LETTER

doi:10.1038/nature14095

Recoded organisms engineered to depend on synthetic amino acids

Alexis J. Rovner[1,2], Adrian D. Haimovich[1,2]*, Spencer R. Katz[1,2]*, Zhe Li[1,2], Michael W. Grome[1,2], Brandon M. Gassaway[2,3], Miriam Amiram[1,2], Jaymin R. Patel[1,2], Ryan R. Gallagher[1,2], Jesse Rinehart[2,3] & Farren J. Isaacs[1,2]

Genetically modified organisms (GMOs) are increasingly used in research and industrial systems to produce high-value pharmaceuticals, fuels and chemicals[1]. Genetic isolation and intrinsic biocontainment would provide essential biosafety measures to secure these closed systems and enable safe applications of GMOs in open systems[2,3], which include bioremediation[4] and probiotics[5]. Although safeguards have been designed to control cell growth by essential gene regulation[6], inducible toxin switches[7] and engineered auxotrophies[8], these approaches are compromised by cross-feeding of essential metabolites, leaked expression of essential genes, or genetic mutations[9,10]. Here we describe the construction of a series of genomically recoded organisms (GROs)[11] whose growth is restricted by the expression of multiple essential genes that depend on exogenously supplied synthetic amino acids (sAAs). We introduced a *Methanocaldococcus jannaschii* tRNA:aminoacyl-tRNA synthetase pair into the chromosome of a GRO derived from *Escherichia coli* that lacks all TAG codons and release factor 1, endowing this organism with the orthogonal translational components to convert TAG into a dedicated sense codon for sAAs. Using multiplex automated genome engineering[12], we introduced in-frame TAG codons into 22 essential genes, linking their expression to the incorporation of synthetic phenylalanine-derived amino acids. Of the 60 sAA-dependent variants isolated, a notable strain harbouring three TAG codons in conserved functional residues[13] of MurG, DnaA and SerS and containing targeted tRNA deletions maintained robust growth and exhibited undetectable escape frequencies upon culturing ~10^{11} cells on solid media for 7 days or in liquid media for 20 days. This is a significant improvement over existing biocontainment approaches[2,3,6–10]. We constructed synthetic auxotrophs dependent on sAAs that were not rescued by cross-feeding in environmental growth assays. These auxotrophic GROs possess alternative genetic codes that impart genetic isolation by impeding horizontal gene transfer[11] and now depend on the use of synthetic biochemical building blocks, advancing orthogonal barriers between engineered organisms and the environment.

The advent of recombinant DNA technologies in the 1970s established genetic cloning methods[14], ushering in the era of biotechnology. Over the past decade, synthetic biology has fuelled the emergence of GMOs with increased sophistication as common and valued solutions in clinical, industrial and environmental settings[1,4,5], necessitating the development of safety and security measures first outlined in the 1975 Asilomar conference on recombinant DNA[15]. While guidelines for physical containment and safe use of organisms have been widely adopted, intrinsic biocontainment—biological barriers limiting the spread and survival of microorganisms in natural environments—remains a defining challenge. Existing biocontainment strategies employ natural auxotrophies or conditional suicide switches where top safeguards meet the 10^{-8} NIH standard (http://osp.od.nih.gov/office-biotechnology-activities/biosafety/nih-guidelines) for escape frequencies (that is, one escape mutant per 10^8 cells), but can be compromised by metabolic cross-feeding or genetic mutation[9,10]. We hypothesized that engineering dependencies on synthetic biochemical building blocks would enhance existing containment strategies by establishing orthogonal barriers not feasible in organisms with a standard genetic code.

Our approach to engineering biocontainment used a GRO lacking all instances of the TAG codon and release factor 1 (terminates translation at UAA and UAG), eliminating termination of translation at UAG and endowing the organism with increased viral resistance, a common form of horizontal gene transfer (HGT). The TAG codon was then converted to a sense codon through the introduction of an orthogonal translation system (OTS) containing an aminoacyl-tRNA synthetase (aaRS):tRNA pair, permitting site-specific incorporation of sAAs into proteins without impairing cellular fitness[11]. Leveraging these unique properties of the GRO, we sought to reintroduce the TAG codon into essential genes to restrict growth to defined media containing sAAs. We also eliminated the use of multi-copy plasmids, which reduce viability and growth[16], impose biosynthetic burden, persist poorly in host cells over time[17], and increase the risk of acquiring genetic escape mutants[3], by manipulating native chromosomal essential genes and integrating the OTS into the genome. To engineer synthetic auxotrophies, we chose essential genes of varying expression levels (Methods), many of whose functions (for example, replication or translation) cannot be complemented by cross-feeding of metabolites. Genes dispersed throughout the genome were selected to prevent a single HGT event from compromising containment.

We pursued three strategies to engineer dependence on non-toxic, membrane-permeable, and well-characterized sAAs through the introduction of TAG codons into essential genes: (1) insertion at the amino terminus; (2) substitution of residues with computationally predicted tolerances[18]; and (3) substitution of conserved[13] residues at functional sites (Fig. 1a). We initially pursued the first two strategies in a GRO containing an OTS optimized for the sAA p-acetyl-L-phenylalanine (pAcF, α; see Methods for a detailed explanation of the nomenclature used). Using multiplex automated genome engineering (MAGE)[12], we targeted 155 codons for TAG incorporation via 4 pools of oligonucleotides (Supplementary Tables 1 and 2) in permissive media containing pAcF and L-arabinose (aaRS induction) (Fig. 1b). After replica plating on non-permissive media lacking pAcF and L-arabinose, we isolated eight pAcF auxotrophs with one strain containing two TAGs in essential genes (Fig. 1c and Supplementary Table 3). To determine whether our strategy was capable of creating synthetic auxotrophs dependent on other sAAs, MAGE was used to mutagenize annotated residues in the sAA binding pocket of the pAcF aaRS (Supplementary Table 4) to accommodate p-iodo-L-phenylalanine (pIF, β) or p-azido-L-phenyalanine (pAzF, γ) in two strains. After MAGE-based incorporation of TAGs and selections on permissive and non-permissive solid media, we obtained 8 pIF and 23 pAzF auxotrophs harbouring 1–4 TAGs at 30 distinct loci across 20 essential genes (Supplementary Tables 3 and 5). Together, these data

[1]Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut 06520, USA. [2]Systems Biology Institute, Yale University, West Haven, Connecticut 06516, USA. [3]Department of Cellular and Molecular Physiology, Yale University, New Haven, Connecticut 06520, USA.
*These authors contributed equally to this work.

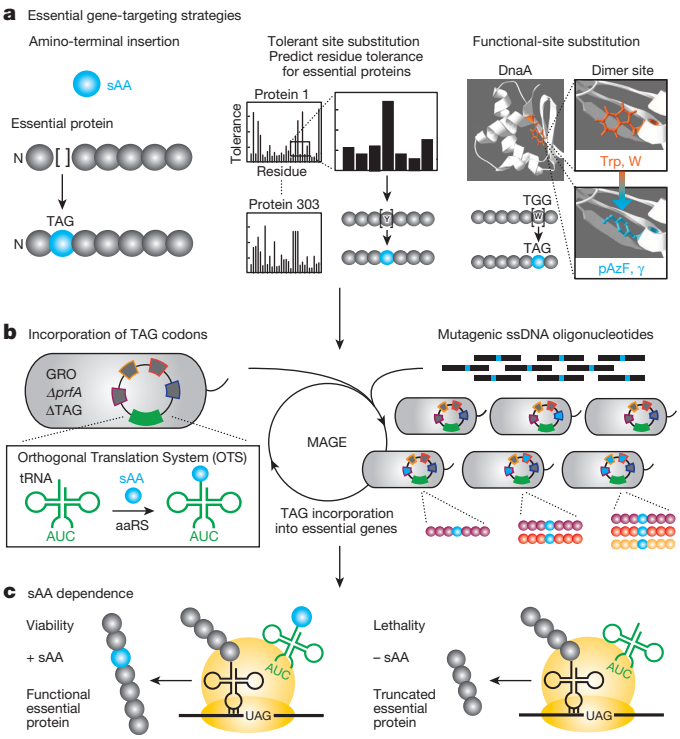


Figure 1 | Strategy used to engineer GROs to depend on sAAs for growth. **a**, Approaches used to identify suitable loci within essential proteins for sAA (blue) incorporation. **b**, MAGE was used for site-specific incorporation of TAG codons into essential genes of a GRO lacking all natural TAG codons (ΔTAG) and release factor 1 (*ΔprfA*), and containing an OTS (green) consisting of the *M. jannaschii* aaRS and cognate UAG-decoding tRNA. **c**, Synthetic auxotrophs that depend on sAAs for growth were isolated.

demonstrate the modularity of our approach and that synthetic auxotrophs can be engineered across many essential genes using multiple sAAs (Extended Data Fig. 1).

Measurements of doubling time in permissive media revealed minimal or no fitness impairment of synthetic auxotrophs relative to their non-contained ancestors with a genomically integrated OTS (Fig. 2a and Supplementary Table 6). To quantify the degree of containment, we measured the ratio of colony-forming units (c.f.u.) on non-permissive to permissive solid media and observed a range of escape frequencies spanning 10^{-3} to 10^{-7} (Fig. 2b). One notable strain DnaX.Y113α preserved the doubling time of its non-contained ancestor (Fig. 2a) while maintaining an escape frequency of 6.7×10^{-7} (Fig. 2b). We directly investigated pAcF incorporation in DnaX.Y113α using mass spectrometry and identified peptides containing pAcF at Y113 (Fig. 2c).

To investigate escape mechanisms of escape mutants derived from synthetic auxotrophs with one essential TAG codon, we performed targeted sequencing and observed transition mutations (A•T to G•C and G•C to A•T) commonly observed in mismatch-repair-deficient strains (*ΔmutS*)[19]. All isolated DnaX.Y113α escape mutants incorporate tryptophan by mutation of the TAG codon to TGG. SecY.Y122α escape mutants incorporate glutamine by mutation of *glnV* to form a glutamine amber suppressor or mutation of the *secY.Y122* TAG codon to CAG (Supplementary Table 7). One of three SecY.Y122α escape mutants was wild type at the *secY.Y122* TAG codon and putative amber-suppressor loci[20], but whole-genome sequencing (Supplementary Table 8) revealed a Q54D missense mutation in *rpsD* (30S ribosomal subunit S4). This site is implicated in ribosome fidelity[21,22] and is the causal mutation leading to escape in this mutant (Extended Data Fig. 2).

These escape mechanisms informed two sets of experiments to engineer strains with lower escape frequencies. First, we sought to create synthetic auxotrophs with an increased numbers of TAGs (Fig. 2d) by combining TAGs from strains possessing the lowest escape frequencies

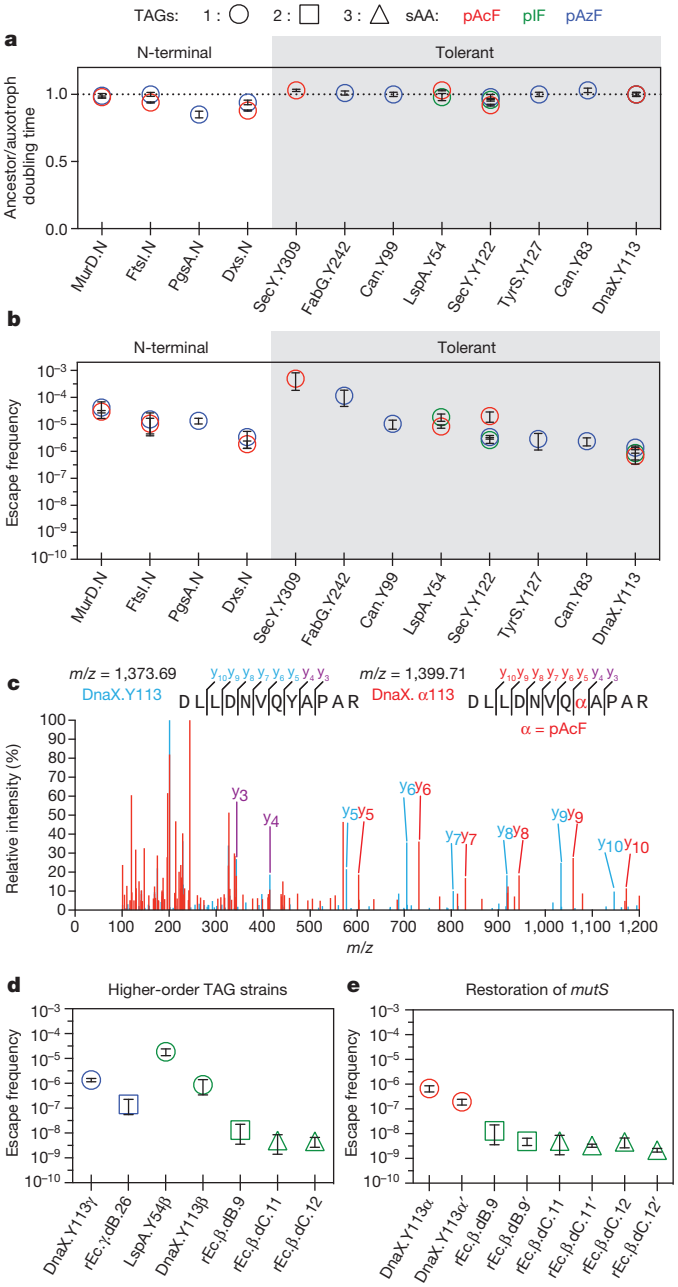


Figure 2 | Characterization of strains dependent on sAA incorporation in essential proteins. **a**, Doubling time ratios for the non-contained ancestor to synthetic auxotroph containing one TAG. **b**, Escape frequencies of strains from **a**. **c**, Superimposed MS/MS spectra for DnaX peptides from DnaX.Y113α (red) and the non-contained ancestor rEc.α (blue). Overlapping peaks are purple and a mass shift relative to rEc.α identifies Y113 as the pAcF incorporation site in DnaX.Y113α; see Methods. **d**, **e**, Escape frequencies for strains with multiple TAG codons (**d**) and/or functional mismatch repair (**e**) (prime, *mutS*$^{+}$). For all plots, average values of three technical replicates are plotted with error bars representing ±s.d. Reported results repeated at least three times in independent experiments.

(that is, *dnaX.Y113*, *lspA.Y54* and *secY.Y122*) into a single strain. In strains containing two TAGs, the escape frequency was reduced to 1.4×10^{-7} (rEc.γ.dB.26) and 1.4×10^{-8} (rEc.β.dB.9) (strain annotations are listed in Supplementary Table 6 and a complete description of our nomenclature can be found in the Methods). In strains containing three TAGs, escape frequencies were further reduced to 5.0×10^{-9} (rEc.β.dC.11) and 4.7×10^{-9} (rEc.β.dC.12). We used MAGE to assess quantitatively the effects of non-synonymous mutations at individual TAG codons in

strains incorporating pIF at SecY.Y122, DnaX.Y113, and LspA.Y54 by mutating the TAG site to sense codons for all 20 natural amino acids. Strains containing multiple TAGs were less likely to survive when one TAG was compromised (Extended Data Fig. 3). In a second set of experiments, we restored *mutS* (prime symbol (′) denotes a *mutS*$^{+}$ strain) and observed a decreased escape frequency in strains by 1.5- to 3.5-fold (Fig. 2e and Supplementary Table 6). Escape mutants derived from *mutS*$^{+}$ higher-order TAG strains exhibited impaired fitness with 1.14- to 1.28-fold greater doubling times than their contained ancestors. Whole-genome sequencing was performed on these escape mutants and revealed mutations of tyrosine tRNAs to form tyrosine amber (UAG) or ochre (UAA) suppressors (Supplementary Table 9).

To reduce escape frequencies below $\sim 10^{-9}$ and eliminate rescue by natural amino acids, we pursued a third strategy to replace conserved and functional residues in essential proteins with sAAs (Fig. 1a). Using the Conserved Domain Database[13], we searched all essential proteins for tyrosine, tryptophan and phenylalanine residues involved in protein–protein interactions (for example, dimerization) or located within active sites to identify candidates suitable for replacement with phenylalanine-derived sAAs. After targeted insertion of TAG codons using MAGE, we isolated four synthetic auxotrophs with pAzF incorporated at GlyQ.Y226 (glycyl-tRNA synthetase α subunit, dimer interface), Lnt.Y388 (apolipoprotein *N*-acyltransferase, active site), MurG.F243 (*N*-acetylglucosaminyl transferase, active site), and DnaA.W6 (chromosomal replication initiator protein, oligomerization site[23]) in strains with minor fitness impairments (Fig. 3a) and escape frequencies spanning 10^{-5} to 10^{-7} (Fig. 3b). Identical experiments to incorporate pAcF and pIF failed to generate synthetic auxotrophs, suggesting that the targeted residues are recalcitrant

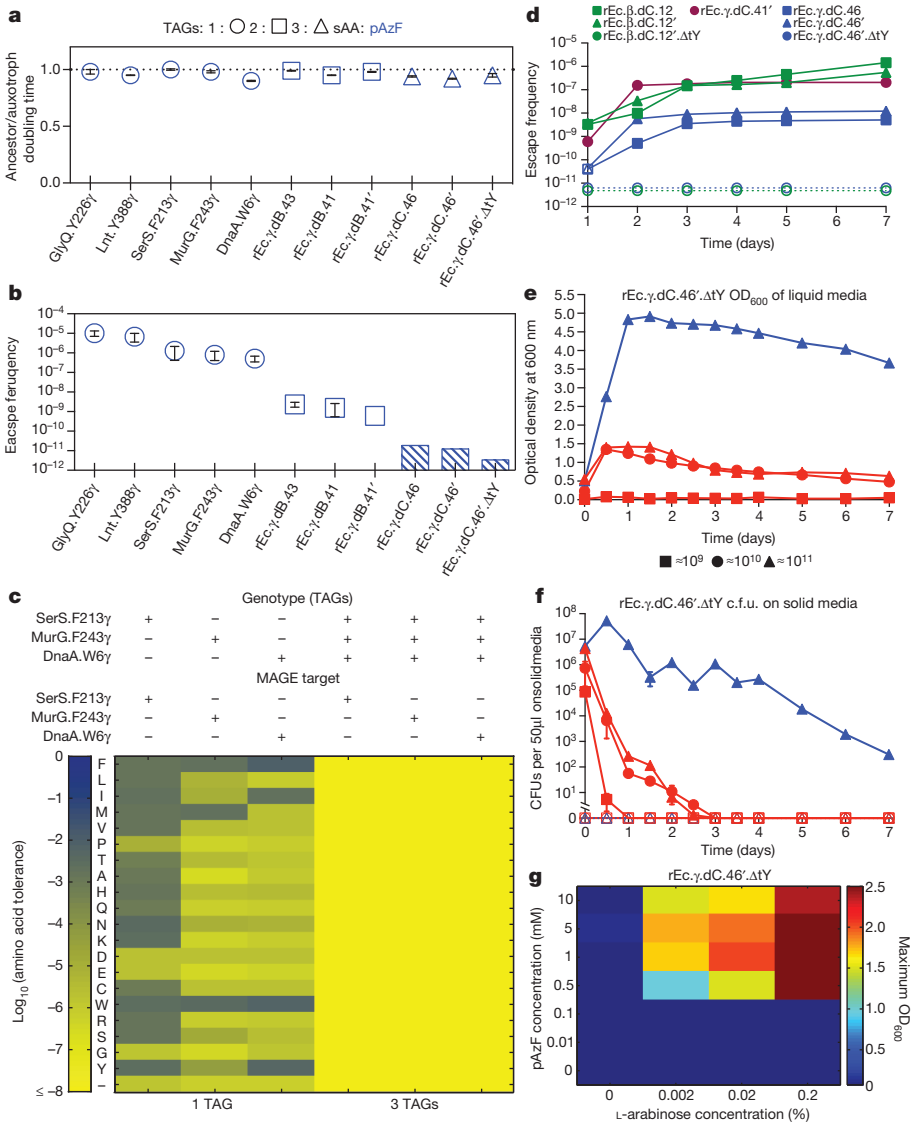


Figure 3 | Characterization of strains dependent on sAA incorporation at active and dimerization sites in essential proteins. **a**, Doubling time ratios for the non-contained ancestor to pAzF auxotroph with one or more TAGs at functional loci calculated from growth in 5 mM pAzF and 0.2% L-arabinose. **b**, Escape frequencies of strains in **a**; bars represent escape frequencies below the detection limit; average escape frequencies are plotted. **c**, Representative assay surveying tolerance of TAG loci to 20 amino acids in different synthetic auxotrophs and expressed as $\log_{10}$ of total cell survival. A + symbol indicates a TAG codon at the locus in the background strain; – indicates the wild-type codon; blue and yellow indicate high and low tolerance to substitution, respectively; see Methods. **d**, Representative escape assay monitoring escape frequencies up to 7 days after plating on solid non-permissive media; hollow symbols/dashed lines, no observed escape mutants; see Methods. **e**, Temporal monitoring of permissive (P, blue) and non-permissive (NP, red) cultures inoculated with $\sim 10^{9}$, 10^{10} or 10^{11} cells of rEc.γ.dC.46′.ΔtY by OD_{600}. **f**, Associated c.f.u. from **e** as sampled on permissive (solid lines) or non-permissive (dashed lines) solid media; c.f.u. were never observed on non-permissive solid media; hollow data points indicate no observed c.f.u. **g**, Maximum OD_{600} values during growth in LB media across a concentration gradient of pAzF and L-arabinose. For all plots, average values of three technical replicates are plotted with error bars representing ±s.d. Reported results repeated at least three times in independent experiments.

to replacement by pAcF and pIF. Since pAzF was able to replace conserved and functional tyrosine, phenylalanine and tryptophan residues across several essential proteins, we hypothesized that engineering strains to contain higher-order TAG combinations would limit escape by mutations that cause incorporation of natural amino acids at multiple TAG codons. Escape frequencies of 1.6×10^{-9} and 2.3×10^{-9} were observed for strains containing two TAG codons: rEc.γ.dB.41 (DnaA.W6 and MurG.F243) and rEc.γ.dB.43 (DnaA.W6 and SerS.F213), respectively. Upon restoring *mutS*, the escape frequency of rEc.γ.dB.41′ fell to 6.0×10^{-10} (Fig. 3b). Merging all three sites into one strain (rEc.γ.dC.46) and its *mutS*$^{+}$ derivative (rEc.γ.dC.46′) led to escape frequencies of $<7.9 \times 10^{-11}$ and $<4.4 \times 10^{-11}$ (below the detection limit of our plate-based assay), respectively (Supplementary Table 6).

Temporal monitoring of rEc.γ.dC.46′ revealed the emergence of growth-impaired escape mutants 2 days post-plating on non-permissive solid media (Fig. 3d). Sequencing of escape mutants derived from strains rEc.γ.dC.41′ and rEc.γ.dC.46′ revealed amber-suppressor-forming mutations at one of three tyrosine tRNAs (*tyrT*, *tyrV*, *tyrU*) with growth impairments spanning 1.61- to 2.10-fold increases in doubling time relative to contained ancestors (Supplementary Tables 9 and 10). Given that *E. coli* contains three tyrosine tRNAs, we hypothesized that deletion of *tyrT* and *tyrV*[24] would prevent acquisition of amber-suppressor-forming mutations at *tyrU*, as preservation of this single remaining copy of $tRNA^{Tyr}$ would be required to maintain fidelity of tyrosine incorporation during protein synthesis. We used λ-Red recombination to delete *tyrT* and *tyrV* in rEc.β.dC.12′, rEc.β.dC.12′.E7 (escape mutant of rEc.β.dC.12′), and rEc.γ.dC.46′ with a chloramphenicol resistance gene. Deletion of *tyrT* and *tyrV* restored containment of the escape mutant, establishing the causal escape mechanism (Extended Data Fig. 4). Moreover, *tyrT*/*V* deletions in rEc.β.dC.12′.ΔtY and rEc.γ.dC.46′.ΔtY decreased escape frequencies below detectable levels ($<4.9 \times 10^{-12}$ and $<6.3 \times 10^{-12}$, respectively) over the 7-day observation period (Fig. 3d and Supplementary Table 11).

To challenge strains rEc.β.dC.12′.ΔtY and rEc.γ.dC.46′.ΔtY with natural amino acids and mimic a potential HGT event, we introduced constructs containing phenylalanine or tryptophan amber-suppressor tRNAs. While growth of suppressor-containing strains was equivalent to the cognate-contained ancestor in permissive liquid media, severely impaired growth or no growth was observed in non-permissive media (Extended Data Fig. 5). Such findings are further supported in experiments where a large ($\sim 10^{11}$) inoculum of cells challenged on solid or in liquid (see below) non-permissive media do not yield escape mutants, providing ample opportunity for natural formation of a phenylalanine amber suppressor via mutation of one of two native copies of $tRNA^{Phe}$. These data support our hypothesis that synthetic auxotrophs containing higher-order TAG combinations depend on the sAA and limit growth from natural amino acids.

To interrogate the long-term stability of synthetic auxotrophs where escape mutant formation is not limited by a colony growth environment, temporal monitoring of rEc.γ.dC.46′.ΔtY was performed on large cell populations in liquid culture (1 litre of Luria–Bertani (LB) media) for 7 days with frequent OD_{600} measurements to track cell growth (Fig. 3e). Inoculation of $\sim 10^{11}$ cells in permissive media led to a confluent culture of contained cells within 24 h. Inoculation of $\sim 10^{9}$, $\sim 10^{10}$ and $\sim 10^{11}$ cells into non-permissive media revealed transient growth, which we propose is due to residual pAzF and L-arabinose from large inoculums, followed by a sustained decrease in cell density and growth termination. Cell survival and escape from liquid cultures was monitored by quantifying c.f.u. on permissive and non-permissive solid media, respectively (Fig. 3f). Plating on permissive solid media revealed a drop in c.f.u. to below the limit of detection within 1 day from the non-permissive flask inoculated with $\sim 10^{9}$ cells and 3 days from non-permissive flasks inoculated with $\sim 10^{10}$ or $\sim 10^{11}$ cells (Fig. 3f). No c.f.u. were observed from any culture plated on non-permissive solid media. To confirm the absence of a single escape mutant following an extended 20-day growth period (Extended Data Fig. 6), the non-permissive and permissive cultures inoculated with $\sim 10^{11}$ cells were plated across 30 non-permissive plates. Escape mutants were not observed and escape frequencies remained below the detection limit after 7 days, which is comparable to the solid media results. These results demonstrate that rEc.γ.dC.46′.ΔtY depends on pAzF, maintains long-term stability of biocontainment in permissive liquid media and exhibits termination of growth in non-permissive media.

To further investigate the dependency on sAAs, liquid growth profiles were collected for synthetic auxotrophs across sAA and L-arabinose concentration gradients. Growth of rEc.γ.dC.46′.ΔtY was not observed below 0.002% L-arabinose and 0.5 mM pAzF (Fig. 3g and Extended Data Fig. 7). Growth increased in a dose-dependent manner with increasing concentrations of pAzF and L-arabinose, where 5 mM pAzF and 0.2% L-arabinose was optimal for fitness (that is, maximum OD_{600} and minimum doubling time). In an equivalent experiment with rEc.β.dC.12′.ΔtY, 1 mM pIF and 0.2% L-arabinose was optimal for fitness (Extended Data Fig. 8). Since growth was not observed in media lacking either L-arabinose or the sAA, these data further support the dependency of synthetic auxotrophs on sAAs.

To determine whether a synthetic auxotroph could be rescued by metabolic cross-feeding, we evaluated the viability of strains on diverse media types. We grew wild-type *MG1655 Escherichia coli*, a biotin auxotroph (EcNR2[12]), a non-contained GRO (rEc.γ), and the pAzF synthetic auxotroph (rEc.γ.dC.46′) on solid media containing both pAzF/L-arabinose and biotin, either pAzF/L-arabinose or biotin, and on plates lacking small molecules (Fig. 4). Despite biotin auxotrophy, growth of EcNR2 on rich defined media without biotin was rescued in close proximity to wild-type *E. coli*, suggesting cross-feeding of essential metabolites (Extended Data Fig. 9). Blood agar and soil extracts without biotin or pAzF/L-arabinose supplementation supported growth of all strains except the synthetic auxotroph, which only grew on media supplemented with pAzF and L-arabinose. These data suggest that synthetic auxotrophies could lead to a more viable containment strategy for clinical (for example, blood) and environmental (for example, soil) settings, where metabolic auxotrophies can be overcome by proximal, metabolically competent strains.

Synthetic auxotrophs utilize unnatural biochemical building blocks necessary for essential proteins with activities that cannot be complemented by naturally occurring small molecules. We have previously shown that genomic recoding interferes with HGT from viruses[11], and have now extended orthogonal barriers by engineering two synthetic auxotrophs using two unique sAAs that exhibit escape frequencies below our detection limit ($<6.3 \times 10^{-12}$). These synthetic auxotrophs possess three essential TAGs at loci dispersed throughout the genome (0.84, 0.86 and 2.9 megabases apart), thereby limiting the likelihood that a single HGT event could compromise containment (Extended Data Fig. 1). These orthogonal barriers can be expanded further by incorporation of additional TAG sense codons across more than three essential

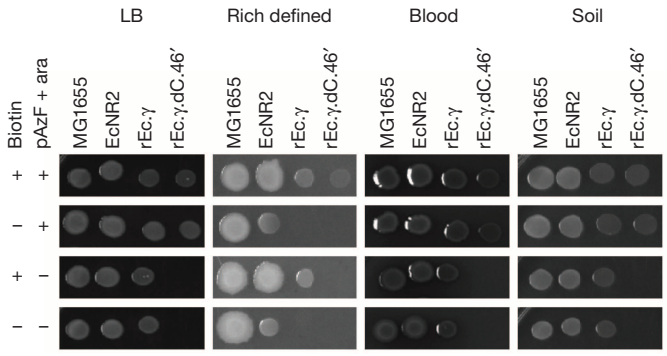


Figure 4 | Investigating the viability of synthetic auxotrophs on diverse media types. Rescue by cross-feeding shown through spotting on diverse media types with or without pAzF/L-arabinose (pAzF + ara) and biotin supplementation; EcNR2, rEc.γ, and rEc.γ.dC.46′ are auxotrophic for biotin (*ΔbioA*/*B*) and rEc.γ.dC.46′ is also a pAzF auxotroph.

genes, but will probably require concurrent advances in OTS performance to maintain fitness and viability (for example, enhanced activity and specificity of aaRS:tRNA pairs[25]). Our modular approach to biocontainment limits growth to synthetic environments containing unnatural biochemical building blocks with diverse chemistries. We anticipate that further genome recoding efforts[11,26,27] will enable auxotrophies for multiple sAAs that could be enhanced by other synthetic components including unnatural nucleotides and extended genetic alphabets[28–30]. Orthogonal biological systems employing multi-level containment mechanisms are uniquely suited to provide safe GMOs for clinical, environmental and industrial applications[1].

Despite the breadth of genomic diversity found in nature, all species utilize the same biochemical foundation to sustain life. The semantic architecture of the GRO employs orthogonal translational components, establishing the basis for a synthetic molecular language that relieves limitations on natural biological functions by depending on the incorporation of sAAs with exotic chemistries. This work sets the stage for future experiments to probe the optimality of the natural genetic code and to explore the plasticity of proteins and whole organisms capable of sampling new evolutionary landscapes.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 April; accepted 17 November 2014.
Published online 21 January 2015.

1. Way, J. C., Collins, J. J., Keasling, J. D. & Silver, P. A. Integrating biological redesign: where synthetic biology came from and where it needs to go. *Cell* **157,** 151–161 (2014).
2. Schmidt, M. & de Lorenzo, V. Synthetic constructs in/for the environment: managing the interplay between natural and engineered biology. *FEBS Lett.* **586,** 2199–2206 (2012).
3. Moe-Behrens, G. H., Davis, R. & Haynes, K. A. Preparing synthetic biology for the world. *Front. Microbiol.* **4,** 5 (2013).
4. Pieper, D. H. & Reineke, W. Engineering bacteria for bioremediation. *Curr. Opin. Biotechnol.* **11,** 262–270 (2000).
5. Steidler, L. Genetically engineered probiotics. *Best Pract. Res. Clin. Gastroenterol.* **17,** 861–876 (2003).
6. Kong, W. *et al.* Regulated programmed lysis of recombinant *Salmonella* in host tissues to release protective antigens and confer biological containment. *Proc. Natl Acad. Sci. USA* **105,** 9361–9366 (2008).
7. Szafranski, P. *et al.* A new approach for containment of microorganisms: dual control of streptavidin expression by antisense RNA and the T7 transcription system. *Proc. Natl Acad. Sci. USA* **94,** 1059–1063 (1997).
8. Steidler, L. *et al.* Biological containment of genetically modified *Lactococcus lactis* for intestinal delivery of human interleukin 10. *Nature Biotechnol.* **21,** 785–789 (2003).
9. Ronchel, M. C. & Ramos, J. L. Dual system to reinforce biological containment of recombinant bacteria designed for rhizoremediation. *Appl. Environ. Microbiol.* **67,** 2649–2656 (2001).
10. Jensen, L. B., Ramos, J. L., Kaneva, Z. & Molin, S. A substrate-dependent biological containment system for *Pseudomonas putida* based on the *Escherichia coli gef* gene. *Appl. Environ. Microbiol.* **59,** 3713–3717 (1993).
11. Lajoie, M. J. *et al.* Genomically recoded organisms expand biological functions. *Science* **342,** 357–360 (2013).
12. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460,** 894–898 (2009).
13. Marchler-Bauer, A. *et al.* CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res.* **33,** D192–D196 (2005).
14. Cohen, S. N., Chang, A. C., Boyer, H. W. & Helling, R. B. Construction of biologically functional bacterial plasmids *in vitro*. *Proc. Natl Acad. Sci. USA* **70,** 3240–3244 (1973).
15. Berg, P., Baltimore, D., Brenner, S., Roblin, R. O. & Singer, M. F. Summary statement of the Asilomar conference on recombinant DNA molecules. *Proc. Natl Acad. Sci. USA* **72,** 1981–1984 (1975).
16. Betenbaugh, M. J., Beaty, C. & Dhurjati, P. Effects of plasmid amplification and recombinant gene expression on the growth kinetics of recombinant *E. coli*. *Biotechnol. Bioeng.* **33,** 1425–1436 (1989).
17. Sørensen, S. J., Bailey, M., Hansen, L. H., Kroer, N. & Wuertz, S. Studying plasmid horizontal transfer *in situ*: a critical review. *Nature Rev. Microbiol.* **3,** 700–710 (2005).
18. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols* **4,** 1073–1081 (2009).
19. Schaaper, R. M. & Dunn, R. L. Spectra of spontaneous mutations in *Escherichia coli* strains defective in mismatch correction: the nature of *in vivo* DNA replication errors. *Proc. Natl Acad. Sci. USA* **84,** 6220–6224 (1987).
20. Eggertsson, G. & Soll, D. Transfer ribonucleic acid-mediated suppression of termination codons in *Escherichia coli*. *Microbiol. Rev.* **52,** 354–374 (1988).
21. Maisnier-Patin, S., Berg, O. G., Liljas, L. & Andersson, D. I. Compensatory adaptation to the deleterious effect of antibiotic resistance in *Salmonella typhimurium*. *Mol. Microbiol.* **46,** 355–366 (2002).
22. Holberger, L. E. & Hayes, C. S. Ribosomal protein S12 and aminoglycoside antibiotics modulate A-site mRNA cleavage and transfer-messenger RNA activity in *Escherichia coli*. *J. Biol. Chem.* **284,** 32188–32200 (2009).
23. Abe, Y. *et al.* Structure and function of DnaA N-terminal domains: specific sites and mechanisms in inter-DnaA interaction and in DnaB helicase loading on oriC. *J. Biol. Chem.* **282,** 17816–17827 (2007).
24. Bösl, M. & Kersten, H. A novel RNA product of the tyrT operon of *Escherichia coli*. *Nucleic Acids Res.* **19,** 5863–5870 (1991).
25. O'Donoghue, P., Ling, J., Wang, Y. S. & Soll, D. Upgrading protein synthesis for synthetic biology. *Nature Chem. Biol.* **9,** 594–598 (2013).
26. Isaacs, F. J. *et al.* Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333,** 348–353 (2011).
27. Lajoie, M. J. *et al.* Probing the limits of genetic recoding in essential genes. *Science* **342,** 361–363 (2013).
28. Bain, J. D., Switzer, C., Chamberlin, A. R. & Benner, S. A. Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code. *Nature* **356,** 537–539 (1992).
29. Pinheiro, V. B. *et al.* Synthetic genetic polymers capable of heredity and evolution. *Science* **336,** 341–344 (2012).
30. Hammerling, M. J. *et al.* Bacteriophages use an expanded genetic code on evolutionary paths to higher fitness. *Nature Chem. Biol.* **10,** 178–180 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to D. Söll and Y. S. Wang for discussion and for providing pTech-*supU*. We thank N. Carriero and R. Bjornson at the Yale Biomedical High Performance Computing Cluster for assistance with SIFT. Funding received was from the Defense Advanced Research Projects Agency (N66001-12-C-4020, N66001-12-C-4211); NIH-MSTP-TG-T32GM07205 (A.D.H.); DuPont Inc. and the Arnold and Mabel Beckman Foundation (F.J.I.).

Author Contributions A.J.R. and F.J.I. conceived the study, designed experiments and wrote the paper with assistance from A.D.H. and S.R.K.; A.J.R. conducted experiments with assistance from S.R.K., A.D.H., Z.L., M.W.G., M.A., J.R.P. and R.R.G.; B.M.G. and J.R. conducted mass spectrometry. All authors commented on the paper and F.J.I. supervised all aspects of the study.

Author Information Genome sequences have been deposited in GenBank under the accession numbers CP010455 (rEc.γ.dC.46) and CP010456 (rEc.β.dC.12). Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.J.I. (farren.isaacs@yale.edu).

METHODS

Reagents. Oligonucleotide synthesis was performed by Integrated DNA Technologies (IDT) and Keck Foundation Biotechnology Resource Laboratory at Yale University (Supplementary Table 1). Unless otherwise stated, all cultures were grown in LB media. The following selective agents and inducers were used at the specified concentrations: ampicillin (amp, 50 μg ml^{-1}), carbenicillin (carb, 50 μg ml^{-1}), zeocin (zeo, 10 μg ml^{-1}), spectinomycin (spec, 95 μg ml^{-1}) and sodium dodecyl sulphate (SDS, 0.005% w/v), isopropyl-β-D-1-thiogalactopyranoside (IPTG, 100 μM), 5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside (X-Gal, 40 μg ml^{-1}), and L-arabinose (ara, 0.2% w/v unless otherwise indicated). sAAs were used at 1 mM unless otherwise indicated and purchased from PepTech (pAcF, AL624-2), BaChem (pIF, F-3075.0005) and Chem-Impex International (pAzF, 03376).

Plasmids. All tRNAs used to assess tolerance for tryptophan and phenylalanine at TAG codons were contained within the pTech plasmid backbone and driven by the *lpp* promoter[31]. Isothermal assembly[32] was used to replace the chloramphenicol acetyltransferase (*cat*) gene with the *sh ble* gene for resistance to zeocin.

The *supU* amber suppressor tRNA[33] was used to assess tolerance for tryptophan and a phenylalanine amber suppressor[31] was used to assess tolerance for phenylalanine. pTech-*supU* was provided by the laboratory of D. Söll and *supPhe* was synthesized by IDT and isothermally assembled into the pTech plasmid backbone to obtain pTech-*supPhe*.

Conversion of aminoacyl-tRNA synthetase specificity. The pAcF OTS was integrated into the genome of the GRO linked to a counter-selectable gene *tolC*. Co-selection multiplex automated genome engineering (CoS-MAGE[34]) was used as described previously to introduce annotated mutations[35] to the sAA binding pocket of the aaRS for specificity towards pAzF or pIF (Supplementary Table 4). Sanger sequencing was used to verify these mutations. Conversion of sAA-specificity was assessed in sequence-verified clones upon growth in the presence of sAA incorporation and episomally-expressed GFP containing an in-frame TAG codon at residue 151 within the protein product. OTS-mediated suppression of this codon with the sAA (that is, pAzF, pIF) generated a full-length fluorescent product, indicating that sAA incorporation had occurred and specificity was achieved.

TAG codon incorporation into essential genes. We applied three unique strategies to identify permissive sites in essential genes for TAG codon incorporation (Fig. 1). In our first strategy, a subset of essential genes[36] were chosen for the incorporation of one or more TAG codons immediately after the start codon to encode a sAA at the amino terminus. To explore a diverse library of incorporation targets within the *E. coli* proteome, in our second strategy we applied the sorting intolerant from tolerant (SIFT) algorithm[18] (downloaded on the Yale Biomedical High Performance Computing Cluster, http://sift.jcvi.org/) to the entire panel of essential *E. coli* proteins[36]. SIFT is an algorithm that uses sequence homology to predict the tolerance of amino acid substitutions at different indices. In our workflow, genes were first split into three categories on the basis of wild-type expression level[37], and a further four subgroups by genomic location with the goal of targeting essential genes dispersed throughout the *E. coli* chromosome. Next, genes shown to be essential by multiple studies[36] were passed through SIFT. For each essential gene, two high, medium, and low tolerance sites were targeted for TAG incorporation by MAGE. By this approach, we were able to sample diverse residue types in proteins with varying wild-type expression levels.

In our third strategy, using the conserved domain database[13], we searched within all annotated essential proteins for tyrosine, tryptophan and phenylalanine residues predicted to participate in essential enzymatic reactions or protein–protein interactions (e.g., dimerization). To minimize the probability that the added functionality of the sAA would perturb protein function, we targeted sites that were observed to occur as tyrosine or tryptophan in different homologues.

GROs containing an OTS integrated into the chromosome were grown to mid-log phase in liquid permissive LB media and four cycles of MAGE were performed per pool of mutagenic oligonucleotides (oligonucleotide concentration ≤ 15 μM) as described previously[12,38]. To isolate synthetic auxotrophs, mutagenized cultures were plated on solid media and replica plated onto non-permissive media. To identify TAG incorporation loci, multiplex allele-specific colony (MASC) PCR was used to interrogate pools of up to eleven targeted loci as previously described[38], followed by verification using Sanger sequencing.

Genotyping. Sanger sequencing was performed by the Keck DNA Sequencing Facility at Yale University or by GENEWIZ, Inc. Genomic DNA for whole genome sequencing was prepared using a Qiagen Genomic DNA purification kit. Illumina libraries were prepared by the Yale Center for Genomic Analysis or the Dana Farber Cancer Institute. Illumina HiSeq or MiSeq sequencing systems were used for whole genome sequencing to generate 50- or 150-base-pair (bp) paired-end reads, respectively.

Whole-genome sequencing was used to analyse three escape mutants per background. In all cases, the direct ancestor to the escape mutant was also analysed. SNPs in escape mutants were identified relative to the reference genome *E. coli* C321.ΔA (CP006698.1, GI:54981157) using a previously described[11] software pipeline. SNPs listed in Supplementary Table 8 and Supplementary Table 10 were called by Freebayes in escape mutants.

Strains. All GROs used in this study are derived from C321.ΔA (CP006698.1, GI:54981157)[38] which lacks all TAG codons and release factor 1. This strain is derived from strain EcNR2 (*ΔmutS*:*catΔ*(*ybhB-bioAB*):[*cI857Δ*(*cro-ea59*):*tetR-bla*]), modified from *E. coli K-12 substr. MG1655*. In all synthetic auxotrophs, the *M. jannaschii*-derived OTS was genomically integrated into the GRO fused to the counter-selectable gene *tolC*. The OTS consists of an L-arabinose-inducible aaRS driven by the *araBAD* promoter, and a constitutively expressed cognate amber-decoding tRNA driven by the *proK* promoter. All genome modifications that required incorporation of dsDNA (for example, modifications to the *mutS* gene or incorporation of antibiotic selectable markers) were performed via λ-Red recombination[38].

Nomenclature of genomically recoded organisms and synthetic auxotrophs. To succinctly name strains, we have introduced a new one-letter amino acid code for sAAs using Greek lettering (pAcF = α, pIF = β, and pAzF = γ). Non-contained GROs lacking essential TAG codons are named according to the one letter sAA code for the specific OTS present in the organism. For example, a ΔTAG GRO with a genomically integrated pAcF OTS is rEc.α.

Biocontained GROs containing essential TAG codons are named according to two conventions based on the number of essential TAG codons in the auxotroph: (1) Strains with one essential TAG are named by the essential protein containing the sAA and the position and identity of the residue substituted therein (for example, a strain containing pAcF at residue 113 in DnaX is DnaX.Y113α); (2) Strains containing more than one essential TAG are named using the one letter sAA code for which the organism is auxotrophic. This is followed by a dependency code, d, indicating the presence of two (dB), three (dC) or four (dD) essential TAG codons, and then by a TAG combination number that uniquely identifies the specific combination of TAGs in the strain. Combinations are numbered from one through 46 and are listed in Supplementary Table 6.

Mismatch repair. The presence of a prime symbol (′) following the TAG combination number indicates that *mutS* has been restored at its native locus, imparting functional mismatch repair to the organism.

tRNA redundancy. Following the TAG combination number, Δt indicates the amino acid for which tRNA redundancy has been eliminated and is followed by the relevant amino acid (for example, a strain in which two of three total tyrosine tRNAs were deleted is ΔtY).

Escape mutant identity. At least three escape mutants were characterized per strain background that permitted an escape mutant. An escape mutant is designated by a number following the letter 'E' (for example, E1).

The summary of synthetic auxotrophs generated in this study illustrated in Extended Data Fig. 1 was constructed using the Circos[39] software.

Strains were grown at 34 °C in flat-bottomed 96-well plates containing 150 μL of LB media permissive for sAA incorporation, unless otherwise indicated. Strains were washed twice with sterile dH_2O before assessing growth in non-permissive media. Kinetic growth (OD_{600}) was monitored on a BioTek plate reader at ten-minute intervals in triplicate. Raw OD_{600} data from the plate reader were normalized to standard absorbance (OD_{600} at 1 cm path length) values using an empirically derived calibration curve ($Y = 1.9704x - 0.1183$, where $y = OD_{600}$ at 1 cm path length and $x = OD_{600}$ from plate reader; $R^2 = 0.998$). DTs were calculated in MATLAB using custom code. Reported values are the average between three technical replicates where error bars represent ± s.d. All reported results repeated at least three times in independent experiments. Maximum OD_{600} values were obtained after 24 h of growth and represent the average of three technical replicates. Reported results repeated at least three times in independent experiments.

Mass spectrometry. Histidine-tagged proteins were purified on NiNTA resin (Qiagen). Resolution of purity was assessed via SDS–PAGE. In-gel digestion was performed similarly to previously described methods[40]. Proteins were stained and imaged within the gel using Coomassie blue (R-250). A band corresponding to the molecular weight of DnaX was excised. Gel slices were processed into 1-mm cubes, washed in 1:1 (v/v) 50% CH_3CN/50 mM NH_4HCO_3, and then washed in 1:1 (v/v) 50% CH_3CN/10 mM NH_4CO_3. 13.33 ng μl^{-1} trypsin solution in 9:1 (v/v) 50 mM NH_4CO_3/50% CH_3CN was added and samples were incubated overnight at 37 °C. Peptides were extracted with 1:2 (v/v) 5% formic acid/50% CH_3CN and dried. Peptides were desalted by reconstitution in 3:8 (v/v) 70% formic acid/0.1% TFA, followed by loading onto a custom-made stage tip (2 × 1.06 mm punches of Empore C18 extraction disks [3 M] in a 200 μl pipette tip)[41] activated with 80% CH_3CN and 0.1% TFA. Tips were washed twice with 0.1% TFA and peptides eluted with 80% CH_3CN and 0.1% TFA. Peptides were dried and reconstituted for LC/MS/MS analysis. Capillary LC/MS/MS was carried out using an LTQ Orbitrap Velos (Thermo Scientific) with a nanoAcquity uHPLC (Waters) system as described previously[42]. The data were processed as described previously[11]. MASCOT scores were above

the identity or extensive homology threshold and representative spectra are illustrated in Fig. 2c.

Quantitative assessment of amino acid tolerance. The following workflow was performed to assess the tolerance for natural amino acid substitution at residues chosen for sAA incorporation. Strains were grown to mid-log phase in 1 ml of permissive LB media, and MAGE was performed as described[12] with modifications described here. Post induction of λ-Red proteins, cells were transferred to individual wells of a 96-well, V-bottomed plate, and washed twice at 4 °C with sterile deionized water (dH_2O). Cells were re-suspended in 50 µl of water or 1 µM mutagenic single-stranded DNA to convert a single in-frame essential TAG codon to one of 20 sense codons, and electroporated in a 96-well plate. Cells were electroporated using the Harvard BTX electroporation system (2.4 kV, 750 Ω, 25 µF). Electroporated cells were recovered in 1.5 ml of fresh permissive media in a 96-well plate for 4 h at 34 °C shaking. Cells were pelleted, washed twice with sterile dH_2O, and re-suspended in 200 µl of 1× PBS. Serial dilutions were made in 1× PBS and 50 µl each of non-diluted and 100-fold diluted samples were plated on solid non-permissive LB media. 50 µl each of higher dilutions were plated on permissive solid media and all plates were incubated for 20 h at 34 °C.

Colony-forming units counted on non-permissive media were expressed as a ratio of total c.f.u. on permissive media. Since the frequency of MAGE-mediated recombination (~0.3)[12] exceeds the escape frequencies of these background strains ($\leq 10^{-5}$), we directly correlated these ratios to amino acid tolerance. MATLAB was used to calculate the $\log_{10}$ of this ratio. Where no c.f.u. were observed on non-permissive media, indicative of a highly intolerant substitution, a ratio could not be calculated and these values were defaulted to NaN within MATLAB. A heat map was used to compare representative data for one experiment, where blue indicates a tolerated substitution and yellow, a non-tolerated substitution.

Twenty-one separate MAGE experiments were performed as described above for each strain, per essential genomic TAG, to assess tolerance for each of the 20 natural amino acids at each TAG site, plus a negative control (water). Strains with one TAG codon (SecY.Y122β, DnaX.Y113β, LspA.Y54β, DnaA.W6γ, SerS.F213γ, and MurG.F243γ) were assessed across 21 (including the negative control) experiments per strain, strains with two TAG codons (rEc.β.dB.9) were assessed across 42 (including two negative control) experiments per strain, and strains with three TAG codons (rEc.β.dC.12 and rEc.γ.dC.46) were assessed across 63 (including three negative control) experiments per strain. Reported results repeated at least three times in independent experiments.

Escape assays. Strains were grown in triplicate to late-log phase in 2 ml of permissive LB media, pelleted, washed twice with sterile dH_2O, and re-suspended in 200 µl 1× PBS. To obtain total and escape mutant c.f.u., serial dilutions were made and equal volumes were plated on permissive and non-permissive solid media plates (100 × 15 mm). Plates were incubated at 34 °C and escape frequency was calculated as the total number of escape mutant c.f.u. observed per total cells plated. Reported escape frequencies are means of three technical replicates where error bars represent ±s.d. To isolate escape mutants in strains with lower escape frequencies, ~10^{10}–10^{11} cells were plated and the resulting escape frequency from a representative escape assay is reported. When escape mutants were not detected upon plating ~10^{10}–10^{11} cells, the escape frequency is described to be below the limit of detection and reported as less than a frequency of one over the total number of cells plated. In all cases, reported results repeated at least three times in independent experiments. Where temporal monitoring of escape frequencies on solid media is reported (Fig. 3d), representative escape assays are plotted and results repeated at least three times in independent experiments.

Liquid escape assays. Long-term liquid growth was assessed for two strain backgrounds: the pAzF-dependent strain, rEc.γ.dC.46′.ΔtY, and its non-contained ancestor, rEc.γ. Growth of rEc.γ.dC.46′.ΔtY was separately assessed in permissive (+sAA/+L-arabinose) and non-permissive (−sAA/−L-arabinose) media and growth of rEc.γ was assessed in non-permissive media. In all cases, flasks contained carbenicillin to prevent contamination.

Strains were grown in 100 ml of LB media overnight. Cultures were then pelleted and washed twice with the same volume of sterile dH_2O. Washed pellets were re-suspended in LB media plus or minus small molecules and this slurry was then added to shake flasks containing 1 l of LB media plus or minus small molecules, at time zero. At this first time point, a 1 ml sample was obtained from each flask, from which the OD_{600} was measured and 50 µl was plated on both permissive (+sAA/+L-arabinose/+carbenicillin) and non-permissive (−sAA/−L-arabinose/+carbenicillin) solid LB media, in three technical replicates. Average c.f.u. counts are reported and error bars represent ±s.d. In all cases, c.f.u. on solid media were counted after 24 h of incubation at 34 °C. $OD_{600\,nm}$ readings and c.f.u. counts were collected in this manner for all subsequent time points for the following 20 days.

After 20 days of growth in liquid media, the two 1-l cultures of rEc.γ.dC.46′.ΔtY grown in non-permissive and permissive media were interrogated for the presence of a single escape mutant. The entire culture was pelleted, re-suspended in 7 ml of 1× PBS, and plated across 30 large non-permissive solid media plates that were subsequently monitored for c.f.u. formation over the following 7-day period. All reported results repeated at least three times in independent experiments.

Environmental challenges. Wild-type *E. coli K-12 substr. MG1655*, and additional *E. coli* strains EcNR2[12], rEc.γ.dC.46′ (a pAzF auxotroph), and rEc.γ (non-contained ancestor to rEc.γ.dC.46′) were grown to mid-log phase in 1 ml of LB media supplemented with small molecules, where necessary. Cultures were washed three times with sterile dH_2O and re-suspended in 1 ml 1× PBS. A total of 16, two-fold serial dilutions were made and spotted on the following solid media types: (1) LB, (2) EZ rich defined[43] (with modifications by Teknova) and containing 100× carbon source (40% glycerol), (3) blood agar (Teknova), or (4) soil extract agar (HiMedia). Prior to spotting, plates were topically supplemented with pAzF, L-arabinose, and/or biotin, and dried for at least 1 h. Spotted plates were incubated for 1 day at 34 °C and photographed in a Gel Doc XR+ running ImageLab v4.0.1 (BioRad).

Selectable markers used in this study. *cat* (1,015 bp)

cctgtgacggaagatcacttcgcagaataaataaatcctggtgtccctgttgataccgggaagccctgggccaacttttg
gcgaaaatgagacgttgatcggcacgtaagaggttccaactttcaccataatgaaataagatcactaccgggcgtattt
tttgagttgtcgagatttcaggagctaaggaagctaaaatggagaaaaaaatcactggatataccaccgttgatatatc
ccaatggcatcgtaaagaacattttgaggcatttcagtcagttgctcaatgtacctataaccagaccgttcagctggata
ttacggccttttaaagaccgtaaagaaaaataagcacaagttttatccggcctttattcacattcttgcccgcctgatga
atgctcatccggaattacgtatggcaatgaaagacggtgagctggtgatatgggatagtgttcacccttgttacaccgtt
ttccatgagcaaactgaaacgttttcatcgctctggagtgaataccacgacgatttccggcagtttctacacatatattcg
caagatgtggcgtgttacggtgaaaacctggcctatttccctaaagggtttattgagaatatgtttttcgtctcagccaat
ccctgggtgagtttcaccagttttgatttaaacgtggccaatatggacaacttcttcgcccccgttttcaccatgggcaaa
tattatacgcaaggcgacaaggtgctgatgccgctggcgattcaggttcatcatgccgtttgtgatggcttccatgtcgg
cagaatgcttaatgaattacaacagtactgcgatgagtggcagggcggggcgtaatttttttaaggcagttattggtgc
ccttaaacgcctggttgctacgcctgaataagtgataataagcggatgaatggcagaaattcgaaagcaaattcgacc
cggtcgtcggttcagggcagggtcgttaaatagccgcttatgtctattgctggtt

kanR (1,165 bp)

cctgtgacggaagatcacttcgcagaataaataaatcctggtgtccctgttgataccgggaagccctgggccaacttttg
gcgaaaatgagacgttgatcggcacgtaagaggttccaactttcaccataatgaaataagatcactaccgggcgtattt
tttgagttgtcgagattttcaggagctaaggaagctaaaatgagccatattcaacgggaaacgtcgaggccgcgatta
aattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcg
cttgtatgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgag
atggtcagactaaactggctgacggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatgg
ttactcaccactgcgatccccggaaaaacagcattccaggtattagaagaatatcctgattcaggtgaaaatattgttga
tgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgtaattgtccttttaacagcgatcgcgtatttcgtctcg
ctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgtaatggctggcctgttgaa
caagtctggaaagaaatgcataaacttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataac
cttatttttgacgaggggaaattaataggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgc
catcctatggaactgcctcggtgagttttctccttcattacagaaacggcttttttcaaaaatatggtattgataatcctgata
tgaataaattgcagtttcatttgatgctcgatgagtttttctaatttttttaaggcagttattggtgcccttaaacgcctggtt
gctacgcctgaataagtgataataagcggatgaatggcagaaattcgaaagcaaattcgacccggtcgtcggttcagg
gcagggtcgttaaatagccgcttatgtctattgctggtt

spec^R (1,201 bp)

cagccaggacagaaatgcctcgacttcgctgctgcccaaggttgccgggtgacgcacaccgtggaaacggatgaagg
cacgaacccagtggacataagcctgttcggttcgtaagctgtaatgcaagtagcgtatgcgctcacgcaactggtcca
gaaccttgaccgaacgcagcggtggtaacggcgcagtggcggttttcatggcttgttatgactgttttttggggtacag
tctatgcctcgggcatccaagcagcaagcgcgttacgccgtgggtcgatgtttgatgttatggagcagcaacgatgtta
cgcagcagggcagtcgcccataaacaaagttaaacatcatgagggaagcggtgatcgccgaagtatcgactcaacta
tcagaggtagttggcgtcatcgagcgccatctcgaaccgacgttgctggccgtacatttgtacggctccgcagtggatg
gcggcctgaagccacacagtgatattgatttgctggttacggtgaccgtaaggcttgatgaaacaacgcggcgagcttt
gatcaacgacctttttggaaacttcggcttcccctggagagagcgagattctccgcgctgtagaagtcaccattgttgtg
cacgacgacatcattccgtggcgttatccagctaagcgcgaactgcaatttggagaatggcagcgcaatgacattcttg
caggtatcttcgagccagccacgatcgacattgatctggctatcttgctgacaaaagcaagagaacatagcgttgcctt
ggtaggtccagcggcggaggaactctttgatccggttcctgaacaggatctatttgaggcgctaaatgaaaccttaacg
ctatggaactcgccgcccgactgggctggcgatgagcgaaatgtagtgcttacgttgtcccgcatttggtacagcgca
gtaaccggcaaaatcgcgccgaaggatgtcgctgccgactgggcaatggagcgcctgccggcccagtatcagcccg
tcatacttgaagctagacaggcttatcttggacaagaagaagatcgcttggcctcgcgcgcagatcagttggaagaat
ttgtccactacgtgaaaggcgagatcaccaaggtagtcggcaaataaagctttactgagctaataacaggactgctgg
taatcgcaggccttttttatttctgca

tolC (1,746 bp)

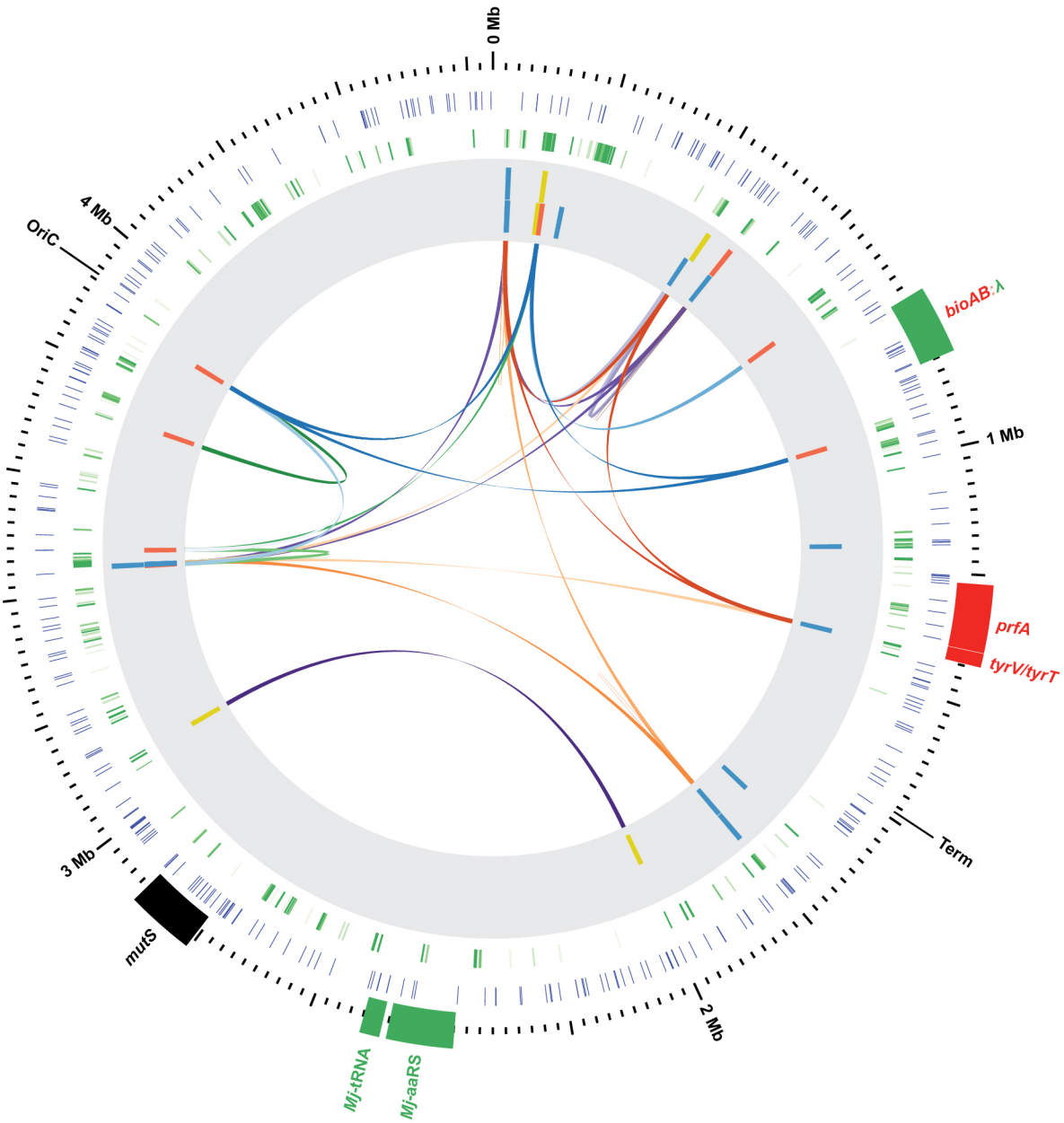
ttgaggcacattaacgcctatggcacgtaacgccaacctttgcggtagcggcttctgctagaatccgcaataatttttac
agtttgatcgcgctaaatactgcttcaccacaaggaatgcaaatgaagaaattgctccccattcttatcggcctgagcc
tttctgggttcagttcgttgagccaggccgagaacctgatgcaagtttatcagcaagcacgccttagtaacccggaatt
gcgtaagtctgccgccgatcgtgatgctgcctttgaaaaaattaatgaagcgcgcagtccattactgccacagctagg
tttaggtgcagattacacctatagcaacggctaccgcgacgcgaacggcatcaactctaacgcgaccagtgcgtcctt
gcagttaactcaatccatttttgatatgtcgaaatggcgtgcgttaacgctgcaggaaaaagcagcagggattcaggac
gtcacgtatcagaccgatcagcaaaccttgatcctcaacaccgcgaccgcttatttcaacgtgttgaatgctattgacg
ttctttcctatacacaggcacaaaaagaagcgatctaccgtcaattagatcaaaccacccaacgttttaacgtgggcct
ggtagcgatcaccgacgtgcagaacgcccgcgcacagtacgatacgtgctggcgaacgaagtgaccgcacgtaat
aaccttgataacgcggtagagcagctgcgccagatcaccggtaactactatccggaactggctgcgctgaatgtcgaa
aactttaaaaccgacaaaccacagccggttaacgcgctgctgaaagaagccgaaaaacgcaacctgtcgctgttaca

ggcacgcttgagccaggacctggcgcgcgagcaaattcgccaggcgcaggatggtcacttaccgactctggatttaa
cggcttctaccgggatttctgacacctcttatagcggttcgaaaacccgtggtgccgctggtacccagtatgacgatag
caatatgggccagaacaaagttggcctgagcttctcgctgccgatttatcagggcggaatggttaactcgcaggtgaa
acaggcacagtacaactttgtcggtgccagcgagcaactggaaagtgcccatcgtagcgtcgtgcagaccgtgcgttc
ctccttcaacaacattaatgcatctatcagtagcattaacgcctacaaacaagccgtagtttccgctcaaagctcattag
acgcgatggaagcgggctactcggtcggtacgcgtaccattgttgatgtgttggatgcgaccaccacgttgtacaacg
ccaagcaagagctggcgaatgcgcgttataactacctgattaatcagctgaatattaagtcagctctgggtacgttgaa
cgagcaggatctgctggcactgaacaatgcgctgagcaaaccggtttccactaatccggaaaacgttgcaccgcaaa
cgccggaacagaatgctattgctgatggttatgcgcctgatagcccggcaccagtcgttcagcaaacatccgcacgca
ctaccaccagtaacggtcataacccttttccgtaactgatgacgacgacggggaagcttaattagctgatctagaggca
tcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgctctcctgagtag
gacaaatccgccgccctaga

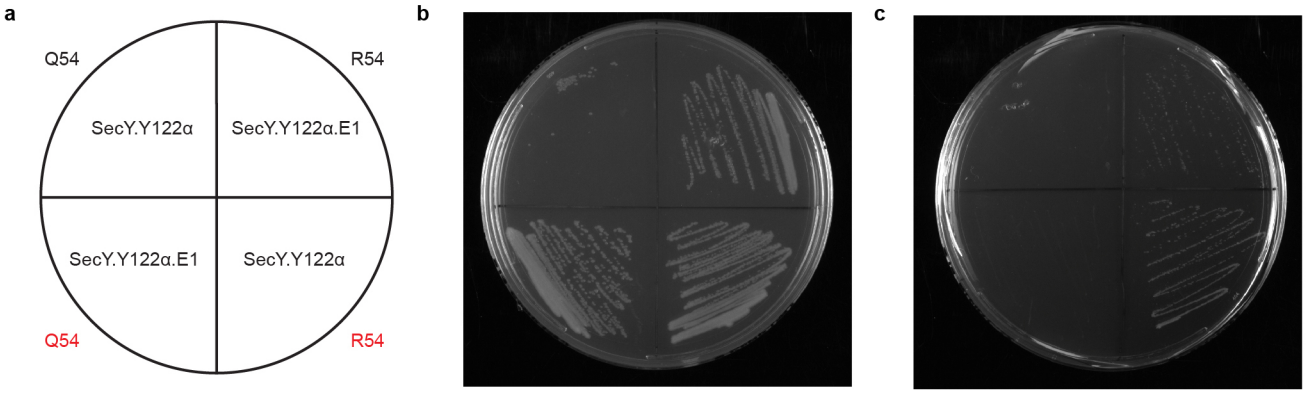
zeo^R (762 bp)

Ggtgttgacaattaatcatcggcatagtatatcggcatagtataatacgacaaggtgaggaactaaaccatggccaag
ttgaccagtgccgttccggtgctcaccgcgcgcgacgtcgccggagcggtcgagttctggaccgaccggctcgggttc
tcccgggacttcgtggaggacgacttcgccggtgtggtccgggacgacgtgaccctgttcatcagcgcggtccaggac
caggtggtgccggacaacaccctggcctgggtgtgggtgcgcggcctggacgagctgtacgccgagtggtcggagg
tcgtgtccacgaacttccgggacgcctccgggccggccatgaccgagatcggcgagcagccgtgggggcgggagtt
cgccctgcgcgacccggccggcaactgcgtgcacttcgtggccgaggagcaggactgacacgtccgacggcggccc
acgggtcccaggcctcggagatccgtccccctttccttttgtcgatatcatgtaattagttatgtcacgcttacattcacg
ccctcccccacatccgctctaaccgaaaaggaaggagttagacaacctgaagtctaggtccctatttatttttttatag
ttatgttagtattaagaacgttatttatatttcaaatttttctttttttttctgtacagacgcgtgtacgcatgtaacattatact
gaaaaccttgcttgagaaggttttgggacgctcgaaggctttaatttgcaagct

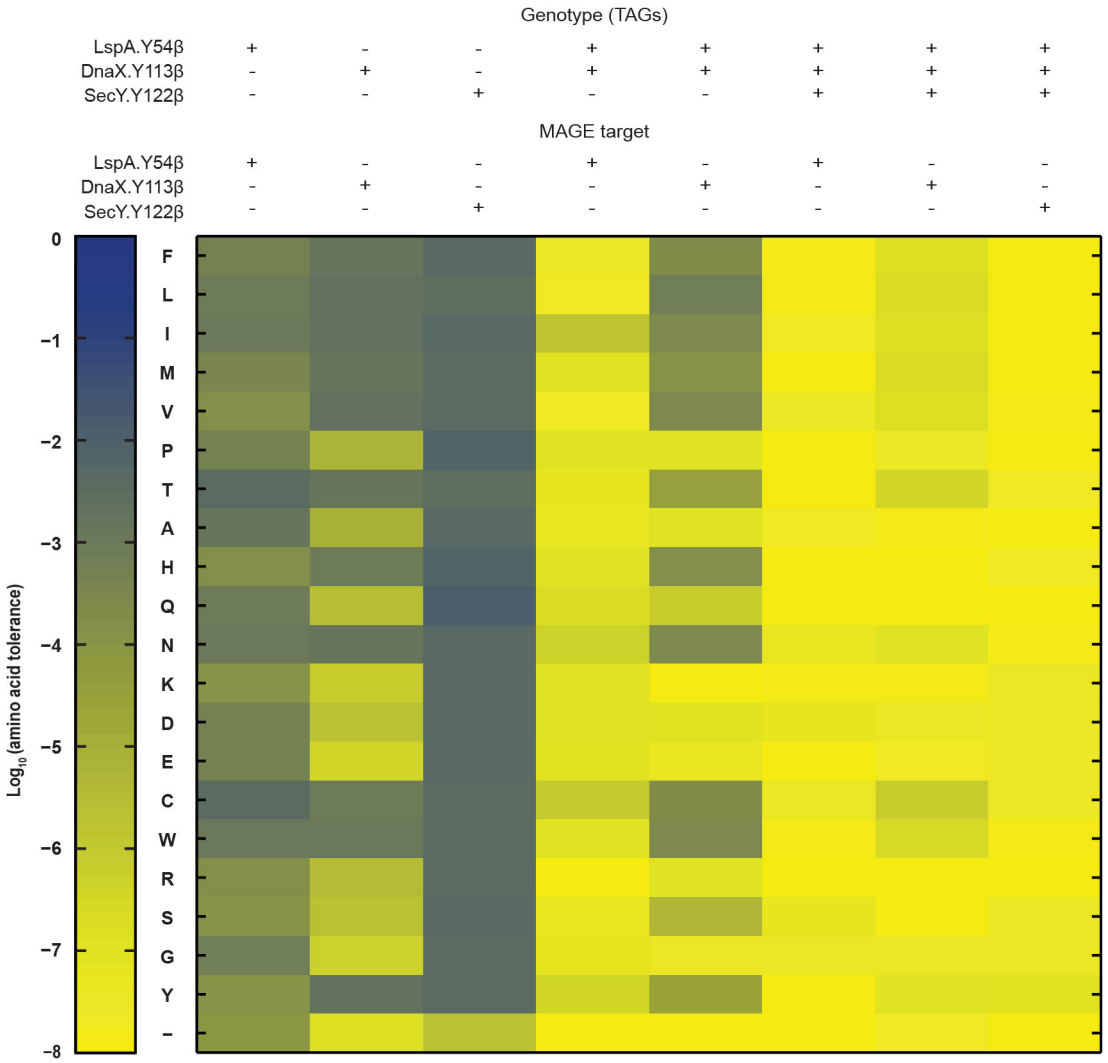
31. Normanly, J., Masson, J. M., Kleina, L. G., Abelson, J. & Miller, J. H. Construction of two *Escherichia coli* amber suppressor genes: tRNAPheCUA and tRNACysCUA. *Proc. Natl Acad. Sci. USA* **83,** 6548–6552 (1986).
32. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6,** 343–345 (2009).
33. Fan, C., Ho, J. M., Chirathivat, N., Soll, D. & Wang, Y. S. Exploring the substrate range of wild-type aminoacyl-tRNA synthetases. *ChemBioChem* **15,** 1805–1809 (2014).
34. Carr, P. A. *et al.* Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection. *Nucleic Acids Res.* **40,** e132 (2012).
35. Young, T. S., Ahmad, I., Yin, J. A. & Schultz, P. G. An enhanced system for unnatural amino acid mutagenesis in *E. coli*. *J. Mol. Biol.* **395,** 361–374 (2010).
36. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2,** 2006.0008 (2006).
37. Richmond, C. S., Glasner, J. D., Mau, R., Jin, H. & Blattner, F. R. Genome-wide expression profiling in *Escherichia coli* K-12. *Nucleic Acids Res.* **27,** 3821–3835 (1999).
38. Sharan, S. K., Thomason, L. C., Kuznetsov, S. G. & Court, D. L. Recombineering: a homologous recombination-based method of genetic engineering. *Nature Protocols* **4,** 206–223 (2009).
39. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19,** 1639–1645 (2009).
40. Rinehart, J. *et al.* WNK2 kinase is a novel regulator of essential neuronal cation-chloride cotransporters. *J. Biol. Chem.* **286,** 30171–30180 (2011).
41. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75,** 663–670 (2003).
42. Heinemann, I. U. *et al.* Enhanced phosphoserine insertion during *Escherichia coli* protein synthesis via partial UAG codon reassignment and release factor 1 deletion. *FEBS Lett.* **586,** 3716–3722 (2012).
43. Neidhardt, F. C., Bloch, P. L. & Smith, D. F. Culture medium for enterobacteria. *J. Bacteriol.* **119,** 736–747 (1974).



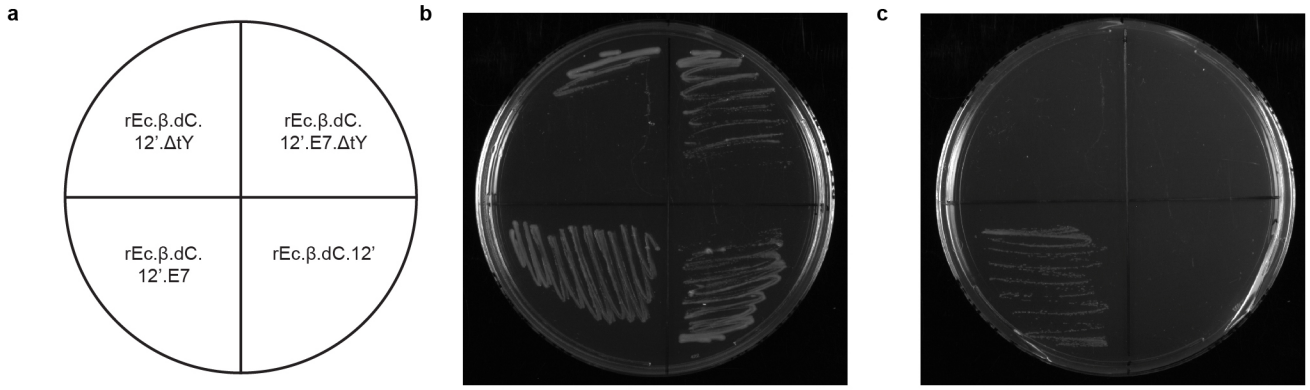
Extended Data Figure 1 | Comprehensive map of synthetic auxotrophs. Circos plot summarizing synthetic auxotrophs generated in this study. Red and green genes reflect knockouts and insertions, respectively. Outermost ticks indicate genomic location, inner blue ticks indicate locations where TAG codons were converted to TAA in the GRO, and green ticks reflect the locations of 303 *E. coli* essential genes. The shaded grey inner circle contains essential TAG loci in synthetic auxotrophs, where yellow ticks represent amino-terminal insertions, blue ticks represent tolerant substitutions, and red ticks represent functional-site substitutions. Innermost links represent unique combinations of TAGs in higher-order synthetic auxotrophs. Links of a single colour correspond to a single strain.



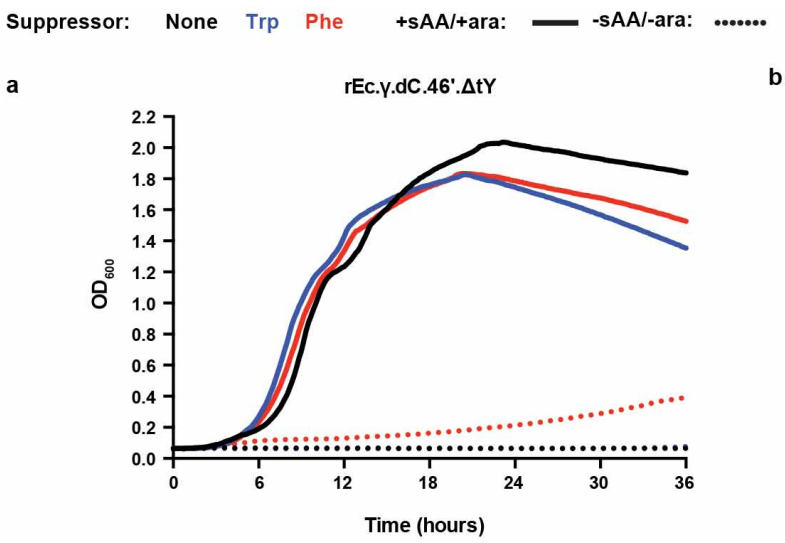
Extended Data Figure 2 | *rpsD.Q54R* is sufficient for loss of pAcF-dependence in SecY.Y122α. **a**, Plate map with genotypes of strains shown in **b** and **c**. On the top half of the plate SecY.Y122α.E1 (top right quadrant) contains the *rpsD.Q54R* mutation and is an escape mutant of pAcF-auxotroph, SecY.Y122α (top left quadrant). On the bottom half of the plate the *rpsD.Q54R* mutation was introduced into SecY.Y122α (bottom right quadrant), resulting in a loss of pAcF-dependence, and reverted to wild type in SecY.Y122α.E1 (bottom left quadrant), restoring pAcF-dependence. The amino acid present at residue 54 within RpsD is indicated at the perimeter of the plate, where red signifies that the given mutation was introduced into the genotype by MAGE to demonstrate the causal mechanism of escape. **b**, Growth on solid permissive media demonstrates growth of all strains. **c**, Growth on solid non-permissive media. Introduction of the *rpsD.Q54R* mutation into the synthetic auxotroph SecY.Y122α results in loss of containment (bottom right quadrant). Reverting the mutation to wild type in SecY.Y122α.E1 results in restoration of containment (bottom left quadrant). Together, these data demonstrate that the *rpsD.Q54R* mutation is sufficient for loss of pAcF-dependence in SecY.Y122α.



Extended Data Figure 3 | Quantitative assessment of amino acid tolerance in higher-order pIF auxotrophs. Representative assay surveying tolerance of one of three essential TAG loci to the twenty amino acids in different synthetic auxotrophs and expressed as $\log_{10}$ of total cell survival. The + symbol indicates the presence of a TAG codon at the specified locus in the background strain and – indicates the wild-type codon. Blue and yellow indicate high and low tolerance to substitution, respectively. Substitutions DnaX.Y113W and SecY.Y122Q are tolerated but yielded a lower percentage of survival on non-permissive media in a background with two TAGs, an effect that was pronounced in a background with three TAGs. While DnaX.Y113, SecY.Y122 and LspA.Y54 are permissive for most natural amino acids, strains with more than one of these essential TAGs are less prone to survive in the event that any one TAG is compromised. SecY.Y122Q and DnaX.Y113W were tolerated substitutions also observed in real escape mutants of these strains (Supplementary Table 7). Reported results repeated at least three times in independent experiments. Refer to the Methods for a complete description of this experiment.



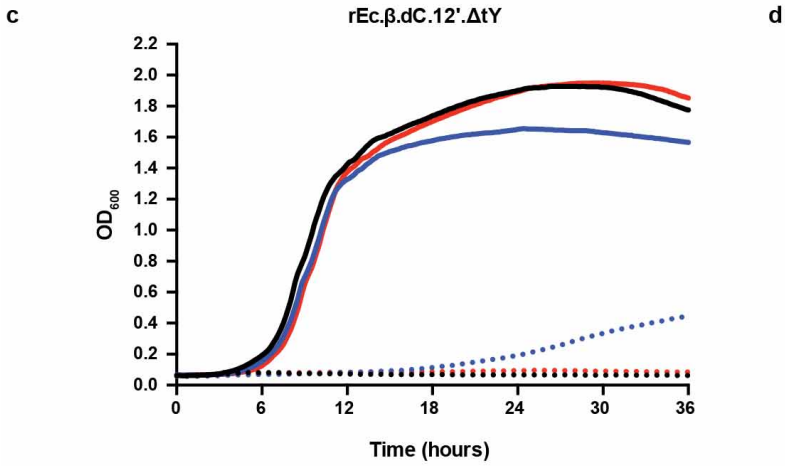
Extended Data Figure 4 | Deletion of *tyrT* and *tyrV* restores pIF-dependence and fitness of rEc.β.dC.12′.E7. **a**, Plate map with genotypes of strains shown in **b** and **c**. rEc.β.dC.12′.E7 is an escape mutant of its sAA-dependent ancestor (rEc.β.dC.12′) and contains a *tyrT* ochre suppressor mutation (*supC*). The fitness of rEc.β.dC.12′.E7 in permissive media is impaired relative to rEc.β.dC.12′, with doubling times of 91.74 (±1.49) and 61.81 (±0.65) minutes, respectively. Tyrosine tRNA redundancy was eliminated (ΔtY) in both strains by λ-Red mediated replacement of *tyrT* and *tyrV* with chloramphenicol acetyltransferase (*cat*), rendering the resulting strains (rEc.β.dC.12′.ΔtY and rEc.β.dC.12′.E7.ΔtY) dependent on *tyrU* for tyrosine incorporation during normal protein synthesis. Elimination of tyrosine redundancy reduced the escape frequency of rEc.β.dC.12′ from 2.17×10^{-9} (Fig. 2e) to $<4.85 \times 10^{-12}$ (no escape mutants were observed upon plating 2.06×10^{11} cells) and restored pIF-dependence in rEc.β.dC.12′.E7 to $<4.73 \times 10^{-12}$ (no escape mutants were observed upon plating 2.12×10^{11} cells). Escape mutants were not detected for either strain up to 7 days after plating on non-permissive media (Fig. 3d and Supplementary Table 11). Tyrosine tRNA deletion also restored the fitness of the escape mutant to approximately that of its sAA-dependent ancestor (60.66 ± 0.12 min). Taken together, these results establish *tyrT* as the causal mechanism of escape in rEc.β.dC.12′.E7. **b**, Growth on solid permissive LB media. **c**, Growth on solid non-permissive LB media. All reported doubling times are averages, where $n = 3$ technical replicates, and error bars represent ±s.d. Refer to the Methods for a complete description of escape frequencies.



b

rEc.γ.dC.46'.ΔtY

Suppressor	Condition	DT (mins)
None	+sAA/+ara	64.91
supU (Trp)	-sAA/-ara	No growth
supPhe (Phe)	-sAA/-ara	252.86

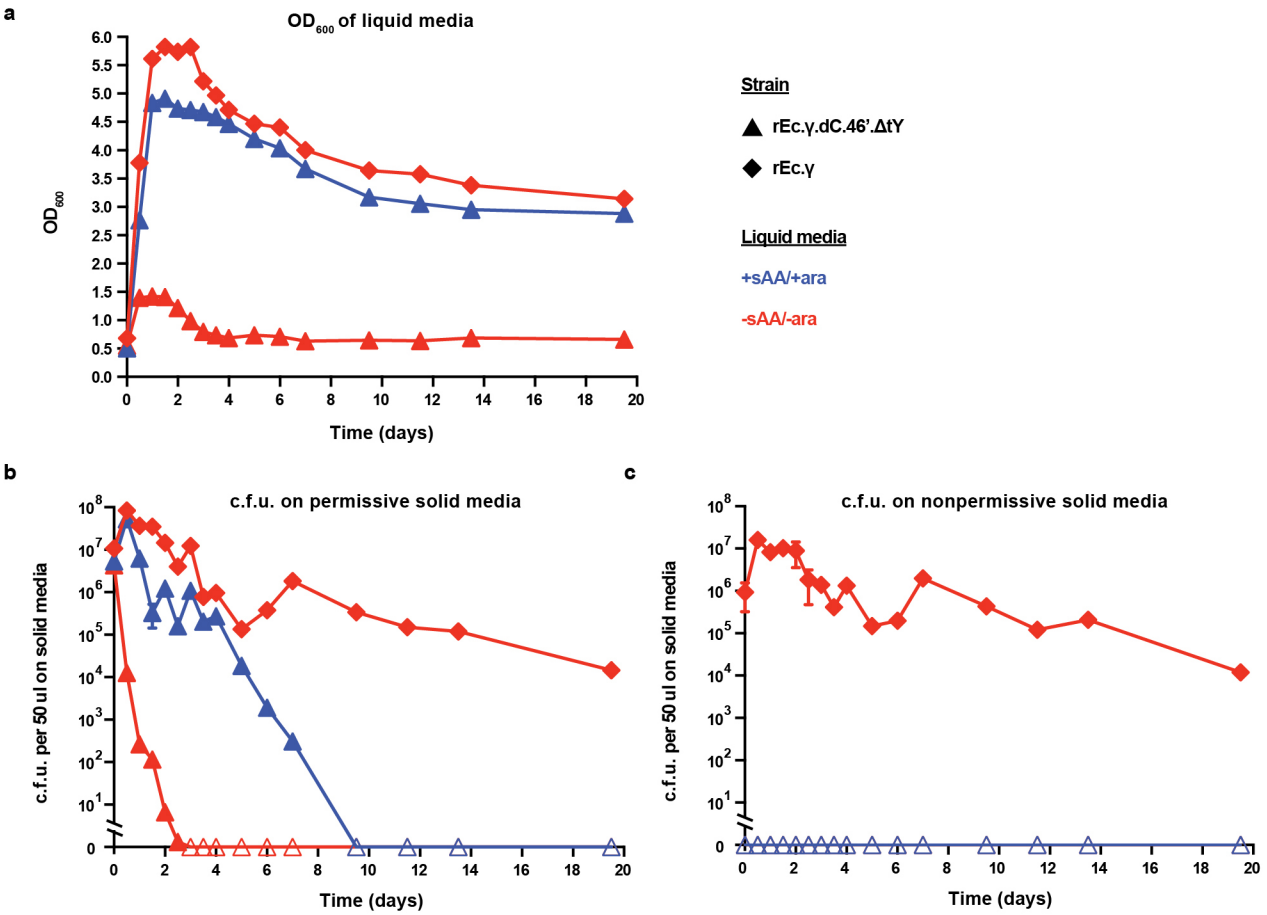


d

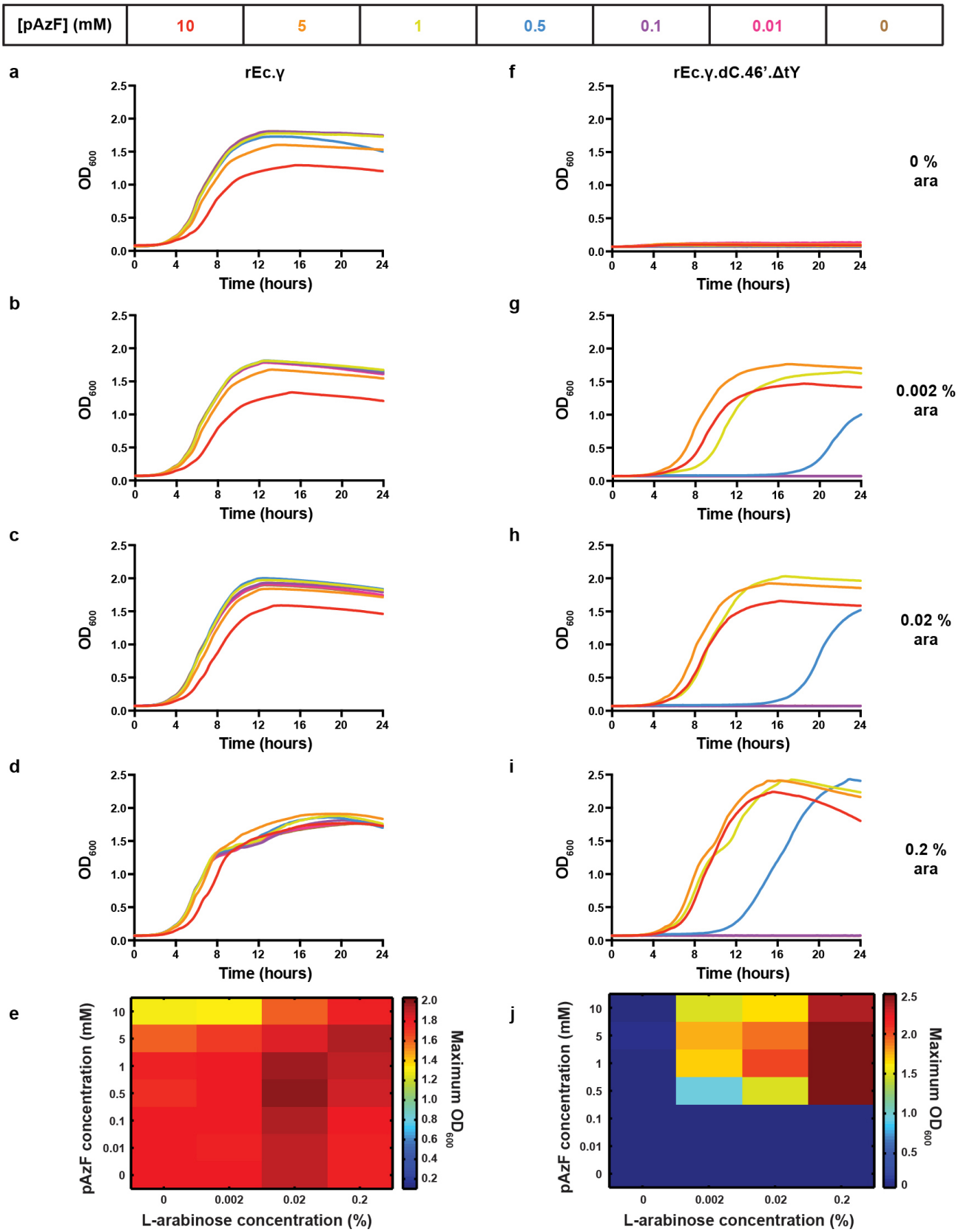
rEc.β.dC.12'.ΔtY

Suppressor	Condition	DT (mins)
None	+sAA/+ara	60.99
supU (Trp)	-sAA/-ara	380.82
supPhe (Phe)	-sAA/-ara	No growth

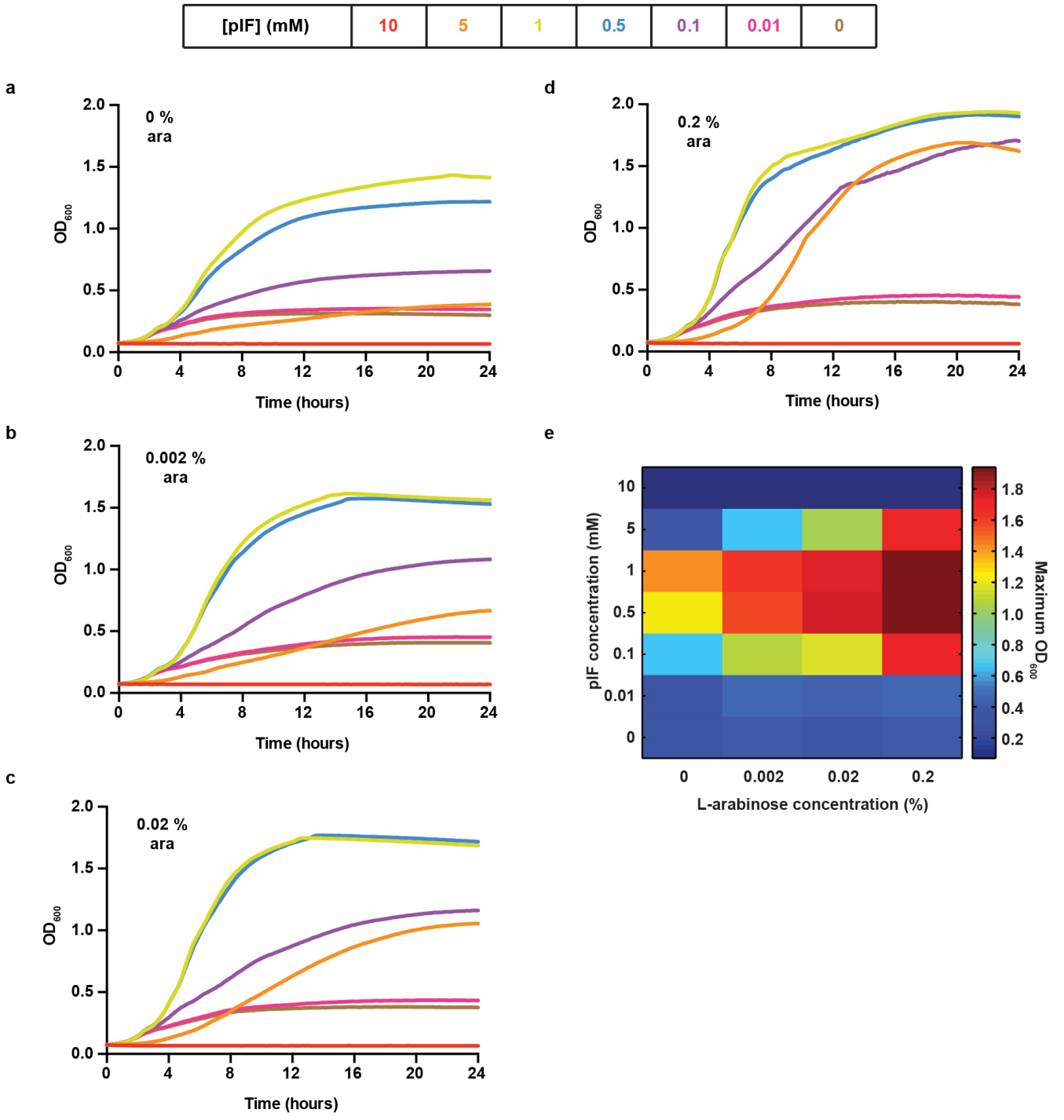
Extended Data Figure 5 | Growth profiles of strains expressing phenylalanine or tryptophan amber-suppressor tRNAs. a–d, Growth was assessed for rEc.γ.dC.46′.ΔtY and rEc.β.dC.12′.ΔtY in the presence of amber suppression by either pTech-*supU* (blue), pTech-*supPhe* (red), or in the absence of plasmid-based amber suppression (black). Cells were washed twice with dH_2O and re-suspended in the same volume of 1× PBS. Washed cells were normalized by OD_{600} to inoculate roughly equal numbers of cells per well. Growth profiles are shown for rEc.γ.dC.46′.ΔtY (**a**, **b**) and rEc.β.dC.12′.ΔtY (**c**, **d**) in permissive (+sAA/+L-arabinose, solid lines) and non-permissive (−sAA/−L-arabinose, dashed lines) LB liquid media. Doubling times are shown for the ancestral strain (black) in permissive media and suppressor-containing strains (red and blue) in non-permissive media where growth was observed. Plasmid containing strains were always grown in the presence of zeocin for plasmid maintenance. Growth was never observed for the contained ancestors in non-permissive media (black, dashed lines). In the presence of tryptophan suppression, growth of rEc.γ.dC.46′.ΔtY was not observed and growth of rEc.β.dC.12′.ΔtY was severely impaired (380 min doubling time), with a 6.24-fold increase in doubling time relative to the contained ancestor grown in permissive media. In the presence of phenylalanine suppression, growth of rEc.β.dC.12′.ΔtY was not observed and growth of rEc.γ.dC.46′.ΔtY was severely impaired (252 min doubling time), with a 3.90-fold increase in doubling time relative to the contained ancestor grown in permissive media. Representative growth profiles and doubling times are reported. These results repeated at least three times in individual experiments.



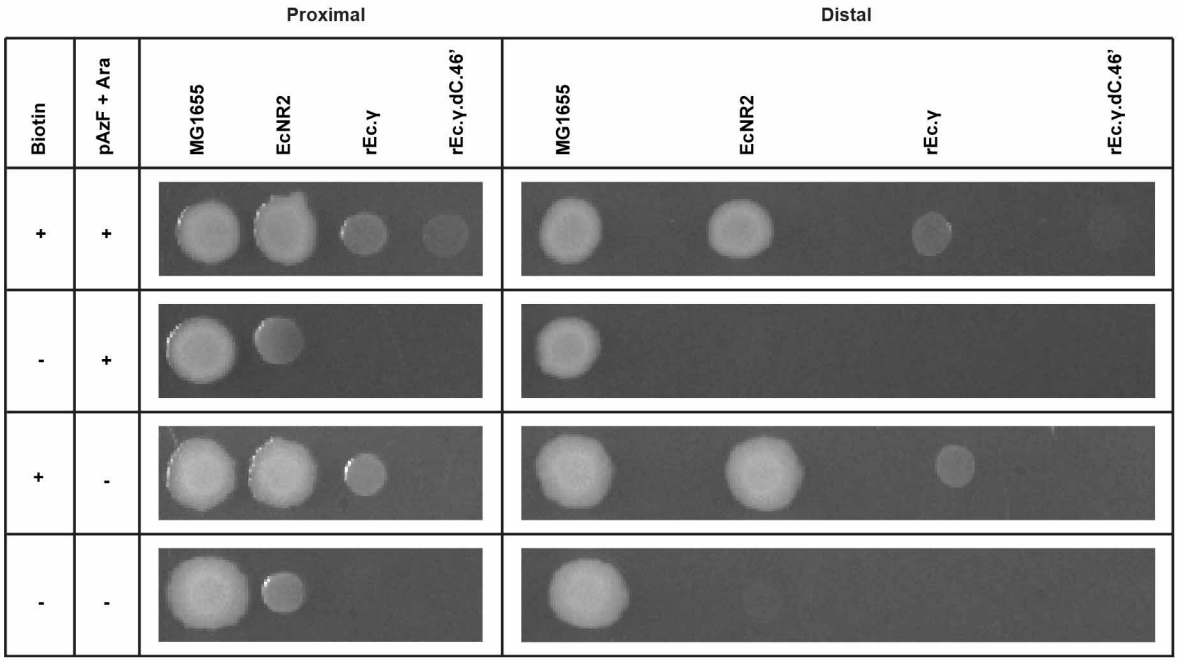
Extended Data Figure 6 | Long-term growth of rEc.γ.dC.46′.ΔtY in liquid LB media relative to rEc.γ. a–c, Approximately 10^{11} cells of strain rEc.γ.dC.46′.ΔtY (indicated with triangles) were inoculated into 1 l of permissive (+sAA/+L-arabinose, blue) or non-permissive (−sAA/−L-arabinose, red) LB media and incubated with agitation at 34 °C for 20 days. Results from the equivalent experiment with the non-contained ancestor rEc.γ (indicated with diamonds) are also shown. Cultures were frequently monitored by OD_{600} (**a**) and quantification of c.f.u. on solid permissive (+sAA/+L-arabinose) (**b**) and non-permissive (−sAA/−L-arabinose) (**c**) LB media. C.f.u. are plotted as the average of three replicates. Open symbols indicate that no c.f.u. were observed. Symbols for rEc.γ.dC.46′.ΔtY are not visible because c.f.u. were never observed from either permissive or non-permissive liquid cultures plated on non-permissive solid media. At the end of the 20-day growth period, both cultures containing rEc.γ.dC.46′.ΔtY were interrogated for the presence of a single escape mutant by plating each 1 l of culture across 30 non-permissive solid media plates. C.f.u. were not observed and remained below the limit of detection for the following 7-day observation period. We hypothesize that the decrease in c.f.u. counts obtained on permissive solid media for the permissive culture of rEc.γ.dC.46′.ΔtY reflects pAzF degradation at ≥6 days. Reported c.f.u. values are averages, where $n = 3$ technical replicates, and error bars are ±s.d. Reported results repeated at least three times in independent experiments. Refer to the Methods for a complete description of this experiment.



Extended Data Figure 7 | Dose-dependent growth of rEc.γ.dC.46′.ΔtY in pAzF and L-arabinose compared to the non-contained ancestor. Growth in LB media supplemented with different concentrations of pAzF and L-arabinose. **a**–**d**, Growth profiles for rEc.γ across a gradient of pAzF concentrations in the presence of 0% (**a**), 0.002% (**b**), 0.02% (**c**) and 0.2% (**d**) L-arabinose. **f**–**i**, Growth profiles for rEc.γ.dC.46′.ΔtY across a gradient of pAzF concentrations in the presence of 0% (**f**), 0.002% (**g**), 0.02% (**h**) and 0.2% (**i**) L-arabinose. **e**, **j**, Growth profiles illustrated in **a**–**d** and **f**–**i** are depicted as heat maps in **e** and **j**, respectively, where the maximum OD_{600} was obtained from the average of three replicates and plotted in MATLAB. Reported growth profiles and heat map values are averages, where $n = 3$ technical replicates. Reported results repeated at least three times in independent experiments.



Extended Data Figure 8 | Dose-dependent growth of rEc.β.dC.12′.ΔtY in pIF and L-arabinose. Growth in LB media supplemented with different concentrations of pIF and L-arabinose. **a**–**d**, Growth profiles for rEc.β.dC.12′.ΔtY across a gradient of pIF concentrations in the presence of 0% (**a**), 0.002% (**b**), 0.02% (**c**) and 0.2% (**d**) L-arabinose. **e**, Growth profiles illustrated in parts **a**–**d** are depicted as a heat map, where the maximum OD_{600} was obtained from the average of three replicates and plotted in MATLAB. Reported growth profiles and heat map values are averages, where $n = 3$ technical replicates. Reported results repeated at least three times in independent experiments.



Extended Data Figure 9 | Proximity-dependent complementation of biotin auxotrophy. Wild-type *E. coli K-12 substr. MG1655* and three strains auxotrophic for biotin, EcNR2, rEc.γ (a non-contained GRO with an integrated pAzF OTS) and rEc.γ.dC.46′ (also a synthetic auxotroph) were grown either adjacent or separately on rich-defined solid media. EcNR2 grew on biotin-deficient media when plated in close proximity to wild-type *E. coli*, suggesting cross-feeding of the essential metabolite. The pAzF auxotroph only grew on media supplemented with biotin, pAzF and L-arabinose.