

SCIENTIFIC REPORTS



OPEN

Evolving Bacterial Fitness with an Expanded Genetic Code

Drew S. Tack^{1,2}, Austin C. Cole², Raghav Shroff², Barrett R. Morrow² & Andrew D. Ellington²

Since the fixation of the genetic code, evolution has largely been confined to 20 proteinogenic amino acids. The development of orthogonal translation systems that allow for the codon-specific incorporation of noncanonical amino acids may provide a means to expand the code, but these translation systems cannot be simply superimposed on cells that have spent billions of years optimizing their genomes with the canonical code. We have therefore carried out directed evolution experiments with an orthogonal translation system that inserts 3-nitro-L-tyrosine across from amber codons, creating a 21 amino acid genetic code in which the amber stop codon ambiguously encodes either 3-nitro-L-tyrosine or stop. The 21 amino acid code is enforced through the inclusion of an addicted, essential gene, a beta-lactamase dependent upon 3-nitro-L-tyrosine incorporation. After 2000 generations of directed evolution, the fitness deficit of the original strain was largely repaired through mutations that limited the toxicity of the noncanonical. While the evolved lineages had not resolved the ambiguous coding of the amber codon, the improvements in fitness allowed new amber codons to populate protein coding sequences.

Since the fixation of the genetic code evolution has been confined to the 20 canonical amino acids, with some incursions by selenocysteine and pyrrolysine. Alternative codon tables (e.g. mitochondrial genomes) are likely evolved from the standard codon table and provide evidence that the canonical genetic code can evolve¹. A number of theories for the evolution of codon assignment and re-assignment have been proposed²⁻⁴, and directed evolution experiments have demonstrated the code is not as frozen as once believed⁵⁻⁷. However, a full accounting of how a cell might adapt to an expanded genetic code has yet to be presented.

Expanding the standard set of proteinogenic amino acids can be accomplished through changes to the underlying translational machinery. Orthogonal translation systems (OTSs) comprising aminoacyl-tRNA synthetase (aaRS)/suppressor tRNA pairs have been developed that do not significantly interact with the host translational machinery or interfere with already occupied portions of the genetic code⁸⁻¹⁰. Typically, these OTSs allow the incorporation of noncanonical amino acids (ncAAs) by suppressing the amber stop codon (UAG).

Unsurprisingly, cells containing an active OTS often exhibit fitness deficits¹¹, possibly because any protein terminated by an amber codon can be unnaturally extended. Efforts to knockout the protein responsible for termination at amber codons, release factor 1 (*prfA*), support this claim: some strains lacking *prfA* were found to be viable only when essential genes terminating with an amber stop codon were recoded to terminate with an alternative stop codon¹². In order to avoid these fitness impacts upon adopting a new code that would otherwise result in its rejection, previous studies with expanded genetic codes have either relied on bacteriophage, where the fitness of the host organism is irrelevant^{7,13}, or have relied on strains that entirely lack amber codons¹⁴, allowing ready capture of the eliminated codon to create a 21 amino acid genetic code¹⁵.

Here we utilize an engineered β -lactamase (*bla*) that is structurally dependent on OTS incorporation of the ncAA 3-nitro-L-tyrosine (3nY)⁹. This 'addicted' *bla* has allowed us to overcome fitness deficits and carry out long term evolution experiments with an ambiguous amber codon without loss of the underlying OTS. This system models the 'ambiguous intermediate' hypothesis of genetic code evolution, which proposes that translation of a specific codon can change by first becoming ambiguously translated before losing ambiguity and gaining specificity for a different amino acid. Here we demonstrate that our system stably incorporates an ncAA for 2,000 generations of evolution, and for the first time identify the entire complement of genomic mutations that lead to improved fitness in the presence of an enforced 21 amino acid code.

¹National Institute for Standards and Technology, Gaithersburg, Maryland, USA. ²Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, Texas, USA. Correspondence and requests for materials should be addressed to D.S.T. (email: drew.scott.tack@gmail.com)

Received: 27 November 2017

Accepted: 6 February 2018

Published online: 19 February 2018

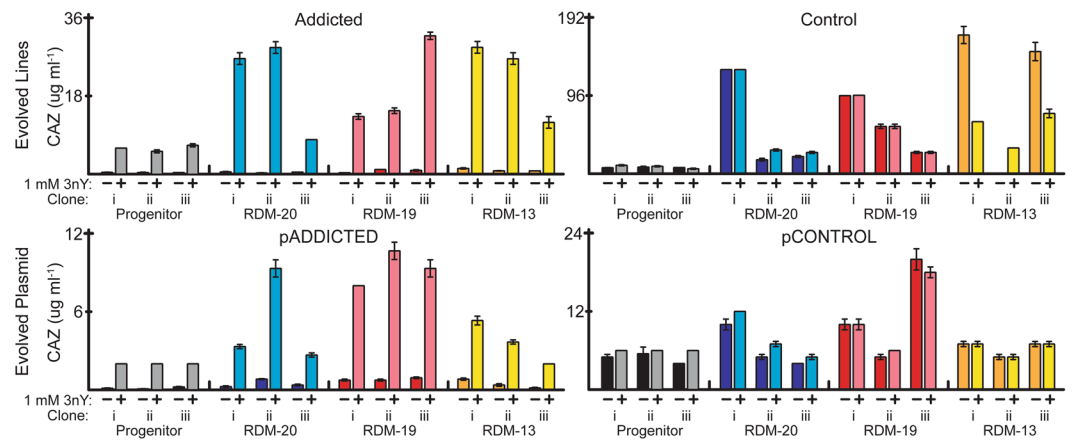


Figure 1. Ceftazidime MICs. MICs of progenitor cells (gray and black) and evolved lines in the absence and presence of 1 mM 3nY. All lines increased MICs during evolution (upper panels, compare gray/black to colored bars). Lines addicted to 3nY remained dependent on 3nY for ceftazidime resistance after 2000 generations (upper left) while control lines never required 3nY for ceftazidime resistance (upper right). Plasmids extracted from evolved lines and transferred to wild-type *E. coli* strain MG1655 showed smaller increases in ceftazidime resistance (lower graphs). Values are the average of biological triplicates, error bars represent s.e.m.

Results

Experimental set-up. We wished to examine the long-term adaptation and evolution of *E. coli* addicted to a nCAA, 3nY. We assembled an OTS for the incorporation of 3nY comprised of a *Methanocaldococcus jannaschii* tyrosyl-aaRS variant that had previously been engineered to be specific for 3-iodo-L-tyrosine¹⁰ but was also compatible with 3nY¹⁶, and the corresponding *M. jannaschii* tyrosyl-tRNA in which the anticodon was complementary to the UAG amber stop codon. This OTS enabled 'addiction' via a β -lactamase variant ($bla_{TEM-1,B9}$) that had been previously selected to be dependent upon 3nY incorporation at amino acid position 162¹⁶. However, since $bla_{TEM-1,B9}$ with 3nY already conferred resistance to high levels of ampicillin we further engineered $bla_{TEM-1,B9}$ to use ceftazidime (CAZ) as a substrate^{17,18}. The new β -lactamase, $bla_{Addicted}$, conferred moderate resistance to CAZ in a 3nY-dependent manner at concentrations commonly used in bacterial cultures, with a measured minimal inhibitory concentration (MIC) of ceftazidime of 3–10 $\mu\text{g mL}^{-1}$ (see Fig. 1). This lower MIC allowed us to both retain and challenge the 3nY incorporation by progressively increasing CAZ concentrations during culture.

The OTS and $bla_{Addicted}$ were assembled together on a plasmid (pADDICTED). We also constructed a control plasmid (pCONTROL) by replacing the 3nY codon (UAG) at position 162 in $bla_{Addicted}$ with a phenylalanine codon (UUU), generating $bla_{Control}$. Phenylalanine is the only canonical amino acid that produces a functional β -lactamase when replacing 3nY162 in $bla_{TEM-1,B9}$ ¹⁶. As expected, pCONTROL conferred CAZ resistance in a 3nY-independent manner (Fig. 1).

As a chassis for evolution, we chose to use *E. coli* strain MG1655 because it is well-characterized, with a sequenced and annotated genome¹⁹. MG1655 is autotrophic for all 20 canonical amino acids allowing for robust growth in amino acid knockout media. MG1655 was transformed with pADDICTED or pCONTROL, and lines were passaged in three different mixtures of amino acids in MOPS-EZ Rich Defined Medium (RDM). The first mixture contained all 20 standard amino acids (RDM-20), the second mixture lacked tyrosine (RDM-19), and the third mixture lacked seven amino acids; serine, leucine, tryptophan, glutamine, tyrosine, lysine, and glutamate (RDM-13) (Fig. 2). These seven amino acids represent all amino acids encoded by codons accessible through single nucleotide mutations from the UAG stop codon; by limiting the charging of the tRNAs for these amino acids, it should prove more difficult for any single mutation in a codon to be readily suppressed by mutations to tRNA anticodons or by mis-pairing. The RDM-13 media condition also proved a more stringent challenge to growth and evolution. Each media condition was supplemented with 10 mM 3nY, matching the concentration of L-serine, the most abundant amino acid in RDM.

Experimental evolution of bacteria in the presence of 3nY. We initiated selections with six independent clones, three containing pADDICTED, and three containing pCONTROL. These lines were denoted as Addicted(i), Addicted(ii), Addicted(iii), and Control(i), Control(ii), Control(iii) (Fig. 2). Each clone was passaged in the three different amino acid environments described above, and evolved lines are identified by the evolutionary media condition and the progenitor clone from which it was initiated (e.g. Addicted-20(i), Addicted-19(i)...Control-13(iii)). Growth cultures were passaged daily by inoculating 5 mL of RDM with 1 μL of overnight growth. This resulted in approximately 12.5 generations per daily passage. While passaging, CAZ concentration was increased at a rate of 1 $\mu\text{g mL}^{-1}$ per 100 generations to a final concentration of 22 $\mu\text{g mL}^{-1}$ to provide evolutionary pressure and ensure enforcement of 3nY dependence. While wild-type MG1655 was capable of growth in all media conditions, the progenitor clones transformed with pADDICTED or pCONTROL proved largely incapable of growth in RDM-13 when supplemented with 3nY, even in the absence of CAZ (Fig. 3, Supplementary Figure 1), so for the first 125 generations (10 passages) in RDM-13 the media was supplemented with 25% RDM-19. After these initial 10 passages, all lines were capable of growth in RDM-13.

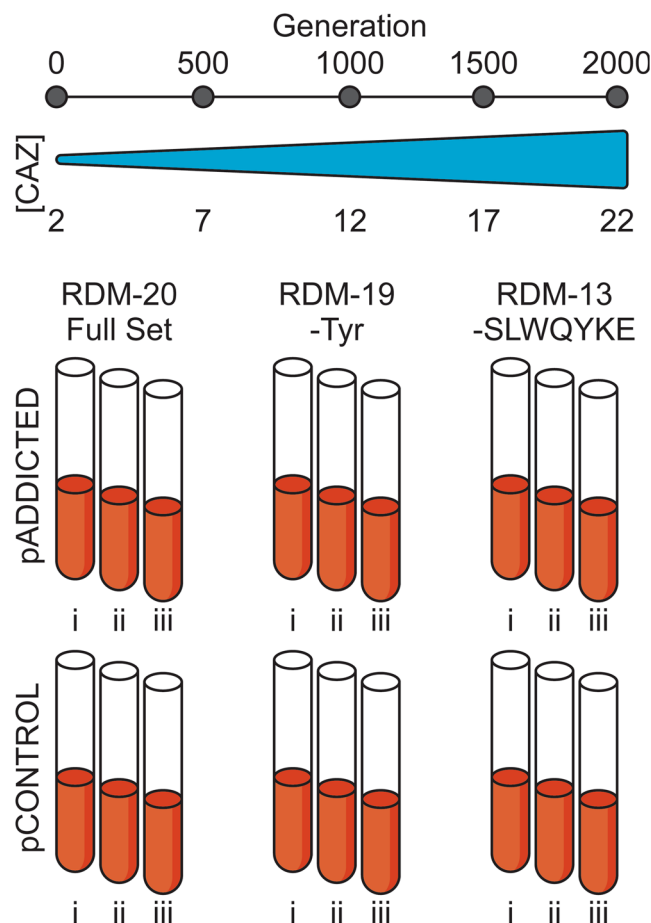


Figure 2. Evolution set up. Wild-type *E. coli* strain MG1655 containing either pADDICTED or pCONTROL was evolved in biological triplicates (i, ii, and iii) for the first 2000 generations in one of three rich defined media conditions each supplemented with 10 mM 3nY. The first (RDM-20) contained all 20 canonical amino acids, the second (RDM-19) lacked tyrosine, and the third (RDM-13) lacked seven amino acids (serine, leucine, tryptophan, glutamine, tyrosine, lysine, and glutamate). During evolution ceftazidime concentration was increased to provide a fitness burden and enforce OTS activity.

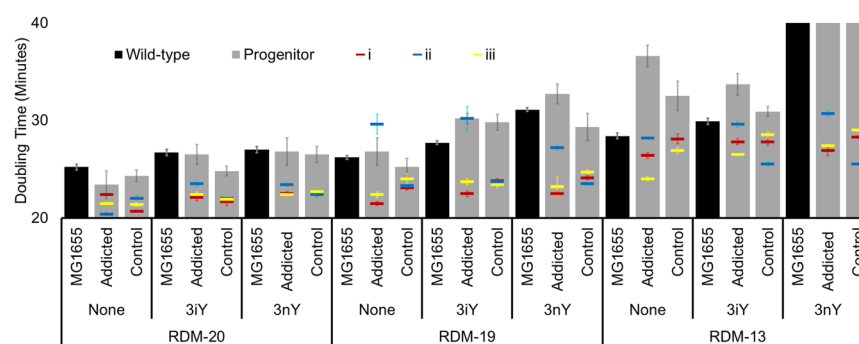


Figure 3. Doubling time (in minutes) of wild-type *E. coli* strain MG1655 (black bars), and MG1655 containing the plasmid pADDICTED or pCONTROL (gray bars). Doubling times were measured in all three media conditions (RDM-20, 19, and 13) and with no ncAA (none), with 10 mM 3-iodo-tyrosine (3iY), or with 10 mM 3-nitro-tyrosine (3nY) before evolution (gray bars). Doubling times of each evolved lineage (i, ii, and iii for each condition) were measure after 2000 generations of evolution (red, blue, yellow, respectively) in the media in which they were evolved, without ncAA, with 10 mM 3iY, or with 10 mM 3nY. Reported values are based on a minimum of two growth curves from biological replicates, error bars represent s.e.m.

At the conclusion of 160 passages, corresponding to approximately 2000 generations, we selected a single clone from each evolved line and sequenced the bacterial genomes of the single clone as well as the entire bacterial culture using the HiSeq. 4000 Illumina platform. Selected cells were, in general, genotypically representative of the average bacterial population. We further characterized phenotypic parameters of the selected evolved cells, as well as the progenitor cells.

Growth Rates. We used growth rates, evaluated as doubling times, as a general indicator of cellular fitness both before and after evolution. Doubling times were determined for the progenitor cells in all three media conditions, and the doubling time of the evolved lines was calculated in the media in which they were evolved. Doubling time calculations were automated using the software program *GrowthRates*²⁰. In general, doubling times were measured in the absence of ceftazidime, and in three different amino acid environments: without ncAA, with 3nY, or with 3iY. Additionally, doubling times were measured with 2 or 22 $\mu\text{g mL}^{-1}$ CAZ (progenitor and evolved cells, respectively), both with and without 3nY.

We first wanted to evaluate the effect amino acid supplements had on *E. coli* in the absence of 3nY. We determined the doubling time of wild-type *E. coli* strain MG1655 in RDM-20, which compared well to previously reported doubling times of MG1655 in rich media²¹. We found statistically significant increases in the doubling time of MG1655 with the removal of amino acids (compare the doubling times of wild-type MG1655 in RDM-20, RDM-19, and RDM-13, in Fig. 3, p-value calculations in Supplementary Table 1). We then evaluated the effect the OTS had on fitness in the absence of ncAAs and found that doubling times were not significantly affected by the addition of the OTS (compare doubling times of wild-type MG1655 to MG1655 containing pADDICTED or pCONTROL, within RDM-20 or RDM-19). Somewhat surprisingly, the addition of the OTS in RDM-13 did show a statistically significant increase in doubling time when encoded on pADDICTED, but not pCONTROL. This increase in doubling time may be a fitness effect due to *bla* truncation at the UAG codon of *bla*_{Addicted}, as doubling times of MG1655 with pADDICTED are generally higher than MG1655 with pCONTROL in all conditions, despite identical plasmid sequence except for the single amber codon in the *bla* gene (Fig. 3, compare pADDICTED and pCONTROL before evolution in each RDM/ncAA combination). Together, these results suggest that the amino acid environment directly affects growth rates, and that encoding of the OTS on a plasmid on its own does not have a profound effect on fitness.

We next wanted to evaluate the effect ncAAs had on *E. coli* fitness. We evaluated wild-type MG1655 in RDM-20, -19, and -13 with 10 mM 3iY or 10 mM 3nY and found significant increases in doubling times in all three conditions when compared to media without ncAAs (Fig. 3). The addition of an OTS did not significantly increase doubling times of MG1655 in ncAA enriched RDM-20 or RDM-19 when compared to the doubling time of wild-type MG1655 in the same media. However, in RDM-13 the addition of 3nY significantly increased the doubling time of wild-type MG1655, and the addition of the OTS further exacerbated this fitness challenge, preventing growth entirely. Surprisingly, the addition of 3iY to RDM-13 did not have a significant effect, with doubling times similar to growth in RDM-13 without ncAA. These results suggest 3nY is mildly toxic and 3iY less so, but the fitness burden is significantly enhanced in minimal amino acid environments (RDM-13), particularly when the OTS is encoded and suppression is possible. These latter results are consistent with our findings in the absence of ncAAs.

Over the course of evolution, the doubling times of all lines decreased in all conditions (Fig. 3), indicating that minimal media itself is a fitness burden for the strains, including the parental strain, and thus that many of the genomic mutations we observed (below) were generally improving the fitness of the strains for supplemented minimal media. The relative fitness effects associated with media conditions remained consistent even after evolution; doubling times were fastest in RDM-20, and slowed with greater metabolic stresses with decreasing numbers of amino acid supplements. Surprisingly, the large growth deficits in RDM-13 were largely overcome during evolution, with evolved cells reaching doubling times near that of wild-type MG1655 grown in richer conditions (i.e., compare with RDM-20 without ncAA to doubling time of evolved lines in RDM-13 with 3nY in Fig. 3).

There were two exceptions to these general observations: Control-13(ii), which showed no growth in media lacking ncAA at the conclusion of the experiment (Fig. 3, Supplementary Figure 1), and Addicted-19(ii), which showed reduced growth (increased doubling time) in the absence of 3nY, even when supplemented with the nominally less toxic 3iY.

Evolution of antibiotic resistance. During the course of passaging, CAZ concentrations were increased to levels beyond the MIC of the progenitor cells (the initial MIC of lines was approximately 2–6 $\mu\text{g mL}^{-1}$, while the final CAZ concentration challenge was 22 $\mu\text{g mL}^{-1}$). Bacterial survival at increasingly higher CAZ concentrations indicated that CAZ resistance was evolving in the lines. We measured the MIC of each evolved clone with and without 3nY (Fig. 1). The addicted lines remained dependent on 3nY for improved ceftazidime resistance, while the MICs of control lines increased in a 3nY-independent manner.

All nine of the addicted cell lines acquired at least a single mutation in *bla*_{Addicted} during evolution (Table 1). Several of these mutations are known or expected to be stabilizing mutations, while others are known to expand substrate specificity of *bla*_{TEM-1}^{22–27}. Other mutations are specific to *bla*_{TEM-1.B9}, which originally included a number of substitutions relative to the wild-type *bla*_{TEM-1} that addicted it to 3nY. The substitution T139I in Addicted-20(iii) and Control-19(ii), is more similar to the original wild-type residue, leucine, but does not reduce the 3nY dependence of the enzyme. In contrast, only five control lines acquired mutations in *bla*_{Control}: Control-20(i), -20(ii), -20(iii), -19(i), and -19(iii). The higher rate of mutation for *bla*_{Addicted} relative to *bla*_{Control} ($p = 0.0054$) indicates that the enzyme dependent upon the ncAA was not initially as fit as its non-addicted counterpart, especially in the presence of increasing antibiotic concentrations.

Mutation	Lines	Effect	Previously Identified
V33I	Addicted-13(ii)	—	22,58
Q39K	Control-20(ii)	Expanded Substrate Activity	27
G92D	Addicted-20(i), -20(ii), -19(ii)	Stabilizing	22,59
T139I	Addicted-20(iii), Control-19(ii)	<i>bla</i> _{TEM-1.B9} Specific	—
T140K	Addicted-19(ii), -19(iii), -13(iii)	—	—
M152I	Control-20(i)	<i>bla</i> _{TEM-1.B9} Specific	—
H153R	Addicted-20(ii)	Stabilizing	60,61
M155I	Control-20(iii)	—	22,62,63
M182T	Addicted-19(i)	Stabilizing	25–27
A184V	Addicted-13(i)	Expanded Substrate Activity	64,65
T200P	Addicted-20(i)	—	—
A224V	Control-19(i)	Stabilizing	25

Table 1. Mutations found in *bla*_{Addicted} and *bla*_{Control} in evolved lines.

The phenotypic impact of the *bla* mutations was further examined by purifying the plasmids pADDICTED and pCONTROL from the evolved lines and transforming them into wild-type MG1655. These evolved plasmids generally conferred higher MICs than the progenitor plasmids (Fig. 1). Notably, pADDICTED from lines Addicted-20(ii), -19(i), -19(ii), -19(iii), and -13(i) yielded MIC increases greater than 2-fold the original MIC. These plasmids all contained *bla* mutations known or suggested to have stabilizing effects (Table 1). However, *bla* mutations cannot fully explain MIC changes, as the original lines are generally more resistant to CAZ than are wild-type MG1655 cells with the evolved plasmids transformed (compare upper panels of Fig. 1 with lower panels). Moreover, *bla*_{Addicted} in lines Addicted-19(iii) and Addicted-13(iii) each have only a single, identical mutation (T138K) yet confer different MICs when transformed into wild-type *E. coli* strain MG1655.

Beyond mutations in the *bla* gene itself, increases in MIC may have also been the result of other plasmid or genomic mutations that altered antibiotic resistance through other mechanisms. For example, the plasmid from Addicted-13(iii) has a mutation in *repC*, which can affect copy number and in turn antibiotic resistance^{28,29}. In addition, genomic mutations occurred in genes that are known to have a role in antibiotic tolerance. Four lines (Addicted-20(i), -20(iii), -19(iii), and Control-20(iii)) had mutations in *envZ* (Table 2), a histidine kinase that regulates *ompF* and *ompC* expression, which in turn alter membrane porosity and have been tied to β -lactam resistance^{30,31}. In the same vein, Addicted-13(iii) has a mutation directly in *ompF*. Mutations to *cyoA* or *crp*, two related proteins directly or indirectly involved in transcriptional regulation, occurred in eight evolved lines (Addicted-20(ii), -19(ii), -19(iii), -13(i), -13(iii) and Control-19(i), -19(iii), -13(iii)), and inactivation of these genes has been shown to produce resistance to β -lactams^{32,33}. The *opgG* gene was mutated in Addicted-20(iii), -19(i), and Control-20(ii), and is involved in conferring resistance to antibiotics³⁴. Finally, lipopolysaccharide (LPS) expression has been tied ceftazidime treatments, especially at or near MIC levels, and several lines had mutations in LPS biosynthesis and maintenance genes. Notably, seven lines had IS insertions in the *waa* operon that encodes the core oligosaccharide of LPS, including *waaO* (Addicted-20(ii), -19(i), and Control-20(ii), -20(iii)), *waaQ* (Control-20(i)), *waaP* (Addicted-13(i)), and *waaB* (Control-19(ii)). Also, three lines had mutations in the LPS-related gene *galU*, (Control-19(i), -13(i), -13(iii)), and one line had a mutation in *lptD*, which is involved in the assembly of LPS at the surface of the cell (Control-13(ii))^{35,36}.

Retention of the OTS. One of the primary questions we were addressing was whether lines that were addicted to the OTS would retain and utilize the nCAA differently or better than lines that were not addicted to the OTS. The 3nY dependent MICs (Fig. 1) strongly suggests that a functional OTS has been retained in all addicted lines. To directly evaluate the functionality of the OTS we measured UAG suppression efficiency with a GFP reporter system that contained a tyrosine (TAT), amber (TAG), or ochre (TAA) codon at position Y39 of GFPmut2³⁷. We tested each GFP reporter in progenitor cells as well as each of the evolved lines (Supplementary Figure 2). All of the addicted lines remained capable of efficient amber suppression (>10% 3nY incorporation at UAG codons) (Fig. 4). In contrast, three of the nine control cell lines (Control-20(i), -19(ii), -19(iii)) had lost UAG suppression capability (<1% 3nY incorporation) (Fig. 4, see ‘*’). Surprisingly, UAG suppression efficiency generally dropped in both addicted and control lines relative to their progenitor cells.

Of the addicted lines, only Addicted-19(i) acquired mutations in the OTS, acquiring a T->G mutation in the tRNA processing machinery at position T(-28). In contrast, OTS mutations occurred in Control-20(i), -19(i), -19(ii), -19(iii), three of which had lost OTS activity. Of these, lines Control-20(i) and Control-19(ii) have aaRS mutations: Control-20(i) has a frameshift in the aaRS resulting in a UGA ochre stop codon at amino acid position 123, which leads to the complete loss of UAG suppression, and Control-19(ii) contains an IS-mediated insertion in the aaRS, a mechanism that has previously been observed to inactivate OTS machinery¹¹. Control-19(i) and Control-19(iii) each had mutations in the tRNA. The tRNA of the OTS from Control-19(i) contained a G26C substitution in the hinge between D-stem and anticodon stem (Fig. 5, yellow), while OTS tRNA from Control-19(iii) has a G42A substitution in the anticodon stem of the tRNA (Fig. 5, red). The latter mutation is predicted

Addicted				Control			
Line	Gene	Mutation	Position	Line	Gene	Mutation	Position
20(i)	sdaC	+9 insert (IS1)	925/1290	20(i)	sdaC	+9 insert (IS1)	1192–1200/1290
	yqiG	A > G	815/2439		waaQ	+9 insert (IS1)	581–589/1035
	envZ	M58R			ecpR/ykgL	C > T	311481
20(ii)	ydeT	A > C	1108/1227	20(ii)	opgG	+9 (IS1)	738–746/1656
	waaO	+9 insert (IS1)	531–539/1020		sdaC	W18tag	
	mtdL	A324T		waaO	+4 (IS5)	925–928/1020	
	cyaA	–78	2420–2497/2547	sdaC	+1	798/1290	
20(iii)	dnaA/rpmH	–6	3883825	20(iii)	envZ	S87Y	
	opgG	Q314tag			waaO	+5 (IS2)	100–104/1020
	clsA	+9 (IS1)	1307–1315		yidD	A14G	
	sdaC	A383V		19(i)	ybbP	A580E	
	yghG	G63V			galU	H71L	
	envZ	F107C			uspC-tyrP	–11258 (IS1)	
19(i)	ykgM/ykgR	+9 (IS1)	313109	garD	K343Q		
	opgG	+9 insert (IS1)	385–393/1536	tusD	G43G		
	ycjO	–87	250–336/882	cyaA	D231G		
	waaO	+4 insert (IS5)	533–536/1020	19(ii)	opgH	H417D	
ycaA	–1	2033/2547	uspC-tyrP		–11067		
yahF	K208R		waaB		+4 (IS5)	226–229/1080	
dosP	I282I		prpB/prpC		A > G	349642	
uspC-yecH	–10333 (IS1)		lacZ/lacI		2 bp > TT	366351	
lhgO	R346W		lacZ/lacI		C > T	366409	
yqeK	L30(tga)		lacI/mhpR		G > A	367573	
crp	F137L		ompF/asnS		G > A	987138	
19(iii)	tauA/tauB	G > T	386203	yhfA/crp	A > C	3485942	
	hofC	S27R		19(iii)	caiC	I449S	
	uspC-yecA	–12537 (IS1)			quuD	+5 (IS2)	233–237/384
	ppx	–1	365/1542		tyrP	W357tag	
	rrlD	C > T	2209/2904		cyaA	D184N	
	13(i)	envZ	G85V		yihY	R208L	
		dnaA	–54	251–304/1404	ftsW	L141I	
		cyaA	R160L		galU	+12 (IS4)	695–706/909
rrlD		C > T	3424575	uspC-yecA	–11714 (IS1)		
13(ii)	yihU/yihV	G > T	4073632	tyrA	M117I		
	flgJ	+9 insert (IS1)	157–165/942	kduI	M193I		
	<u>uspC-tyrP</u>	–10976		<u>mtr</u>	–6	712–717/1245	
	rpoD	P97L		malA/yfdC	G > T	2465297	
	mtr	–6	712–717/1245	yhfA/crp	+5 bp (IS2)	3486024	
	waaP	+4	413–416/798	13(ii)	lptD	Q557tag	
cyaA	D231V		uspC-tyrP		–11031 (IS1)		
pykF	I264T		yhcG		S136A		
uspC-tyrP	–11298 (IS1)		tnaA		+5 (IS2)	303–304/1416	
sanA	+9 (IS1)	165–173/720	ompF/asnS		+9 (IS1)	987064	
<u>mtr</u>	–6	391–396/1245	topA/cysB		+4 bp (IS4)	1333763	
<u>ompR</u>	M57L		mtr/deaD		T > C	3305872	
13(iii)	yibA	+9 (IS1)	15–23/843	galU	+9 (IS1)	81–89/909	
	waaQ/waaA	C-T	3808189	uspC-tyrP	–11280		
	yaiT	A > C	682/1458	mtr	L400R		
	ompF	+9 (IS1)	323–331/1089	mtr	–1	1011/1245	
	cysB	–1	610/975	cyaA	C120Y		
	uspC-tyrP	–11018 (IS1)		rpsF	D13G		
	<u>mtr</u>	–6	712/717/1245	lrhA/alaA	A > C	2407094	
	glpD	+9 (IS1)	1109–1117/1506				
cyaA	P301L						
yiiQ	K74taa						

Table 2. Genomic Mutations of Evolved Lines. Genomic mutations which occurred after 2000 generations of evolution. Mutations which appeared after 125 generations in RDM-13 are underlined. The four in-frame amber codons which arose in ORFs are bold.

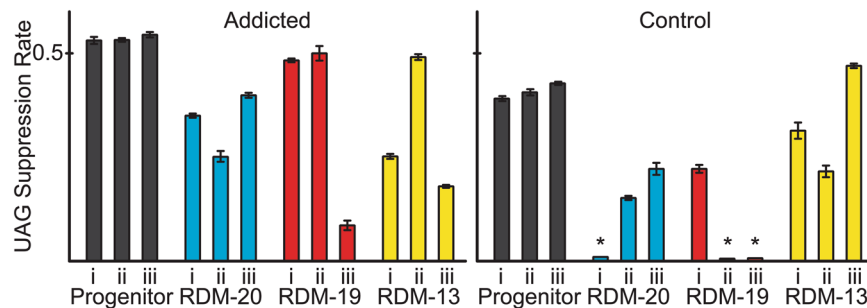


Figure 4. UAG Suppression Efficiency. OTS activity of progenitor cells (black) and evolved lines (colored bars) was determined by measuring UAG suppression efficiency with a GFP reporter. All addicted lines maintained an active OTS (left), while three control lines lost OTS activity over 2000 generations (see '*'). Values are the average of four biological replicates, error bars represent s.e.m.

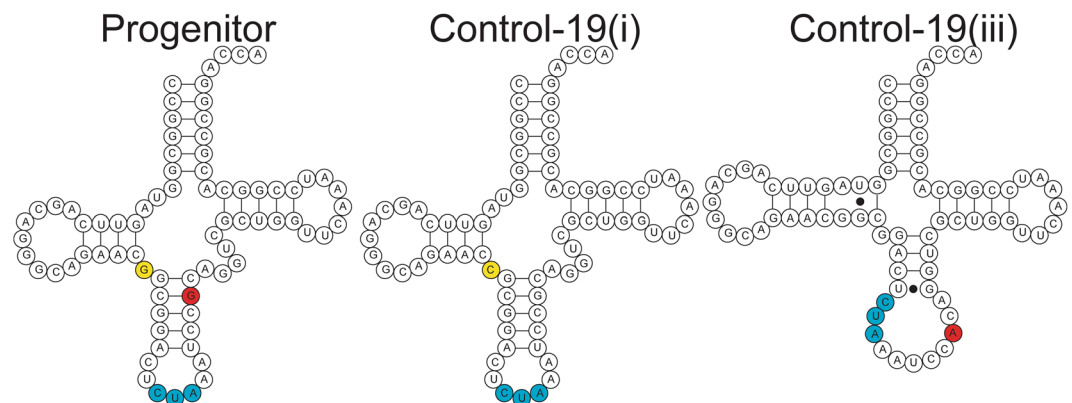


Figure 5. Predicted tRNA structures. Mutations to the OTS-tRNA occurred in two lines during evolution. Computationally predicted tRNA structures from these evolved lines shows substitution G26C (yellow) found in Control-19(i) (center) is not likely to affect general structure or the anticodon (blue), while substitution G42A (red) in line Control-19(iii) is predicted to disrupt tRNA structure (right).

to compromise tRNA structure, as determined using mfold³⁸ and tRNA prediction software³⁹ (Fig. 5). This tRNA mutation likely explains the complete loss of UAG suppression in Control-19(iii) (Fig. 4).

Overall, either suppression in general or the specific incorporation of 3nY puts cells at a fitness disadvantage, and in the absence of addiction to the nCAA this trait can be readily lost. The evidence suggests that suppression in general is less toxic than 3nY incorporation. First, in the most restricting environment (RDM-13) the parental strain can take up and incorporate 3iY into proteins, which affects doubling times less than the incorporation of 3nY (Fig. 3, compare doubling times in RDM-13 of wild-type *E. coli* strain MG1655 in the presence of 3iY with no plasmid, and with pADDICTED or pCONTROL, each carrying an active OTS for 3iY incorporation). Second, in the absence of addiction the OTS is readily lost or compromised during evolution (Fig. 4, compare suppression efficiencies in 'Addicted' and 'Control' panels). Additionally, previous work also supports the hypothesis that suppression is not particularly toxic, given the existence and persistence of natural amber suppressors which have minimal effect on fitness or the transcriptome⁴⁰.

Genomic adaptations. Full genome sequencing and analysis revealed several trends among evolution conditions and the genes affected during evolution. (Table 2). Strikingly, the most commonly mutated or deleted ORFs were amino acid transporters in the hydroxyl and aromatic amino acid permease (HAAAP) family. The most commonly affected HAAAP protein was *tyrP*, a tyrosine-specific permease^{41,42} that was inactivated or deleted in 10 of the evolved lines, including all six evolved in RDM-13, and four of the six lines evolved in RDM-19, including all three control lines, as well as Addicted-19(iii). The most common mechanism (9 of 10 cultures) used to inactivate *tyrP* was an IS1 mediated deletion of a large genomic fragment that excised twelve to fourteen genes, including a portion or the entirety of *tyrP*. The second most commonly mutated HAAAP protein was *mtr*, a tryptophan-specific permease⁴³. The *mtr* mutations were found exclusively in lines evolved in RDM-13, with five of the six lines evolved in RDM-13 having a mutated *mtr* gene, and the final line, Control-13(iii), having an intergenic mutation near the *mtr* ORF. The third most common HAAAP protein that was inactivated was *sdaC*, a serine specific transporter, which was inactivated in four of the six lines evolved in RDM-20; a mutation of unknown consequence appeared in a fifth line (A384V). Surprisingly, *sdaC* was unaffected in all RDM-19 and RDM-13 evolved lines.

We hypothesized that deactivation of HAAAPs may have played a role in reducing the fitness burden of 3nY in the media. Evidence for the importance of these mutations for adaptation to growth comes from an examination of the mutations that arose in the RDM-13 lines between generations 0 and 125, which were initially incapable of growth in RDM-13 and had to be supplemented over their first 125 generations (10 passages) with 25% RDM-19 (Fig. 3). We sequenced the genomes of the six RDM-13 lines after 125 generations of growth in the enriched RDM-13 to determine which mutations had occurred early in experimental evolution, and found that five of the six populations had mutations in *mtr* or the *mtr* promoter. These mutations were fixed in three of the five lines, and for the remaining two lines some 22.8% of the Control-13(i) population had a *mtr* mutation, while 44.2% of the Addicted-19(iii) had mutations in *mtr*; both mutations were fixed by generation 2000. Additionally, the *tyrP* region was excised from the genomes of Addicted-13(i) and Control-13(ii). These mutations were the only identifiable mutations in the population after 125 generations, and were likely responsible for the ability to grow in pure RDM-13 after this point. This data supports the hypothesis that aromatic amino acid transporter deactivation (specifically *mtr* and *tyrP*) reduced the fitness burden imposed by the noncanonical aromatic amino acid, 3nY.

Sampling of new codes. All eighteen lines were initially capable of utilizing 3nY in their proteome, and after 2000 generations of evolution four clones had each acquired a single in-frame amber codon in protein coding sequences. In three of these four instances, the strains did not require 3nY for growth. The clone from Addicted-20(iii) had acquired a SNP resulting in Q314tag in *opgG*, an osmoregulated peptidoglycan biosynthesis protein^{44,45}. The other three in-frame amber codons appeared in clones from the Control lines, including Control-20(ii) and Control-19(iii), which acquired a W18tag mutation in *sdaC* and a W357tag mutation in *tyrP*, respectively. These amber substitutions likely down-regulated the expression of these HAAAP genes, consistent with our findings that these loci were frequently deleted or otherwise compromised. Control-19(iii) had deactivated the OTS entirely, indicating W357tag of *tyrP* was a truncation, while in Control-20(ii) the OTS remained partially active, with only ~15% 3nY incorporation (see Fig. 4).

As mentioned above, two lines showed an unexpected dependence on 3nY for growth. When we assayed the ceftazidime MIC for the clone chosen from the Control-13(ii) line there was minimal to no growth on plates without 3nY, but appropriate growth on 3nY supplemented plates. Later, multiple replicates consistently resulted in a no growth phenotype for line Control-13(ii) in the absence of 3nY during amber suppression assays. Sequencing of this genome revealed an in-frame UAG codon in *lptD*, a Q557tag substitution. *lptD* is an essential gene involved in LPS biosynthesis^{44,46}. Comparison to the full culture sequencing revealed this mutation was not representative of the bacterial population in the evolving culture Control-13(ii). Despite being a minor component of the full population, this mutation demonstrates the additional mutational space available to the evolving bacteria. An amber stop codon in *lptD* in a wild-type *E. coli* would result in a nonviable phenotype, yet with the expanded genetic code this clone has survived.

While the clone from the Control-13(ii) line was found to have an amber stop codon in an essential gene that may explain the lack of growth in the absence of 3nY, we cannot point to a similar rationale for why Addicted-19(ii) shows reduced growth rates in the absence of 3nY. Genome sequencing of the Addicted-19(ii) clone revealed six SNPs had arisen during evolution, four of which caused single amino acid substitutions, a fifth that resulted in a silent mutation, and a sixth that occurred in an intergenic region. As with other strains that had adapted to 3nY toxicity, Addicted-19(ii) contained a 10 kb genomic deletion surrounding the tyrosine permease *tyrP*.

Discussion

Previously, two different approaches have been used to generate organisms with altered or expanded amino acid genetic codes; a bottom-up approach where components of the organism were engineered to function with a ncAA, or a top-down approach, where an organism was allowed to evolve in the presence of a ncAA⁴⁷. Advances in genome editing and protein structural modeling have made the bottom-up approach feasible, with two recent reports of *E. coli* successfully engineered to depend on a ncAA for survival^{48,49}. It should now be possible to make fully synthetic genomes⁵⁰ that are designed to have a chemical dependence on an ncAA throughout the genome using the 'amberless' *E. coli*⁵¹. Alternatively, top-down approaches have previously been used to generate organisms that preferentially function with a ncAA substituting for tryptophan^{5,52}. Using mutagens and selective growth conditions, *Bacillus subtilis* became dependent on a normally toxic tryptophan analog, 4-fluorotryptophan, with only 106 genomic mutations required to change amino acid preference⁵². In contrast, an *E. coli* tryptophan auxotroph evolved in the presence of 4-fluorotryptophan in place of tryptophan⁵³ could tolerate high levels of the ncAA, but in the end still required tryptophan for growth. A similar approach using long-term evolution in *E. coli* allowed for cells to grow using a sulfur-containing tryptophan analog in place of tryptophan⁶. Another recent report shows that the bacteriophage T7 will adopt ncAAs into its genetic code to reach new fitness peaks⁷.

The work described herein can be considered a hybrid approach, resting between bottom-up and top-down. We used a single engineered addiction element, *bla*_{Addicted}, to preserve an active OTS over evolutionary time-frames within an otherwise unmodified organism. Thus during 2000 generations of evolution bacteria were required to retain 3nY incorporation as part of its genetic code and could explore the potential utility of 3nY throughout their proteomes. This process simulates the 'ambiguous intermediate codon' theory of evolution. Over the described 2000 generations, 3nY-addicted bacteria have remained dependent on 3nY for survival in evolutionary conditions, and cells have evolved to overcome the fitness burdens initially seen from the expanded genetic code. This was not a foregone conclusion: in order to re-establish the canonical 20 amino acid code, lines could have evolved *bla*_{Addicted} to no longer require 3nY, or acquired genomic mutations leading the CAZ resistance without *bla*_{Addicted}. Even though 326 genomic amber codons were a single mutational step from stop codons that would not be suppressed by the OTS, in over 2000 generations addicted populations fixed no mutations that led

away from amber stop codons used for translational termination, despite previous evidence that recoding amber codons to alternative stop codons reduces the fitness effects of obligate UAG suppression⁷.

Instead it became clear that the most important adaptations to the ambiguous genetic code we had imposed were ones that alleviated either the toxicity of an unnatural amino acid and/or that better optimized the composition of the 20 amino acids normally used for growth. Growth in 3nY clearly led to fitness deficits, and to compensate if the media lacked tyrosine (RDM-19 and RDM-13), the tyrosine permease *tyrP* was inactivated, and if the media also lacked tryptophan (RDM-13), the tryptophan permease *mtr* was inactivated. In contrast, if all 20 amino acids were present (RDM-20), the serine permease *sdaC* was inactivated. Since serine is the most abundant amino acid in RDM conditions (10 mM), it is possible that the amino acid pool is better balanced through deletion of *sdaC*, while still relying on other serine transporters (*sstT*) for serine uptake.

Once strains had evolved to the point where they could accommodate an ambiguous genetic code that could accept suppression with either tyrosine or 3nY, they were positioned for further evolution to specify a new code. After 2000 generations of evolution with 3nY, the existence of three in-frame amber codons in clones with active OTSs, and the fact that a clone from the Control-13(ii) line had adopted an in-frame UAG codon in the essential gene *lptD*, provides evidence that populations are exploring the 3nY mutational space. Additionally, the reduced doubling time of a second line, Addicted-19(ii), in the absence of 3nY may indicate some preference for the new media condition containing 3nY, even though no genomic in-frame amber codons had arisen. In this regard, the evolution of Addicted-19(ii) may resemble the evolution of a *B. subtilis* strain that could preferentially utilize 4-fluoro-tryptophan in place of tryptophan⁵.

Overall, our method provides one of the first experiments investigating how a new genetic code is adopted by an organism, and evolved lineages may represent evolutionary intermediates to the adoption of a new amino acid. All lineages overcame the fitness burden associated with 3nY toxicity, but nCAA addiction was required to enforce an active OTS. Further experimentation using the method described here, or similar approaches, will provide insight into the recoding of the genetic code during evolution, and may allow the evolution of biochemically unique organisms.

Methods

Evolutionary set up. *Addiction operon.* Plasmid backbone from pMMB67EH⁵⁴ was amplified using DNA oligos (Integrated DNA Technologies) DT01 and DT02 (Supplementary Table 2). The *M. jannaschii* OTS and *bla*_{TEM-1.b9} were amplified from a plasmid described previously¹⁶ using oligos DT03 and DT04. To convert the penicillinase *bla*_{TEM-1.B9} to the cephalosporinase *bla*_{Addicted} residues 165–167 were converted from WEP to YYG using oligo DT05 with either DT06 or DT07 for *bla*_{Addicted} or *bla*_{Control} respectively. Reaction mixtures were transformed into *E. coli* TOP10 and selected on LB-agar with 2 μg mL⁻¹ CAZ for *bla*_{Control} and the same conditions with 10 mM 3nY for *bla*_{Addicted}. Samples were sequence verified at University of Texas core facilities using Sanger sequencing. Properly sequenced plasmid of *bla*_{Control} and *bla*_{Addicted} were used as pCONTROL and pADDICTED respectively, and transformed into *E. coli* MG1655. Three colonies from each were selected as clones i, ii, and iii for passaging.

Media. MOPS-EZ Rich Defined Media (RDM, TEKnova) with the full complement of amino acids (RDM-20), as well as the knockout medias (RDM-19 and RDM-13), were prepared according the manufacturers specification. For media preparation, 3-nitro-L-tyrosine or 3-iodo-L-tyrosine (Sigma-Aldrich) was added to ultrapure deionized water to a final concentration of 17.24 mM. The 3nY or 3iY supplemented water was used to complete RDM, and the entire preparation was filter sterilized with Nalgene Rapid-Flow SFCA filtration units. Prepared media was stored at 4 °C, and moved to room temperature 16–24 hours before use.

Passaging. Selected colonies i, ii, and iii, from each transformation were picked and grown in RDM-20 supplemented with 10 mM 3nY and 2 μg mL⁻¹ CAZ for 16 hours. Each culture was then used to inoculate three subcultures of RDM-20, RDM-19, and RDM13. Initially, cultures were incapable of growth in RDM-13, but were capable of growth when 25% of media was replaced with RDM-19. This supplemented RDM-13 was used for the first 125 generations for RDM-13 evolved cultures, after which cultures were capable of growth in RDM-13. Cultures were passaged every 16–24 hours by transferring 1 μL of culture into 5 mL of fresh media, and grown shaking at 37 °C. A 500 μL sample from each line was preserved at generation 0, 125, 250, and every 250 generations for the duration of evolution, samples are stored in 25% glycerol at –80 °C.

Genome Sequencing and assembly. After the 2000 generations, 1 μL from each line was streaked onto RDM-agar supplemented with 10 mM 3nY. Two single colonies from each line were selected and grown in 3 mL RDM-20 with 10 mM 3nY with 22 μg mL⁻¹ CAZ. Simultaneously, samples from the progenitor cells were grown in similar conditions using 2 μg mL⁻¹ CAZ. After overnight growth, glycerol stocks were made of clonal cultures, and genomic DNA was isolated from one of the two cultures using bacteria genome miniprep kit (Sigma-Aldrich). Genomic DNA from mixed cultures were also prepped. Genomes were sequenced using the HiSeq. 4000 platform, 150 bp paired ends, achieving greater than 100× coverage across all samples. Raw reads were processed through trimming and adapter removal using trimmomatic (v0.32)⁵⁵. Alignment of sequencing reads and variant calling was performed through the breseq workflow (v0.27.2)⁵⁶.

Growth Curves. Glycerol stocks were used to inoculate 2 mL of appropriate RDM and incubated overnight while shaking at 37 °C. Then, 1 μL from overnight cultures were used to inoculate 100 μL of media with appropriate concentrations of CAZ and 3nY. A Tecan Infinite M200 pro microplate reader set to 37 °C monitored absorbance at 600 nm. Data was processed by subtracting background on per well basis and trimming data from early- to mid-log phase (0.05 to 0.2 OD) for optimal cellular concentration estimations⁵⁷. Doubling times were then calculated using GrowthRates program²⁰. All reported growth rates and based on duplicate growth curves (two or more), with multiple internal measurements taken from within each curve, as described²⁰, and reported with standard error.

MICs. Progenitor and evolved cells were grown overnight in 3nY enriched RDM with $2\ \mu\text{g mL}^{-1}$ or $22\ \mu\text{g mL}^{-1}$ of CAZ respectively. Overnight cultures were used to inoculate fresh RDM lacking 3nY and without CAZ, and were grown for five hours at $37\ ^\circ\text{C}$ while shaking. Aliquots of $25\ \mu\text{L}$ from each line (corresponding to $\sim 10^7$ cfu) were plated on LB-agar, with or without $1\ \text{mM}$ 3nY, in triplicate. Plates were allowed 5 minutes to dry in a sterile biosafety cabinet, and ceftazidime E-test MIC strips (Biomérieux) were added to each plate. Plates were incubated 16 hours at $37\ ^\circ\text{C}$. MIC values are reported as the lowest concentration at which no bacterial growth occurred, reported with standard error of the mean^{58–65}.

GFP Assays. The codon for Tyr39 (TAT) in GFPmut2, under control of the *tacI* promoter with *Lac* operator, was changed to the amber codon (TAG), or ochre nonsense codon (TAA) using Gibson cloning with oligo DT10 paired with DT08 (for TAG codon) or DT09 (for TAA codon). Gibson reaction was electroporated into TOP10 *E. coli*, and plated on LB-agar supplemented with $50\ \mu\text{g mL}^{-1}$ kanamycin (KAN). Sequences were verified using Sanger sequencing at the University of Texas Core Facility.

Glycerol stocks of the eighteen evolved clones and six progenitor clones were used to inoculate 2 mL of LB supplemented with $10\ \text{mM}$ 3nY and 2 or $22\ \mu\text{g mL}^{-1}$ ceftazidime (progenitor and evolved cells, respectively), and grown to saturation overnight. A $5\ \mu\text{L}$ aliquot from each was used to inoculate 5 mL of LB, $10\ \text{mM}$ 3nY, and $2\ \mu\text{g mL}^{-1}$ ceftazidime, and grown at $37\ ^\circ\text{C}$ for 3 hours, and then made electrocompetent with serial washes of 10% glycerol. Samples were transformed with GFP variants, recovered with LB media for one hour at $37\ ^\circ\text{C}$, then plated on LB-agar supplemented with $10\ \text{mM}$ 3nY, 0.02% glucose, $50\ \mu\text{g mL}^{-1}$ KAN, and 2 or $22\ \mu\text{g mL}^{-1}$ CAZ, and incubated overnight at $37\ ^\circ\text{C}$.

Four samples from each plate were picked and grown overnight at $37\ ^\circ\text{C}$ in $500\ \mu\text{L}$ LB with $10\ \text{mM}$ 3nY, 0.02% glucose, $50\ \mu\text{g mL}^{-1}$ KAN, and 2 or $22\ \mu\text{g mL}^{-1}$ CAZ. A $2\ \mu\text{L}$ aliquot of each overnight culture was used to inoculate 1 mL LB media supplemented with $50\ \mu\text{g mL}^{-1}$ KAN, and containing either $1\ \text{mM}$ 3nY, $1\ \text{mM}$ 3iY, or with no supplemented nCAA. Cultures were grown at $37\ ^\circ\text{C}$ for 3 hours, then induced with $1\ \text{mM}$ IPTG, and grown for another 5 hours at $37\ ^\circ\text{C}$. Final cultures were rinsed twice with iterative centrifugation at $2000\ \text{g}$ and PBS ($50\ \text{mM}$ phosphate, $300\ \text{mM}$ NaCl, pH 8) at $4\ ^\circ\text{C}$. Samples were resuspended in 1 mL PBS and a $150\ \mu\text{L}$ aliquot was used to measure relative fluorescence by measuring absorbance at $600\ \text{nm}$ (OD600), and fluorescence using a $475\ \text{nm}$ excitation and measuring emission at $525\ \text{nm}$ using Tecan Infinite M200 pro microplate reader.

Data availability. All data generated or analyzed during this study are available from the corresponding author on reasonable request. Sequence data is available as NCBI, BioProject ID PRJNA430697.

References

1. Knight, R. D., Freeland, S. J. & Landweber, L. F. Rewiring the keyboard: evolvability of the genetic code. *Nat. Rev. Genet.* **2**, 49–58 (2001).
2. Osawa, S., Jukes, T. H., Watanabe, K. & Muto, A. Recent evidence for evolution of the genetic code. *Microbiol. Rev.* **56**, 229–264 (1992).
3. Schultz, D. W. & Yarus, M. Transfer RNA Mutation and the Malleability of the Genetic Code. *J. Mol. Biol.* **235**, 1377–1380 (1994).
4. Andersson, S. G. E. & Kurland, C. G. Reductive evolution of resident genomes. *Trends Microbiol.* **6**, 263–268 (1998).
5. Wong, J. T. Membership mutation of the genetic code: loss of fitness by tryptophan. *Proc. Natl. Acad. Sci.* **80**, 6303–6306 (1983).
6. Hoesl, M. G. *et al.* Chemical Evolution of a Bacterial Proteome. *Angew. Chem. Int. Ed Engl.* **54**, 10030–10034 (2015).
7. Hammerling, M. J. *et al.* Bacteriophages use an expanded genetic code on evolutionary paths to higher fitness. *Nat. Chem. Biol.* **10**, 178–180 (2014).
8. Wang, L. & Schultz, P. G. A general approach for the generation of orthogonal tRNAs. *Chem. Biol.* **8**, 883–890 (2001).
9. Maranhao, A. C. & Ellington, A. D. Evolving Orthogonal Suppressor tRNAs To Incorporate Modified Amino Acids. *ACS Synth. Biol.* <https://doi.org/10.1021/acssynbio.6b00145> (2016).
10. Sakamoto, K. *et al.* Genetic Encoding of 3-Iodo-L-Tyrosine in Escherichia coli for Single-Wavelength Anomalous Dispersion Phasing in Protein Crystallography. *Structure* **17**, 335–344 (2009).
11. Wang, Q. *et al.* Response and Adaptation of Escherichia coli to Suppression of the Amber Stop Codon. *ChemBioChem* **15**, 1744–1749 (2014).
12. Mukai, T. *et al.* Codon reassignment in the Escherichia coli genetic code. *Nucleic Acids Res.* **38**, 8188–8195 (2010).
13. Bacher, J. M., Bull, J. J. & Ellington, A. D. Evolution of phage with chemically ambiguous proteomes. *BMC Evol. Biol.* **3**, 24 (2003).
14. Isaacs, F. J. *et al.* Precise Manipulation of Chromosomes *in Vivo* Enables Genome-Wide Codon Replacement. *Science* **333**, 348–353 (2011).
15. Hammerling, M. J. *et al.* Expanded genetic codes create new mutational routes to rifampicin resistance in Escherichia coli. *Mol. Biol. Evol.* [msw094](https://doi.org/10.1093/molbev/msw094), <https://doi.org/10.1093/molbev/msw094> (2016).
16. Tack, D. S. *et al.* Addicting diverse bacteria to a noncanonical amino acid. *Nat. Chem. Biol.* **12**, 138–140 (2016).
17. Stojanoski, V. *et al.* A Triple Mutant in the Ω -loop of TEM-1 β -Lactamase Changes the Substrate Profile via a Large Conformational Change and an Altered General Base for Catalysis. *J. Biol. Chem.* **290**, 10382–10394 (2015).
18. Palzkill, T., Le, Q.-Q., Venkatachalam, K. V., LaRocco, M. & Ocera, H. Evolution of antibiotic resistance: several different amino acid substitutions in an active site loop alter the substrate profile of β -lactamase. *Mol. Microbiol.* **12**, 217–229 (1994).
19. Blattner, F. R. *et al.* The Complete Genome Sequence of Escherichia coli K-12. *Science* **277**, 1453–1462 (1997).
20. Hall, B. G., Acar, H., Nandipati, A. & Barlow, M. Growth Rates Made Easy. *Mol. Biol. Evol.* **31**, 232–238 (2014).
21. Tao, H., Bausch, C., Richmond, C., Blattner, F. R. & Conway, T. Functional Genomics: Expression Analysis of Escherichia coli Growing on Minimal and Rich Media. *J. Bacteriol.* **181**, 6425–6440 (1999).
22. Salverda, M. L. M., De Visser, J. A. G. M. & Barlow, M. Natural evolution of TEM-1 β -lactamase: experimental reconstruction and clinical relevance. *FEMS Microbiol. Rev.* **34**, 1015–1036 (2010).
23. Perilli, M. *et al.* Novel TEM-type extended-spectrum beta-lactamase, TEM-134, in a Citrobacter koseri clinical isolate. *Antimicrob. Agents Chemother.* **49**, 1564–1566 (2005).
24. Bershtein, S., Goldin, K. & Tawfik, D. S. Intense Neutral Drifts Yield Robust and Evolvable Consensus Proteins. *J. Mol. Biol.* **379**, 1029–1044 (2008).
25. Kather, I., Jakob, R. P., Dobbek, H. & Schmid, F. X. Increased Folding Stability of TEM-1 β -Lactamase by *In Vitro* Selection. *J. Mol. Biol.* **383**, 238–251 (2008).
26. Jacquier, H. *et al.* Capturing the mutational landscape of the beta-lactamase TEM-1. *Proc. Natl. Acad. Sci.* **110**, 13067–13072 (2013).
27. Shcherbinin, D. S. *et al.* The study of the role of mutations M182T and Q39K in the TEM-72 β -lactamase structure by the molecular dynamics method. *Biochem. Mosc. Suppl. Ser. B Biomed. Chem.* **11**, 120–127 (2017).

28. San Millan, A. *et al.* Small-Plasmid-Mediated Antibiotic Resistance Is Enhanced by Increases in Plasmid Copy Number and Bacterial Fitness. *Antimicrob. Agents Chemother.* **59**, 3335–3341 (2015).
29. Haring, V. *et al.* Protein RepC is involved in copy number control of the broad host range plasmid RSF1010. *Proc. Natl. Acad. Sci. USA* **82**, 6090–6094 (1985).
30. Jaffé, A., Chabbert, Y. A. & Semonin, O. Role of porin proteins OmpF and OmpC in the permeation of beta-lactams. *Antimicrob. Agents Chemother.* **22**, 942–948 (1982).
31. Couce, A., Rodríguez-Rojas, A. & Blázquez, J. Bypass of genetic constraints during mutator evolution to antibiotic resistance. *Proc R Soc B* **282**, 20142698 (2015).
32. Jaffé, A., Chabbert, Y. A. & Derlot, E. Selection and characterization of beta-lactam-resistant *Escherichia coli* K-12 mutants. *Antimicrob. Agents Chemother.* **23**, 622–625 (1983).
33. Ruiz, C. & Levy, S. B. Use of functional interactions with MarA to discover chromosomal genes affecting antibiotic susceptibility in *Escherichia coli*. *Int. J. Antimicrob. Agents* **37**, 177–178 (2011).
34. Hanouille, X. *et al.* Structural Analysis of *Escherichia coli* OpgG, a Protein Required for the Biosynthesis of Osmoregulated Periplasmic Glucans. *J. Mol. Biol.* **342**, 195–205 (2004).
35. Leying, H. J., Büscher, K. H., Cullmann, W. & Then, R. L. Lipopolysaccharide alterations responsible for combined quinolone and beta-lactam resistance in *Pseudomonas aeruginosa*. *Chemotherapy* **38**, 82–91 (1992).
36. Pagani, L., Landini, P., Luzzaro, F., Debiaggi, M. & Romero, E. Emergence of cross-resistance to imipenem and other beta-lactam antibiotics in *Pseudomonas aeruginosa* during therapy. *Microbiologica* **13**, 43–53 (1990).
37. Cormack, B. P., Valdivia, R. H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33–38 (1996).
38. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
39. Schattner, P., Brooks, A. N. & Lowe, T. M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **33**, W686–W689 (2005).
40. Herring, C. D. & Blattner, F. R. Global Transcriptional Effects of a Suppressor tRNA and the Inactivation of the Regulator frmR. *J. Bacteriol.* **186**, 6714–6720 (2004).
41. Andrews, A. E., Lawley, B. & Pittard, A. J. Mutational analysis of repression and activation of the tyrP gene in *Escherichia coli*. *J. Bacteriol.* **173**, 5068–5078 (1991).
42. Wookey, P. J. & Pittard, A. J. DNA sequence of the gene (tyrP) encoding the tyrosine-specific transport system of *Escherichia coli*. *J. Bacteriol.* **170**, 4946–4949 (1988).
43. Brown, K. D. Formation of Aromatic Amino Acid Pools in *Escherichia coli* K-12. *J. Bacteriol.* **104**, 177–188 (1970).
44. 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).
45. Weissborn, A. C. & Kennedy, E. P. Biosynthesis of membrane-derived oligosaccharides. Novel glucosyltransferase system from *Escherichia coli* for the elongation of beta 1–2-linked polyglucose chains. *J. Biol. Chem.* **259**, 12644–12651 (1984).
46. Braun, M. & Silhavy, T. J. Imp/OstA is required for cell envelope biogenesis in *Escherichia coli*. *Mol. Microbiol.* **45**, 1289–1302 (2002).
47. Bacher, J. M., Hughes, R. A., Tze-Fei Wong, J. & Ellington, A. D. Evolving new genetic codes. *Trends Ecol. Evol.* **19**, 69–75 (2004).
48. Mandell, D. J. *et al.* Biocontainment of genetically modified organisms by synthetic protein design. *Nature* **518**, 55–60N (2015).
49. Rovner, A. J. *et al.* Recoded organisms engineered to depend on synthetic amino acids. *Nature* **518**, 89–93L (2015).
50. Gibson, D. G. *et al.* Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science* **329**, 52–56 (2010).
51. Lajoie, M. J. *et al.* Genomically Recoded Organisms Expand Biological Functions. *Science* **342**, 357–360 (2013).
52. Yu, A. C.-S. *et al.* Mutations Enabling Displacement of Tryptophan by 4-Fluorotryptophan as a Canonical Amino Acid of the Genetic Code. *Genome Biol. Evol.* **6**, 629–641 (2014).
53. Bacher, J. M. & Ellington, A. D. Selection and Characterization of *Escherichia coli* Variants Capable of Growth on an Otherwise Toxic Tryptophan Analogue. *J. Bacteriol.* **183**, 5414–5425 (2001).
54. Fürste, J. P. *et al.* Molecular cloning of the plasmid RP4 primase region in a multi-host-range tacP expression vector. *Gene* **48**, 119–131 (1986).
55. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* btu170, <https://doi.org/10.1093/bioinformatics/btu170> (2014).
56. Deatherage, D. & Barrick, J. Identification of Mutations in Laboratory-Evolved Microbes from Next-Generation Sequencing Data Using bresseq. in *Engineering and Analyzing Multicellular Systems* (eds Sun, L. & Shou, W.) 165–188 (Springer New York, 2014).
57. Stevenson, K., McVey, A. F., Clark, I. B. N., Swain, P. S. & Pilizota, T. General calibration of microbial growth in microplate readers. *Sci. Rep.* **6**, srep38828 (2016).
58. Salverda, M. L. M. *et al.* Initial Mutations Direct Alternative Pathways of Protein Evolution. *PLOS Genet.* **7**, e1001321 (2011).
59. Figliuzzi, M., Jacquier, H., Schug, A., Tenaillon, O. & Weigt, M. Coevolutionary Landscape Inference and the Context-Dependence of Mutations in Beta-Lactamase TEM-1. *Mol. Biol. Evol.* **33**, 268–280 (2016).
60. Blázquez, J., Morosini, M.-I., Negri, M.-C. & Baquero, F. Selection of Naturally Occurring Extended-Spectrum TEM β -Lactamase Variants by Fluctuating β -Lactam Pressure. *Antimicrob. Agents Chemother.* **44**, 2182–2184 (2000).
61. Guthrie, V. B., Allen, J., Camps, M. & Karchin, R. Network Models of TEM β -Lactamase Mutations Coevolving under Antibiotic Selection Show Modular Structure and Anticipate Evolutionary Trajectories. *PLOS Comput. Biol.* **7**, e1002184 (2011).
62. Barlow, M. & Hall, B. G. Experimental prediction of the natural evolution of antibiotic resistance. *Genetics* **163**, 1237–1241 (2003).
63. Thai, Q. K., Bös, F. & Pleiss, J. The Lactamase Engineering Database: a critical survey of TEM sequences in public databases. *BMC Genomics* **10**, 390 (2009).
64. Forcella, C. *et al.* QnrB9 in Association with TEM-116 Extended-Spectrum Beta-Lactamase in *Citrobacter freundii* Isolated from Sewage Effluent: First Report from Italy. *J. Chemother.* **22**, 243–245 (2010).
65. Corvec, S. *et al.* TEM-187, a New Extended-Spectrum β -Lactamase with Weak Activity in a *Proteus mirabilis* Clinical Strain. *Antimicrob. Agents Chemother.* **57**, 2410–2412 (2013).

Acknowledgements

We would like to thank the Genome Sequencing and Analysis Facility (GSAF) at the University of Texas at Austin for genome library preparations and Illumina HiSeq. 4000 and miseq sequencing. We thank Dr. Jeffery Barrick for the thoughtful discussion and insight and for providing the Keio knockout strains used in this study. We would also like to thank Terry Stewart Jr. for diligently preparing laboratory materials. This work was supported by a National Security Science and Engineering Faculty Fellowship (FA9550–10–1–0169), the Air Force Office of Science Research (FA9550–14–1–0089), and Defense Advanced Research Projects Agency (HR0011-15-C0095). As well as the Welch Foundation (F-1654).

Author Contributions

D.T., A.C., and B.M. conducted experimental work. D.T., A.C., R.S., and A.E. wrote the main manuscript text. D.T. and A.C. prepared figures. R.S. assembled genomes. All authors conducted data analysis and reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-21549-w>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018