgttcatcggtgagaaacggatgaatgacactccgccagcagaa
cgcggcgttggtatggtgtttcagtcttacgcgctctatccccacctgtcagtagcagaaaacatgtcatttgg
cctgaaactggctggcgcaaaaaaagaggtgattaaccaacgcgttaaccaggtggcggaagtgctacaactgg
cgcatttgctggatcgcaaaccgaaagcgctctccggtggtcagcgtcagcgtgtggcgattggccgtacgctg
gtggccgagccaagcgtatttttgctcgatgaaccgctctccaacctcgatgctgcactgcgtgtgcaaatgcg
tatcgaaatctcccgtctgcataaacgcctgggccgcacaatgatttacgtcacccacgatcaggtcgaagcga
tgacgctggccgacaaaatcgtggtgctggacgccggtcgcgtggcgcaggttgggaaaccgctggagctgtac
cactatccggcagaccgttttgtcgccggatttatcggttcgccaaagatgaacttcctgccggtaaaagtgac
cgccaccgcaatcgatcaagtgcaggtggagctgccgatgccaaatcgtcagcaagtctggctgccagttgaaa
gccgtgatgtccaggttggagccaatatgtcgctgggtattcgcccggaacatctactgccgagtgatatcgct
gacgtcatccttgagggtgaagttcaggtcgtcgagcaactcggcaacgaaactcaaatccatatccagatccc
ttccattcgtcaaaacctggtgtaccgccagaacgacgtggtgttggtagaagaaggtgccacattcgctatcg
gcctgccgccagagcgttgccatctgttccgtgaggatggcactgcatgtcgtcgactgcataaggagccgggc
gtttaa

Table S33. Sequences of GFP variants containing UAG codons (UAG codons are highlighted in red).

>GFP-NHis-0UAG
atgcaccaccaccaccaccacagtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcaca
aattttctgtcagtggagagggtgaaggtgatgcaacatacggaaaacttacccttaaatttatttgcactactggaaaactacctgttccatgg
ccaacacttgtcactactttctcttatggtgttcaatgcttttcccgttatccggatcacatgaaacggcatgacttttcaagagtgccatgcccg
aaggttatgtacaggaacgcactatatctttcaaagatgacgggaactacaagacgcgtgctgaagtcaagtttgaaggtgataccttgtta
atcgtatcgagttaaaaggtattgattttaaagaagatggaaacattctcggacacaaactcgaatacaactataactcacacaatgtatacatc
acggcagacaaacaaaagaatggaatcaaagctaacttcaaaattcgccacaacattgaagatggatccgttcaactagcagaccattatca
acaaaatactccaattggcgatggccctgtccttttaccagacaaccattacctgtcgacacaatctgccctttcgaaagatcccaacgaaaa
gcgtgaccacatggtccttcttgagtttgtaactgctgctgggattacacatggcatggatgagctctacaaataa

>GFP-NHis-1UAG
atgcaccaccaccaccaccacagtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcaca
aattttctgtcagtggagagggtgaaggtgatgcaacatacggaaaacttacccttaaatttatttgcactactggaaaactacctgttccatgg
ccaacacttgtcactactttctcttatggtgttcaatgcttttcccgttatccggatcacatgaaacggcatgacttttcaagagtgccatgcccg
aaggttatgtacaggaacgcactatatctttcaaagatgacgggaacftacaagacgcgtgctgaagtcaagtttgaaggtgataccttgtta
atcgtatcgagttaaaaggtattgattttaaagaagatggaaacattctcggacacaaactcgaatacaactataactcacacaatgtatagatc
acggcagacaaacaaaagaatggaatcaaagctaacttcaaaattcgccacaacattgaagatggatccgttcaactagcagaccattatca
acaaaatactccaattggcgatggccctgtccttttaccagacaaccattacctgtcgacacaatctgccctttcgaaagatcccaacgaaaa
gcgtgaccacatggtccttcttgagtttgtaactgctgctgggattacacatggcatggatgagctctacaaataa

>GFP-NHis-2UAG
atgcaccaccaccaccaccacagtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcaca
aattttctgtcagtggagagggtgaaggtgatgcaacatagggaaaacttacccttaaatttatttgcactactggaaaactacctgttccatgg
ccaacacttgtcactactttctcttatggtgttcaatgcttttcccgttatccggatcacatgaaacggcatgacttttcaagagtgccatgcccg
aaggttatgtacaggaacgcactatatctttcaaagatgacgggaactacaagacgcgtgctgaagtcaagtttgaaggtgataccttgtta
atcgtatcgagttaaaaggtattgattttaaagaagatggaaacattctcggacacaaactcgaatacaactataactcacacaatgtatacatc
acggcagacaaacaaaagaatggaatcaaagctaacttcaaaattcgccacaacattgaagatggatccgttcaactagcagaccattagc
aacaaaatactccaattggcgatggccctgtccttttaccagacaaccattacctgtcgacacaatctgccctttcgaaagatcccaacgaaaa
gcgtgaccacatggtccttcttgagtttgtaactgctgctgggattacacatggcatggatgagctctacaaataa

>GFP-NHis-3UAG
atgcaccaccaccaccaccacagtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcaca
aattttctgtcagtggagagggtgaaggtgatgcaacatagggaaaacttacccttaaatttatttgcactactggaaaactacctgttccatgg
ccaacacttgtcactactttctcttatggtgttcaatgcttttcccgttatccggatcacatgaaacggcatgacttttcaagagtgccatgcccg
aaggttatgtacaggaacgcactatatctttcaaagatgacgggaactacaagacgcgtgctgaagtcaagtttgaaggtgataccttgtta
atcgtatcgagttaaaaggtattgattttaaagaagatggaaacattctcggacacaaactcgaatacaactataactcacacaatgtatagatc
acggcagacaaacaaaagaatggaatcaaagctaacttcaaaattcgccacaacattgaagatggatccgttcaactagcagaccattagc
aacaaaatactccaattggcgatggccctgtccttttaccagacaaccattacctgtcgacacaatctgccctttcgaaagatcccaacgaaaa
gcgtgaccacatggtccttcttgagtttgtaactgctgctgggattacacatggcatggatgagctctacaaataa

L. References

1. K. Vetsigian, C. Woese, N. Goldenfeld, Collective evolution and the genetic code. *PNAS* **103**, 10696 (2006).
2. D. V. Goeddel, D. G. Kleid, F. Bolivar, H. L. Heyneker, D. G. Yansura, R. Crea, T. Hirose, A. Kraszewski, K. Itakura, A. D. Riggs, Expression in Escherichia coli of Chemically Synthesized Genes for Human Insulin. *PNAS* **76**, 106 (1979).
3. D. C. Krakauer, V. A. A. Jansen, Red queen dynamics of protein translation. *J. Theor. Biol.* **218**, 97 (2002).
4. M. G. Schafer, A. A. Ross, J. P. Londo, C. A. Burdick, E. H. Lee, S. E. Travers, P. K. Van de Water, C. L. Sagers, The Establishment of Genetically Engineered Canola Populations in the U.S. *PLoS One* **6**, e25736 (2011).
5. J. M. Sturino, T. R. Klaenhammer, Engineered bacteriophage-defence systems in bioprocessing. *Nat. Rev. Microbiol.* **4**, 395 (2006).
6. M. Schmidt, V. de Lorenzo, Synthetic constructs in/for the environment: Managing the interplay between natural and engineered Biology. *FEBS Lett.* **586**, 2199 (2012).
7. C. C. Liu, P. G. Schultz, Adding New Chemistries to the Genetic Code. *An. Rev. Biochem.* **79**, 413 (2010).
8. H. Neumann, K. Wang, L. Davis, M. Garcia-Alai, J. W. Chin, Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464**, 441 (2010).
9. T. S. Young, I. Ahmad, J. A. Yin, P. G. Schultz, An Enhanced System for Unnatural Amino Acid Mutagenesis in E. coli. *Journal of Molecular Biology* **395**, 361 (2009).
10. G. Eggertsson, D. Söll, Transfer ribonucleic acid-mediated suppression of termination codons in Escherichia coli. *Microbiological Reviews* **52**, 354 (September 1, 1988, 1988).
11. F. J. Isaacs, P. A. Carr, H. H. Wang, M. J. Lajoie, B. Sterling, L. Kraal, A. C. Tolonen, T. A. Gianoulis, D. B. Goodman, N. B. Reppas, C. J. Emig, D. Bang, S. J. Hwang, M. C. Jewett, J. M. Jacobson, G. M. Church, Precise Manipulation of Chromosomes in Vivo Enables Genome-Wide Codon Replacement. *Science* **333**, 348 (Jul, 2011).
12. D. G. Gibson, J. I. Glass, C. Lartigue, V. N. Noskov, R. Y. Chuang, M. A. Algire, G. A. Benders, M. G. Montague, L. Ma, M. M. Moodie, C. Merryman, S. Vashee, R. Krishnakumar, N. Assad-Garcia, C. Andrews-Pfannkoch, E. A. Denisova, L. Young, Z. Q. Qi, T. H. Segall-Shapiro, C. H. Calvey, P. P. Parmar, C. A. Hutchison, H. O. Smith, J. C. Venter, Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science* **329**, 52 (Jul, 2010).
13. H. H. Wang, F. J. Isaacs, P. A. Carr, Z. Z. Sun, G. Xu, C. R. Forest, G. M. Church, Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894 (Aug, 2009).
14. P. A. Carr, H. H. Wang, B. Sterling, F. J. Isaacs, M. J. Lajoie, G. Xu, G. M. Church, J. M. Jacobson, Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection. *Nucleic Acids Res.*, (May 25, 2012, 2012).
15. T. Mukai, A. Hayashi, F. Iraha, A. Sato, K. Ohtake, S. Yokoyama, K. Sakamoto, Codon reassignment in the Escherichia coli genetic code. *Nucleic Acids Res.* **38**, 8188 (2010).
16. D. B. F. Johnson, J. Xu, Z. Shen, J. K. Takimoto, M. D. Schultz, R. J. Schmitz, Z. Xiang, J. R. Ecker, S. P. Briggs, L. Wang, RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat Chem Biol* **7**, 779 (2011).

17. K. Ohtake, A. Sato, T. Mukai, N. Hino, S. Yokoyama, K. Sakamoto, Efficient Decoding of the UAG Triplet as a Full-Fledged Sense Codon Enhances the Growth of a prfA-Deficient Strain of Escherichia coli. *J. Bacteriol.* **194**, 2606 (May 15, 2012, 2012).
18. P. O'Donoghue, L. Prat, I. U. Heinemann, J. Ling, K. Odoi, W. R. Liu, D. Söll, Near-cognate suppression of amber, opal and quadruplet codons competes with aminoacyl-tRNAPyl for genetic code expansion. *FEBS Lett.*, (2012).
19. I. U. Heinemann, A. J. Rovner, H. R. Aerni, S. Rogulina, L. Cheng, W. Olds, J. T. Fischer, D. Soll, F. J. Isaacs, J. Rinehart, Enhanced phosphoserine insertion during Escherichia coli protein synthesis via partial UAG codon reassignment and release factor 1 deletion. *FEBS Lett.* **586**, 3716 (2012-Oct-19, 2012).
20. J. T. Ngo, D. A. Tirrell, Noncanonical amino acids in the interrogation of cellular protein synthesis. *Accounts of chemical research* **44**, 677 (2011).
21. H.-S. Park, M. J. Hohn, T. Umehara, L.-T. Guo, E. M. Osborne, J. Benner, C. J. Noren, J. Rinehart, D. Söll, Expanding the Genetic Code of Escherichia coli with Phosphoserine. *Science* **333**, 1151 (August 26, 2011, 2011).
22. R. H. Heineman, I. J. Molineux, J. J. Bull, Evolutionary robustness of an optimal phenotype: Re-evolution of lysis in a bacteriophage deleted for its lysin gene. *J. Mol. Evol.* **61**, 181 (Aug, 2005).
23. J. D. Bain, C. Switzer, R. Chamberlin, S. A. Benner, Ribosome-mediated incorporation of a nonstandard amino acid into a peptide through expansion of the genetic code. *Nature* **356**, 537 (APR 9 1992, 1992).
24. J. C. Anderson, N. Wu, S. W. Santoro, V. Lakshman, D. S. King, P. G. Schultz, An expanded genetic code with a functional quadruplet codon. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 7566 (May 18, 2004, 2004).
25. M. J. Lajoie, S. Kosuri, J. A. Mosberg, C. J. Gregg, D. Zhang, G. M. Church, Probing the limits of genetic recoding in essential genes. *Science* **342**, 361 (2013).
26. S. A. Schwartz, D. R. Helinski, Purification and Characterization of Colicin E1. *J. Biol. Chem.* **246**, 6318 (October 25, 1971, 1971).
27. J. A. Mosberg, M. J. Lajoie, G. M. Church, Lambda Red Recombineering in Escherichia coli Occurs Through a Fully Single-Stranded Intermediate. *Genetics* **186**, 791 (Nov, 2010).
28. D. Figurski, R. Meyer, D. S. Miller, D. R. Helinski, Generation in vitro of deletions in the broad host range plasmid RK2 using phage Mu insertions and a restriction endonuclease. *Gene* **1**, 107 (1976).
29. S. Warming, N. Costantino, D. L. Court, N. A. Jenkins, N. G. Copeland, Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res.* **33**, e36 (2005).
30. N. Rohland, D. Reich, Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research* **22**, 939 (May, 2012).
31. W. R. Pearson, T. Wood, Z. Zhang, W. Miller, Comparison of DNA sequences with protein sequences. *Genomics* **46**, 24 (Nov, 1997).
32. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357 (Apr, 2012).
33. M. A. DePristo, E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, M. J. Daly, A

framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43**, 491 (May, 2011).
34. P. Cingolani, A. Platts, L. L. Wang, M. Coon, N. Tung, L. Wang, S. J. Land, X. Lu, D. M. Ruden, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w(1118); iso-2; iso-3. *Fly* **6**, 80 (Apr-Jun, 2012).
35. K. Ye, M. H. Schulz, Q. Long, R. Apweiler, Z. Ning, Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865 (Nov 1, 2009).
36. K. Chen, J. W. Wallis, M. D. McLellan, D. E. Larson, J. M. Kalicki, C. S. Pohl, S. D. McGrath, M. C. Wendl, Q. Zhang, D. P. Locke, X. Shi, R. S. Fulton, T. J. Ley, R. K. Wilson, L. Ding, E. R. Mardis, BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat. Methods* **6**, 677 (Sep, 2009).
37. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841 (Mar 15, 2010).
38. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, P. Genome Project Data, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078 (Aug, 2009).
39. D. G. Gibson, H. O. Smith, C. A. Hutchison, J. C. Venter, C. Merryman, Chemical synthesis of the mouse mitochondrial genome. *Nat. Methods* **7**, 901 (Nov, 2010).
40. R. Lutz, H. Bujard, Independent and Tight Regulation of Transcriptional Units in Escherichia Coli Via the LacR/O, the TetR/O and AraC/I1-I2 Regulatory Elements. *Nucleic Acids Res.* **25**, 1203 (March 1, 1997, 1997).
41. D. Wessel, U. I. Flugge, A Method for the Quantitative Recovery of Protein in Dilute-Solution in the Presence of Detergents and Lipids. *Anal. Biochem.* **138**, 141 (1984).
42. A. N. Kettenbach, S. A. Gerber, Rapid and reproducible single-stage phosphopeptide enrichment of complex peptide mixtures: application to general and phosphotyrosine-specific phosphoproteomics experiments. *Anal. Chem.* **83**, 7635 (Oct 15, 2011).
43. A. J. Alpert, Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. *Anal. Chem.* **80**, 62 (Jan 1, 2008).
44. J. V. Olsen, L. M. de Godoy, G. Li, B. Macek, P. Mortensen, R. Pesch, A. Makarov, O. Lange, S. Horning, M. Mann, Parts per Million Mass Accuracy on an Orbitrap Mass Spectrometer via Lock Mass Injection into a C-trap. *Mol Cell Proteomics* **4**, 2010 (Dec, 2005).
45. I. M. Keseler, J. Collado-Vides, A. Santos-Zavaleta, M. Peralta-Gil, S. Gama-Castro, L. Muñiz-Rascado, C. Bonavides-Martinez, S. Paley, M. Krummenacker, T. Altman, P. Kaipa, A. Spaulding, J. Pacheco, M. Latendresse, C. Fulcher, M. Sarker, A. G. Shearer, A. Mackie, I. Paulsen, R. P. Gunsalus, P. D. Karp, EcoCyc: a comprehensive database of Escherichia coli biology. *Nucleic Acids Res.* **39**, D583 (January 1, 2011, 2011).
46. M. A. Shifman, Y. Li, C. M. Colangelo, K. L. Stone, T. L. Wu, K.-H. Cheung, P. L. Miller, K. R. Williams, YPED: A Web-Accessible Database System for Protein Expression Analysis. *Journal of Proteome Research* **6**, 4019 (2007/10/01, 2007).
47. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nat Meth* **9**, 671 (2012).

48. E. Pennisi, Synthetic Genome Brings New Life to Bacterium. *Science* **328**, 958 (May 21, 2010, 2010).
49. S. Kosuri, N. Eroshenko, E. M. LeProust, M. Super, J. Way, J. B. Li, G. M. Church, Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nat Biotech* **28**, 1295 (2010).
50. M. J. Lajoie, C. J. Gregg, J. A. Mosberg, G. C. Washington, G. M. Church, Manipulating replisome dynamics to enhance lambda Red-mediated multiplex genome engineering. *Nucleic Acids Res.* **40**, e170 (2012-Dec-1, 2012).
51. J. A. Mosberg, C. J. Gregg, M. J. Lajoie, H. H. Wang, G. M. Church, Improving Lambda Red Genome Engineering in *Escherichia coli* via Rational Removal of Endogenous Nucleases. *PLoS One* **7**, e44638 (2012).
52. G. R. Smith, Conjugational Recombination in Escherichia coli: Myths and Mechanisms. *Cell* **64**, 19 (Jan, 1991).
53. R. G. Lloyd, C. Buckman, Conjugational Recombination in Escherichia coli: Genetic Analysis of Recombinant Formation in Hfr X F(-) Crosses. *Genetics* **139**, 1123 (March 1, 1995, 1995).
54. D. B. F. Johnson, C. Wang, J. Xu, M. D. Schultz, R. J. Schmitz, J. R. Ecker, L. Wang, Release Factor One Is Nonessential in Escherichia coli. *ACS Chemical Biology*, (2012).
55. Y. Yamazaki, Niki, H., & Kato, J., Profiling of Escherichia coli Chromosome database. *Methods Mol. Biol.* **416**, 385 (2008).
56. T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, H. Mori, Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 11 (2006).
57. P. Funchain, A. Yeung, J. L. Stewart, R. Lin, M. M. Slupska, J. H. Miller, The consequences of growth of a mutator strain of Escherichia coli as measured by loss of function among multiple gene targets and loss of fitness. *Genetics* **154**, 959 (Mar, 2000).
58. R. M. Schaaper, R. L. Dunn, Spectra of spontaneous mutations in Escherichia coli strains defective in mismatch correction: the nature of in vivo DNA replication errors. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 6220 (1987).
59. E. Bi, J. Lutkenhaus, Cell division inhibitors SulA and MinCD prevent formation of the FtsZ ring. *J. Bacteriol.* **175**, 1118 (February 1, 1993, 1993).
60. Y. Ishihama, T. Schmidt, J. Rappsilber, M. Mann, F. U. Hartl, M. J. Kerner, D. Frishman, Protein abundance profiling of the Escherichia coli cytosol. *BMC Genomics* **9**, (Feb 27, 2008).
61. A. Shevchenko, H. Tomas, J. Havlis, J. V. Olsen, M. Mann, In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protocols* **1**, 2856 (2007).
62. E. S. Miller, E. Kutter, G. Mosig, F. Arisaka, T. Kunisawa, W. Rüger, Bacteriophage T4 Genome. *Microbiology and Molecular Biology Reviews* **67**, 86 (March 1, 2003, 2003).
63. R. Maldonado, A. J. Herr, Efficiency of T4 Gene 60Translational Bypassing. *J. Bacteriol.* **180**, 1822 (April 1, 1998, 1998).
64. C. E. Jones, T. C. Mueser, N. G. Nossal, Interaction of the Bacteriophage T4 Gene 59 Helicase Loading Protein and Gene 41 Helicase with Each Other and with Fork, Flap, and Cruciform DNA. *J. Biol. Chem.* **275**, 27145 (September 1, 2000, 2000).
65. Y. Chadani, K. Ono, S.-i. Ozawa, Y. Takahashi, K. Takai, H. Nanamiya, Y. Tozawa, K. Kutsukake, T. Abo, Ribosome rescue by Escherichia coli ArfA (YhdL) in the absence of trans-translation system. *Mol. Microbiol.* **78**, 796 (2010).

66. Y. Handa, N. Inaho, N. Nameki, YaeJ is a novel ribosome-associated protein in Escherichia coli that can hydrolyze peptidyl-tRNA on stalled ribosomes. *Nucleic Acids Res.*, (2010).
67. A. Fokine, Z. Zhang, S. Kanamaru, V. D. Bowman, A. A. Aksyuk, F. Arisaka, V. B. Rao, M. G. Rossmann, The Molecular Architecture of the Bacteriophage T4 Neck. *Journal of Molecular Biology* **425**, 1731 (2013).
68. I. J. Molineux, in *The Bacteriophages,* R. Calendar, Ed. (Oxford University Press, New York, 2006), pp. 277-301.
69. P. D. Sadowski, C. Kerr, Degradation of Escherichia coli B Deoxyribonucleic Acid After Infection with Deoxyribonucleic Acid-Defective Amber Mutants of Bacteriophage T7. *J. Virol.* **6**, 149 (August 1, 1970, 1970).
70. L. C. You, P. F. Suthers, J. Yin, Effects of Escherichia coli physiology on growth of phage T7 in vivo and in silico. *J. Bacteriol.* **184**, 1888 (Apr, 2002).