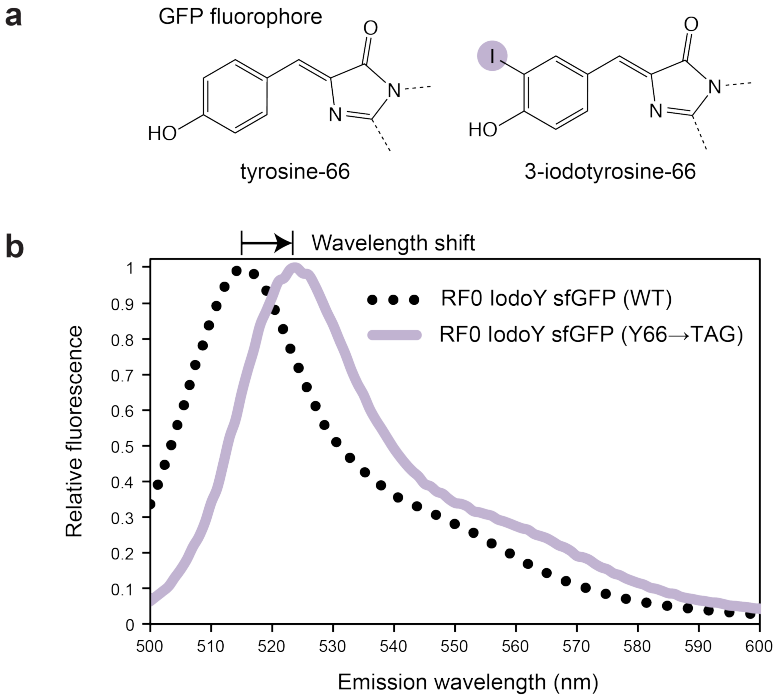
SUPPLEMENTARY INFORMATION

Bacteriophages use an expanded genetic code on evolutionary paths to higher fitness

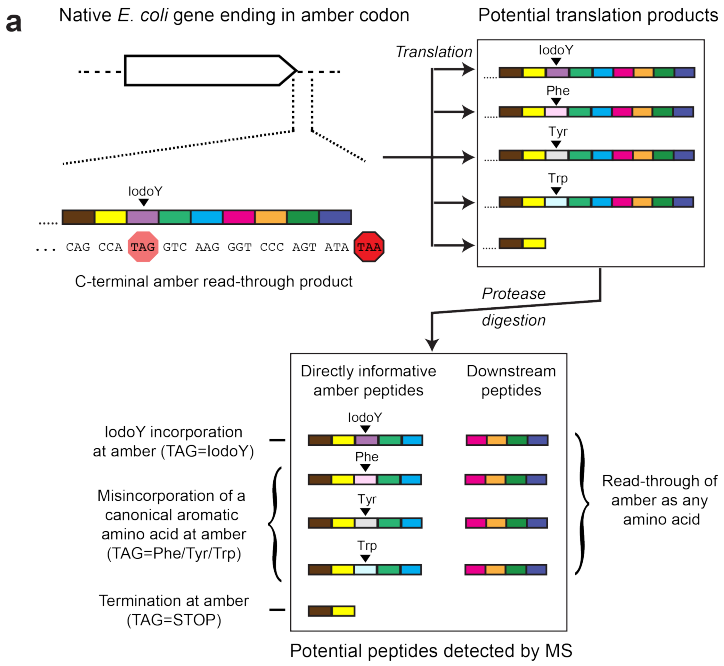
Michael J. Hammerling, Jared W. Ellefson, Daniel R. Boutz, Edward M. Marcotte, Andrew D. Ellington, and Jeffrey E. Barrick*

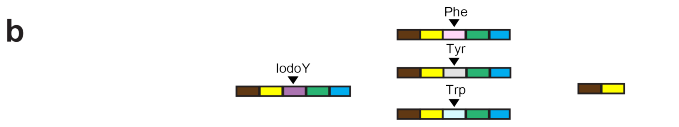
*Author for correspondence. Email: jbarrick@cm.utexas.edu; Phone: +1 512-471-3247

SUPPLEMENTARY RESULTS



Supplementary Figure 1 | Fluorescence wavelength shift from 3-iodotyrosine incorporation in RF0 IodoY. **a**, The fluorophore in normal GFP with tyrosine at position 66 is shown on the left. Incorporation of 3-iodotyrosine at this position results in the modified fluorophore shown at right, which has altered fluorescence properties. **b**, Normalized fluorescence emission spectra with excitation at 480 nm are shown for two *E. coli* RF0 IodoY cell cultures expressing GFP variants. The first contains a plasmid that encodes superfolder GFP (sfGFP) (dashed black curve). The second plasmid encodes the same sfGFP reading frame except with an amber codon substitution at position 66 (solid purple curve). The ~10 nm redshift of the peak intensity in the emission spectrum is characteristic of 3-iodotyrosine incorporation in the GFP fluorophore[28].

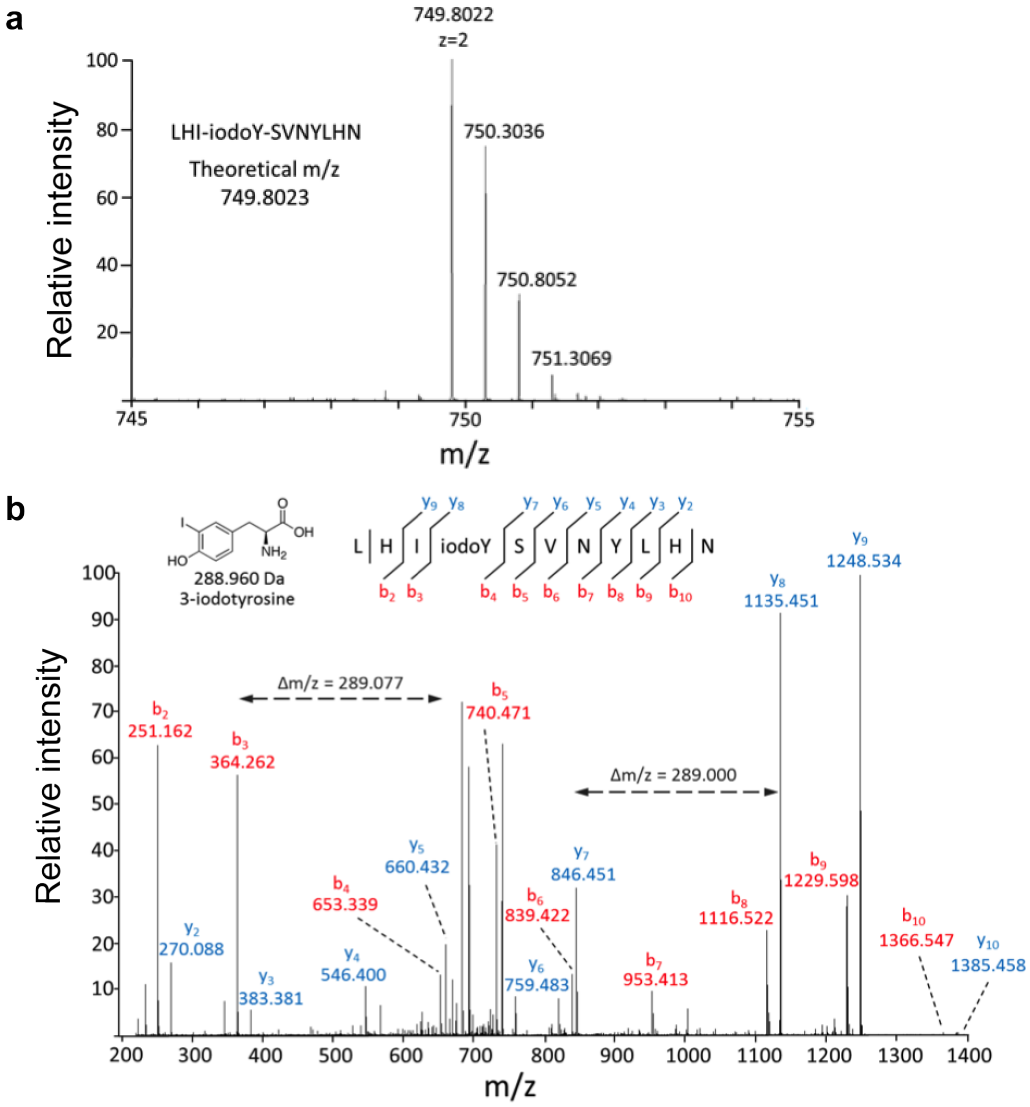




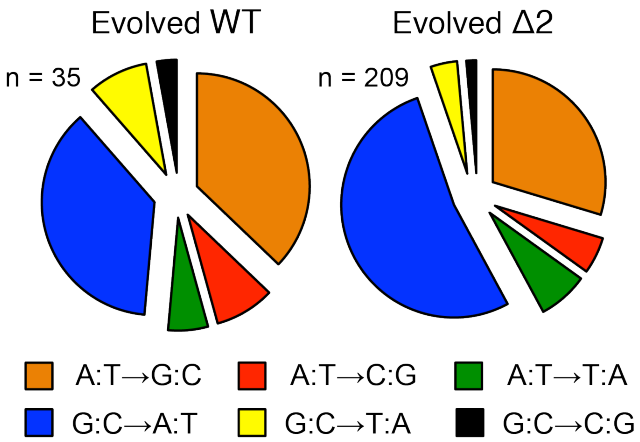
Sample \ Peptide	TAG=IodoY	TAG=Phe/Tyr/Trp	TAG=STOP
RF0 IodoY *E. coli*	62 (4)	5 (5)	1 (1)
RF0 IodoY *E. coli* Holin amber T7 lysate	76 (8)	0	0

Peptide spectral counts (Different proteins detected)

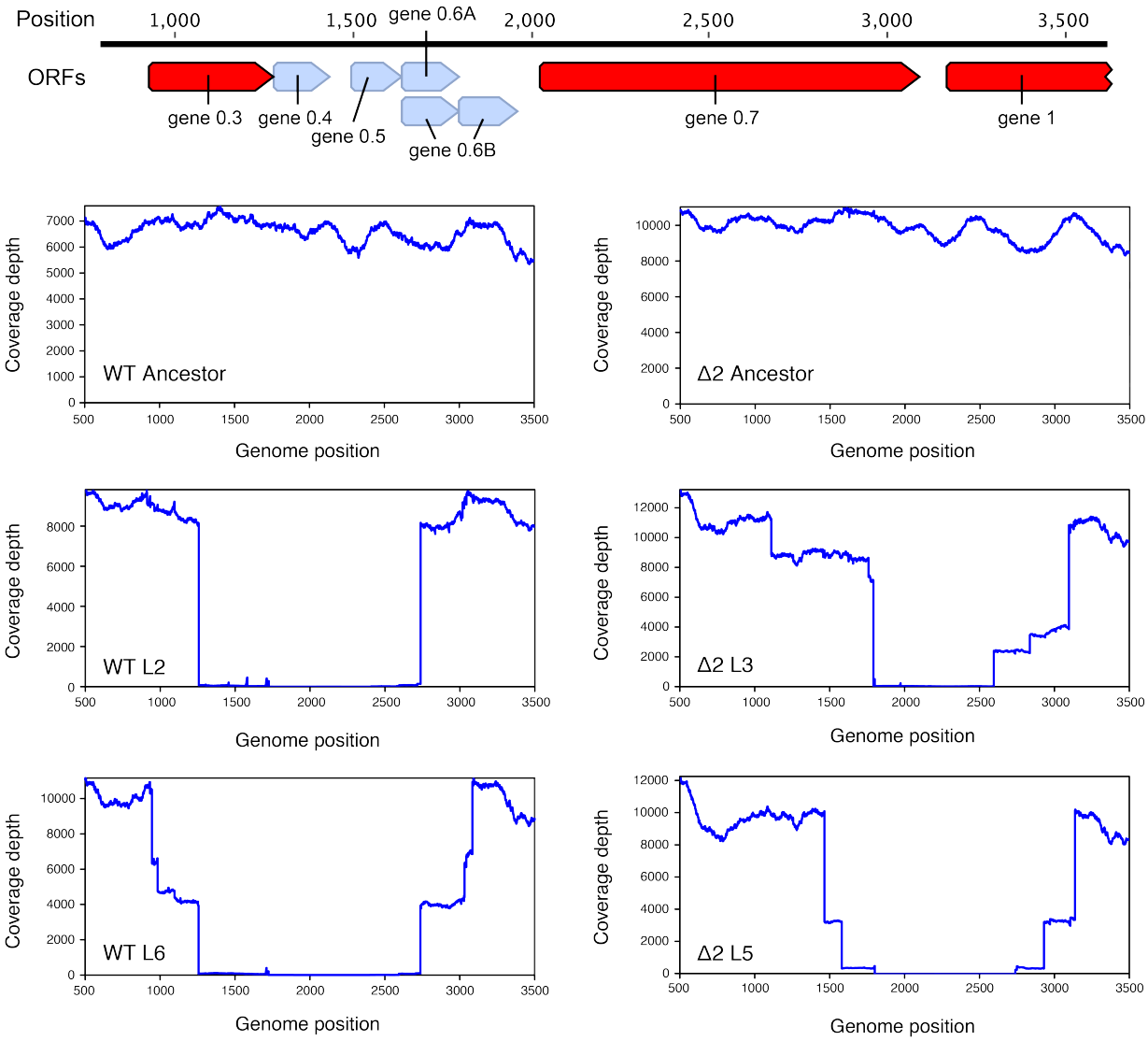
Supplementary Figure 2 | Proteomic characterization of amber read-through in *E. coli* strain RF0 IodoY. a, Schematic showing the possible effects of an amber codon on peptides detected by a mass spectrometry-based proteomics analysis. Termination at the amber codon or incorporation of 3-iodotyrosine or a canonical amino acid results in different masses for the directly informative peptides. For some proteins, read-through of the amber as any amino acid may also result in an additional C-terminal peptide or multiple peptides that do not contain the amber codon but indicate read-through, as long as these peptides are not also present in a downstream gene. **b,** The number of directly informative peptides detected in each category in a sample of RF0 IodoY cells and a sample of these cells lysed by an evolved T7 phage isolate from WT population L6 containing the type II holin amber mutation. All peptides tabulated are derived from *E. coli* genes normally terminated by amber codons. In total, peptides indicating read-through as 3-iodotyrosine were detected for 10 different *E. coli* proteins. Of the spectral counts of directly informative peptides across both samples, 138/144 (96%) match the mass expected if 3-iodotyrosine was incorporated at the amber codon, indicating high coding specificity. Peptide-spectrum matches for incorporation of alternative amino acids at amber codons are generally of lower quality than 3-iodotyrosine matches, so the actual specificity could be even higher. No informative peptides were detected for amber stop codons in the T7 genome except for 17 peptides derived from exonuclease (gene 6) that indicated normal termination of this protein was restored by a mutation to an alternative stop codon that occurred during the evolution experiment (**Supplementary Fig. 6**). No read-through of amber codons as 3-iodotyrosine was detected in a control sample of BL21(DE3) cells lysed with wild-type T7.



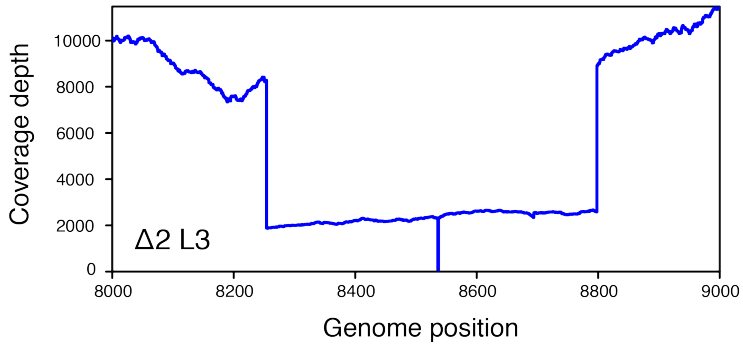
Supplementary Figure 3 | Confirmation of 3-iodotyrosine incorporation at amber codons by LC-MS/MS. Mass spectrometric analysis of RF0 IodoY *E. coli* proteomic samples identified several peptides containing 3-iodotyrosine due to translation of amber codons. MS/MS analysis of one such peptide (LHI-iodoY-SVNYLHN), derived from the *E. coli S*-ribosylhomocysteinase lyase protein (YgaG), is shown. This peptide was identified in the sample consisting of debris from an RF0 IodoY *E. coli* culture lysed with the T7 mutant phage and digested with GluC endonuclease. **a**, The isotopic distribution of the doubly charged precursor ion for this peptide gives an observed mass of 1498.5971 Da, within 0.14 ppm of the theoretical mass for the peptide sequence. **b**, The consensus MS/MS fragmentation spectrum contains b- and y-ions corresponding to nine of eleven residues in the sequence.



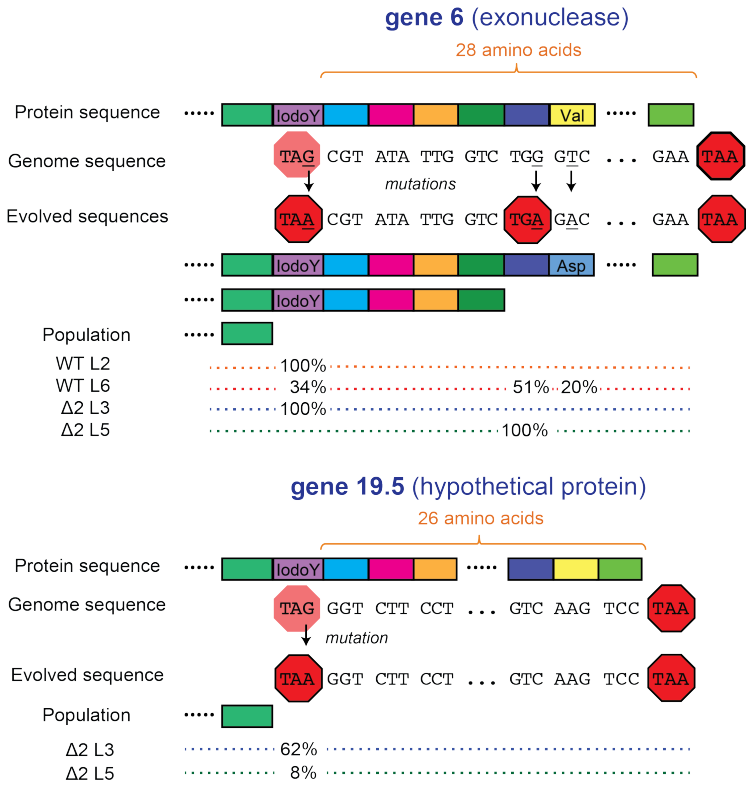
Supplementary Figure 4 | Base substitution mutation spectra. The overall base substitution spectra observed in wild-type (WT) and hypermutator (Δ2) bacteriophage populations are similar. For example, the apparent increase in the fraction of G:C→A:T base pair substitutions from 37% in WT to 57% in Δ2 is not statistically significant (Fisher's Exact Test, $P = 0.10$).



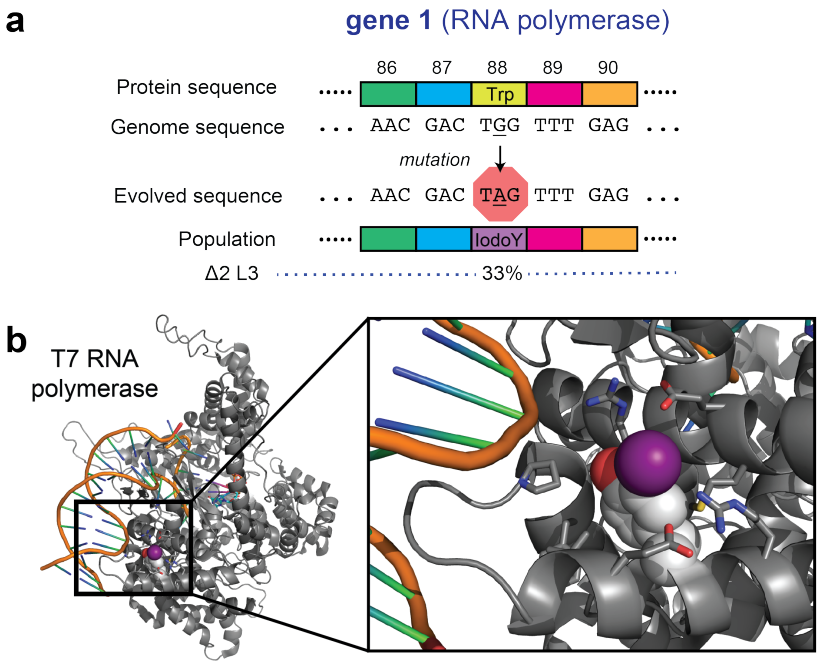
Supplementary Figure 5 | Deletions overlapping gene 0.7 in evolved populations. The top panel shows open reading frames in the region of the T7 bacteriophage genome containing deletions that impacted gene 0.7. Genes with known functions are shown in red. The bottom panels show the coverage of DNA sequencing reads at each position in the reference genome for the two ancestral phage samples and the four evolved mixed-population samples. The staggered steps and cliffs in coverage are due to nested and overlapping deletions of different sizes that occurred in different phage subpopulations. Since coverage of each evolved population as a whole went to zero at some point in this interval, we conclude that every individual in the population contained a deletion that overlapped gene 0.7. The frequency of each deletion with distinct endpoints was estimated from reads spanning the breakpoint it produced (**Supplementary Table 1**).



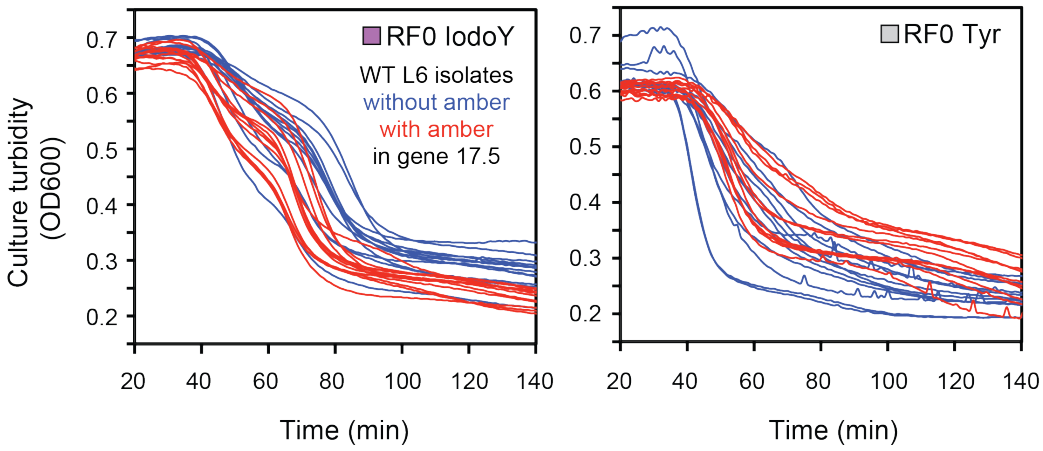
Supplementary Figure 6 | Deletion overlapping hypothetical genes 1.7 and 1.8 in population Δ2 L3. This plot of read coverage versus reference genome position shows a 544-bp deletion overlapping hypothetical genes 1.7 and 1.8 estimated to be present in 74% of the Δ2 L3 population from the proportion of reads spanning the sequence breakpoint this event creates versus reads spanning portions of the reference genome sequence on each side that exist only if the deletion is not present. The spike in the middle that dips down to zero coverage is due to a 1-bp deletion at position 8,536 that is present in the remainder of the population without the larger deletion.



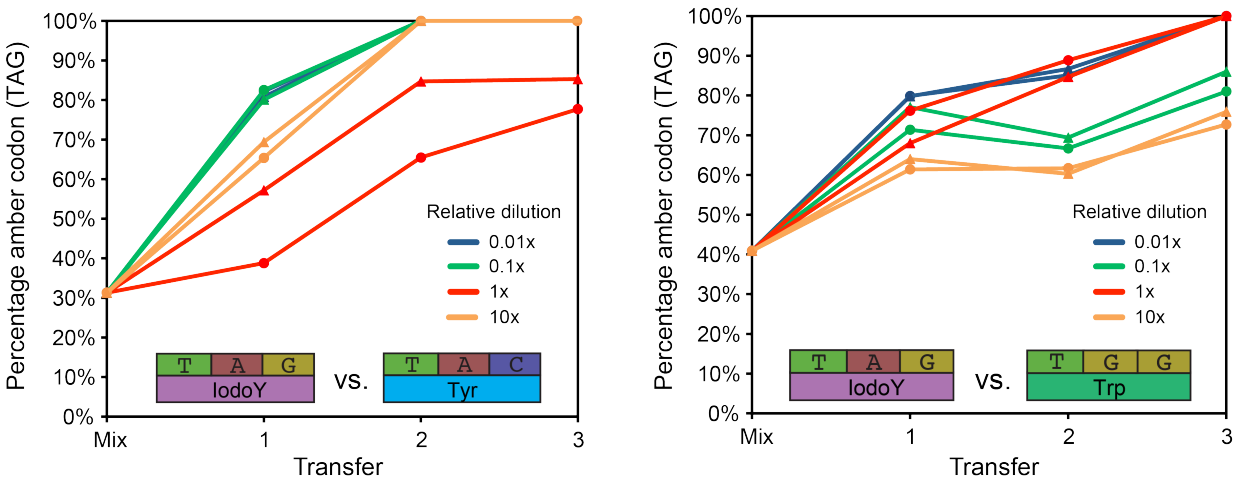
Supplementary Figure 7 | Compensatory evolution of reading frames to restore protein termination. Six open reading frames (ORFs) annotated in the ancestral T7 genome end in amber stop codons (TAG). No mutations that reached an appreciable frequency (>5%) in the sequenced bacteriophage populations were observed to affect translation termination for four of these genes, which have in-frame non-amber stop codons within 8 codons downstream. Read-through of the amber stop codons in the two remaining ORFs, exonuclease (gene 6) and a hypothetical protein (gene 19.5) in the RF0 IodoY host used in the evolution experiment, would result in C-terminal extensions of 29 and 27 new amino acids, respectively. Mutations that may compensate for defects caused by read-through of these amber stop codons evolved in the sequenced T7 population as shown. Most of these mutations change the original amber stop codon to the alternative ochre stop codon (TAA) by a single base substitution or create a new in-frame stop codon in the ORF closer to the original termination site. The valine to aspartic acid substitution observed after gene 6 in the WT L6 population might compensate for a read-through defect by changing the amino acid sequence of the C-terminal extension.



Supplementary Figure 8 | Addiction to an alternative genetic code due to an amber mutation in T7 RNA polymerase. a, The frequency of an amber mutation resulting in a tryptophan to 3-iodotyrosine substitution within T7 RNA polymerase (gene 1) that evolved in population Δ2 L3 can explain why a fraction of this population is unable to form plaques on the standard genetic code host (**Fig. 1c**). Signatures of molecular evolution described in the text suggest that this mutation was neutral with respect to fitness in the context of the evolution experiment. **b**, The 3-iodotyrosine substitution is located at a surface-exposed position of the intercalating β hairpin in the x-ray crystal structure of the T7 RNA polymerase initiation complex[20]. Modeling indicates that the non-canonical amino acid side chain can be accommodated at this site without requiring changes in the surrounding molecular structure. The 3-iodotyrosine side chain (spheres) and nearby amino acid side chains (sticks) are shown in the context of the template DNA and T7 protein backbones.



Supplementary Figure 9 | Lysis curves for T7 phage with and without type II holin amber mutations. Lysis time courses of 24 evolved phage isolates from population WT L6, half with and half without the amber mutation at codon 39 in the type II holin (gene 17.5). On average, phages with the amber mutation lyse the 3-iodotyrosine incorporating *E. coli* host on which they evolved more completely at earlier times than phages without the amber mutation (one-tailed Mann-Whitney U tests for lower OD600 values at ten minute intervals between 40 and 140 minutes, $P \leq 0.05$), but this difference is not found on an *E. coli* host that incorporates tyrosine at amber codons (one-tailed Mann-Whitney U tests for lower OD600 values at ten minute intervals between 40 and 140 minutes, $P > 0.05$). These results suggest that this amber mutation was beneficial with respect to this key trait that determines phage fitness in the context of the expanded genetic code with 3-iodotyrosine.



Supplementary Figure 10 | Evolved phage with the type II holin amber codon outcompete rescue mutants. Phage with the ancestral tyrosine codon (TAC) and a tryptophan codon (TGG) at position 39 of the type II holin (gene 17.5) were selected from an evolved WT L6 phage isolate with the amber codon (TAG) at this position on the basis of their ability to rescue growth on the normal genetic code *E. coli* host BL21(DE3). Each rescue mutant was competed against the original amber isolate by mixing together lysates created by reviving and preconditioning each phage separately on the 3-iodotyrosine incorporating host RF0 IodoY under the conditions of the evolution experiment. Serial transfer of these mixtures to initiate competition experiments and propagate them through a total of three infection cycles on RF0 IodoY were conducted in replicate (pairs of data series in the same color) at a range of different transfer dilutions relative to the 10^4-fold dilution used in the evolution experiment (1×), as indicated. The relative representation of the amber and alternative codons in phage populations at the end of each infection cycle were estimated by PCR and Sanger sequencing followed by analyzing the peak heights of bases corresponding to each allele in Sanger sequencing traces using Phred and polySNP.

Sequenced population	Positions deleted	Estimated frequency	Number of reads spanning junction
WT L2	912–3017	3.9%	323
	933–3043	0.8%	70
	960–2758	1.7%	140
	983–3031	0.8%	67
	1030–3017	1.2%	102
	1096–3027	1.1%	89
	1110–2593	0.5%	43
	1129–2714	0.8%	63
	1257–2737	89.2%	7424
WT L6	911–2914	0.4%	44
	930–2883	0.6%	60
	933–3043	3.2%	324
	945–3083	38.8%	3979
	958–3068	0.3%	33
	960–2758	1.2%	128
	983–3031	16.2%	1664
	1096–3027	2.3%	235
	1110–2593	0.6%	62
	1253–2742	1.1%	110
	1257–2737	34.9%	3578
	1724–3009	0.4%	36
Δ2 L3	1110-2593	21.1%	2252
	1125–2606	0.2%	25
	1257–2738	0.8%	81
	1457–3128	3.0%	321
	1701–3040	0.6%	64
	1723–3009	0.4%	46
	1759–2835	9.6%	1023
	1792–3096	64.2%	6839
Δ2 L5	547–2970	1.5%	156
	579–2748	1.9%	196
	1257–2738	2.4%	239
	1465–3136	63.8%	6461
	1581–2929	27.2%	2752
	1801–3105	3.2%	325

Supplementary Table 1 | Estimated frequencies of deletions overlapping gene 7 in evolved T7 populations. The extent of each deletion was determined from reads with split alignments to the reference genome. Frequencies were estimated by counting the numbers of reads that aligned best to each new sequence junction and assuming that the total frequency of deletions was 100% (**Supplementary Fig. 1**).

Sequenced Population	Amino acid (codon) change	Estimated frequency	Total frequency
WT L2	N7S (AAC→AGC)	40.9%	99.3%
	A13T (GCT→ACT)	17.4%	
	A13V (GCT→GTT)	13.1%	
	V17A (GTT→GCT)	11.5%	
	V21A (GTA→GCA)	11.5%	
	D23A (GAT→GCT)	4.9%	
WT L6	A13V (GCT→GTT)	4.1%	97.5%
	D23A (GAT→GCT)	40.1%	
	Y39* (TAC→TAG)	53.3%	
Δ2 L3	A14T (GCT→ACT)	1.6%	50.6%
	V24A (GTT→GCT)	1.3%	
	R27Q (CGA→CAA)	2.0%	
	F30L (TTT→CTT)	22.9%	
	F30L (TTT→TTA)	1.3%	
	A44T (GCC→ACC)	13.9%	
	V47G (GTG→GGG)	6.2%	
	K66R (AAG→AGG)	1.4%	
Δ2 L5	A14T (GCT→ACT)	1.5%	56.1%
	R27Q (CGA→CAA)	46.5%	
	A44T (GCC→ACC)	5.7%	
	A44V (GCC→GTC)	1.0%	
	K62R (AAG→AGG)	1.4%	

Supplementary Table 2 | Estimated frequencies of point mutations in T7 type II holin (gene 17.5). Mutation frequencies were estimated from the proportions of variant and reference bases in reads aligned to each position using the *breseq* computational pipeline, which takes into account estimates of sequencing errors using re-calibrated base quality scores. Since frequencies are independently estimated at each position and short DNA sequencing reads may overlap one position but not extend to others, the total frequency of mutants estimated in a sample is subject to uncertainty. From inspecting reads that do overlap multiple polymorphisms predicted within a population, there appears to be at most one mutation per individual phage isolate.

Supplementary Dataset 1 | Mutations in bacteriophage samples. This file contains the full results of analyzing the bacteriophage genome re-sequencing data. It lists the genetic differences between the WT and Δ2 phage ancestors and the T7 reference genome, and it provides details for all mutations predicted with frequencies ≥1% in each of the evolved phage populations.

Supplementary Dataset 2 | Informative peptide-spectrum matches. This spreadsheet contains quality information for each of the peptide-spectrum matches from the mass spectrometry proteomics analysis that was included in the estimate of the specificity of amber codon read-through as 3-iodotyrosine in the RF0 IodoY *E. coli* host used for the evolution experiment.