nature

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41586-019-1192-5

In the format provided by the authors and unedited.

Total synthesis of *Escherichia coli* with a recoded genome

Julius Fredens[1,4], Kaihang Wang[1,2,4], Daniel de la Torre[1,4], Louise F. H. Funke[1,4], Wesley E. Robertson[1,4], Yonka Christova[1], Tiongsun Chia[1], Wolfgang H. Schmied[1], Daniel L. Dunkelmann[1], Václav Beránek[1], Chayasith Uttamapinant[1,3], Andres Gonzalez Llamazares[1], Thomas S. Elliott[1] & Jason W. Chin[1]*

[1]Medical Research Council Laboratory of Molecular Biology, Cambridge, UK. [2]Present address: Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. [3]Present address: School of Biomolecular Science and Engineering, Vidyasirimedhi Institute of Science and Technology (VISTEC), Rayong, Thailand. [4]These authors contributed equally: Julius Fredens, Kaihang Wang, Daniel de la Torre, Louise F. H. Funke, Wesley E. Robertson. *e-mail: chin@mrc-lmb.cam.ac.uk

Methods

Recoded genome design

We based our synthetic genome design on the sequence of the *E. coli* MDS42 genome (accession number AP012306.1, released 07-Oct-2016), which has 3547 annotated CDS (**Supplementary Data 1**). We manually curated the starting genome annotation to remove three CDS and add another twelve. The three predicted CDS removed were *htgA*, *ybbV*, and *yzfA*; there is no evidence that these sequences encode proteins[1], and these sequences completely or largely overlap with better characterised genes, which would make it difficult to recode them without disrupting their overlapping genes or creating large repetitive regions. Conversely, the pseudogenes *ydeU*, *ygaY*, *pbl*, *yghX*, *yghY*, *agaW*, *yhiK*, *yhjQ*, *rph*, *ysdC*, *glvG*, and *cybC* were promoted to CDS. To enable negative selection with *rpsL*, we mutated the genomic copy of *rpsL* to *rpsL*^{K43R}. Finally, deep sequencing of our in-house MDS42 revealed a 51 bp insertion between *mrcB* and *hemL* which had not been reported in AP012306.1. We manually introduced and annotated this insertion in our starting genome sequence.

We produced a custom Python script that i) identifies and recodes all target codons, and ii) identifies and resolves overlapping gene sequences that contain target codons (available at https://github.com/TiongSun/genome_recoding). From our curated MDS42 starting sequence, we used the script to generate a new synthetic genome in which all TCG, TCA and TAG codons were replaced with AGC, AGT and TAA respectively. The script reported 91 CDS with overlaps containing target codons. In 33 instances, genes were overlapping tail-to-tail (3', 3') (**Supplementary Data 4**); 12 of these could be recoded by introducing a silent mutation in the overlapping gene, while the remaining 21 were duplicated to separate the genes (**Fig. 1b**). 58 instances

of genes overlapping head-to-tail (5', 3') were resolved by duplicating the overlap plus 20 bp of upstream sequence to allow endogenous expression of the downstream gene (**Fig. 1c**). For overlaps longer than 1 bp, an in-frame TAA was introduced to terminate expression from the original RBS for the downstream gene. *prfB* (release-factor RF-2) was not annotated as a CDS in our starting MDS42 genome due to its regulatory internal stop codon, and we therefore recoded all the target codons in the gene manually, thereby maintaining the internal stop codon. The resulting genome design contained 3556 CDS with 1,156,625 codons of which 18,218 were recoded (**Supplementary Data 2, Supplementary Data 3**).

Retrosynthesis of recoded stretches

We divided the designed genome into 37 fragments of between 91 and 136 kb. We chose the boundary sequences that delimit these fragments so that: i) they consist of a 5'-NGG-3' PAM to allow REXER4 to be used for integration if necessary, ii) the PAM does not sit within 50 bp of a target codon, iii) the PAM is in-between non-essential genes and iv) the PAM does not disturb any annotated features such as promoters. We called the regions ~50-100 bp upstream and downstream of these boundaries '*landing sites*', and these are annotated as Lxx, where xx is the number of the upstream fragment, e.g. L01 is the landing site between fragment 1 and 2 (**Supplementary Data 2)**. In our design, a landing site sequence is contained in the 3' end of a fragment and the 5' end of the next – as a result all 37 fragments contain overlapping homologies of 54-155 bp with their neighbouring fragment.

Each fragment was further broken down to 7-14 stretches of 4-15 kb. We designed the stretches so that they contain overlaps of 80-200 bp with each other, and the overlap

regions were defined at intergenic regions free of any recoding targets. A total of 409 stretches were synthesised (GENEWIZ, USA) and supplied in pSC101 or pST vectors flanked by BsaI, AvrII, SpeI, or XbaI restriction sites. The synthetic stretches naturally did not contain at least one of these restriction sites.

Construction of selection cassettes and plasmids for REXER/GENESIS

The cloning procedures described in this section were performed in *E. coli* DH10b, which is resistant to streptomycin by virtue of an *rpsL*K43R mutation. The plasmid pKW20_CDFtet_pAraRedCas9_tracrRNA used throughout this study encodes Cas9 and the lambda-red recombination components alpha/beta/gamma under the control of an arabinose-inducible promoter, as well as a tracrRNA under its native promoter, as previously described[2].

The protospacers for REXER are encoded in the plasmid pKW1_MB1$_{Amp}$_Spacer (**Supplementary Data 13**), which contains a pMB1 origin of replication, an ampicillin resistance marker and the protospacer array under the control of its endogenous promoter as previously described[2]. From this plasmid we constructed the derivative pKW3_MB1$_{Amp}$ _Tracr^{K}_Spacer (**Supplementary Data 14**), which additionally contains a tracrRNA upstream of the protospacer array. For this we introduced a PCR product containing tracrRNA with its modified endogenous promoter into the BamHI site of pKW1_MB1$_{Amp}$_Spacer via Gibson assembly using the NEBuilder HiFi Master Mix. From this plasmid a derivative that additionally encodes Cas9 was constructed, also by Gibson assembly, and named pKW5_MB1$_{Amp}$_Tracr^{K}_Cas9_Spacer.

For each REXER step, a derivative of one of these three plasmids was constructed to harbour a protospacer/direct repeat array containing 2 (REXER2) or 4 (REXER4) protospacers, corresponding to the target sequences for cutting the BAC and genome. The different protospacer arrays were constructed from overlapping oligos through multiple rounds of PCR – the products were inserted by Gibson assembly between restriction sites AccI and EcoRI in the backbone of pKW1_MB1$_{Amp}$_Spacer, pKW3_MB1$_{Amp}$_Tracr^{K}_Spacer or pKW5_MB1$_{Amp}$_Tracr^{K}_Cas9_Spacer. The protospacer arrays resulting from each assembly were verified to be mutation-free by Sanger sequencing. **Supplementary Data 9** contains a table indicating which backbone was used for each REXER step together with the protospacer sequences they contain.

The positive-negative selection cassettes used in REXER and GENESIS are -1/+1 (*rpsL-Kan*^{R}), -2/+2 (*sacB-Cm*^{R}) and -3/+3 (*pheS*^{T251A_A294G}-*Hyg*^{R}). -1/+1 and -2/+2 are as previously described[2]. In -3/+3, *pheS*^{T251A_A294G} is dominant lethal in the presence of 4-chlorophenylalanine, and *Hyg*^{R} confers resistance to hygromycin. Both proteins are expressed polycistronically under control of the EM7 promoter. The -3/+3 cassette was synthesised *de novo*. The -3/+3 cassette is also referred to as *pheS**-*Hyg*^{R}.

Constructing strains containing double selection cassettes at genomic landing sites.

According to our design, each region of the genome that is targeted for replacement by a synthetic fragment is flanked by an upstream landing site and a downstream landing site; these genomic landing site sequences are the same as the landing site sequences described above. Initiation of REXER/GENESIS requires the insertion of a double selection cassette in the upstream genomic landing site. We inserted double

selection cassettes at the landing sites through lambda-red mediated recombination. Briefly, either the *sacB-Cm*^{R} or the *rpsL-Kan*^{R} cassettes were PCR amplified with primers containing homology regions to the genomic landing sites of interest. For recombination experiments, we prepared electrocompetent cells as described previously[2] and electroporated 3 µg of the purified PCR product into 100 µL of MDS42$^{rpsL\text{K43R}}$ cells harbouring the pKW20_CDFtet_pAraRedCas9_tracrRNA plasmid expressing the lambda-red alpha/beta/gamma genes. The recombination machinery was induced, under control of the arabinose promoter (pAra), with L-arabinose added at 0.5% for 1 hour starting at OD_{600} = 0.2. Pre-induced cells were electroporated and then recovered for 1 hour at 37 °C in 4 mL of super optimal broth (SOB) medium. Cells were then diluted into 100 mL of LB medium with 10 µg/mL tetracycline and grown for 4 hours at 37 °C, 200 rpm. The cells were subsequently spun down, resuspended in 4 mL of H_2O, serially diluted, plated and incubated overnight at 37 °C on LB agar plates containing 10 µg/mL tetracycline, 18 µg/mL chloramphenicol (for *sacB-Cm*^{R}) or 50 µg/mL kanamycin (for *rpsL-Kan*^{R}).

BAC assembly and delivery

We constructed Bacterial Artificial Chromosomes (BACs) shuttle vectors that contained 97-136 kb of synthetic DNA. On the 5' side, the synthetic DNA was flanked by a region of homology to the genome (HR1), and a Cas9 cut site. On the 3' side the synthetic DNA was flanked by a double selection cassette, a region of homology to the genome (HR2), and a second Cas9 cut site. The BAC also contained a negative selection marker, a BAC origin, a URA marker and YAC origin (*CEN6* centromere fused to an autonomously replicating sequence (CEN/ARS)) (**Fig. 2c, Supplementary Data 6-8** provides maps with these features annotated).

BACs were assembled by homologous recombination in *S. cerevisiae.* Each assembly combined i) 7-14 stretches of synthetic DNA, each 6-13 kb in length, with ii) a selection construct (see below) and iii) a BAC shuttle vector backbone (**Supplementary Data 6-8**)[2].

Synthetic DNA stretches were excised by digestion with BsaI, AvrII, SpeI, or XbaI restriction sites from their source vectors provided by GENEWIZ. In the case of AvrII, SpeI, and XbaI, restriction digests were followed by Mung Bean nuclease treatment to remove sticky ends.

Selection constructs contained a region of homology to the 3' most stretch of the fragment, a double selection cassette (*sacB-Cm*^{R} or *rpsL-Kan*^{R}) a region of homology (HR2) to the targeted genomic locus, a negative selection marker (*rpsL, sacB or pheS*- Hyg*^{R}) and YAC. For specific double selection cassettes, negative selection markers, and homology region sequences see **Supplementary Data 9.** We assembled episomal versions of the selection constructs in a pSC101 backbone from 3 PCR fragments with NEBuilder HiFi DNA Assembly Master Mix. The episomal versions were designed so that restriction digestion with BsaI yielded a DNA fragment for BAC assembly.

The BAC backbone containing a BAC origin and a *URA3* marker was amplified by PCR using a previously described BAC[2] as a template, and the PCR product used for BAC assembly. The primers used for these PCR assemblies are listed in **Supplementary Data 9**.

To assemble the stretches, selection construct, and BAC backbone, 30-50 fmol of each piece of DNA was transformed into *S. cerevisiae* spheroplasts; these were prepared as previously described[3]. Following assembly we identified yeast clones potentially harbouring correctly assembled BACs by colony PCR at the junctions of overlapping fragments and vector-insert junctions. Clones that appeared correct by colony PCR were sequence verified by NGS after transformation into *E. coli*, as described below.

The assembled BACs were extracted from yeast with the Gentra Puregene Yeast/Bact. Kit (Qiagen) following the manufacturer's instructions. MDS42^{rpsLK43R} cells were transformed with the assembled BAC by electroporation. Due to the large size of the BACs we sometimes observed inefficient electroporation into target cells. Consequently, we introduced an *oriT*-Apramycin cassette provided as a PCR product with 50 bp homology regions by lambda-red-mediated recombination (as described above) into some BACs post assembly (**Supplementary Data 6-8**). This facilitated transfer of BACs, from *E. coli* that had been successfully transformed, to other strains by conjugation.

Synthesis of recoded sections

We used various genomic and plasmid selection markers for sequential REXER experiments (GENESIS) (**Supplementary Data 11**). We used an *rpsL-Kan*^{R} (-1/+1) or *sacB-Cm*^{R} (-2/+2) cassette at genomic landing sites for selection. We used *rpsL-Kan*^{R}-sacB (-1/+1,-2), *rpsL-Kan*^{R}*-pheS**- *Hyg*^{R} (-1/+1,-3/+3) or *sacB-Cm*^{R}*-rpsL* (-2/+2,-1) cassettes as episomal selection markers.

For each REXER, $MDS42^{rpsLK43R}$ cells containing pKW20_ CDFtet_pAraRedCas9_tracrRNA and a double selection cassette at the relevant upstream genomic landing site were transformed with the relevant BAC. We plated cells on LB agar supplemented with 2% glucose, 5 µg/ml tetracycline and antibiotic selecting for the BAC (i.e. 18 µg/ml chloramphenicol or 50 µg/ml kanamycin). We inoculated individual colonies into LB medium with 5 µg/ml tetracycline and the BAC specific antibiotic and grew cells overnight at 37 °C, 200 rpm. The overnight culture was diluted in LB medium with 5 µg/ml tetracycline, and the BAC specific antibiotic, to OD600 = 0.05 and grown at 37 °C with shaking for about 2 h, until OD600 ≈ 0.2. To induce lambda-red expression we added arabinose powder to the culture to a final concentration of 0.5% and the incubated the culture for one additional hour at 37 °C with shaking. We harvested the cells at OD600 ≈ 0.6, and made the cells electro-competent as described previously[2].

For each REXER experiment a linear dsDNA protospacer array was PCR amplified from pKW1_MB1Amp_Spacers using universal primers (**Supplementary Data 12**). Approximately 5-10 µg of the resulting DpnI digested and purified PCR product was transformed into 100 µL electro-competent and induced cells. Cells were recovered in 4 ml SOB medium for 1 h at 37 °C and then diluted to 100 mL LB supplemented with 5 µg/mL tetracycline and antibiotic selecting for the BAC and incubated for another 4 h at 37 °C with shaking. Alternatively, electrocompetent and induced cells were transformed with 5 µg of circular protospacer array (pKW1_MB1Amp_Spacers or pKW3_MB1Amp_Spacers plasmid) and after 1 h recovery in SOB medium at 37°C transferred into 100 mL LB supplemented with 100 µg/mL ampicillin for another 4 h at 37 °C with shaking (**Supplementary Data 12, 13**). If REXER2 was not sufficient

we performed REXER4 using pKW5_MB1Amp_Spacers plasmid as previously described[2].

We spun down the culture and resuspended it in 4 ml Milli-Q filtered water and spread in serial dilutions on selection plates of LB agar with 5 µg/ml tetracycline, an agent selecting against the negative selection marker and an antibiotic selecting for the positive marker originating from the BAC. The plates were incubated at 37 °C overnight. Multiple colonies were picked, resuspended in Milli-Q filtered water, and arrayed on several LB agar plates supplemented with 50 µg/ml kanamycin, 18 µg/ml chloramphenicol, 200 µg/ml streptomycin, 7.5% sucrose or 2.5 mM 4-chloro-phenylalanine. Colony PCR was also performed from resuspended colonies using both a primer pair flanking the genomic locus of the landing site and the position of the newly integrated selection cassette from the BAC. REXER-mediated recombination results in an approximately 500 bp band at the upstream genomic locus with a 2.5 kb (rK-landing site) or 3.5 kb (sC-landing site) band for the control $MDS42^{rk}$/ $MDS42^{sC}$ strain indicating successful removal of the landing site from the genome. Primer pairs flanking the 3' end of the replaced DNA generate an approximately 2.5 kb (rK selection cassette on pBAC) or 3.5 kb (sC selection cassette on pBAC) band and a 500 bp band for the control $MDS42^{rk}$/ $MDS42^{sC}$ strain indicating successful integration of the selection markers.

If a plasmid based circular protospacer array was used in the previous REXER experiment the plasmid had to be lost before the next experiment. Thus, a successful clone from the first REXER experiment was grown in LB supplemented with 2% glucose, 5 µg/mL tetracycline and antibiotic selecting for the positive marker in the genome to a dense culture at 37 °C with shaking. 2 µL of the culture were then

streaked out on an LB agar plate with the same supplements and incubated at 37°C overnight. Several colonies were arrayed in replica on LB agar plate and LB agar plate supplemented with 100 µg/mL ampicillin to screen for the loss of the plasmid.

BAC editing

When encountering loss-of-function mutations in a selection cassette on BACs in *E. coli*, the faulty cassette was replaced with a suitable double selection cassette provided (**Supplementary Data 9**) as a PCR-product flanked by 50 bp homology regions and integrated by lambda-red-mediated recombination.

Changes in the synthetic, recoded sequence of a BAC, either to correct spontaneous mutations or change recoded codons, were introduced by a two-step replacement approach; For BACs containing the selection cassettes -2/+2 and -1 in the end of the recoded sequence, the -3/+3 cassette was provided as a PCR-product flanked by 50bp-homology regions targeting the desired locus and integrated by lambda-red-mediated recombination followed by selection for +3. Due to the homology between the recoded DNA and the genome, some of the resulting clones would contain -3/+3 on the BAC and some on the genome. To identify clones with the cassette on the BAC, clones were plated in replica on agar plates selecting (1) for +3, (2) against -3, and (3) for +2 and against -3; Only clones surviving on plate (1) and (2) but not on (3) have the -3/+3 cassette integrated on the BAC. The location of the cassette was verified by purifying the BAC using QIAprep Spin Miniprep Kit followed by genotyping. In a second step, the -3/+3 cassette was replaced by providing a PCR-product of the desired sequence flanked by 50 bp-homology regions and integrated by lambda-red-mediated recombination followed by selection for +2 and against -3. The BAC was genotyped as above and sequence-verified by NGS.

Preparing a non-transferable F' plasmid and conjugative transfer of episomes

We created the version of the F' plasmid used for conjugation of genomic DNA, as well as transfer of BACs between strains, to enable transfer of sequences bearing *oriT* without transfer of the F' plasmid itself (**Supplementary Data 17**). We achieved this by deleting the nick-site in the origin of transfer (*oriT*) within the F' plasmid itself, a related approach was previously reported[4]. The F' plasmid derivative, pRK24 (addgene #51950), was modified by integrating desired markers as PCR-products flanked by 50 bp-homology regions and integration was performed by lambda-red-mediated recombination using a variant of pKW20 carrying Kan^R instead of Tet^R. First, the β-lactamase gene, conferring ampicillin resistance in pRK24, was replaced with the artificial T5-*luxABCDE* operon[5], which generates bioluminescence that allows visual identification of infected bacterial cells. Next, Tet^R was replaced with T3-*aac3* that produces aminoglycoside 3-N-acetyltransferase IV for selection with 50 μg/mL apramycin. Finally, a 24 bp deletion of the nick-site in *oriT* was made by integrating EM7-*bsd* that expresses blasticidin-S deaminase, and can be selected for with 50 μg/mL blasticidin in low-salt TYE/LB. The resulting F'-plasmid called pJF146 (**Supplementary Data 17**), was extracted using QIAprep Spin Miniprep Kit (QIAgen) and transformed by electroporation into donor strains for subsequent conjugation.

Transfer of episomal DNA containing *oriT* was performed by conjugation[6,7]. A donor strain was double transformed with pJF146 and an assembled BAC with *oriT* (see above). A recipient strain was transformed with pKW20. 5 ml of donor and recipient culture were grown to saturation overnight in selective LB media and subsequently

washed 3 times with LB media without antibiotics. The resuspended donor and recipient strains were combined in a 4:1 ratio, spotted on TYE agar plates and incubated for 1h at 37°C. The cells were washed off the plate and spread in serial dilutions on LB agar plates with 2% glucose, 5 µg/ml tetracycline selecting for the recipient strain and antibiotic selecting for the BAC. Successful transfer of the BAC was confirmed by colony PCR of the BAC-vector insert junctions.

Assembling a synthetic genome from recoded sections

Transfer of genomic DNA was combined with subsequent *recBCD*-mediated recombination to assemble partially synthetic *E. coli* genomes into a synthetic genome. In preparation of the donor and recipient strains a *rpsL-HygR-oriT* or Gm^R-*oriT* cassette was supplied as PCR product and integrated into the donor strain genome via lambda-red-mediated recombination (**Supplementary Data 15, 16**). Separately, a $pheS^*$-Hyg^R cassette was integrated approximately 3 kb downstream of the synthetic DNA in the donor strains. This provided a template genomic DNA for PCR amplification of a 3 kb synthetic DNA segment with 3' $pheS^*$- Hyg^R selection cassette. This PCR product was provided to the recipient strains to replace the WT DNA in a lambda-red-mediated recombination. Thereby, the selection marker at the 3' end of the synthetic segment was replaced and a 3 kb homology region to the donor synthetic DNA was generated. This strategy served to systematically generate recipient strains with 3 kb of homology with their respective donors, always with a *pheS*-Hyg^R at the 3' end. Additionally, the donor strains were transformed with pJF146 and sensitivity to tetracycline was confirmed. In contrast, pKW20 was maintained in the donor strains to confer tetracycline resistance.

For conjugation, donor and recipient strain were grown to saturation overnight in LB medium with 2% glucose, 5 μg/ml tetracycline and 50 μg/ml kanamycin or 20 μg/ml chloramphenicol (donor) and 50 μg/ml apramycin and 200 μg/mL hygromycin B (recipient). The overnight cultures were diluted 1:10 in the same selective LB medium and grown to $OD_{600} = 0.5$. 50 ml of both donor and recipient culture were washed 3 times with LB medium with 2% glucose and then each resuspended in 400 μl LB medium with 2% glucose. 320 μl of donor was mixed with 80 μl of recipient, spotted on TYE agar plates and incubated at 37°C. The incubation time depended on the length of transferred synthetic DNA and doubling time of the recipient strain and varied from 1h to 3h. Cells were washed off the plate and transferred into 100 ml LB medium with 2% glucose and 5 μg/ml tetracycline and incubated at 37°C for 2h with shaking. Subsequently 50 μg/ml kanamycin or 20 μg/ml chloramphenicol (selecting for the transferred positive selection marker of the donor) was added, followed by another 2 h incubation at 37°C. The culture was spun down and resuspended in 4 ml Milli-Q filtered water and spread in serial dilutions on selection plates of LB agar with 2% glucose, 5 μg/ml tetracycline, 2.5 mM 4-chloro-phenylalanine and 50 μg/ml kanamycin or 20 μg/ml chloramphenicol. Successful DNA transfer and recombination was determined by colony PCR for the loss of the $pheS^{*}$- Hyg^{R} cassette, integration of the donor's selection cassette and absence of the *Gm-oriT* cassette.

We performed a convergent synthesis of a genome recoded through sections A-E (**Extended Data Fig. 7**). We then used the A-E strain as a recipient for F, generating a recoded strain, A-F. A-F was then used as a recipient for F-G, generating A-G; this conjugation used a much longer shared recoded sequence (0.4 Mb) between the donor and recipient strains to increase conjugation efficiency.

To create a completely recoded genome we first created a recipient strain by introducing 37a and 37b into A-G to create A-G-37ab (providing a 115 kb homology region with the final donor). We created the final donor strain by conjugation between strain H and strain AB, which yielded strain H-A-09, in which H, A and fragment 9 from section B are recoded. The additional sequence from A and B was added to H to ensure that we did not erase the recoding in A in the final conjugation. The final conjugation between the H-A-09 donor strain and A-G-37ab recipient strain led to the synthesis of *E. coli*, which we name *E.coli* Syn61, in which all 1.8 $x10^4$ target codons in the genome are recoded.

Preparation of whole-genome and BAC libraries for next-generation sequencing

E. coli genomic DNA was purified using the DNEasy Blood and Tissue Kit (QIAgen) as per manufacturer's instructions. BACs were extracted from cells with the QIAprep Spin Miniprep Kit (QIAgen) as per manufacturer's instructions. We found that this kit was suitable for purification of BACs in excess of 130 kb. We avoided vigorous shaking of the samples throughout purification so as to reduce DNA shearing.

Paired-end Illumina sequencing libraries were prepared using the Illumina Nextera XT Kit as per manufacturer's instructions. Sequencing data was obtained in the Illumina MiSeq, running 2 x 300 or 2 x 75 cycles with the MiSeq Reagent kit v3.

Sequencing data analysis

The standard workflow for sequence analysis in this work is compiled in the iSeq package, available at https://github.com/TiongSun/iSeq. In short, sequencing reads were aligned to a reference recoded or wild-type genome using bowtie2 with soft-

clipping activated[8]. Aligned reads were sorted and indexed with samtools[9]. A customised Python script combines functionalities of samtools and igvtools to yield a variant calling summary. This script was used to assess mutations, indels and structural variations, in combination with visual analysis in the Integrative Genomics Viewer[10].

We produced a custom Python script to generate recoding landscapes across a target genomic region (available at https://github.com/TiongSun/recoding_landscapes). Briefly, the script takes a BAM alignment file, a reference in fasta and a GeneBank annotation file as inputs. It identifies the target codons for recoding, and compiles the reads that align to these target codons in the alignment file. It then outputs the frequency of recoding at each target codon, and plots these frequencies across the length of the genomic region of interest.

Growth rate measurement and analysis

Bacterial clones were grown overnight at 37 °C in LB with 2 % glucose and 100 μg/mL streptomycin. Overnight cultures were diluted 1:50 and monitored for growth while varying temperature (25 °C, 37 °C, or 42 °C) and media conditions (LB, LB with 2 % glucose, M9 minimal media, 2XTY). Measurements of OD_{600} were taken every 5 min for 18 h on a Biomek automated workstation platform with high speed linear shaking.

To determine doubling times, the growth curves were log2-transformed. At a linear phase of the curve during exponential growth, the first derivative was determined (d(log2(x))/dt) and ten consecutive time-points with the maximal log2-derivatives

were used to calculate the doubling time for each replicate. A total of 10 independently grown biological replicates were measured for the recoded Syn61 strain and wt MDS42$^{rpsL\text{K43R}}$. The mean doubling time and standard deviation from the mean were calculated for all n=10 replicates.

Microscopy and cell size measurement

Cells were grown with shaking in LB supplemented with 100 µg/mL streptomycin to approximately OD_{600}=0.2. A thin layer of bacteria was sandwiched between an agarose pad and a coverslip. A standard microscope slide was prepared with a 1% agarose pad (Sigma-Aldrich A4018-5G). A sample of 2 µl to 4 µl of bacterial culture was dropped onto the top of the pad. This was covered by a #1 coverslip supported on either side by a glass spacer matched to the ~1 mm height of the pad. Samples were imaged on an upright Zeiss Axiophot phase contrast microscope using a 63X 1.25NA Plan Neofluar phase objective (Zeiss UK, Cambridge, UK). Images were taken using an IDS ueye monochrome camera under control of ueye cockpit software (IDS Imaging Development Systems GmbH, Obersulm, Germany). 10 fields were taken of each sample. Images were loaded in to Nikon NIS Elements software for further quantitation (Nikon Instruments Surrey UK). The General analysis tool was used to apply an intensity threshold to segment the bacteria. A one micron lower size limit was imposed to remove background particulates and dust. Length measurements were subsequently made on the segmented bacteria using the General Analysis quantification tools.

Mass Spectrometry

Three biological replicates were performed for each strain. Proteins from each Escherichia coli lysates were solubilized in a buffer containing 6 M urea in 50 mM ammonium bicarbonate, reduced with 10 mM DTT, and alkylated with 55 mM iodoacetamide. After alkylation, proteins were diluted to 1 M urea with 50 mM ammonium bicarbonate, digested with Lys-C (Promega, UK) at a protein to enzyme ratio of 1:50 for 2 hours at 37 °C, followed by digestion with Trypsin (Promega, UK) at a protein to enzyme ratio of 1:100 for 12 hours 37 °C. The resulting peptide mixtures were acidified by the addition formic acid to a final concentration of 2% v/v. The digests were analysed in duplicate (1 ug initial protein/injection) by nano-scale capillary LC-MS/MS using a Ultimate U3000 HPLC (ThermoScientific Dionex, San Jose, USA) to deliver a flow of approximately 300 nL/min. A C18 Acclaim PepMap100 5 µm, 100 µm x 20 mm nanoViper (ThermoScientific Dionex, San Jose, USA), trapped the peptides prior to separation on a C18 Acclaim PepMap100 3 µm, 75 µm x 250 mm nanoViper (ThermoScientific Dionex, San Jose, USA). Peptides were eluted with a 100 minute gradient of acetonitrile (2% to 60%). The analytical column outlet was directly interfaced via a nano-flow electrospray ionisation source, with a hybrid dual pressure linear ion trap mass spectrometer (Orbitrap Velos, ThermoScientific, San Jose, USA). Data dependent analysis was carried out, using a resolution of 30,000 for the full MS spectrum, followed by ten MS/MS spectra in the linear ion trap. MS spectra were collected over a m/z range of 300–2000. MS/MS scans were collected using a threshold energy of 35 for collision induced dissociation. All raw files were processed with MaxQuant 1.5.5.1[11] using standard settings and searched against an *Escherichia coli* strain K-12 with the Andromeda search engine[12] integrated into the MaxQuant software suite. Enzyme search specificity was Trypsin/P for both endoproteinases. Up to two missed cleavages for each peptide

were allowed. Carbamidomethylation of cysteines was set as fixed modification with oxidized methionine and protein N-acetylation considered as variable modifications. The search was performed with an initial mass tolerance of 6 ppm for the precursor ion and 0.5 Da for CID MS/MS spectra. The false discovery rate was fixed at 1% at the peptide and protein level. Statistical analysis was carried out using the Perseus (1.5.5.3) module of MaxQuant. Prior to statistical analysis, peptides mapped to known contaminants, reverse hits and protein groups only identified by site were removed. Only protein groups identified with at least two peptides, one of which was unique and two quantitation events were considered for data analysis. For proteins quantified at least once in each strain, the average abundance of each protein across replicates of Syn61 was divided by the abundance in MDS42 replicates, and then log2-transformed. A P-value for the difference in abundance between strains was calculated by two-sample T-test (Perseus).

Toxicity of CYPK incorporation using orthogonal aminoacyl-tRNA synthetases tRNA$_{XXX}$s

We used a variant of stochastic orthogonal recoding of translation (SORT) to investigate the toxicity of a non--canonical amino acid using tRNAs with different anticodons[13-15]. Electrocompetent MDS42 and Syn61 cells were transformed with plasmid pKW1_*Mm*PylS_PylT$_{XXX}$ for expression of PylRS and tRNA$^{Pyl}_{XXX}$, where XXX is the indicated anticodon. Three variants of this plasmid were used, with the anticodon of tRNA^{Pyl} mutated to CGA (pKW1_*Mm*PylS_PylT$_{CGA}$), UGA (pKW1_*Mm*PylS_PylT$_{UGA}$) or GCU (pKW1_*Mm*PylS_PylT$_{GCU}$). Cells were grown over night in LB medium with 75 µg/ml spectinomycin. Overnight cultures were diluted 1:100 into LB supplemented with *Nε*-(((2-methylcycloprop-2-en-1-yl)

methoxy) carbonyl)-L-lysine (CYPK) at 0 mM, 0.5 mM, 1 mM, 2.5 mM and 5 mM and growth was measured as described above. "% Max Growth" was determined as the final OD_{600} in the presence of the indicated concentration of CYPK divided by the final OD_{600} in the absence of CYPK. Final OD_{600}s were determined after 600 min.

Deletion of prfA, serU and serT by homologous recombination

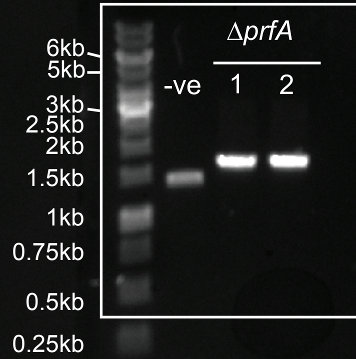
Recoded versions of the *pheS***-Hyg*^{R} and *rpsL-Kan*^{R} cassettes, according to the recoding scheme described in **Fig. 1a**, were synthesised *de novo,* so that expression of the selection proteins would not rely on decoding by *serU* or *serT*. For deleting *prfA*, the recoded *rpsL-Kan*^{R} was amplified with oligos containing ~50 bp homology to the *prfA* flanking genomic sequences. The same was done for *serU* and *serT* with recoded selection cassette *pheS***-Hyg*^{R}. Oligonucleotide sequences are provided in **Supplementary Data 21**. Syn61 cells harbouring the plasmid pKW20_CDFtet_pAraRedCas9_tracrRNA were made competent as described above, using 2xTY instead of LB. Cells were electroporated with ~8 µg of PCR product, and recovered for 1 hour in 4 mL SOB, then transferred to 100 mL 2xTY supplemented with 5 µg/ml tetracycline. After 4 hours cells were spun down, resuspended in 500 µL H_2O and plated in serial dilutions in 2xTY agar plates supplemented with 5 µg/ml tetracycline and 200 µg/ml hygromycin B (for *pheS***-Hyg*^{R}) or 50 µg/ml kanamycin (for *rpsL-Kan*^{R}). Deletions were verified in each case by colony PCR with primers flanking the locus of interest.

References for Methods

49 Strand, T. A., Lale, R., Degnes, K. F., Lando, M. & Valla, S. A new and improved host-independent plasmid system for RK2-based conjugal transfer. *PLoS One* **9**, e90372, doi:10.1371/journal.pone.0090372 (2014).

50. Bryksin, A. V. & Matsumura, I. Rational design of a plasmid origin that replicates efficiently in both gram-positive and gram-negative bacteria. *PLoS One* **5**, e13244, doi:10.1371/journal.pone.0013244 (2010).
51. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
52. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
53. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-192, doi:10.1093/bib/bbs017 (2013).
54. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372, doi:10.1038/nbt.1511 (2008).
55. Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* **10**, 1794-1805, doi:10.1021/pr101065j (2011).

Extended Data Fig 9e



Extended Data Fig 9f

Figure 4c