**Supporting Information Appendix**

# Adaptive evolution of genomically recoded *Escherichia coli*

Timothy M. Wannier[1†], Aditya M. Kunjapur[1†], Daniel P. Rice[2], Michael J. McDonald[2], Michael M. Desai[2], George M. Church[1*]

1 Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, NRB 238, Boston, MA 02115

2 Department of Organismic and Evolutionary Biology, Harvard University, 457 Northwest Labs, 52 Oxford Street, Cambridge, MA 02138

† These authors contributed equally to this work.

* Corresponding author: George M. Church, 77 Avenue Louis Pasteur, NRB 238, Boston, MA 02115, 617-432-7562, gchurch@genetics.med.harvard.edu

**SI Discussion**

**Mutations found to broadly affect all evolved populations**

We identified genes mutated multiple times across independent populations as putative targets of selection. For each lineage and for all lineages combined, we compared the distribution of the number of independent mutations observed per gene to the expectation under multinomial resampling (**Fig. S9).** In each comparison, the observed distribution has a much heavier right tail than the expected distribution, indicating enrichment in multiply mutated genes. Table S3 lists the 50 genes with at least four independent mutations in any lineage or at least six mutations across all lineages, while Table 1 lists a subset of these that are of interests to the discussion in this work. Many of these genes are common genetic targets in previous ALE studies or have been noted as being adaptive to minimal media or biofilm formation. In particular, 10 of these 50 genes were found previously during adaptive evolution of *E. coli* K-12 strains to minimal media. These include pykF*, rph*, *rpoB*, and *rpoC*, found in three studies from the Palsson laboratory (1–3); *fis*, *gltB*, *kup*, and again *pykF* found by the Lenski laboratory (4); *fimH* found by a Bolívar lab study (5), and *folA* found during C321 passaging by the Barrick Lab (6).

It has been previously documented that a frameshift mutation in the C-terminus of *rph* in wild-type K-12 strains causes pyrimidine starvation in minimal media. This is due to low *pyrE* expression because a premature *rph* stop codon impedes transcription/translation coupling, which supports optimal levels of *pyrE* expression past the intercistronic *pyrE* attenuator (7). Evolutionary studies from the Palsson lab identified a recurring 82-bp deletion in the regulatory region between *rph* and *pyrE*. In contrast, we in no cases see this same deletion, but instead see thirteen separate mutations to a 233-bp stretch extending from the C-terminus of *rph* through some intergenic space, and into the first 15-bp of the *pyrE* gene. The five most frequent mutations make up 86% of the total hits. Four of these are in the first attenuator stem, weakening it (**Fig. S10**), and the other, found in eight clonal isolates from four populations, is an exact reversion (3815879 TCC to TCCC) to restore the original wild-type locus found in the K-12 ancestor. Other mutations found are frameshifts in *rph* to extend its reading frame either by 10 amino acids to its wild-type UGA or a further 14 amino acids to a UAG stop codon, which we presume also relieves *pyrE* attenuation.

Of the other genes that we found in common with previous ALE studies, mutations to RNA polymerase subunits B and C (*rpoB* and *rpoC*) have been particularly well-documented. In one study, Conrad *et al.* characterized three *rpoB* mutations that were found during ALE to glucose minimal media, and determined that all variants were fitter in defined media, but less fit in rich media (1). Conrad *et al.* additionally found that these variants exhibited decreased open-complex longevity and greater processivity, which they proposed led to greater expression of metabolic genes, and lower rRNA expression. In this work we found eleven unique *rpoB* variants and 20 unique *rpoC* variants, all of which are mis-sense mutations, and none of which were documented in previous ALE studies. This suggests to us that the landscape for adaptive fitness gain in the two polymerase subunits is quite large, with many routes to switching between adaptation to LB and glucose minimal media.

Eight of the remaining 43 genes, including many of the most heavily-mutated genes, are involved in biofilm formation. A study by *Danese et al.* showed that gene *flu* (or *agn43*) is an abundant outer-membrane protein that is important for biofilm formation in glucose minimal media but not in rich media (8). The same study demonstrated that *fimH* is important for biofilm formation in both poor and rich media and showed the role of *oxyR* as a repressor of *flu*. A gene expression study found that *gltB* and *ompT* were expressed at levels 12- and 39-fold higher levels in biofilm growth compared to stationary phase (9). Purine biosynthesis has been shown to be important for biofilm formation in *E. coli* and other Gram-negative bacteria, with a study in *Burkholderia* specifically implicating *purL* and *purT*. Finally, *mdtJ* null strains were shown to have enhanced biofilm growth (10).

**Differences in stop codon effectiveness at translation termination**

The effectiveness of stop codons at coding for translation termination is affected by their identity (UAG, UGA, or UAA) and by sequence context (11–13). Suppression experiments suggest that UAA codons, which are recognized by both RF1 and RF2, are read more often by RF1 in *E. coli* K12 (14). Analyses in *E. coli* show that codon adaption index correlates positively with increasing UAA preference (15) and that most highly expressed genes end in UAA (16). In the absence of RF1, insufficient RF2 activity in C321.ΔA strains may impair fitness due to ribosome stalling at UAA or UGA stop codons. This would result in a smaller pool of free ribosomes, lower translational capacity, and increase the requirement for ribosome rescue by tmRNA-mediated trans-translation (17, 18) or by alternative ribosome-rescue factor (19, 20).

**Previous observations of *prfB* and *prfC* mutations resulting from defective translation termination at the ribosomal level**

Mutations to *prfB* and *prfC* also occur in response to translation termination impairment at the ribosomal level, especially where interactions occur with RFs. 23S ribosomal RNA (rRNA) is modified by the pseudouridine synthase RluD at positions that form a bridge between the 20S and 50S ribosomal subunits and interact with translation factors and tRNAs throughout protein synthesis (21–23). In Δ*rluD* K-12 strains, suppressor-containing derivatives rapidly arose from slow growing cultures (22, 23). One study identified and characterized one suppressor strain containing PrfB_E172K and found that wild-type levels of termination were restored at UAA codons after *rluD* inactivation (22). Another study investigated 21 additional suppressor strains (23). Of these strains, 15 contained point mutations in *prfB* (resulting in amino acid changes of D131Y, E157K, E167K, E172A, Q290K, A293T, and A293V) and 5 contained point mutations in *prfC* (T30I, T99I, V119F, V119G, and D313Y). Interestingly, none of these positions were found to be effected in our adaptive evolution study. However, deliberate replacement of the K-12 *prfB* gene with the *E. coli* B *prfB* (PrfB_T246A) and subsequent inactivation of *rluD* resulted in negligible impairment to growth and translation termination (23).

**Effect of media composition on strains containing release factor mutations**

Minimal media can enhance the severity of release factor mutations. One set of investigations found that PrmC (an $N^5$-glutamine methyltransferase that enhances the activity of RF1 and RF2) inactivation in *E. coli* imparts growth defects, particularly in minimal media or with poorer carbon sources, due in part to translational read-through of stop codons (24–26).

Strategies to improve protein expression in evolved recoded strains

Our data indicate that recoded strain capacity for protein overproduction in glycerol minimal media was significantly impaired, but that ALE here restored this deficit. Evolution for successive passaging fitness, should however not have selected for protein production, but for improved growth rate, and therefore we believe that rational genome engineering can be used to further improve protein expression in these strains. Well-documented targets include proteases encoded by *lon* (27) and *ompT* (28), as well as nucleases encoded by *rnb* (29), *rne* (30), and *endA* (31).

Others have attempted to engineer some of these targets in C321.ΔA and observed increased protein expression (32). While the expression of nsAA-containing proteins may also benefit from these approaches, nsAA incorporation faces other limitations that ought to be addressed, such as limited activity and selectivity in the engineered orthogonal aminoacyl-tRNA synthetases. An unusually selective but not highly active synthetase was used for nsAA incorporation in this study. Rational engineering or directed evolution approaches can be used to generate synthetases with improved attributes in either or both categories.

**SI Figures**

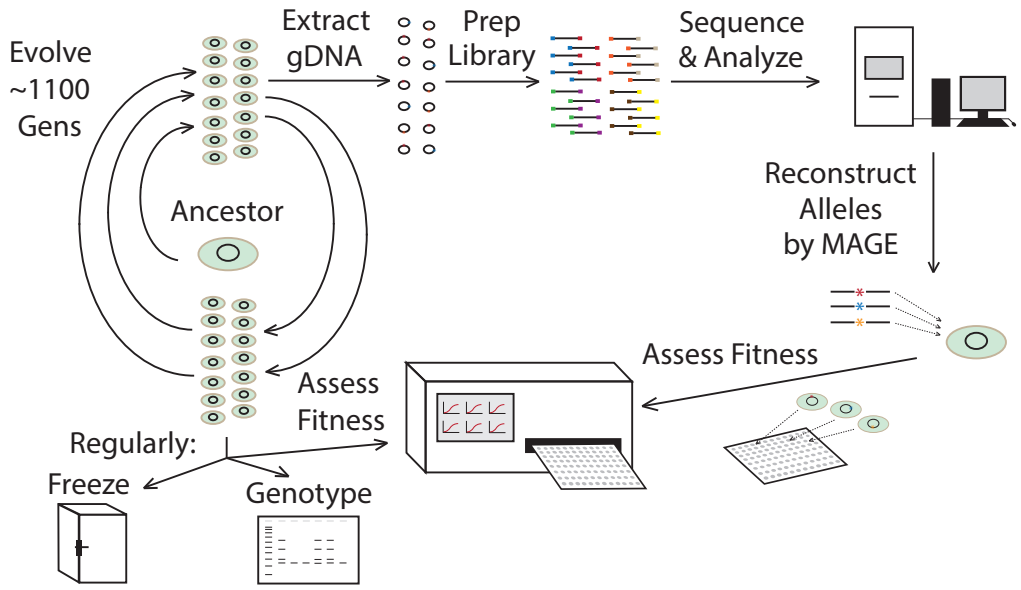


**Figure S1.** Project workflow for long-term adaptive laboratory evolution (ALE) of recoded *Escherichia coli*. Briefly, strains were passaged in serial batch cultures as 14 independent replicates for 1,000+ generations with routine genotyping (via allele-specific PCR and gel electrophoresis), storage as glycerol stocks, and growth rate analysis by plate reader. Two clones from each population at the final time point were used for genomic DNA preparation and Illumina sequencing library preparation. Sequencing was performed, data was analyzed to identify variants, and MAGE was used to reconstruct alleles in ancestral strains.

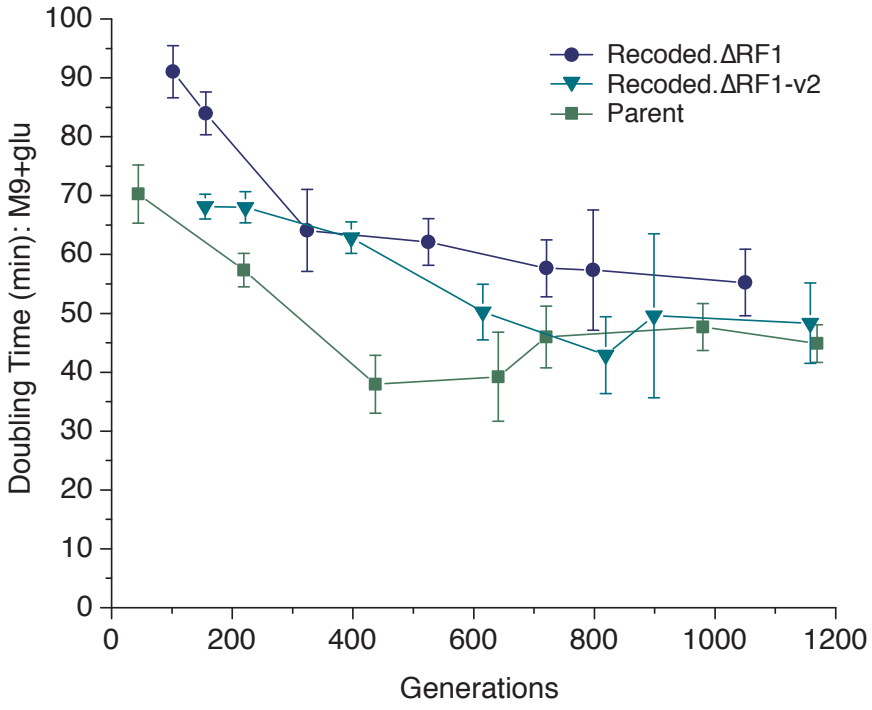


**Figure S2.** Trajectories of strain doubling time as a function of generation during evolution. This data was obtained by sampling individual colonies from all 14 replicate populations for each strain.

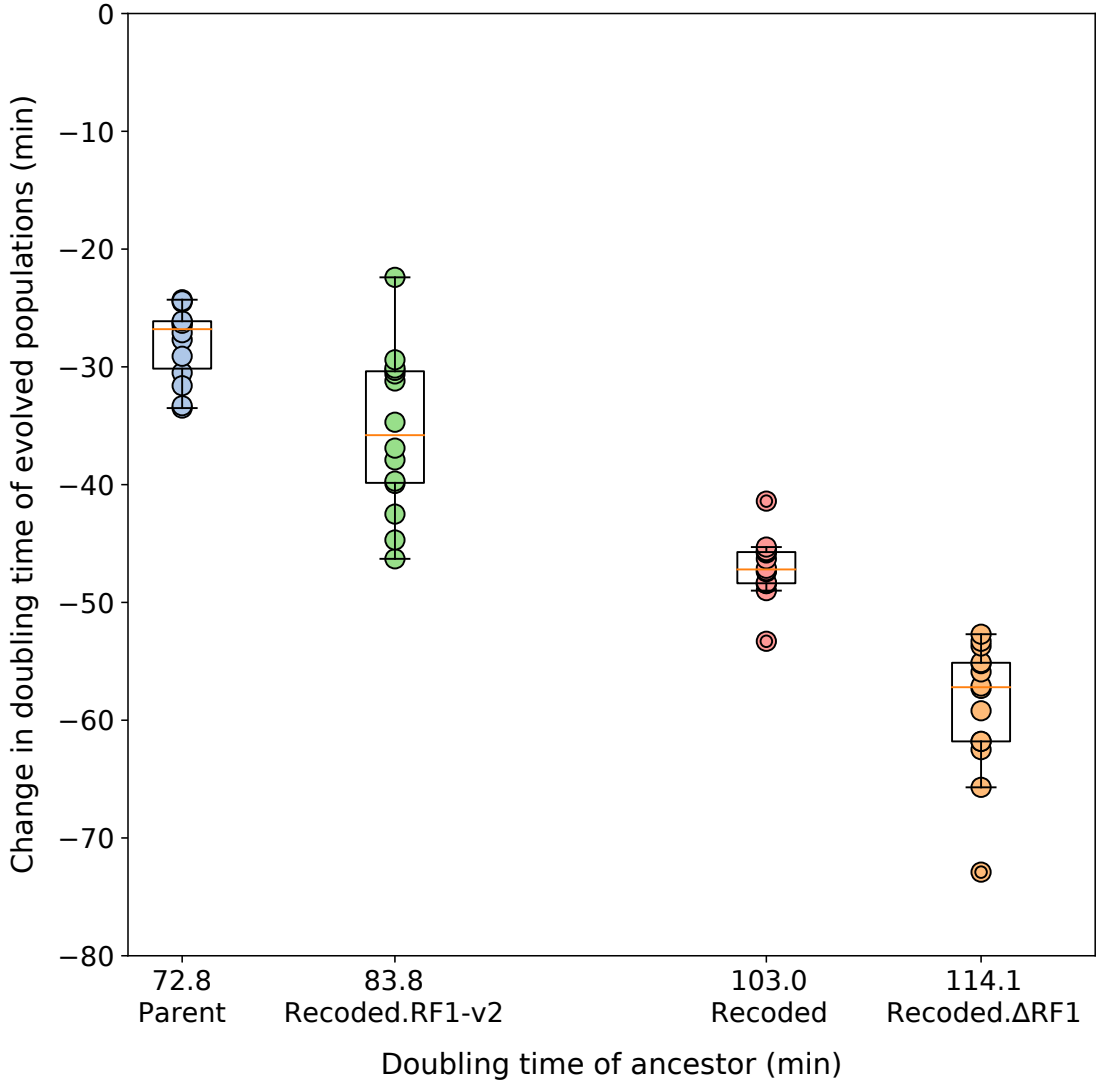


**Figure S3.** The change in doubling time of evolved populations in M9 + 1% glucose inversely correlates with the doubling time of the variants' ancestors.

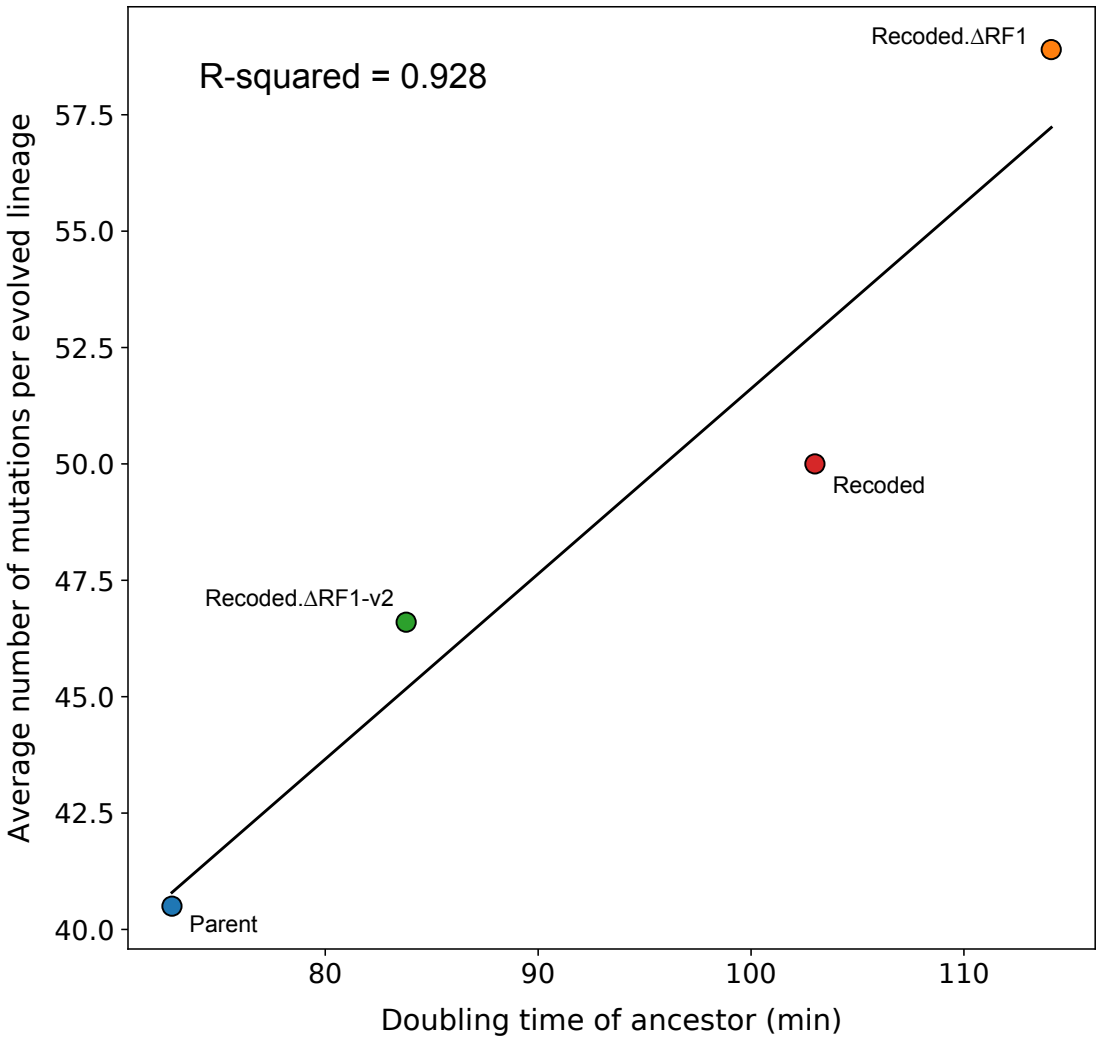


**Figure S4.** Relationship between the doubling time of the ancestral strain and the average number of mutations incorporated in the evolved lineages.

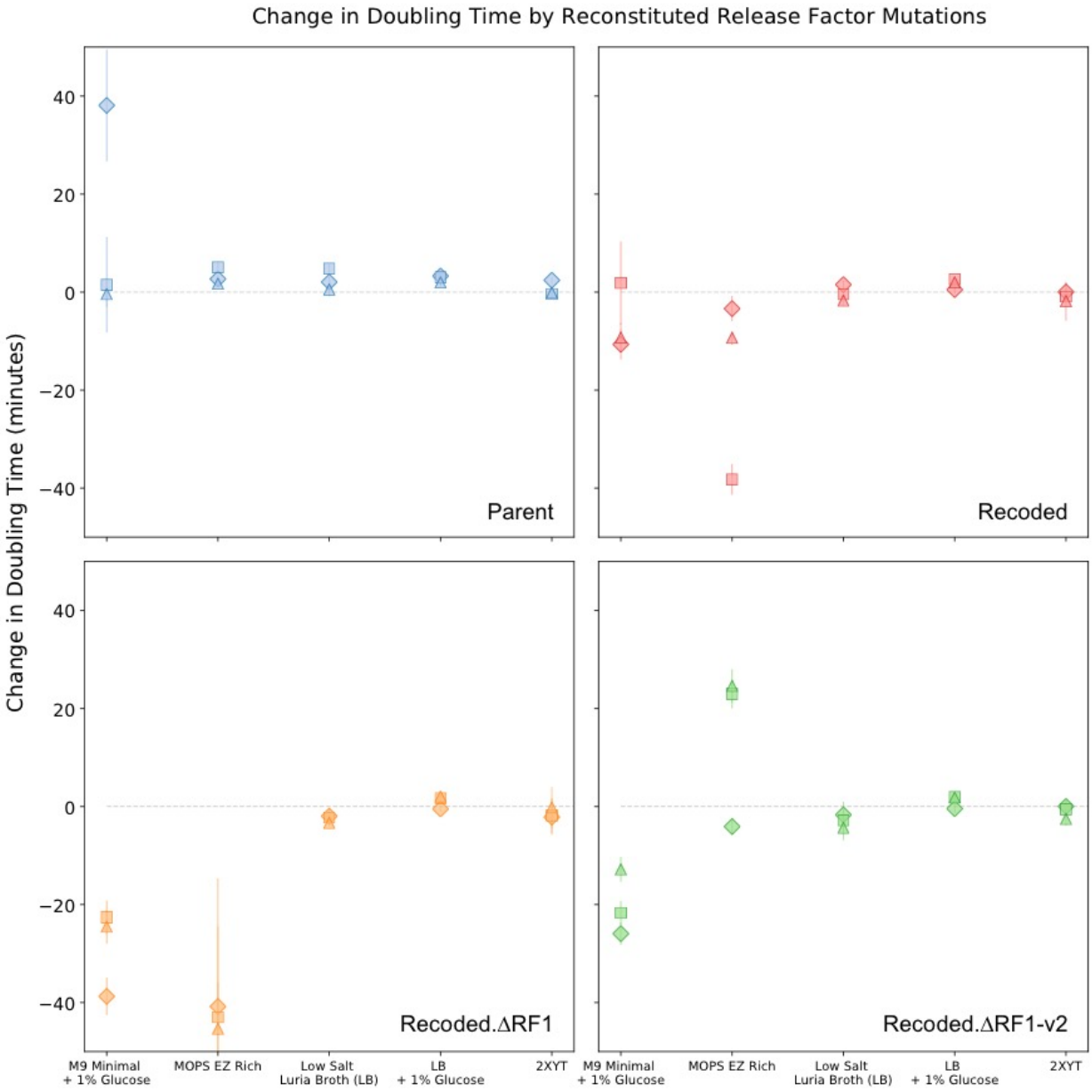


**Figure S5.** Change in the doubling time of ancestral lineages with various RF mutations in five different media.

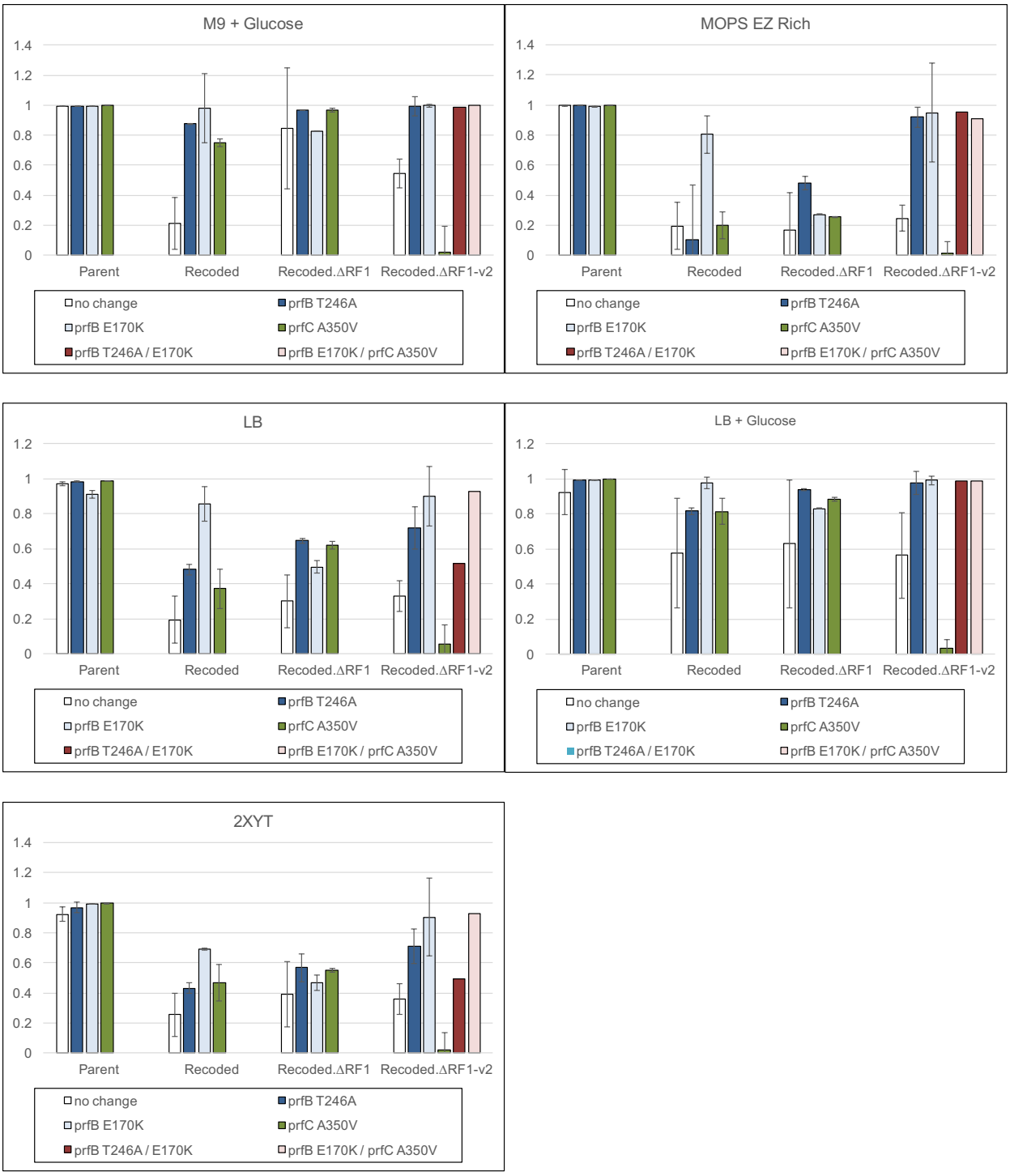


**Figure S6.** Results from a competition assay of strains containing release factor variants. The percentages in the figure indicate the percent of cells that were non-fluorescent after a 24-hour head-to-head competition, in which each strain was competed against a fluorescent reference strain. These measurements were made in triplicate.

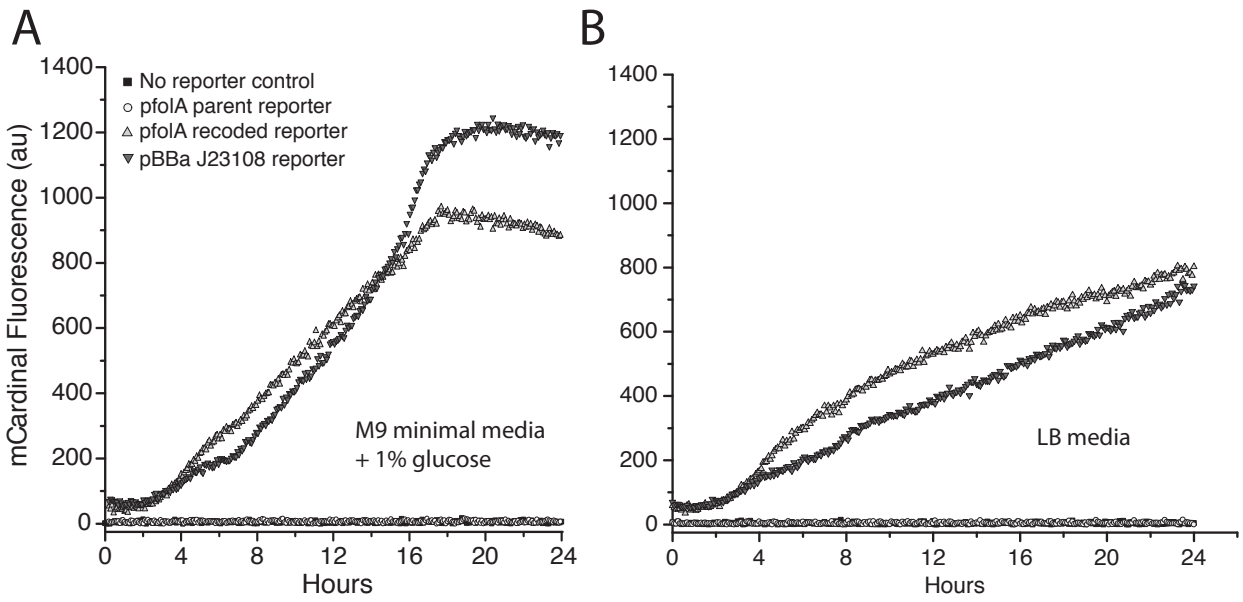


**Figure S7.** Effect of the *folA* promoter mutation on fluorescence resulting from fluorescent protein gene expression in (A) M9 minimal media + 1% glucose or (B) LB media.

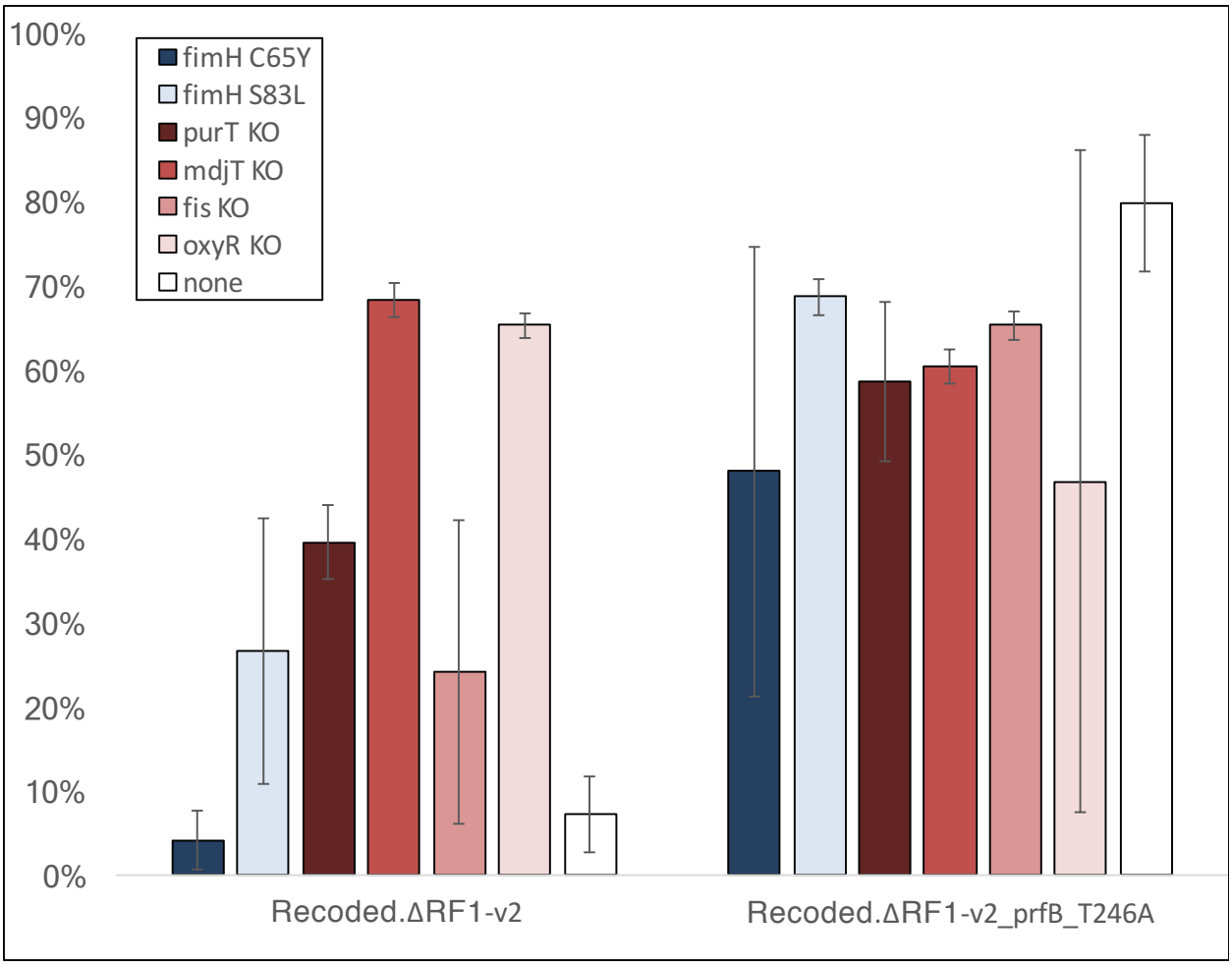


**Figure S8.** Head to head competition assay results for strains containing other mutations of interest. The percentages in the figure indicate the percent of cells that were non-fluorescent after a 24-hour head-to-head competition, in which each strain was competed against a fluorescent reference strain. These measurements were made in triplicate.

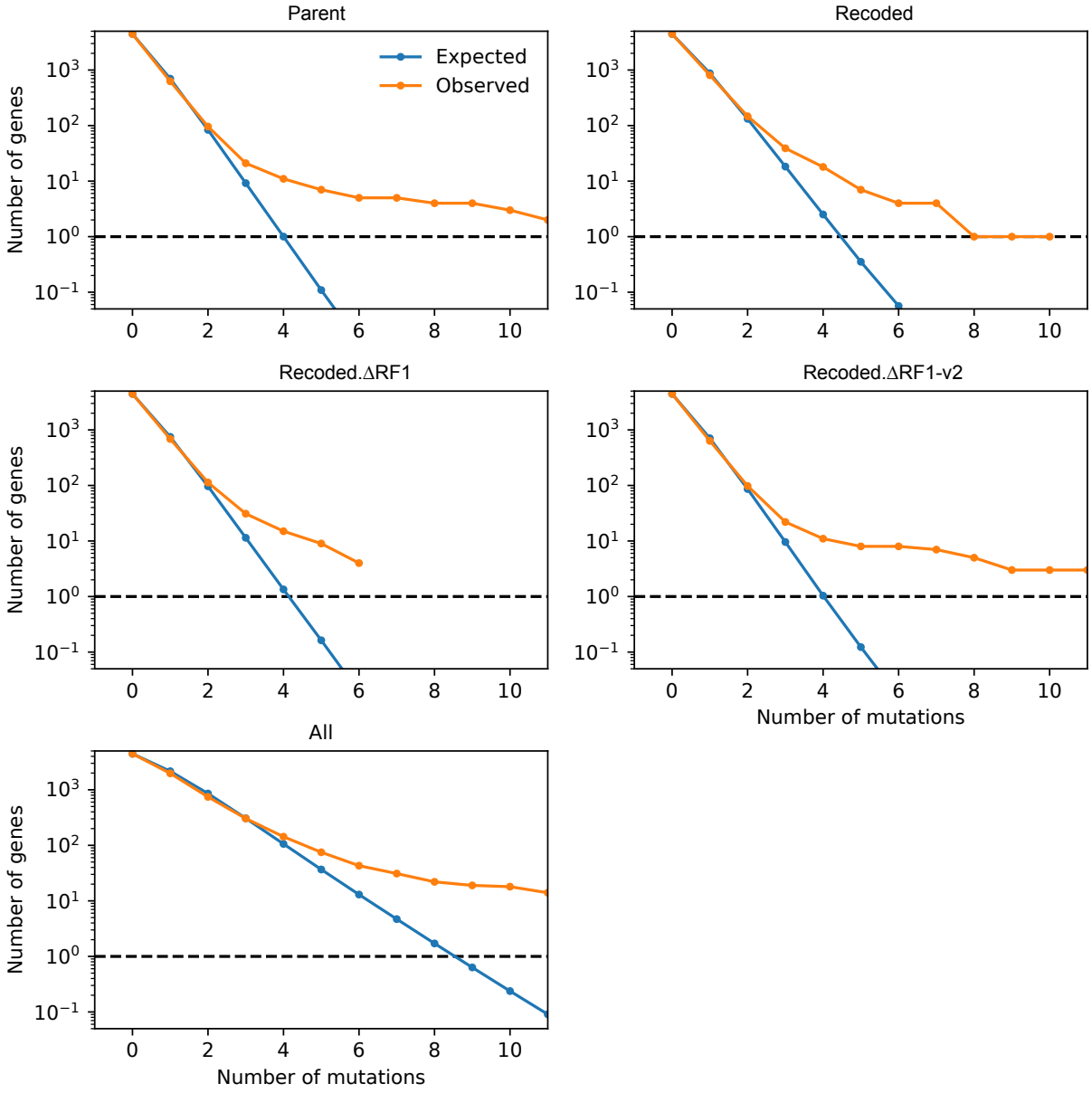


**Figure S9.** Complementary cumulative distributions of the number of mutations per gene. The orange curves show the number of genes in which we observed at least n independent mutations at generation 1,100 as a function of n, within each lineage and across all populations. The blue curves show the average of 10,000 multinomial re-samplings with the total number of mutations fixed and expected number of mutations per gene proportional to gene length.

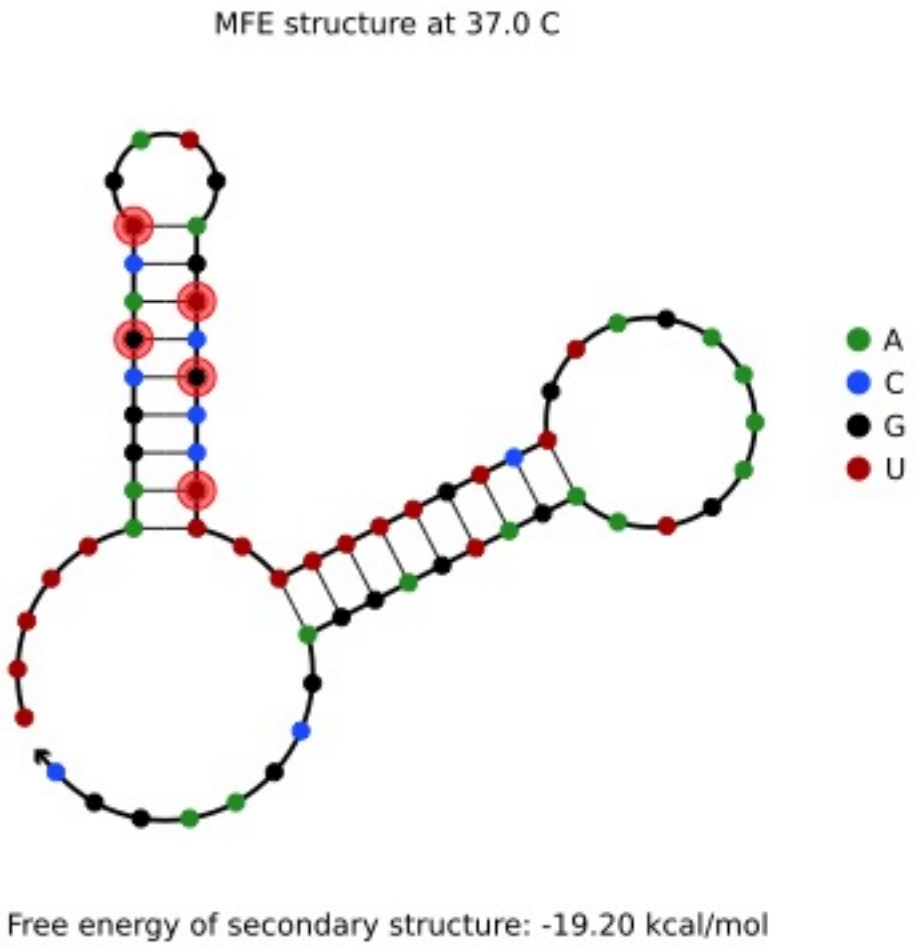


**Figure S10.** Predicted secondary structure of the attenuator loop between *rph* and *pyrE*. Sites that were observed to mutate in the evolved populations are highlighted in red. Nupack was used to simulate the RNA structure at 37°C in normal physiological conditions.

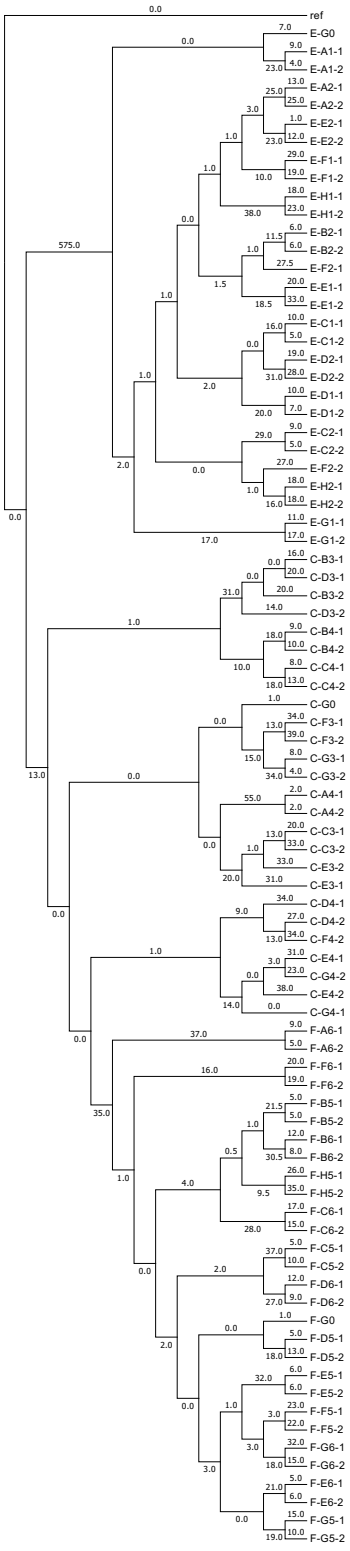


**Figure S11.** Maximum parsimony phylogenetic tree detailing all potential well-well cross-seeding events.

**SI Materials and Methods**

Culture Conditions:

A 1X M9 salt medium (M9+glucose) containing 6.78 g/L $Na_2HPO_4 \cdot 7H_2O$, 3 g/L $KH_2PO_4$, 1 g/L $NH_4Cl$, and 0.5 g/L NaCl, supplemented with 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 1% glucose, trace elements, biotin, and carbenicillin was used as the culture medium for evolution and most experiments in this study. The trace element solution (100X) used contained 5 g/L EDTA, 0.83 g/L $FeCl_3 \cdot 6H_2O$, 84 mg/L $ZnCl_2$, 10 mg/L $CoCl_2 \cdot 6H_2O$, 13 mg/L $CuCl_2 \cdot 2H_2O$, 1.6 mg/L $MnCl_2 \cdot 2H_2O$ and 10 mg/L $H_3BO_3$ dissolved in water (33, 34). Trace element solution was added to a concentration of 1X to decrease variability due to metal content in water.

The following media are also used in this study: (i) M9+glycerol medium, which is identical to M9+glucose except with 1% glycerol substituted for 1% glucose; (ii) MOPS EZ Rich medium, which is a commercially available defined medium (Teknova, Cat. No. M2105); (iii) Low Salt (0.5g/l NaCl) LB Lennox (LB) medium (; (iv) Low Salt LB supplemented with 1% glucose (LB + glucose).

Adaptive evolution:

For adaptive evolution, ancestral strains were streaked on LB agar plates, and colonies were used to inoculate 14 independent 1 mL culture volumes in glucose minimal media. Cells were typically passaged during late exponential phase at a dilution of 1:1,000 in 1 mL culture volumes incubated in deep 96-well plates at 34 C and 400 rpm. The dilution and passaging rate varied with the strains' growth, meaning that strains were initially diluted between 1:100 and 1:500 once per day (~7-9 generations per day), and then gradually ramped up to a maximum passaging rate of two 1:1,000 dilutions each day (~20 [$\log_2(2{,}000)$] generations per day as their fitness increased.

Sequencing library preparation:

After evolution, two clones were isolated from each population for genomic DNA purification and for next-generation sequencing using an adapted form of the Illumina transposase library preparation (35). Barcoded libraries were pooled and run in an Illumina HiSeq 2500.

Genome resequencing and analysis:

Genome resequencing and analysis was performed in parallel with two analytical pipelines to focus on multiply-mutated genes or on individual single nucleotide polymorphisms (SNPs). These analyses were conducted respectively (i) using Millstone (36) and Python scripting; or (ii) using a group of widely-distributed sequence-analysis tools and Python scripting. (i) Millstone is a custom software suite built by our lab for the purpose of whole genome mutational analysis. Illumina reads are aligned to a reference genome (NC_000913.3), and a list of variants is output that was then analyzed with custom scripting in Python 3.7. Code-base is available online at http://churchlab.github.io/C321evolve/. In brief, variants were culled based on minimum read depth and allele frequencies, multiply-hit genes were identified, cross-contamination events were sorted out, and a final list of genes was assembled that saw at least three unique mutational episodes in any of the four strains. (ii) A custom analytics pipeline was run in parallel to the Millstone analysis, using bowtie2 version 2.1.0 (37) to align reads to the reference genome and Picard version 1.44 to mark duplicate reads. A permissive list of candidate SNPs and indels was generated by applying the GATK UnifiedGenotyper (38) version 2.3 to all samples at once, with the minimum phred-scaled confidence threshold set to its most permissive value, 4.0. The list of candidate mutations was then filtered by excluding all sites with more than one alternate allele and all indels longer than one base pair because these sites are less likely to contain true variants and harder to call confidently. To ensure adequate coverage, it was required that the site: (1) have an average coverage depth of at least 10x across samples, (2) have zero coverage depth in no more than one sample, and (3) have at least ten total reads supporting the alternate allele. Because the samples are haploid, we required the alternate allele to be supported by at least 85% of the reads in at least one sample. Finally, read were filtered on GATK's quality score and strand bias fields (requiring QUAL > 50 and FS < 40).

Gene ontology analysis:

The gene ontology (GO) of all mutations was analyzed with reference to a set of standard categorizations for *E. coli* (39).

Neutral mutation rate analysis:

We briefly investigated whether faster adaptation of less-fit strains could be due to an elevated mutation rate. If this were the case, we would expect the C321 lineages to have an elevated mutation rate compared to the

ECNR2 lineages. According to theory, the number of *neutral* mutations in a clone relative to its ancestor is Poisson distributed with mean equal to the neutral mutation rate times the number of generations. In order to use this Poisson model, we have to restrict ourselves to independently-evolved clones, i.e. clones whose most recent common ancestor is the ancestor of all strains. This means that we (1) can use no more than one clone per population, and (2) can use no more than one clone per set of populations that are connected by cross-contamination/re-seeding events.

With these considerations, we selected six clones from the C321 (C) populations and six clones from the ancestral ECNR2 (E) populations according to the following rules. Using the maximum parsimony tree, we identified all clades supported by more than one mutation, as well as one C clade supported by a single prfB mutation. Then, within each of these clades, we picked the clone whose ID comes first in alphanumeric order. This ensures that we picked clones at random with respect to how many mutations they had. For each of these twelve clones, we then counted the number of single-hit synonymous mutations along the lineage from Generation 0 to the clone. The clones and mutation counts are included below.

Adding the counts of synonymous mutations from the clones from each lineage, we count 64 mutations per clone per 6K generations in the ECNR2 lineage and 78 mutations per clone per 6K generations in the C321 lineage. If synonymous mutations accumulate as a Poisson process average rate of 71 mutations per clone per 6K generations, we'd expect the difference between the counts in the two lineages to have a standard deviation of ~12 mutations, compared to an observed difference of 14 mutations. This suggests, that the data is consistent with the hypothesis that the two lineages are accumulating synonymous mutations at the same rate.

| Clone | Synonymous mutations |
|---|---|
| C-A4-1 | 13 |
| C-B3-1 | 12 |
| C-C3-1 | 18 |
| C-D4-1 | 10 |
| C-E4-1 | 14 |
| C-F3-1 | 11 |
| E-A1-1 | 11 |
| E-A2-1 | 14 |

| | |
|---|---|
| E-C1-1 | 7 |
| E-F1-1 | 11 |
| E-G1-1 | 7 |
| E-H1-1 | 14 |

Stop codon analysis:

To account for all possible stop codon changes, we considered three potential events: (i) direct conversion from one stop codon to another; (ii) read-through effects of mutating a stop codon into a sense codon; and, (iii) frameshift mutations close to the end of a protein-coding gene which would have the effect of switching termination from the wild-type stop codon to a new position. We only considered frameshift mutations that both occurred within 100bp (~33 amino acids (AA)) of the 3' end of a gene and that did not add more than 200bp (~66 AA) of additional length to the end of a gene. Based on our observations, we decided to exclude certain stop codons from further analysis because of when or where they occurred. We saw only three instances of direct UAA to UAG reversions: to *hemA* in C321 during the re-introduction of *prfA* to the strain in preparation for long-term passaging, and to the pseudogenes *eaeH* and *ychE* during passaging of the C321 (*prfA*+) and C321.ΔA-v2 lineages. We exclude these two latter reversions from our downstream analysis because they both occur in pseudogenes, and so are unlikely to be of great significance. Using the broader criteria described above, we catalogue 39 unique instances of stop codon changes (**Table S1**). Three changes are excluded from further analysis because they occur in pseudogenes, and five changes to *rph* and one to *wzzE* are ignored because they are either direct or indirect repairs of prior frameshift mutations occurring in either the wild-type K-12 strain, in the case of *rph*, or the C321 ancestor during *prfA re-introduction*, in the case of *wzzE*, This leaves 30 total termination codon changes, 9 of which are due to a point mutation to the termination codon itself, and 21 of which are due to frameshift mutations.

Analysis of parallel evolution:

Because there were several cross-contamination events between wells over the course of the evolution, we used phylogenetic methods to estimate the number of independent occurrences of each mutation. A sequence

alignment of all single base pair substitutions in all samples was constructed, and from this we found the maximum parsimony sample genealogy using MEGA7 version 7.0.16 (40) **(Fig. S11)**. This genealogy was then used to identify the minimum number of distinct mutational events required to explain each sample genotype. This method is conservative with regard to identifying parallel evolution because the maximum parsimony genealogy may incorrectly group populations that acquired the same mutations into clades. Table S4 shows the mutations that we infer occurred more than one time within the c321 and/or c321-v2 populations, but which we did not observe in the ECNR2 populations. To identify parallel evolution at a higher level of biological organization, we counted the number of independent mutational events in genes in each GO process category. Table S5 shows the number of non-synonymous mutations in each GO category for populations descended from each of the c321, c321-v2, and ECNR2.

Allele reconstructions:

MAGE was used to reconstruct mutations of interest in the ancestral strains. Saturated overnight cultures were diluted 100-fold into 3 mL LB containing appropriate antibiotics and grown at 34 °C until mid-log. The integrated Lambda Red cassette in C321.ΔA derived strains was induced in a shaking water bath (42 °C, 300 rpm, 15 minutes), followed by incubation on ice for at least two minutes. The cells were then made electrocompetent at 4 °C by pelleting 1 mL of culture (16,000x rcf, 20 seconds) and washing twice with 1 mL ice cold deionized water (dH2O). Electrocompetent pellets were resuspended in 50 µL of dH2O containing the desired DNA. For MAGE oligonucleotides, 5 µM of each oligonucleotide was used. Please see Table 2 for a list of all oligonucleotides used in this study. For integration of dsDNA cassettes, 50 ng was used. Allele-specific colony PCR was used to identify desired colonies as previously described (41). Colony PCR was performed using Kapa 2G Fast HotStart ReadyMix following manufacturer protocols and Sanger sequencing was performed by Genewiz to verify strain engineering.

Doubling time measurement:

To assess fitness by measuring doubling times, strains were grown in triplicate in transparent flat-bottom 96-well plates (150 µL, 34 °C, 300 rpm). Kinetic growth ($OD_{600}$) was monitored on Eon plate readers at 5 minute intervals. After blanking reads by subtracting wells containing only media, 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 $\ln(OD_{600})$. The slope was calculated based on the linear regression of $\ln(OD_{600})$ through 21 contiguous time points (60 minutes).

Head-to-head competition assays:

Fitness was also evaluated using head-to-head competition assays featuring a fluorescent reference strain. To construct the fluorescent reference strain we transformed wild-type MG1655 with a pET-plasmid vector containing mCardinal (42) under control of a strong synthetic promoter (Anderson: BBa_J23108). Competition assays were set up by growing cultures separately overnight, then mixing an equal amount of fluorescent reference strain with the strain whose fitness is to be measured at dilutions of 1:100 for both strains. This was done in plate-format, grown 24 hours in a shaker at 30°C / 300RPM, and assayed in a BD LT Fortessa flow cytometer, using the laser and filter settings for the Cy5 dye. The percentage of fluorescent cells was calculated after analysis in FloJo, and used as an inverse measurement of the fitness of the assayed strain.

nsAA incorporation assays:

We used the bipyridylalanine aminoacyl-tRNA synthetase (BipyARS) and *Methanococcus jannaschii* $tRNAopt_{CUA}$ as our OTS (43), which was recently shown to be more orthogonal to the native 20 amino acids than related engineered *M. jannaschii* tyrosyl-tRNA synthetase variants (44). As reporter proteins, we used ubiquitin fused N-terminally to super-folder GFP. The positive expression control contained no UAG sites (0-UAG), while the other second construct contained two UAG codons (2-UAG), the first between the two domains and the second internal to the GFP at position 151 (44). We measured GFP fluorescence normalized to culture optical density (FL/OD) in the presence and absence of p-Acetyl-phenylalanine (pAcF). To avoid catabolite repression of the arabinose inducible promoter system that regulates BipyARS, we used glycerol minimal media for these experiments.

Strains harboring GFP reporters and BipyARS/tRNA plasmids were inoculated from frozen stocks in biological triplicate and grown to confluence overnight in deep well plates. Experimental cultures were inoculated at 1:10 dilution in glycerol minimal media supplemented with chloramphenicol, arabinose, and pAcF. Cultures were incubated at 34 °C to an $OD_{600}$ of 0.5–0.8 in a shaking plate incubator at 400 rpm (~3-4 h). GFP expression was

induced by addition of anhydrotetracycline, and cells were incubated at 34 °C for an additional 16-20 h before measurement.

Assays were performed in 96-well plate format. Cells were centrifuged at 5,000*g* for 3 min, washed with PBS, and resuspended in PBS after a second spin. GFP fluorescence was measured on a Biotek spectrophotometric plate reader using excitation and emission wavelengths of 485 and 525 nm (Gain = 80). Fluorescence signals were corrected for autofluorescence as a linear function of $OD_{600}$ using the parent C321.ΔA strain that does not contain a reporter. Fluorescence was then normalized by the $OD_{600}$ reading to obtain FL/OD.

**SI Table 1.** All mutations that affect stop codons.

| Position | Occurrence | Type* | Gene | REF | ALT | Ancestral codon | New codon | Δ Tail length | Lineage** | Reason Excluded |
|---|---|---|---|---|---|---|---|---|---|---|
| 115797 | evolution | fs | hofB | CAAAAAG | CAAAAG | TAA | TAG | -64 | P-5-2 | - |
| 183606 | C321 ancestor | fs | cdaR | GAT | GT | TAA | TAG | 55 | CFP | - |
| 232548 | evolution | p | yafE | A | G | TAA | TGA | 0 | P-12-1 | - |
| 315244 | evolution | p | pseudogene | G | - | TAA | TAG | 0 | F-F6-1 | pseudogene |
| 474302 | evolution | p | tesB | T | A | TAA | TGA | 18 | C-D4-1 | - |
| 578096 | evolution | fs | rrrD | TGGGGGC | TGGGGGGC | TGA | TAG | 460 | E-E2-1 | nonsense tail |
| 635369 | evolution | fs | ybdN | TCCCCCA | TCCCCCCA | TAA | TAG | 600 | P-9-1 / C-A4-1 / C-A4-2 | nonsense tail |
| 675600 | C321 ancestor | fs | ybeQ | CTG | CG | TAA | TGA | 106 | CFP | - |
| 821493 | evolution | fs | ybhM | GTTTTTTC | GTTTTTTTC | TAG | TAA | 32 | E-B2-1 | - |
| 1006412 | C321 ancestor | fs | zapC | TGG | TG | TAA | TGA | -61 | CFP | - |
| 1084580 | evolution | fs | efeB | ATTTTTTG | ATTTTTTTG | TAA | TAA | 17 | C-D3-1 | no change |
| 1189846 | evolution | fs | phoP | ATTTTTTG | ATTTTTTTG | TGA | ATC | 931 | C-C4-2 / F-B6-1 | nonsense tail |
| 1199759 | evolution | fs | intE | GTTTTTG | GTTTTTTG | TGA | TAA | -69 | F-G5-2 / F-G5-1 | - |
| 1208533 | C321 ancestor | fs | tfaE | TCA | TA | TAA | TAA | 279 | CFP | nonsense tail |
| 1264970 | prfA+ ancestor | p | hemA | G | - | TAA | TAG | 0 | P | - |
| 1417538 | evolution | fs | racC | CTTTTTTTTC | CTTTTTTTC | TGA | TAA | -31 | E-G1-1 | - |
| 1423525 | evolution | fs | rzoR | TCCCCCCG | CCCCCCCG | TAA | TAA | 0 | P-13-1 / P-12-2 / P-3-1 / P-6-2 | rzpR fix |
| 1423525 | prfA+ ancestor | fs | rzoR | TCCCCCCG | TCCCCCCCG | TAA | TAA | 60 | P | no change |
| 1639535 | evolution | fs | rrrQ | TCCCCCA | TCCCCCCA | TGA | TAG | 492 | C-F3-2 | nonsense tail |
| 1655355 | F11 ancestor | fs | ynfA | GCCCCCA | GCCCCA | TAA | TAA | 23 | F | no change |
| 1660484 | evolution | fs | ynfE | GAAAAAAT | GAAAAAAAT | TAA | TAA | 52 | P-14-2 | no change |
| 1667344 | evolution | p | mlc | T | C | TAA | TAG | 0 | E-F1-1 | - |
| 1757108 | evolution | p | pykF | T | C | TAA | TAA | 81 | E-A2-1 / E-A2-2 | no change |
| 1757751 | C321 ancestor | fs | ldtE | CTCTGGT | CT | TAA | TAG | 80 | CFP | - |
| 1931971 | evolution | fs | purT | TGGGGGT | TGGGGGGT | TAA | TGA | -52 | F-D5-1 / F-D5-2 / F-G5-1 / F-G5-2 | - |
| 1967130 | evolution | fs | cheY | GCCCCCG | GCCCCCCG | TGA | TGA | -17 | E-A2-1 | no change |
| 1968506 | evolution | p | cheR | A | C | TAA | TGA | 6 | P-11-2 | - |
| 2024571 | evolution | fs | rcsA | ATTTTTG | ATTTTG | TAA | TGA | 10 | F-F6-2 | - |
| 2186661 | evolution | fs | rcnA | GCCCCCT | GCCCCCCT | TAA | TAA | 559 | F-C5-1 | nonsense tail |
| 2241812 | evolution | p | yeiB | A | G | TAA | TAA | 36 | C-C4-1 | no change |
| 2297107 | evolution | fs | ccmA | TCCCCCCCT | TCCCCCCCCT | TGA | TGA | -48 | F-E6-1 | no change |
| 2367984 | evolution | fs | arnC | GCCCCCG | GCCCCCCG | TAA | TGA | -4 | P-11-2 / P-8-1 / E-F1-2 | - |
| 2370019 | C321 ancestor | p | arnA | T | C | TGA | TAG | 9 | CFP | - |
| 2770672 | evolution | fs | ypjK | GAAAAAAT | GAAAAAT | TAA | TAA | 9 | P-11-1 | no change |
| 2824495 | C321 ancestor | fs | mltB | TGT | TT | TAA | TAA | 66 | CFP | no change |
| 2984447 | evolution | fs | yqeF | GCCCCCG | GCCCCG | TAG | TGA | -20 | E-E1-1 | - |
| 3003473 | evolution | fs | xdhB | CGGGGGGA | CGGGGGGGA | TGA | TAA | 476 | C-B3-1 / P-13-2 | nonsense tail |
| 3019150 | evolution | fs | ygfK | GAG | GAAG | TAA | CGC | 994 | C-D4-2 | nonsense tail |
| 3092826 | evolution | p | gshB | T | C | TAA | TGA | 36 | C-G4-2 / C-E4-1 | - |
| 3134810 | C321 ancestor | fs | yghT | GTG | GG | TAA | TAA | 50 | CFP | no change |
| 3269600 | evolution | p | yhaC | T | C | TAG | TAA | 6 | E-G1-1 | - |
| 3330353 | C321 ancestor | fs | dacB | GTT | GTTT | TAA | TGA | -37 | CFP | - |
| 3365664 | evolution | p | pseudogene | G | - | TAA | TAG | 0 | P-5-1 / P-8-2 | pseudogene |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 3411546 | evolution | fs | fis | GAAAAAAT | GAAAAAT | TAA | TGA | -6 | C-G3-2 / C-F3-1 / C-G3-1 / C-F3-2'] | - |
| 3411565 | evolution | p | fis | T | C | TAA | TGA | 102 | C-B3-2 / C-B3-1 / C-D3-2 / C-D3-1 | - |
| 3418135 | evolution | p | acrF | T | C | TAA | TAA | 63 | E-H2-2 | no change |
| 3598557 | C321 ancestor | p | livJ | A | G | TAA | TAA | 39 | CFP | no change |
| 3799382 | evolution | fs | waaZ | ATTTTTTTTC | ATTTTTTTC | TAA | TAA | 90 | F-G6-2 / P-12-2 / F-G6-1 / P-3-2 | no change |
| 3815879 | evolution | fs | pseudogene | TCC | TCCC | TAA | TGA | 28 | C-G4-1 / C-G4-2 / P-14-1 / P-10-2 / C-E4-1 / W72 / C-E4-2 / P-8-1 / W62 / P-4-2 | rph fix |
| 3815879 | evolution | fs | pseudogene | TCC | TC | TAA | TAG | 69 | F-C6-2 / F-C6-1 | rph fix |
| 3815907 | evolution | fs | pseudogene | GTA | GA | TAA | TAG | 68 | F-E6-1 / F-E6-2 | rph fix |
| 3815916 | evolution | fs | pseudogene | CAT | CT | TAA | TAG | 68 | F-F5-2 / F-F5-1 | rph fix |
| 3815943 | evolution | fs | pseudogene | TCC | TC | TAA | TAG | 68 | F-G5-1 / F-G5-2 | rph fix |
| 3862515 | C321 ancestor | fs | pseudogene | GCC | GC | TAA | TGA | -8 | CFP | pseudogene |
| 3862523 | C321 ancestor | fs | pseudogene | TGCCC | TC | TAA | TAG | -2 | CFP | redundant |
| 3867032 | evolution | fs | ibpA | GTTTTTTC | GTTTTTTTC | TAA | TAA | 6 | C-F4-2 | no change |
| 3902308 | evolution | fs | bglB | ATTTTTTTTA | ATTTTTTTA | TAA | TAA | 39 | E-B2-2 / F-B6-2 | no change |
| 3938217 | evolution | fs | rbsK | GAG | GAAG | TGA | TAA | 75 | C-C3-2 / C-C3-1 / C-E3-1 / C-E3-2 | - |
| 3950463 | evolution | fs | ilvX | CGGGGGT | CGGGGT | TAA | TAA | -67 | E-C1-2 / E-C1-1 | no change |
| 3950463 | evolution | fs | ilvX | CGGGGGT | CGGGGGTT | TAA | TGA | -79 | E-F1-2 / E-E1-2 | - |
| 3970026 | evolution | fs | wzzE | CGGGGGGC | CGGGGGGGGC | TGA | TGA | -39 | P-13-1 / P-4-2 / P-1-2 / F-H5-1 / C-E4-1 | wzzE fix |
| 3970026 | evolution | fs | wzzE | CGGGGGGC | CGGGGGGGGGGGC | TGA | TGA | -36 | C-C3-2 / C-C3-1 | wzzE fix |
| 3970026 | evolution | fs | wzzE | CGGGGGGC | CGGGGGGGGGGC | TGA | TGA | 21 | C-E3-1 / C-E3-2 | wzzE fix |
| 3970026 | evolution | fs | wzzE | CGGGGGGC | - | TGA | TAG | 0 | C-E4-2 | wzzE fix |
| 3970026 | C321 ancestor | fs | wzzE | CGGGGGGC | CGGGGGGGC | TAA | TGA | 18 | CFP | - |
| 3980619 | evolution | fs | wecG | CTT | CT | TGA | TAA | 82 | P-14-2 | - |
| 4293099 | evolution | fs | nrfE | TGGGGGGGA | TGGGGGGGGA | TAA | TGA | -2 | E-E1-2 / P-6-2 | - |
| 4415881 | evolution | fs | aidB | CGGGGGGA | CGGGGGGGA | TAA | TGA | 14 | P-2-1 / E-A1-1 / E-A1-2 / P-11-2 | - |
| 4424542 | C321 ancestor | fs | yjfY | GCA | GA | TAA | TAA | 143 | CFP | no change |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 4447866 | C321 ancestor | fs | tamB | ATT | AT | TAA | TAA | 95 | CFP | no change |
| 4472418 | evolution | fs | pyrL | CAAAAAAAAG | CAAAAAAAG | TAA | TAA | 42 | P-9-1 / C-D4-2 / P-7-2 / P-12-1 | no change |
| 4488572 | evolution | fs | yjgR | TCCCCCCCA | TCCCCCCCCA | TAA | TGA | 24 | P-13-2 / E-F1-1 / F-C6-1 | - |
| 4619469 | C321 ancestor | fs | deoA | LONG:11bp | CG | TAA | TAA | -10 | CFP | no change |

*** fs = frameshift; p = point mutation**

**** C = c321.ΔA; P = c321; E = ECNR2.1; F = c321.ΔA-v2**

**SI Table 2.** All observed mutations to release factors *prfB* and *prfC*.

| Position | Lineage | REF | ALT | Read Depth | Allele Frequency | Gene | Amino Acid change |
|---|---|---|---|---|---|---|---|
| 3035437 | E-H2-2 | T | C | 77.0 | 0.97 | prfB | p.Lys282Arg/c.845A>G |
| 3035437 | E-H2-1 | T | C | 49.0 | 0.98 | prfB | p.Lys282Arg/c.845A>G |
| 3035546 | C-D4-1 | T | C | 60.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | C-C4-1 | T | C | 73.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | C-B3-1 | T | C | 99.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | C-B4-1 | T | C | 77.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | C-C4-2 | T | C | 82.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | C-B4-2 | T | C | 87.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | C-D3-2 | T | C | 43.0 | 0.98 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | C-B3-2 | T | C | 52.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | C-D3-1 | T | C | 41.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035774 | C-F4-1 | C | T | 7.0 | 1.0 | prfB | p.Glu170Lys/c.508G>A |
| 3035774 | C-A4-1 | C | T | 44.0 | 1.0 | prfB | p.Glu170Lys/c.508G>A |
| 3035774 | C-D4-2 | C | T | 61.0 | 0.98 | prfB | p.Glu170Lys/c.508G>A |
| 3035774 | C-F4-2 | C | T | 37.0 | 1.0 | prfB | p.Glu170Lys/c.508G>A |
| 3035774 | C-A4-2 | C | T | 46.0 | 0.98 | prfB | p.Glu170Lys/c.508G>A |
| 3035546 | F-G5-1 | T | C | 73.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | F-E6-2 | T | C | 75.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | F-D6-1 | T | C | 84.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | F-D5-1 | T | C | 56.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | F-D5-2 | T | C | 66.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | F-F5-2 | T | C | 34.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | F-E5-2 | T | C | 51.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | F-B5-1 | T | C | 36.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | F-G6-2 | T | C | 29.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | F-F6-1 | T | C | 90.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | F-F5-1 | T | C | 75.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | F-B6-2 | T | C | 65.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | F-G5-2 | T | C | 70.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | F-B6-1 | T | C | 45.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | F-C6-2 | T | C | 115.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | F-F6-2 | T | C | 75.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | F-C6-1 | T | C | 61.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | F-B5-2 | T | C | 46.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | F-E5-1 | T | C | 60.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | F-D6-2 | T | C | 94.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | F-G6-1 | T | C | 90.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 3035546 | F-E6-1 | T | C | 89.0 | 1.0 | prfB | p.Thr246Ala/c.736A>G |
| 4610462 | F-C5-1 | C | T | 40.0 | 0.98 | prfC | p.Ala350Val/c.1049C>T |
| 4610462 | F-C5-2 | C | T | 84.0 | 0.94 | prfC | p.Ala350Val/c.1049C>T |
| 4610477 | F-F6-2 | A | G | 46.0 | 1.0 | prfC | p.His355Arg/c.1064A>G |
| 4610816 | F-B6-2 | T | C | 82.0 | 1.0 | prfC | p.Val468Ala/c.1403T>C |
| 4610816 | F-B6-1 | T | C | 55.0 | 1.0 | prfC | p.Val468Ala/c.1403T>C |

**SI Table 3 -** Summary of the most frequently-mutated genes in each evolved population. Only genes for which at least four instances are observed from any given starting population, or for which at least six instances are observed across all populations, are shown in this table.

| | | Instances of mutation to evolved lineages | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Gene | Gene Length (bp) | ECNR2 | C321 | C321.ΔA | C321.ΔA-v2 | Total | Brief Remarks | Previously Observed in ALE |
| *aceE* | 2,664 | - | 1 | 5 | 1 | 7 | | |
| *cyaA* | 2,547 | - | 2 | 2 | 2 | 6 | | |
| *fdnG* | 3,048 | - | - | 4 | - | 4 | | |
| *fimH* | 903 | 1 | 1 | 2 | 11 | 15 | Important for biofilm formation in M9 + glucose and LB | ¶ |
| *fis* | 297 | - | 4 | 6 | - | 10 | | * |
| *flu* | 3,120 | 14 | - | 5 | 13 | 32 | Important for biofilm formation in M9 + glucose but not in LB | |
| *folA* | 480 | - | 10 | 5 | 2 | 17 | Hitchhiker mutation to promoter during C321 recoding | ‖ |
| *folM* | 723 | - | 4 | 1 | - | 5 | | |
| *glrR* | 1,335 | - | 4 | - | - | 4 | | |
| *gltB* | 4,461 | - | 4 | 3 | - | 7 | Highly expressed in biofilms in Gram-negative bacteria | * |
| *hyfB* | 2,019 | 3 | 1 | - | 2 | 6 | | |
| *insH1* | 1,017 | 2 | 2 | 3 | 1 | 8 | | |
| *kup* | 1,869 | 1 | 4 | 2 | - | 7 | | * |
| *leuP* | 87 | 4 | 1 | 6 | 4 | 15 | Long strings of guanines were mutated in these tRNAs, a type of mutation that is a commonly found in mutS- hyper-mutator strains. | |
| *leuQ* | 87 | 2 | - | 3 | 3 | 8 | | |
| *leuV* | 87 | 5 | - | - | 4 | 9 | | |
| *mdtJ* | 366 | - | - | 1 | 4 | 5 | Knockout led to increased biofilm mass in minimal media | |
| *ompT* | 954 | 5 | 1 | - | 1 | 7 | Highly expressed in biofilms in Gram-negative bacteria | |
| *oxyR* | 918 | - | - | 5 | 12 | 17 | Null mutation upregulates *flu* | |
| *phoP* | 672 | 4 | - | 1 | 6 | 11 | | |
| *prfB* | 1,099 | 1 | - | 6 | 7 | 14 | Release factor 2, discussed in text | |
| *prfC* | 1,590 | - | - | 3 | 3 | 6 | Release factor 3, discussed in text | |
| *pta* | 2,145 | - | 5 | 1 | - | 6 | | |
| *ptrB* | 2,061 | 3 | 1 | 2 | - | 6 | | |
| *ptsP* | 2,247 | 2 | 2 | 1 | 2 | 7 | | |
| *purL* | 3,888 | 1 | 4 | 2 | - | 7 | Purine biosynthesis, specifically *purL* and *purT*, are important for biofilm formation in many Gram-negative bacteria | |
| *purT* | 1,179 | - | 3 | 1 | 7 | 11 | | |
| *pykF* | 1,413 | 9 | 4 | 1 | 2 | 16 | | *,§ |
| *pyrC* | 1,047 | 4 | 1 | - | - | 5 | | |
| *pyrE* | 642 | 16 | 7 | 6 | 8 | 37 | Mutation in MG1655 founder strain | †,‡,§ |
| *rcsD* | 2,673 | 2 | - | 1 | 3 | 6 | | |
| *rhsC* | 4,194 | 2 | 2 | - | 3 | 7 | | |
| *rlmB* | 732 | 3 | 3 | 1 | 3 | 10 | | |
| *rph* | 717 | 2 | 4 | 2 | 8 | 16 | Mutation in MG1655 founder strain | †,‡,§ |
| *rpoB* | 4,029 | 7 | 5 | 2 | - | 14 | Commonly seen in adaptation to minimal media | †,‡,§ |
| *rpoC* | 4,224 | 10 | 7 | 4 | 2 | 23 | Commonly seen in adaptation to minimal media | †,‡ |
| *rsxC* | 2,223 | 1 | 7 | - | 2 | 10 | | |
| *speA* | 1,977 | 1 | 1 | 2 | 2 | 6 | | |
| *tktA* | 1,992 | 2 | 1 | 2 | 1 | 6 | | |
| *xdhA* | 2,259 | 3 | 1 | 1 | 2 | 7 | | |
| *yccS* | 2,154 | 2 | 5 | 3 | - | 10 | | |
| *ycgB* | 1,533 | - | 1 | 4 | - | 5 | | |
| *ydbA* | 8,622 | 1 | 2 | 2 | 1 | 6 | | |
| *ydfJ* | 1,368 | 1 | - | 4 | 1 | 6 | | |
| *yeeJ* | 7,077 | - | - | 4 | 3 | 7 | | |
| *yegE* | 3,318 | 2 | 4 | - | - | 6 | | |
| *yehB* | 2,481 | 1 | - | 4 | - | 5 | | |
| *yfhM* | 4,962 | 2 | 4 | 1 | 1 | 8 | | |
| *yfiH* | 732 | 4 | 3 | 5 | - | 12 | | |
| *yneO* | 5,323 | - | 4 | 2 | - | 6 | | |

* Barrick et al.; † Conrad et al.; ‡ Herring et al.; § LaCroix et al.; ‖ Monk et al.; ¶ Aguilar et al.

**SI Table 4.** SNPs seen in more than one independent recoded lineage missing *prfA*, but not found in ECNR2 lineages. C-lineages refer to C321.ΔA, F-lineages refer to C321.ΔA-v2.

| Position | Ref | Alt | Independent origins | Clades | Gene | Amino acid change |
|---|---|---|---|---|---|---|
| 40229 | A | G | 2 | (C-E4-1;C-E4-2;C-G4-1;C-G4-2);(F-A6-1;F-A6-2) | caiA | NA |
| 49768 | A | G | 2 | (C-A4-1;C-A4-2);(F-B5-1;F-B5-2;F-B6-1;F-B6-2;F-C6-1;F-C6-2;F-H5-1;F-H5-2) | NA | NA |
| 107375 | GT | G | 2 | (C-D4-1);(F-A6-1;F-A6-2) | NA | NA |
| 224738 | G | A | 2 | (F-F5-1);(F-E6-2) | rrsH | NA |
| 716166 | A | AG | 2 | (C-D3-2);(C-G4-2) | potE | frameshift |
| 875520 | GC | G | 2 | (C-G3-1;C-G3-2);(F-C6-1) | gsiA | frameshift |
| 1196946 | A | AT | 2 | (C-C4-2);(F-B6-1) | phoP | frameshift |
| 1243369 | A | AG | 2 | (C-B3-1;C-B3-2;C-D3-1;C-D3-2);(C-D4-2;C-F4-2) | ycgB | frameshift |
| 1270725 | A | AG | 2 | (C-F3-2);(C-A4-1;C-A4-2) | NA | NA |
| 1425603 | A | AG | 2 | (F-D6-1;F-D6-2);(F-E5-1;F-E5-2) | ydaT | frameshift |
| 1678153 | A | AT | 2 | (C-F3-1;C-F3-2);(F-G5-1;F-G5-2) | mdtJ | frameshift |
| 1678604 | T | C | 2 | (C-G3-1;C-G3-2);(F-H5-1) | NA | NA |
| 1936778 | T | TG | 2 | (F-D5-1;F-D5-2);(F-G5-1;F-G5-2) | purT | frameshift |
| 2048923 | G | A | 2 | (C-G3-1;C-G3-2);(F-H5-2) | yeeJ | NA |
| 2382800 | T | TC | 2 | (C-B3-1);(F-C5-2) | menD | frameshift |
| 2771523 | T | TC | 2 | (F-H5-2);(F-G5-2) | NA | NA |
| 3001978 | A | AG | 2 | (C-D3-1);(F-B6-1;F-B6-2) | ygeR | frameshift |
| 3037849 | T | C | 8 | (C-B3-1;C-B3-2;C-B4-1;C-B4-2;C-C4-1;C-C4-2;C-D3-1;C-D3-2);(C-D4-1);(F-F6-1;F-F6-2);(F-B5-1;F-B5-2;F-B6-1;F-B6-2);(F-C6-1;F-C6-2);(F-D6-1;F-D6-2);(F-D5-1;F-D5-2);(F-E5-1;F-E5-2;F-E6-1;F-E6-2;F-F5-1;F-F5-2;F-G5-1;F-G5-2;F-G6-1;F-G6-2) | prfB | T->A |
| 3038077 | C | T | 2 | (C-A4-1;C-A4-2);(C-D4-2;C-F4-2) | prfB | E->K |
| 3410910 | C | CA | 2 | (C-E4-1;C-G4-2);(F-C6-1;F-C6-2) | NA | NA |
| 3762046 | A | AG | 3 | (C-B3-1;C-B3-2;C-B4-1;C-B4-2;C-C4-1;C-C4-2;C-D3-1;C-D3-2);(C-A4-1;C-A4-2;C-C3-1;C-C3-2;C-E3-1;C-E3-2;C-F3-1;C-F3-2;C-G0;C-G3-1;C-G3-2);(C-D4-1;C-D4-2;C-E4-1;C-E4-2;C-F4-2;C-G4-1;C-G4-2) | selA | frameshift |
| 3817740 | A | G | 3 | (C-B4-1;C-B4-2;C-C4-1;C-C4-2);(F-D5-1;F-D5-2);(F-E5-1;F-E5-2) | rph | NA |
| 3817764 | A | G | 2 | (F-F6-1;F-F6-2);(F-G6-1;F-G6-2) | rph | NA |
| 3924987 | G | A | 2 | (F-B6-1;F-B6-2);(F-H5-1;F-H5-2) | rsmG | A->V |
| 3967466 | CT | C | 2 | (F-A6-1;F-A6-2);(F-F5-1) | rhlB | frameshift |
| 4096334 | T | G | 2 | (C-B3-1;C-B3-2;C-D3-1;C-D3-2);(F-E5-1;F-E5-2) | NA | NA |
| 4160505 | A | G | 2 | (F-B5-1;F-B5-2);(F-G5-1;F-G5-2) | oxyR | S->G |
| 4160511 | C | T | 2 | (C-F4-2);(F-F6-1;F-F6-2) | oxyR | Q->* |
| 4160536 | T | C | 2 | (F-H5-2);(F-F5-1) | oxyR | L->P |
| 4160969 | AG | A | 2 | (F-D5-2);(F-E5-1;F-E5-2) | oxyR | frameshift |
| 4160973 | GA | G | 2 | (F-D5-1;F-D5-2);(F-E5-1) | oxyR | frameshift |
| 4550911 | G | A | 2 | (F-F5-1);(F-E6-1;F-E6-2) | fimH | C->Y |
| 4550965 | C | T | 2 | (C-C4-1;C-C4-2);(F-D6-1;F-D6-2) | fimH | S->L |
| 4550971 | A | G | 2 | (F-A6-1;F-A6-2);(F-E5-1;F-E5-2) | fimH | Y->C |
| 4612372 | C | T | 2 | (C-E4-1);(F-C5-1;F-C5-2) | prfC | A->V |
| | | | | | | |

**SI Table 5.** Gene ontology categorization of all mutations in the evolved populatinos.

| GO process description | GO process ID | C321 mutations | ECNR2 mutations | C321-v2 mutations | Fold increase in Recoded vs. Ecnr2 |
|---|---|---|---|---|---|
| glycoprotein biosynthetic process | 9101 | 2 | 1 | 12 | 7 |
| DNA recombination | 6310 | 11 | 1 | 1 | 6 |
| tRNA aminoacylation for protein translation | 6418 | 9 | 1 | 3 | 6 |
| peptidoglycan biosynthetic process | 9252 | 6 | 1 | 3 | 4.5 |
| glutamate biosynthetic process | 6537 | 5 | 1 | 2 | 3.5 |
| regulation of transcription DNA-dependent | 6355 | 8 | 2 | 5 | 3.25 |
| translation | 6412 | 12 | 4 | 12 | 3 |
| amine catabolic process | 9310 | 6 | 2 | 4 | 2.5 |
| pyridoxine biosynthetic process | 8615 | 4 | 1 | 1 | 2.5 |
| polysaccharide catabolic process | 272 | 2 | 1 | 3 | 2.5 |
| asparagine biosynthetic process | 6529 | 1 | 1 | 4 | 2.5 |
| purine nucleotide biosynthetic process | 6164 | 2 | 2 | 6 | 2 |
| pentose-phosphate shunt non-oxidative branch | 9052 | 3 | 1 | 1 | 2 |
| DNA methylation | 6306 | 1 | 1 | 3 | 2 |
| O antigen biosynthetic process | 9243 | 1 | 1 | 3 | 2 |
| DNA-dependent DNA replication | 6261 | 7 | 4 | 8 | 1.875 |
| RNA catabolic process | 6401 | 2 | 2 | 5 | 1.75 |
| fatty acid oxidation | 19395 | 6 | 3 | 4 | 1.667 |
| flagellum assembly | 9296 | 6 | 3 | 3 | 1.5 |
| phosphorus metabolic process | 6793 | 3 | 2 | 3 | 1.5 |
| SOS response | 9432 | 3 | 1 | 0 | 1.5 |
| enterobacterial common antigen biosynthetic process | 9246 | 2 | 1 | 1 | 1.5 |
| nucleotide-sugar biosynthetic process | 9226 | 2 | 1 | 1 | 1.5 |
| betaine biosynthetic process | 6578 | 1 | 1 | 2 | 1.5 |
| response to temperature stimulus | 9266 | 2 | 2 | 3 | 1.25 |
| DNA repair | 6281 | 3 | 3 | 4 | 1.167 |
| taxis | 42330 | 9 | 7 | 7 | 1.143 |
| nitrogen compound metabolic process | 6807 | 5 | 4 | 4 | 1.125 |
| None | NA | 239 | 232 | 236 | 1.024 |
| carbohydrate catabolic process | 16052 | 41 | 40 | 40 | 1.013 |
| response to osmotic stress | 6970 | 5 | 4 | 3 | 1 |
| cytochrome complex assembly | 17004 | 3 | 4 | 5 | 1 |
| polyamine biosynthetic process | 6596 | 4 | 3 | 2 | 1 |
| methionine biosynthetic process | 9086 | 3 | 2 | 1 | 1 |
| ATP synthesis coupled proton transport | 15986 | 3 | 2 | 1 | 1 |
| amino sugar biosynthetic process | 46349 | 3 | 2 | 1 | 1 |
| tryptophan biosynthetic process | 162 | 0 | 2 | 4 | 1 |
| glycolate metabolic process | 9441 | 2 | 1 | 0 | 1 |
| aspartate biosynthetic process | 6532 | 1 | 1 | 1 | 1 |
| glyoxylate cycle | 6097 | 0 | 1 | 2 | 1 |
| RNA modification | 9451 | 5 | 6 | 6 | 0.917 |
| phospholipid biosynthetic process | 8654 | 3 | 3 | 2 | 0.833 |
| protein folding | 6457 | 9 | 14 | 14 | 0.822 |
| transcription | 6350 | 38 | 51 | 45 | 0.814 |
| anaerobic respiration | 9061 | 23 | 28 | 19 | 0.75 |
| glucose metabolic process | 6006 | 3 | 2 | 0 | 0.75 |
| methylglyoxal metabolic process | 9438 | 2 | 2 | 1 | 0.75 |
| peptidoglycan metabolic process | 270 | 2 | 2 | 1 | 0.75 |
| thiamin biosynthetic process | 9228 | 2 | 2 | 1 | 0.75 |
| response to drug | 42493 | 6 | 10 | 8 | 0.7 |
| glycolysis | 6096 | 7 | 9 | 5 | 0.667 |
| xenobiotic metabolic process | 6805 | 5 | 6 | 3 | 0.667 |
| DNA catabolic process | 6308 | 2 | 5 | 4 | 0.6 |
| sulfur metabolic process | 6790 | 6 | 7 | 2 | 0.571 |
| nucleobase nucleoside and nucleotide interconversion | 15949 | 4 | 10 | 7 | 0.55 |

| | | | | | |
|---|---|---|---|---|---|
| cellular amino acid catabolic process | 9063 | 9 | 14 | 6 | 0.536 |
| aerobic respiration | 9060 | 4 | 7 | 3 | 0.5 |
| gluconeogenesis | 6094 | 1 | 2 | 1 | 0.5 |
| galactose metabolic process | 6012 | 1 | 1 | 0 | 0.5 |
| Mo-molybdopterin cofactor biosynthetic process | 6777 | 1 | 1 | 0 | 0.5 |
| K antigen biosynthetic process | 9248 | 0 | 1 | 1 | 0.5 |
| lipopolysaccharide biosynthetic process | 9103 | 0 | 1 | 1 | 0.5 |
| homoserine biosynthetic process | 9090 | 0 | 1 | 1 | 0.5 |
| glycerol metabolic process | 6071 | 0 | 1 | 1 | 0.5 |
| protein modification process | 6464 | 6 | 22 | 14 | 0.455 |
| colanic acid biosynthetic process | 9242 | 3 | 7 | 2 | 0.357 |
| tricarboxylic acid cycle | 6099 | 3 | 7 | 2 | 0.357 |
| NAD biosynthetic process | 9435 | 2 | 6 | 2 | 0.333 |
| pyrimidine nucleotide biosynthetic process | 6221 | 1 | 3 | 1 | 0.333 |
| fermentation | 6113 | 5 | 17 | 6 | 0.324 |
| glycine biosynthetic process | 6545 | 1 | 2 | 0 | 0.25 |
| purine ribonucleotide biosynthetic process | 9152 | 1 | 2 | 0 | 0.25 |
| proline biosynthetic process | 6561 | 0 | 2 | 1 | 0.25 |
| translational attenuation | 9386 | 1 | 3 | 0 | 0.167 |
| response to desiccation | 9269 | 1 | 3 | 0 | 0.167 |
| histidine biosynthetic process | 105 | 0 | 3 | 0 | 0 |
| response to starvation | 42594 | 0 | 3 | 0 | 0 |
| response to stress | 6950 | 0 | 2 | 0 | 0 |
| macromolecule catabolic process | 9057 | 0 | 1 | 0 | 0 |
| cellular amino acid metabolic process | 6520 | 0 | 1 | 0 | 0 |
| iron ion transport | 6826 | 5 | 0 | 3 | 0 |
| folic acid biosynthetic process | 46656 | 4 | 0 | 3 | 0 |
| enterobactin biosynthetic process | 9239 | 3 | 0 | 1 | 0 |
| glyoxylate catabolic process | 9436 | 3 | 0 | 1 | 0 |
| chorismate biosynthetic process | 9423 | 1 | 0 | 3 | 0 |
| lysine biosynthetic process via diaminopimelate | 9089 | 1 | 0 | 3 | 0 |
| acetyl-CoA biosynthetic process from pyruvate | 6086 | 3 | 0 | 0 | 0 |
| nonribosomal peptide biosynthetic process | 19184 | 2 | 0 | 1 | 0 |
| coenzyme A biosynthetic process | 15937 | 2 | 0 | 1 | 0 |
| alanine biosynthetic process | 6523 | 2 | 0 | 1 | 0 |
| lipopolysaccharide core region biosynthetic process | 9244 | 1 | 0 | 2 | 0 |
| response to radiation | 9314 | 1 | 0 | 2 | 0 |
| lipoprotein biosynthetic process | 42158 | 1 | 0 | 2 | 0 |
| leucine biosynthetic process | 9098 | 1 | 0 | 2 | 0 |
| glutathione biosynthetic process | 6750 | 2 | 0 | 0 | 0 |
| arginine biosynthetic process | 6526 | 2 | 0 | 0 | 0 |
| phosphoenolpyruvate-dependent sugar phosphotransferase system | 9401 | 1 | 0 | 1 | 0 |
| carnitine catabolic process | 42413 | 1 | 0 | 1 | 0 |
| lipid A biosynthetic process | 9245 | 1 | 0 | 0 | 0 |
| DNA metabolic process | 6259 | 1 | 0 | 0 | 0 |
| S-adenosylmethionine biosynthetic process | 6556 | 1 | 0 | 0 | 0 |
| lipoate biosynthetic process | 9107 | 1 | 0 | 0 | 0 |
| L-phenylalanine biosynthetic process | 9094 | 1 | 0 | 0 | 0 |
| coenzyme and prosthetic group biosynthetic process | 46138 | 0 | 0 | 1 | 0 |
| pentose-phosphate shunt oxidative branch | 9051 | 0 | 0 | 1 | 0 |
| putrescine catabolic process | 9447 | 0 | 0 | 1 | 0 |
| pyruvate catabolic process | 42867 | 0 | 0 | 1 | 0 |
| glutamine biosynthetic process | 6542 | 0 | 0 | 1 | 0 |
| putrescine transport | 15847 | 0 | 0 | 1 | 0 |
| biotin biosynthetic process | 9102 | 0 | 0 | 1 | 0 |
| cysteine biosynthetic process | 19344 | 0 | 0 | 1 | 0 |
| glucan biosynthetic process | 9250 | 0 | 0 | 1 | 0 |
| polysaccharide biosynthetic process | 271 | 0 | 0 | 1 | 0 |
| gamma-aminobutyric acid catabolic process | 9450 | 0 | 0 | 1 | 0 |
| pilus assembly | 9297 | 0 | 0 | 1 | 0 |

**SI Table 6.** Oligonucleotides used in this study.

| Oligo Name | Sequence |
|---|---|
| P_C321D_4102449 | CCGGTTTTGCCAGACTCCG |
| P_C321D_1263523 | GCAGATGACGCGCCACCG |
| P_C321D_1511492 | CATATAGAGATCGCGCGGGGA |
| P_C321D_3092256 | GCGCTATTTGAATTGCTGATTAACA |
| P_C321D_3990077 | AGATTGTATCGTTTGGTCTCGATCC |
| P_C321D_322579 | GGATATGGCTCTATTTATCCAGGGC |
| P_FIX_4102449 | CCGGTTTTGCCAGACTCCA |
| P_FIX_1263523 | GCAGATGACGCGCCACCA |
| P_FIX_1511492 | CATATAGAGATCGCGCGGGGG |
| P_FIX_3092256 | GCGCTATTTGAATTGCTGATTAACG |
| P_FIX_3990077 | AGATTGTATCGTTTGGTCTCGATCT |
| P_FIX_322579 | GGATATGGCTCTATTTATCCAGGGT |
| P_COM_4102449 | AGCAGTTCTCAATCCGATTTCATTAAC |
| P_COM_1263523 | AAATCTTTCTCTGTCGCGAAAC |
| P_COM_1511492 | ACGCAGGATTACATTGCCATTATG |
| P_COM_3092256 | GGCGCTCAACAAAGGCAAA |
| P_COM_3990077 | CGCGGTTATACAGATACCAACC |
| P_COM_322579 | CTCACAGCGCGTCAACATAAC |
| prfA-F | CAACCAACCAATGCTGTTG |
| prfA-R | GTCGCTTTGCAGAATGTG |
| prfA-gRNA1-F | CTTTTTTTACAGGGTGGAGGGTTTTAGAGCTAGAAATAGCAAG |
| prfA-gRNA2-F | ACAGGGTGGAGGAGGAATAAGTTTTAGAGCTAGAAATAGCAAG |
| gRNA-seq-F | AGTAAGGCAGCGGTATCATC |
| MAGE_prfBT246A | G*C*AGATTCGGTACGGTTAACGTGCTGACCGCCCGCGCCGGACGcGCGATAAACGTCAATGCGCAGATCCGC CGGGTTGATTTCGATATCA |
| MAGE_prfBE170K | G*G*AGATTTTGATCGTCACGGATTTAATACCCGCCACTTCACCTTtCGACTCTTCGATGATTTCAGTTTTGAAAC CACGCGATTCTGCCCA |
| MAGE_prfCA350V | G*G*ATATCGCCCGGATACGCTTCTTCAACGTGCGAACGGTCACCCaCCATAAAGGTCAGCGCGTCGGAGATCA CCACATCTTTCGCAGTGC |
| asPCR-prfBT246A-mut-F | CGCATTGACGTTTATCGCG |
| asPCR-prfBT246A-WT-F | CGCATTGACGTTTATCGCA |
| asPCR-prfBE170K-mut-F | CTGAAATCATCGAAGAGTCGA |
| asPCR-prfBE170K-WT-F | CTGAAATCATCGAAGAGTCGG |
| asPCR-prfB-R | CTTCGCTTTCATCTGCTTC |
| asPCR-prfCA350V-mut-F | GCGCTGACCTTTATGGt |
| asPCR-prfCA350V-WT-F | GCGCTGACCTTTATGGC |
| asPCR-prfC-R | CTTCGGAAAGCTGTACCAG |
| prfB-removeFS-F | ACGTTCTTAGGGGGTATCTTGACTACGACGCCAAGAAAG |
| prfB-removeFS-R | CTCTTTCTTGGCGTCGTAGTCAAGATACCCCCTAAGAACGTC |
| prfB_vars-seq | GTATGTTCTCTGGCGAATATG |
| prfC_vars-seq | GATACCCGTACCGTAGAAG |
| MAGE-KO-fis | A*A*ATTTTGCGTAAACAGAAATAAAGAGCTGACAGAACTATGTaatAAtgACGCGTAAATTCTGACGTACTGAC CGTTTCTACCGTTAACT |
| MAGE-KO-mdtJ | A*C*CGGTAATTTCTGTAGCAATAGCCAGACCTAATAAAATCCAtTAtcaATACATTGTCCTTCTCCTGCAAGAGA ATTATTTTAATTTTCG |
| MAGE-KO-oxyR | C*C*GCACGCCGAAAATGGCGGTGTTCAGCCAATGCCACCAGtTAtTatcaATCACGAATATTCATTATCCATCCT CCATCGCCACGATAGT |
| MAGE-KO-purT | A*C*ACACGCAAACGTTTTCGTTTATACTGCGCGCGGAATTAATCAGGGGATATTCGTTATGtaaTaATgAGGCA CTGCGCTGCGTCCGGCA |
| fimH_C65Y-MAGE | G*C*AGTGTGACATAGTCTGTAATGGTTTCCGGATAATCGTTATGGTAAAAGATTTGCGTCGAAAGATCCACGA CCAGGTTTTGCCCCACAT |
| fimH_S83L-MAGE | A*T*TTTACGGTCCCGGAAAAATTAGATAACACGCCGCCATAAGCCAAGCCTCGTTGCAGTGTGACATAGTCTG TAATGGTTTCCGGATAA |
| fimH_65-ASPCR-WT | GGATCTTTCGACGCAAATCTTTTG |
| fimH_65-ASPCR-mut | GGATCTTTCGACGCAAATCTTTTA |
| fimH_83-ASPCR-WT | GTCACACTGCAACGAGGCTC |
| fimH_83-ASPCR-mut | GTCACACTGCAACGAGGCTT |
| fimH-ASPCR-R | CTGCCACTATATTTTACGGTCCC |
| fis-ASPCR-WT | GTAAACAGAAATAAAGAGCTGACAGAACTATGTTC |

| | |
|---|---|
| fis-ASPCR-mut | GTAAACAGAAATAAAGAGCTGACAGAACTATGTAA |
| folA-ASPCR-WT | TCTCAATGATCAGTCTGATTGCGGCGTTAG |
| folA-ASPCR-mut | TCTCAATGATCAGTCTGATTGCGGCGTAAT |
| mdtJ-ASPCR-WT | GTAATTTCTGTAGCAATAGCCAGACCTAATAAAATCCAA |
| mdtJ-ASPCR-mut | GTAATTTCTGTAGCAATAGCCAGACCTAATAAAATCCAT |
| oxyR-ASPCR-WT | GATGGAGGATGGATAATGAATATTCGTGATC |
| oxyR-ASPCR-mut | GATGGAGGATGGATAATGAATATTCGTGATT |
| purT-ASPCR-WT | CGGAATTAATCAGGGGATATTCGTTATGA |
| purT-ASPCR-mut | GGAATTAATCAGGGGATATTCGTTATGT |
| fis-ASPCR-R | GCAGCGTACCACGGTTGATGCC |
| folA-ASPCR-R | CAGTCATCCGGCTCGTAATCCGGG |
| mdtJ-ASPCR-R | GACGATTTTTAGGCCATTGAGCGTGATGATCG |
| oxyR-ASPCR-R | TGTTGCGCAGAGTTTCCAGGCTGG |
| purT-ASPCR-R | TAATGACATGGGAGCGATGCGCGA |
| pfolA-ins-F | cttacattaattgcgttgcgcTTCCTCAACATCATCCTCGC |
| pfolA-ins-R | gtggtgatgatggtgcatTGAGATTTCCCGATAAAAAAAATTGTC |
| mCard-bbone-F | atgcaccatcatcaccac |

**SI References**


1. Conrad TM, et al. (2010) RNA polymerase mutants found through adaptive evolution reprogram Escherichia coli for optimal growth in minimal media. *Proc Natl Acad Sci U S A* 107(47):20500–5.

2. Herring CD, et al. (2006) Comparative genome sequencing of Escherichia coli allows observation of bacterial evolution on a laboratory timescale. *Nat Genet* 38(12):1406–1412.

3. LaCroix RA, et al. (2015) Use of adaptive laboratory evolution to discover key mutations enabling rapid growth of Escherichia coli K-12 MG1655 on glucose minimal medium. *Appl Environ Microbiol* 81(1):17–30.

4. Barrick JE, et al. (2009) Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature* 461(7268):1243–1247.

5. Aguilar C, et al. (2012) Genetic changes during a laboratory adaptive evolution process that allowed fast growth in glucose to an Escherichia coli strain lacking the major glucose transport system. *BMC Genomics* 13. doi:10.1186/1471-2164-13-385.

6. Monk JW, et al. (2016) Rapid and Inexpensive Evaluation of Nonstandard Amino Acid Incorporation in Escherichia coli. *ACS Synth Biol* 6(1):45–54.

7. Jensen KF (1993) The Escherichia coli K-12 “wild types” W3110 and MG1655 have an rph frameshift mutation that leads to pyrimidine starvation due to low pyrE expression levels. *J Bacteriol* 175(11):3401–7.

8. Danese PN, Pratt LA, Dove SL, Kolter R (2000) The outer membrane protein, Antigen 43, mediates cell-to-cell interactions within Escherichia coli biofilms. *Mol Microbiol* 37(2):424–432.

9. Schembri MA, Kjaergaard K, Klemm P (2003) Global gene expression in Escherichia coli biofilms. *Mol Microbiol* 48(1):253–267.

10. Bay DC, Stremick CA, Slipski CJ, Turner RJ (2017) Secondary multidrug efflux pump mutants alter Escherichia coli biofilm growth in the presence of cationic antimicrobial compounds. *Res Microbiol* 168(3):208–221.

11. Bossi L (1983) Context effects: Translation of UAG codon by suppressor tRNA is affected by the sequence following UAG in the message. *J Mol Biol* 164(1):73–87.

12. Martin R, Weiner M, Gallant J (1988) Effects of release factor context at UAA codons in Escherichia coli. *J Bacteriol* 170(10):4714–7.

13. Poole ES, Brown CM, Tate WP (1995) The identity of the base following the stop codon determines the efficiency of in vivo translational termination in Escherichia coli. *EMBO J* 14(1):151–158.

14. Nakamura Y, Kawakami K, Mikuni O (1990) Alternative Translation and Functional Diversity of Release Factor 2 and Lysyl-tRNA Synthetase. *Post-Transcriptional Control of Gene Expression* (Springer Berlin Heidelberg, Berlin, Heidelberg), pp 455–464.

15. Brown CM, Stockwell PA, Trotman CNA, Tate WP (1990) The signal for the termination of protein synthesis in procaryotes. *Nucleic Acids Res* 18(8):2079–2086.

16. Korkmaz G, Holm M, Wiens T, Sanyal S (2014) Comprehensive analysis of stop codon usage in bacteria and its correlation with release factor abundance. *J Biol Chem* 289(44):30334–42.

17. Keiler KC, Waller PRH, Sauer RT (1996) Role of a Peptide Tagging System in Degradation of Proteins Synthesized from Damaged Messenger RNA. *Science (80- )* 271(5251). Available at: http://science.sciencemag.org/content/271/5251/990 [Accessed May 11, 2017].

18. Dulebohn D, Choy J, Sundermeier T, Okan N, Karzai AW (2007) Trans-Translation: The tmRNA-Mediated Surveillance Mechanism for Ribosome Rescue, Directed Protein Degradation, and Nonstop mRNA Decay.

*Biochemistry* 46(16):4681–4693.

19. Chadani Y, et al. (2010) Ribosome rescue by Escherichia coli ArfA (YhdL) in the absence of trans-translation system. *Mol Microbiol* 78(4):796–808.

20. James NR, Brown A, Gordiyenko Y, Ramakrishnan V (2016) Translational termination without a stop codon. *Science (80- )* 354(6318). Available at: http://science.sciencemag.org.libproxy.mit.edu/content/354/6318/1437/tab-pdf [Accessed March 27, 2017].

21. Yusupov MM, et al. (2001) Crystal Structure of the Ribosome at 5.5 Å Resolution. *Science (80- )* 292(5518). Available at: http://science.sciencemag.org/content/292/5518/883 [Accessed May 11, 2017].

22. Ejby M, Sørensen MA, Pedersen S (2007) Pseudouridylation of helix 69 of 23S rRNA is necessary for an effective translation termination. *Proc Natl Acad Sci U S A* 104(49):19410–5.

23. O'Connor M, Gregory ST (2011) Inactivation of the RluD pseudouridine synthase has minimal effects on growth and ribosome function in wild-type Escherichia coli and Salmonella enterica. *J Bacteriol* 193(1):154–62.

24. Heurgue-Hamard V, Champ S, Engström A, Ehrenberg M, Buckingham RH (2002) The hemK gene in Escherichia coli encodes the N5-glutamine methyltransferase that modifies peptide release factors. *EMBO J* 21(4):769–778.

25. Nakahigashi K, et al. (2002) HemK, a class of protein methyl transferase with similarity to DNA methyl transferases, methylates polypeptide chain release factors, and hemK knockout induces defects in translational termination. *Proc Natl Acad Sci U S A* 99(3):1473–8.

26. Mora L, Heurgué-Hamard V, de Zamaroczy M, Kervestin S, Buckingham RH (2007) Methylation of Bacterial Release Factors RF1 and RF2 Is Required for Normal Translation Termination in Vivo. *J Biol Chem* 282(49):35638–35645.

27. SaiSree L, Reddy M, Gowrishankar J (2001) IS186 insertion at a hot spot in the lon promoter as a basis for lon protease deficiency of Escherichia coli B: identification of a consensus target sequence for IS186 transposition. *J Bacteriol* 183(23):6943–6.

28. Baneyx F, Georgiou G (1990) In vivo degradation of secreted fusion proteins by the Escherichia coli outer membrane protease OmpT. *J Bacteriol* 172(1):491–4.

29. Zilhão R, Cairrão F, Régnier P, Arraiano CM (1996) PNPase modulates RNase II expression in Escherichia coli: implications for mRNA decay and cell metabolism. *Mol Microbiol* 20(5):1033–1042.

30. Lopez PJ, Marchand I, Joyce SA, Dreyfus M (1999) The C-terminal half of RNase E, which organizes the Escherichia coli degradosome, participates in mRNA degradation but not rRNA processing in vivo. *Mol Microbiol* 33(1):188–199.

31. Michel-Reydellet N, Woodrow K, Swartz J (2005) Increasing PCR fragment stability and protein yields in a cell-free system with genetically modified Escherichia coli extracts. *J Mol Microbiol Biotechnol* 9(1):26–34.

32. Hong SH, et al. (2015) Improving Cell-Free Protein Synthesis through Genome Engineering of Escherichia coli Lacking Release Factor 1. *ChemBioChem* 16(5):844–853.

33. Brockman IM, Prather KLJ (2015) Dynamic knockdown of E. coli central metabolism for redirecting fluxes of primary metabolites. *Metab Eng* 28:104–113.

34. Kunjapur AM, Hyun JC, Prather KLJ (2016) Deregulation of S-adenosylmethionine biosynthesis and regeneration improves methylation in the E. coli de novo vanillin biosynthesis pathway. *Microb Cell Fact* 15(1):1.

35. Baym M, et al. (2015) Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes. *PLoS One* 10(5):e0128036.

36. Goodman DB, et al. (2017) Millstone: software for multiplex microbial genome analysis and engineering. *Genome Biol* 18(1):101.

37. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–9.

38. DePristo MA, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5):491–8.

39. Riley M, et al. (2006) Escherichia coli K-12: a cooperatively developed annotation snapshot--2005. *Nucleic Acids Res* 34(1):1–9.

40. Kumar S, et al. (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 33(7):1870–1874.

41. Isaacs FJ, et al. (2011) Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science (80- )* 333(6040):348–353.

42. Chu J, et al. (2014) Non-invasive intravital imaging of cellular differentiation with a bright red-excitable fluorescent protein. *Nat Methods* 11(5):572–8.

43. Xie J, Liu W, Schultz PG (2007) A Genetically Encoded Bidentate, Metal-Binding Amino Acid. *Angew Chemie* 119(48):9399–9402.

44. Kunjapur AM, et al. (2017) Engineering post-translational proofreading to discriminate non-standard amino acids. *bioRxiv*. Available at: http://www.biorxiv.org/content/early/2017/06/30/158246 [Accessed July 1, 2017].