Science

RESEARCH ARTICLES

Cite as: W. E. Robertson *et al.*, *Science* 10.1126/science.ady4368 (2025).

Escherichia coli with a 57-codon genetic code

Wesley E. Robertson*†, Fabian B. H. Rehm†, Martin Spinck†, Raffael L. Schumann†, Rongzhen Tian†, Wei Liu‡, Yangqi Gu‡, Askar A. Kleefeldt, Cicely F. Day, Kim C. Liu, Yonka Christova, Jérôme F. Zürcher, Franz L. Böge, Jakob Birnbaum, Linda van Bijsterveldt, Jason W. Chin*

Medical Research Council Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge, UK.

***Corresponding author. Email: chin@mrc-lmb.cam.ac.uk (J.W.C.); wesr@mrc-lmb.cam.ac.uk (W.E.R.)**
†These authors contributed equally to this work. ‡These authors contributed equally to this work.

The near-universal genetic code uses 64 codons to encode the 20 canonical amino acids and protein synthesis. Here we designed and generated *Escherichia coli* with a 4 Mb synthetic genome in which we replaced known occurrences of six sense codons and a stop codon with synonymous codons. The resulting organism, Syn57, uses 55 codons to encode the 20 canonical amino acids.

All forms of life on earth use the near-universal genetic code (*1*), or its minimal variations (*2–4*). The 64 triplet codons, in the genetic code, encode the 20 canonical amino acids and the initiation and termination of protein synthesis, and 18 of the canonical amino acids are encoded by more than one synonymous codon (*5*). Synonymous codon choice is known to influence mRNA folding (*6*), gene expression (*7–9*), co-translational folding (*10*), and protein levels (*6*, *11*, *12*), and has distinct effects in distinct sequence contexts within genes and genomes (*13*).

Reducing the number of codons used for genomically-encoded protein synthesis has provided a basis for generating virus resistant organisms (*14–19*); it has also enabled: the more efficient synthesis of proteins containing non-canonical amino acids (*14–16*), the encoded synthesis of proteins containing multiple distinct non-canonical amino acids (*15*, *16*), and the genetically encoded synthesis of non-canonical polymers (*16*, *20*).

Stop codons have been removed from the *E. coli* genome by mutagenesis methods (*14*, *15*). These methods typically introduce large numbers of off target mutations (*14*, *21*), and are not generally scalable to the removal of all occurrences of targeted sense codons, which are commonly orders of magnitude more abundant than stop codons (*22*, *23*).

Emerging methods for the total synthesis of genomes (*24–27*) provide opportunities to explore genome sequences that cannot be accessed by editing. Synthetic genomes may be radically different from those accessed by natural evolution, and genome synthesis provides a route to generating sequence and function that is out-of-distribution with respect to extant life. To date, functional genomes from two bacteria have been synthesized (*28–30*), and efforts to synthesize other genomes (*31–34*) and strategies to recode (*14*, *21*, *26*, *31*, *35*), minimize (*30*, *31*), or rearrange genomes (*30*, *36–39*) are under way.

We previously reported the generation of Syn61 (*29*), a strain of *E. coli* with a synthetic genome in which we replaced annotated occurrences of the TCG and TCA sense codons, which normally encode serine, with defined synonyms, and the TAG stop codon with the TAA stop codon. Syn61 contains 18,214 codon changes in its genome with respect to the parental strain and uses 61 codons to make proteins. This strain has provided a foundation for creating virus-resistant cells, orthogonal genetic systems, and genetically programmed protein, polymer and macrocycle synthesis with up to three non-canonical monomers (*16*, *18–20*). Despite the success of this strain, it remained unclear whether living organisms could tolerate much deeper sense codon compression that take them further away from natural sequence.

Here we report the total synthesis of a recoded *E. coli* genome, which implements the genome-wide recoding of six sense codons and one stop codon, to deeply compress the genetic code. The resulting organism, named Syn57, has more than 10^5 codon changes with respect to the parental strain, replaces annotated occurrences of two alanine codons and four serine codons with synonyms, and uses a 57-codon genetic code.

Results

Defined codon compression schemes

We empirically determined allowed defined synonymous codon recoding schemes (r.s.) for code compression across a 20-kb region of the *E. coli* genome rich in essential genes and target codons (*24*) (allowed schemes are those for which the synonymous substitutions introduced through the synthetic DNA are tolerated over the targeted region of the genome; defined schemes are ones in which a target codon is systematically replaced with a defined synonym). For each scheme we assembled the synthetic recoded DNA in a bacterial artificial chromosome (BAC), introduced it into the genome by REXER (Replicon EXcision Enhanced Recombination) (fig. S1), and assessed the genomic recoding allowed at each position by sequencing individual post-REXER clones (fig. S1, A and B) (*24*).

We previously used this approach to test recoding

schemes in which we replaced two sense codons from the serine codon box (vS1, vS2, and vS3), or two sense codons from the alanine codon box, (vA7 and vA8) with defined synonyms across the 20-kb test region (Fig. 1, A and B, and data S1) (*24*). Alanine and serine codons (along with leucine codons, which were challenging to recode by the three schemes previously tested (*24*) and are not considered further here) were chosen because they are decoded by tRNAs in which the anticodon is not recognized by its cognate synthetase; this property simplifies subsequent codon reassignment (*16*, *29*). In previous experiments, and all experiments reported herein, we also replaced the TAG stop codon with TAA; these experiments aimed to test genetic codes that are compressed to use three fewer codons (3-codon compression schemes). We were able to recode the 20-kb region with 2 serine schemes (vS2, vS3) and the alanine vA7 scheme, and we define these schemes as allowed recoding schemes on this region.

In the present work, we created five new recoding schemes. In these schemes, two serine codons and two alanine codons were recoded (vS3A7, combining recoding schemes that were allowed individually), or four serine codons were recoded (vS33-vS36). We were able to recode the 20-kb region with all five schemes (Fig. 1B and fig. S2A), thereby demonstrating five allowed defined recoding schemes, each of which compressed the genetic code used on this region to use five fewer codons (5-codon compression schemes).

Next, we combined each of the serine recoding schemes from the 5-codon compression schemes (vS33-vS36) with the alanine recoding scheme (vA7) to generate four distinct 7-codon compression schemes (Fig. 1, B and C, and fig. S2B). We were able to recode the 20-kb region tested with all 7-codon compression schemes (Fig. 1C and fig. S2B). From amongst these four schemes, which appeared functionally equivalent, we chose the vS33A7 scheme to generate our genome design.

Design of a codon compressed genome

We designed a 4 Mb *E. coli* genome using the vS33A7 recoding scheme. Our design recoded 101,605 target codons to defined synonyms throughout the open reading frames of the genome (Fig. 1, C and D, and data S2 and S3). Our genome design also included 169 refactoring events that duplicated the region of overlapping genes containing target codons, essentially as described for Syn61 (*29*), to enable their independent recoding (Fig. 1D). We performed a retrosynthesis on the designed genome (*29*), by dividing the designed genome into 38 fragments of ~100 kb each, and dividing each fragment into 9-14 stretches of ~10 kb each.

Testing defined code compression genome-wide

We obtained 10-kb stretches of synthetic DNA (from a commercial vendor) that implement our design and cover the genome. We then used homologous recombination in *S. cerevisiae* to assemble BACs containing each ~100 kb synthetic fragment from the corresponding synthetic stretches.

We investigated the replacement of each 100-kb fragment of the genome with the corresponding synthetic fragment in 38 parallel strains. We did this using uREXER, a variant of REXER (*24*) employing universal CRISPR/Cas9 spacers (*25*) to excise the synthetic DNA from the BAC, and lambda-red-mediated recombination to facilitate its genomic integration in a Δ*recA* background (fig. S1, C and D).

27 of the 38 100-kb fragments led to complete replacement of the corresponding 100-kb fragment of the genome with synthetic DNA (Fig. 2A). These 27 fragments comprise 2.96 Mb of the entire 3.98 Mb genome and 75,880 of 101,605 total recoding events (Fig. 2B).

Regions of synthetic DNA that do not support cell growth as the sole copy in the genome are crossed out in post-uREXER clones; this generates chimeras between synthetic DNA and wildtype DNA. Compiled recoding landscapes (*24*, *25*, *40*), generated from the sequencing of post-uREXER clones, allowed us to localize the regions that maintain wildtype sequence (non-recoded) in all clones; these regions encompassed regions where the synthetic sequence tested is disallowed. Within ten of the eleven remaining fragments for which we did not directly identify fully recoded clones by uREXER (01, 02, 10, 13, 14, 24, 29, 31, 36, and 37), the compiled recoding landscapes allowed us to identify regions in the genome where wildtype DNA was maintained (figs. S3 to S12).

For fragments 02, 10, 13, 14, 24 we further refined the region where the originally designed synthetic sequence is disallowed. We achieved this by performing uREXERs on the wildtype genome using (mini)BACs that contained recoded synthetic sequence covering the region of non-recoded sequence identified in the compiled recoding landscapes from the original 100-kb uREXERs. These experiments generated higher resolution compiled recoding landscapes covering each region and narrowed the regions where non-recoded DNA was maintained and where the original design for synthetic, recoded, DNA was disallowed (figs. S4 to S8).

For fragment 27 (100k27) we did not obtain any correctly phenotyping clones following uREXER. To map the problematic regions in the recoded 100k27 sequence, we placed negative selection markers at several positions along the genomic sequence of 100k27 in parallel strains. We introduced the uREXER BAC into these strains in the absence of the lambda-red machinery, excised the 100k27 synthetic DNA with Cas9 and integrated it into the genome, using the endogenous the RecABCD machinery, while selecting against the genomic negative selection marker (fig. S13). We compiled a recoding landscape from these experiments which clearly defined a 21-kb region containing ribosomal operons, in the 3′ half of

100k27, that was never recoded. We split the 100k27 sequence across two uREXER BACs: 27A contains 61 kb of synthetic sequence from the 5′ half of 100k27 and, 27B contains 64 kb of synthetic sequence from the 3′ half of 100k27. uREXER experiments with the 27A BAC led to complete recoding of the corresponding genomic region (fig. S14).

Overall, we recoded 3.92 Mb of the genome across 39 strains. This left 60 kb of the genome, across the 11 fragments, that was recalcitrant to replacement with the original synthetic genome design (Fig. 2B).

Recoding the whole genome in 100-kb fragments

Next, we focused on discovering allowed synthetic sequences, which are consistent with our codon compression scheme, for the regions that were recalcitrant to replacement with the original synthetic genome design. We used a number of approaches, individually or in concert, to address this challenge within the eleven fragments where these regions occur; these approaches included: selecting alternative N-terminal coding sequences (NCS), using alternative recoding schemes which are consistent with code compression, altering promoter and regulatory sequences potentially disrupted by recoding, and adjusting initial refactoring designs (Fig. 2B and figs. S3 to S12 and S15) (*41*).

N-terminal coding sequences are important for gene expression (*6*, *13*, *42*–*45*). We hypothesized that using alternate N-terminal coding sequences (encoded in the first eight to ten codons of genes), while maintaining recoding and code compression, might favor recoding of otherwise non-recoded sequences. We generated synthetic DNA libraries which varied the N-terminal regions of all genes, or a rationally chosen subset of genes (e.g., essential genes) within previously non-recoded regions. We aimed to replace wildtype genomic sequence with the corresponding library of variants. The NCS libraries were recombined into the (commonly wildtype) genome to identify NCS variants that enabled full recoding across previously non-recoded regions. This strategy was used to recode the previously non-recoded region in the genomic fragment corresponding to fragment 24 (fig. S8), and contributed to the full recoding of the genomic fragments 01, 02, 10, 14, 27B, 29, and 36 (figs. S3 to S5, S7, S15, S9, and S11).

The recoding scheme in our genome design, vS33A7, is one of several schemes that replace the same 7 target codons with synonyms, and several alternate synonymous codon compression schemes that replace the same seven codons were viable in the original 20-kb test region (fig. S2B). We used an alternate recoding scheme (vS34A7) to recode the region of the genome in fragment 13 that was not recoded with the vS33A7 scheme used in the original design. Alternative recoding schemes were also used as part of strategies that led to the full recoding of fragments 14 and 27B in the genome (figs. S7 and S15).

In some cases, recoding events within annotated open reading frames overlap with the sequences of annotated promoters or other sequences that regulate gene expression. Within fragments 02, 10, 27B, and 37, such regulatory sequences are found within the regions that cannot be recoded using the original genome design. These regulatory sequences were targeted for mutagenesis as part of strategies that led to the full recoding, with code compression, of fragments 02, 10, 27B, and 37 in the genome (figs. S4, S5, S15, and S12).

In four cases (located in fragments 01, 10, 27B and 31) the refactoring in our original genome design was not sufficient to enable recoding. By either extending or omitting the refactoring, we were able to fully recode these regions of the genome that could not be recoded by the original design (Fig. 2B and figs. S3, S5, S15, and S10).

We found that the initial strain containing fragment 31 was not sufficiently fit for subsequent steps (Fig. 3). Mapping experiments demonstrated that recoding of a 4.2-kb region encompassing the ATP operon was associated with slow growth (figs. S10 and S16). The fitness of this strain was improved, using NCS libraries covering the ATP operon and altered regulatory sequences (Fig. 3B and figs. S17 and S18). The fitness improved strain was used for further conjugation-based assemblies.

Overall, we recoded the entirety of each 100-kb fragment (Fig. 2A), such that the sequence of the entire genome was recoded across 39 strains.

Synthesis of recoded sections

We used a version of genome stepwise interchange synthesis (GENESIS, GENomE Stepwise Interchange Synthesis: iterations of uREXER) to compile recoded fragments into single strains (Fig. 3A). This generated 13 recoded strains, containing larger recoded genomic sections, that cover the genome (Fig. 3, A and C). In the process of generating these recoded sections we fixed 100k33 (fig. S19) to enable complete recoding of 100k32-35 (r32-35) and, improved the growth of r03-04 and r16-21, as described below.

Fixing and enhancing recoded sections from GENESIS

The compiled recoding landscape from post-uREXER clones for adding 100k33 to 100k32 in the genome identified an 8-kb region within 100k33 that was not recoded in the resulting genome (fig. S19A); NCS fixes within this region enabled complete genome recoding within r32-35 (Fig. 3B and fig. S19E).

The complete recoding of 100k03 and 100k04 (r03-04) in a single genome and the complete recoding of r16-21 led to slow growing cells (fig. S16). We therefore wanted to identify regions of recoded sequence that limit growth of these cells, and to use this information to target these regions for fixing.

To pinpoint sequences within larger recoded sections that

are associated with fitness defects, we generated linkage maps (*46*, *47*) between recoding and fitness. We used bacterial conjugation to transfer a genome section from a parental, non-recoded (donor) strain into a (recipient) strain harbouring the recoded section of interest. After conjugative transfer of the wildtype DNA, RecABCD recombination generated genomes which contain chimeras between wildtype sequence and recoded sequence. These chimeras may have varying degrees of recoding and varying degrees of fitness (fig. S20).

We then subjected the strains with chimeric genomes to growth selection, to enrich faster growing cells. At regular intervals we isolated individual clones and measured their growth curves and determined the extent of their recoding by NGS. We collected this data for numerous clones with diverse fitness.

The data from all clones was compiled into a recoding-fitness linkage map (fig. S20). For each recoded codon position in the recipient genome we plotted the limit value for a fitness metric when recoding is maintained at that position in chimeric genomes; for example, we plotted the lowest doubling time for all clones that maintained recoding at each position. This revealed the location and severity of defects within the recoded sequence.

For the strain with r03-04 recoded we compiled sequencing and growth data from 88 unique clones into a recoding-fitness linkage map. The linkage map revealed a 3-kb region within which recoded positions limited growth. NCS and regulatory fixes within this region substantially enhanced growth (figs. S16 and S21).

In the case of the strain r16-21, we compiled sequencing and growth data from 128 unique clones into a recoding-fitness linkage map. The linkage map highlighted two genomic regions, of about 30 kb each; these regions were centered on the border between fragments 18 and 19, and the border between fragments 19 and 20 (fig. S22). Recoding in these regions was linked to slower growth. This linkage map therefore pinpointed two distinct regions where recoding limited growth (fig. S22). The structure of the growth associated *nuo* operon, within the growth limiting region between fragments 18 and 19, suggests that it utilizes the termination-reinitiation mechanism for operon translation. We predicted that a refactoring event in our genome design, which inserted 20 bases between *nuoJ* and *nuoK*, might disrupt translation of genes downstream of *nuoJ*; we therefore altered this refactoring event by deleting the 20 base insertion to generate a sequence where the stop codon of *nuoJ* and the start codon of *nuoK* are adjacent in DNA sequence, in order to re-instate the desired termination-reinitiation (fig. S22). Within the growth limiting region between fragments 19 and 20 we identified two essential genes (*ligA* and *zipA*) and targeted these for NCS fixes. The NCS fixes and refactoring alterations within these two regions enhanced growth (figs. S16 and S22).

The fixed and improved growth strains with r32-35, r03-04, and r16-21 recoded, along with the other strains generated by GENESIS (Fig. 3C), were used for conjugative assembly of larger recoded genomic sections.

Assembly of a recoded genome

We used conjugative transfer and recombination (conjugative assembly) to recombine recoded genome sections, generated by GENESIS, from donor strains into adjacent recoded genome sections in recipient strains (*29*, *48*) (fig. S23). Origin of transfer (*oriT*) cassettes were installed at one end of the recoded sequence in donor strains, such that they transferred the recoded sequence to the recipient stain. To ensure efficient recombination, recipient cells contained at least 3 kb of recoded sequence that overlapped with the recoded sequence adjacent to *oriT* in the donor cells; this sequence was followed by a double selection cassette (data S6). We selected for recipient cells which had lost the negative marker at the end of the recoded recipient sequence and gained the positive marker from the end of the donated recoded sequence.

We used a convergent series of conjugative assembly steps to compile the recoded sequence into a single genome (Fig. 4A). At several points in the series, we mapped and/or evolved the strains to generate conjugative donors or recipients that were competent for the next assembly step (Fig. 4A and figs. S24 to S28). For example, the evolution of r03-31(-27B) (fig. S28) was crucial for enabling the conjugation to generate a robust Syn57(-27B) strain; this strain was then conjugated with r24-30, which was derived from r24-30(-27B) via uREXER of a 27B sequence (fig. S29), to generate the final strain.

We named the final synthetic *E. coli* 'Syn57'. In this strain all 1.01×10^5 target codons from the parental genome are recoded (data S8). 106 target codons are recoded to codons distinct from those in the original design, but consistent with code compression; 101 of these codon changes were intentionally programmed (of which 100 were synonymous), and five were non-programmed (spontaneous mutations) to non-synonymous codons (data S9). 5 of the 169 refactoring events in the original design are deleted in Syn57 (fig. S30). Thus, 99.9% of the target codons were recoded according to the original design and 97% of the refactoring events are maintained from the original design.

Beyond target codon changes, Syn57 contains 286 mutations with respect to the original design; the vast majority of these mutations (199/286) were intentionally programmed into the genome, while 87/286 of these mutations were non-programmed (fig. S30 and data S9). The strain contained 20 indels, 5 of which were programmed (the 4 refactoring deletions plus 1 refactoring insertion) and 15 of which were non-programmed (12 deletions, 3 insertions) (fig. S30 and data S9).

The final strain grew in 2xYT on agar and in liquid culture (Fig. 4, B and C, and fig. S31). Syn57 grew about 4 times slower than the parental strain, and showed a distinct RNA-seq profile (fig. S31 and data S10); we expect it will be possible to accelerate the growth of this strain by mutagenesis and passaging, essentially as previously described (*16*, *49*).

Discussion

We have demonstrated the high-fidelity total synthesis of a functional genome for an organism with a 57-codon genetic code. The synthetic genome contains more than 10^5 codon changes with respect to the parental genome and replaces annotated occurrences of seven codons – four serine codons, two alanine codons, and the amber stop codon – with synonyms for code compression. TCN codons for serine were converted to AGT/C codons and GCR codons for alanine were converted to GCT/C codons; these changes converted both the six-codon box for serine and the four-codon box for alanine to two-codon boxes. This work exemplifies how genome synthesis can move the genome sequences of organisms into new regions of sequence space that may not have been accessed by natural life.

The synthesis of Syn57 proceeded through the recoding of 100-kb fragments (that cover the whole genome) across 39 strains by uREXER, the assembly of these fragments into sections by GENESIS, the conjugative assembly of these sections into larger sections and, ultimately the completion of the recoded genome. At each stage of the synthesis, we developed and implemented strategies to map the locations of disallowed sequences in the original design. Disallowed sequences within 100-kb synthetic fragments were mapped using uREXER and RecABCD mediated recombination, while disallowed sequences arising in larger sections were identified using variants of linkage mapping that we developed. These mapping experiments provided the targets for fixing the synthetic sequences. We developed a battery of methods for fixing potential defects in the regions identified by mapping, including: combinatorially varying the N-terminal coding sequences of genes in the region and selecting for variant sequences that improved growth, implementing alternative recoding schemes which are consistent with code compression, altering promoter and regulatory sequences potentially disrupted by recoding, and adjusting initial refactoring designs to reinstate regulatory mechanisms that may have been disrupted by refactoring.

Mapping and fixing at each stage of the synthesis was often crucial to enabling the next step of the synthesis. These experiments provide a paradigm for integrating 'just in time' (*50*) defect mapping and fixing of initial designs into synthetic schemes, such that local defects are identified and fixed early in the synthesis and longer range, potentially epistatic or synthetic lethal, defects are identified and fixed as they emerge in the assembly process.

In future work we will build on the generation of the deeply recoded strain we have created to explore the generation of deeply orthogonal genetic codes, enhanced virus resistance, the genetic encoding of up to seven distinct non-canonical monomers into proteins and the encoded cellular synthesis of non-canonical macrocycles and polymers composed of up to seven new monomers.

Materials and methods

Combining codon compression schemes

To empirically determine deep codon compression schemes, we combined codon compression schemes targeting the serine, alanine, and stop codon boxes. In our approach, additive recoding requires combining recoding schemes which target different 2-codon boxes (Fig. 1B). As input 3-codon compression schemes, we combined schemes targeting the TCR (serine), TCY (serine), and GCR (alanine) 2-codon boxes with the vST1 (STOP) scheme. To generate 5-codon compression schemes, we combined 3-codon compression schemes targeting the TCR serine box with schemes targeting the TCY serine box (TCR + TCY) or the GCR alanine box (TCR + GCR). Before proceeding with further codon compression, we experimentally determined the viability of individual 5-codon compression schemes by performing REXER and compiling a recoding landscape across a 20-kb region of the *E. coli* genome rich in essential genes and target codons (*24*). We used MDS42$^{rpsL(K43R)}$ cells harbouring a *rpsL-kanR* double-selection cassette as a genomic landing site at the 5′ end of the 20-kb region and the pKW20_CDFtet_pAraRedCas9_tracrRNA plasmid expressing the lambda-red alpha/beta/gamma genes along with Cas9 (*24*). To assemble the synthetic DNA harbouring the recoded sequence, we performed yeast assemblies to generate the YAC/BAC using the following input pieces: i) a YAC origin/ARS alongside a BAC origin, ii) a *sacB-cat* double-selection cassette with homology to the 3′ end of the 20-kb region, and iii) two 10-kb stretches of synthetic DNA encoding the recoded 20-kb region (see below in the section '*BAC assembly and delivery*' for details on yeast assembly). After transforming an assembled BAC harbouring the recoded 20-kb region, we performed REXER as described previously (*24*) and took 16 post-REXER clones forward for next-generation sequencing (NGS). Schemes with at least 75% of clones with a fully recoded 20-kb region were deemed as successful codon compression schemes and were taken forward for further codon compression (data S1). This process was repeated for the 7-codon compression schemes to yield an empirically validated design for a 57-codon genetic code.

Codon compressed genome design

We based our synthetic genome design on the sequence of the *E. coli* MDS42 genome (accession number AP012306.1),

which has 3547 annotated CDSs. From this initial genome template, we then updated our annotation based upon i) identifying discrepancies between the MDS42 genome annotation and its parental MG1655 genome annotation, and ii) comparing nucleotide sequence-based CDS length from the MDS42 genome annotation to the proteomics-based CDS length from the Uniprot database (*51*). In total we updated 33 CDS annotations, mostly to include N-terminal extensions that increased the length of the CDS (data S3). Our synthetic genome annotation includes the manual downgrading of 3 CDS annotations to pseudogenes (*htgA, ybbV, yzfA*) and the manual promotion of 12 pseudogenes to CDS annotations (*ydeU*, *ygaY*, *pbl*, *yghX*, *yghY*, *agaW*, *yhiK*, *yhjQ*, *rph*, *ysdC*, *glvG*, *cybC*) – these updates were based upon proteomic evidence or lack thereof (*51*). To enable negative selection with *rpsL*, we mutated the genomic copy of *rpsL* to $rpsL^{K43R}$.

To recode our updated MDS42 genome annotation, we used a custom Python script (*29*) that i) identifies and recodes all target codons to defined synonyms, and ii) identifies and resolves overlapping gene sequences that contain target codons. From our curated MDS42 starting sequence, we used the script to generate a recoded synthetic genome in which all TCG, TCA, TCT, TCC, GCG, GCA and TAG codons were replaced with AGC, AGT, AGC, AGC, GCT, GCT, and TAA codons, respectively (data S2 and S3).

Retrosynthesis of recoded stretches

We divided the ~4 Mb designed genome into 38 fragments of between 60 and 136 kb (data S2). We designated the boundary sequences between these fragments as 'landing sites.' For each fragment, the landing site is located at its 3′ end and overlaps by 50-100 bp of homologous sequence to the 5′ end of the downstream, neighboring fragment.

Each of the ~100-kb fragments were broken down into ~10 stetches of ~10 kb each (data S2). To facilitate downstream assembly by yeast, we designed each stretch to overlap with an adjacent stretch by 80-200 bp of homologous sequence. We ordered a total of 409 stretches to be synthesized and delivered in vector format (Twist Biosciences, USA), which we subsequently linearized by virtue of flanking BsaI, AvrII, SpeI, or XbaI restriction digest sites.

Construction of E. coli *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 a landing site. The location of genomic landing sites was chosen primarily to avoid disrupting annotated genome features such as promoters or CDSs, therefore avoiding potentially deleterious effects from double-selection cassette integration. The replacement of wild-type (WT) genomic DNA with recoded synthetic DNA requires the presence of a double-selection cassette at the 5′ landing site, for which we used the following negative and positive selection markers: *rpsL* (−1, streptomycin sensitivity), *kanR* (+1, kanamycin resistance), *sacB* (−2, sucrose sensitivity), *cat* (+2, chloramphenicol resistance), $pheS^{T251A\_A294G}$ (*pheS**; −3, 4-Chlorophenylalanine (4-CP) sensitivity), *hygR* (+3, hygromycin resistance), *gentR* (+4, gentamycin resistance), *tetR* (+5, tetracycline resistance), *ampR* (+6, ampicillin resistance), *nrsR* (+7, nourseothricin resistance), *aprR* (+8, apramycin resistance), and *specR* (+9, spectinomycin resistance). We primarily integrated *rpsL-kanR, pheS*-hygR*, or *sacB-cat* double-selection cassettes as the upstream genomic landing site for the region of interest (e.g., LS01 *sacB-cat* for 100k02 recoded fragment integration). As described previously in detail (*40*), to integrate a double-selection cassette at a given genomic landing site via lambda-red-based recombination, we first transformed $MDS42^{rpsL(K43R)/\Delta recA}$ cells (*25*) harbouring either the pKW20_CDFtet_pAraRedCas9_tracrRNA (*tetR*) or pKW20_CDFApr_pAraRedCas9_tracrRNA (*aprR*) plasmid expressing the lambda-red alpha/beta/gamma genes, induced them with 0.5% L-arabinose, and prepared them as electrocompetent. We then transformed PCR products of double-selection cassettes flanked by 50 bp homology arms to the genomic landing site of interest, and following recovery we plated cells on LB agar plates supplemented with either kanamycin (50 μg/mL) for *rpsL-kanR*, hygromycin (200 μg/mL) for *pheS*-hygR*, or chloramphenicol (20 μg/mL) for *sacB-cat* cassettes.

BAC assembly and delivery

We constructed Bacterial Artificial Chromosome (BAC) shuttle vectors that contained 60-136 kb of synthetic DNA and all the functional components required for uREXER. The 5′ side of synthetic DNA was flanked by a region of homology to the genome (HR1), and a Cas9 cut site. The 3′ side of synthetic DNA was flanked by a double-selection cassette, a region of homology to the genome (HR2), a second Cas9 cut site, and the expression cassette for two universal spacers targeting the two Cas9 cut sites flanking the HR1 and HR2 regions. The BAC also contained a negative selection marker, a BAC origin, a URA marker and a YAC origin (*CEN6* centromere fused to an autonomously replicating sequence (CEN/ARS)).

We assembled BACs by homologous recombination in *S. cerevisiae* B4741 cells. Each assembly combined i) 7-14 stretches of synthetic DNA, each 6-13 kb in length; with ii) a selection construct (see below), iii) a universal spacers construct, and iv) a BAC/YAC shuttle vector backbone (*24*).

Synthetic DNA stretches were either excised by digestion with BsaI, AvrII, SpeI, or XbaI restriction sites, or amplified by PCR from commercially-sourced 10 kb vectors (Twist Biosciences, USA).

Selection constructs contained a region of homology to

the *3′* most stretch of the fragment, a double-selection cassette (*sacB-cat* or *rpsL-kanR)*, a region of homology (HR2) to the targeted genomic locus, a negative selection marker (*rpsL, sacB or pheS*)*, and the YAC origin CEN/ARS. We amplified selection constructs by PCR from previously assembled BACs (*29*) corresponding to the homologous fragment in the genome. For examples of even and odd BAC selection cassettes along with negative selection markers and homology region sequences, see 'yeast piece 1' in data S4.

Universal spacer constructs contain two universal spacers, targeting either even or odd BACs, and enable the conversion of BACs from circular to linear dsDNA in the presence of Cas9 in cells. We generated universal spacer construct pieces with homologies for yeast assembly by PCR from a previously generated template – see 'yeast piece 2' in data S4.

For the BAC backbone, we PCR amplified the sequence containing a BAC origin and a *URA3* marker from a previously assembled BAC (*24*). For the nucleotide sequence of this PCR-generated yeast assembly piece, see 'yeast piece *3′* in data S4.

For BAC assembly, we utilized the natural recombinogenic ability of *S. cerevisiae*, as described previously (*40*, *52*). To program recombination, we designed each input piece with 60-180 bp of homology to adjacent pieces. To initiate yeast assembly, we transformed *S. cerevisiae* BY4741 spheroplasts with 30-50 fmol of the following input DNA pieces: the BAC selection construct, a BAC/YAC backbone, and each piece of synthetic DNA. Following selection on ΔURA plates, we identified yeast clones harbouring correctly assembled BACs by either i) colony PCR followed by NGS, or ii) direct NGS with yeast total nucleic acid extracts (as described below).

To extract assembled BACs from yeast, we adapted the Zymo Yeast Extract Kit for a 96-well plate format. Briefly, we inoculated single yeast colonies into 300 μL of YM4/2% glucose/ΔURA media into individual wells of 1.2 mL 96-well plates (AB-1127) and incubated them overnight at 30°C while shaking at 700 rpm (Grant Bio PHMP-4). We then spun down overnight cultures and digested them with 3.5 μL of zymolyase mixed with 80 μL of YD digestion buffer (Zymo Research) – digestion proceeded for 1 hour at 37°C while shaking at 500 rpm. We then isolated DNA by adding 30 μL of AMPure Beads (Beckman Coulter) followed by the manufacturer's recommended protocol for washing with 70% ethanol and eluting with 50 μL of water. For yeast extracts from each well, we assessed the fidelity of BAC assembly by next-generation sequencing (NGS) (see section '*Illumina NGS data analysis'* for details). Following NGS confirmation, we electroporated the assembled BAC yeast extract into DH10b and/or MDS42$^{rpsL(K43R)/\Delta recA}$ cells and selected for the positive selection marker on the BAC when plating. We isolated the BACs from overnight cultures using a commercial miniprep kit (Qiagen) where we mixed very gently after addition of the lysis and the neutralization buffers, and we decreased the spinning speed from ~17,900 rcf to 5,000 rcf in the column binding and washing steps to avoid shearing of the BAC. For higher yields and better transformation efficiencies, we alternatively isolated the BAC from a 20 mL overnight culture via isopropanol precipitation where we followed the steps of the miniprep kit (Qiagen) but precipitated the DNA from the supernatant with isopropanol and the obtained pellet was washed with 70% ethanol. The DNA was then resuspended in a final volume of 50-100 μL of water.

uREXER to integrate 100-kb recoded fragments

We generated universal REXER (uREXER) BACs by adapting the design of REXER (Replicon Excision for Enhanced Recombination) BACs by adding Cas9-based universal spacers to the BAC backbone. We PCR amplified universal spacer cassettes for odd and even BACs from episomes harbouring the relevant cassettes (*25*), and we placed them downstream of selection constructs in uREXER BAC designs. Utilizing universal spacers cassettes fused into the uREXER BAC construct obviates the need for: i) bespoke spacer plasmid cloning (*24*), and ii) supplementation of spacers in trans during synthetic DNA integration experiments - this expedited the synthesis of our recoded genome.

To initiate uREXER, we transformed MDS42$^{rpsL(K43R)/\Delta recA}$ cells containing pKW20_ CDFtet_pAraRedCas9_tracrRNA (*24*) and a double-selection cassette at the relevant upstream genomic landing site with the relevant isolated BAC (e.g., LS01 *sacB-cat* with 100k02 BAC). Following recovery for 1 hour in 1 mL of SOB, we plated transformed cells on 2xYT agar supplemented with 2% glucose, 5 μg/ml tetracycline and antibiotic selecting for the BAC (i.e., 20 μg/ml chloramphenicol for odd BACs or 50 μg/ml kanamycin for even BACs). Following restreaking of individual colonies, we inoculated colonies into 2xYT medium (5 mL) with 5 μg/ml tetracycline, 2% glucose (w/v) and the BAC-specific antibiotic, and we grew the cultures overnight at 37°C, 220 rpm. To initiate the expression of the lambda-red components and Cas9, we then spun down the overnight cultures and resuspended them into 200 mL of 2xYT medium with 0.5% L-arabinose (w/v), 5 μg/ml tetracycline, and the BAC specific antibiotic. We incubated uREXER cultures grown at 37°C at 220 rpm until the OD_{600} reached 0.3, at which point cultures were spun down and resuspended into 500 mL of fresh 2xYT media supplemented with 5 μg/ml tetracycline and the BAC specific antibiotic. We then grew cells until the OD_{600} reached 0.5, harvested the cells by centrifugation, and resuspended them in 1 mL of H_2O prior to plating them in serial dilutions on selection plates (245 × 245 × 25 mm, Thermo Scientific). For uREXER selection plates, we supplemented 2xYT agar with 5 μg/ml tetracycline, a negative selection agent against uncut

BAC (e.g., 1000 μg/mL streptomycin for odd BACs, or 7.5% (w/v) sucrose for even BACs), a negative selection agent against the upstream genomic landing site (e.g., 1000 μg/mL streptomycin for even genomic landing sites, or 7.5% (w/v) sucrose for odd genomic landing sites), and an antibiotic for the positive downstream marker from the BAC (e.g., 50 μg/mL kanamycin for even BACs, or 20 μg/mL chloramphenicol for odd BACs). We then incubated plates at 37°C overnight, picked multiple colonies, resuspended them in 50 μL of Milli-Q filtered H_2O, and restreaked them onto selective plates. We phenotyped them for uREXER marker swap by arraying 3.5 μL of resuspended colonies on several 2xYT agar plates supplemented with either 50 μg/ml kanamycin, 20 μg/ml chloramphenicol, 1000 μg/ml streptomycin, 7.5% sucrose or 2.5 mM 4-chloro-phenylalanine either on the same day as restreaking, or phenotyped from restreaked colonies. Additionally, the restreaked colonies were again resuspended in 50 μL of Milli-Q filtered H_2O and we genotyped for uREXER marker swap by colony PCR on resuspended colonies using both a primer pair flanking the genomic locus of the upstream landing site and the newly introduced selection cassette from the BAC. Recombinants with successful marker swap yielded a ~500 bp amplicon from the upstream genomic locus (i.e., loss of landing site) and a ~2-3 kb amplicon (i.e., gain of landing site) from the downstream genomic locus. For clones which passed phenotyping and genotyping, we then prepared genomic DNA using the QuickExtract DNA Extraction solution (LGC Biosearch Technologies), according to manufacturer's recommendations, which served as input material for subsequent NGS analyses. All uREXERs were performed as described above, except that the 100k31 full uREXER and the 100k24 NCS library uREXER were performed in a *recA*$^+$ MDS42$^{rpsL(K43R)}$ genomic background. For experiments which required additional negative selection pressure on the genome, we performed REXER4, where four Cas9-based cuts were generated during a synthetic DNA integration experiment – 2 cuts on the BAC to liberate the synthetic DNA as a linear dsDNA substrate, and 2 cuts on the genome flanking the targeted region for integration (fig. S29). Additional gRNAs for REXER4 were provided in trans, as described previously (*24*).

Synthesis of recoded walk strains by sequential uREXER/GENESIS

We alternated genomic selection markers for sequential uREXER experiments (GENESIS (*24*)), enabling the successive integration of 100-kb recoded fragments in a clockwise manner across the genome to thus generate increasingly recoded strains (fig. S1). For even BAC uREXERs, we selected against the maintenance of a *sacB-cat* genomic landing site and selected for the integration of a BAC-encoded *rpsL-kanR* cassette. For odd BAC uREXERs, we selected against maintenance of a *rpsL-kanR* genomic landing site and selected for the integration of a BAC-encoded *sacB-cat* cassette.

Following the NGS confirmation of a fully recoded uREXER clone, we inoculated a scrape from the corresponding glycerol stock into 5 mL of 2xYT supplemented with 2% glucose (w/v), 5 μg/mL of tetracycline, and either 50 μg/mL of kanamycin (for even fragments) or 20 μg/mL of chloramphenicol (for odd fragments). After overnight incubation at 37°C with shaking at 220 rpm, we diluted cultures into larger volumes of 2xYT with the same antibiotic/glucose combination and prepared electrocompetent cells prior to transformation of the next BAC. As described above in the section '*uREXER to integrate 100-kb recoded fragments*', we then proceeded to perform uREXER to integrate the next 100-kb fragment to increasingly recode the walk strain. Typically, GENESIS proceeded for four to five steps (400 to 500 kb) before characterizing recoded strain fitness.

Diagnostic recABCD integration of BACs

As an additional method beyond uREXER to define the non-recodable regions of 100-kb fragments, we performed diagnostic integrations of 100-kb BACs by using our recABCD recombination approach (described below, fig. S13). Prior to BAC transformation, into separate strains we first genomically integrated different landing sites (*sacB-cat, or rpsL-kanR*) in different positions across the region of interest. For each landing site strain, we then transformed the BAC (Kan or Cm) and pKW20_CDFApr_pAra_RecA_Cas9_tracrRNA (*aprR*) via electroporation and selected for co-transformants on 2xYT agar plates containing 50 μg/mL of kanamycin, 20 μg/mL chloramphenicol, 50 μg/mL apramycin and 2% glucose. We then inoculated three single colonies per landing site strain into 2 mL of 2xYT containing 50 μg/mL apramycin, 0.2% arabinose and grew these cultures at 37°C shaking at 220 rpm for ~6-8 hours to an OD_{600} of ~0.5-1.0. We harvested the cells by centrifugation and then spotted 3 μL of several dilutions (10^{-0} to 10^{-7}) and/or plated onto different selective plates (2xYT agar, 50 μg/mL apramycin, 400 μg/mL streptomycin or 7.5% sucrose and 2.5 mM 4-chloro-phenylalanine). We confirmed the phenotype of colonies from each landing site by the loss of the genomic marker (*sacB-cat*, or *rpsL-kanR*) and prepared an aliquot of the cells for subsequent NGS analyses. We compiled NGS data from multiple clones into a genomic recoding landscape as described above to generate diagnostic landscapes.

χ-Rec integration of 100-kb recoded fragments

Induced endogenous Chi-recABCD recombination (χ-Rec) utilizes the recBCD pathway for genomic integration of recoded 100-kb fragments. Briefly, we first reassembled the fixed recoded section in yeast with a 3χBAC backbone containing: 3 kb of homology on each side, universal spacers for

excision of the synthetic insert, an extended replication origin (*rfs-ccdAB*-resD-ori2-repE-SopABC*), an *oriT* and a negative selection marker (*pheS**) (data S7). Following NGS confirmation of the assembled 3χBAC, we then transformed a yeast extract of the BAC by electroporation into DH10b cells. The BAC was isolated from DH10b via isopropanol precipitation as described in '*BAC assembly and delivery*' and transformed via electroporation into strain harbouring a landing site (e.g., *sacB-cat*) upstream of the genomic fragment of interest and the pKW20_CDFApr_pAra_RecA_Cas9_tracrRNA (*aprR*) plasmid. After obtaining single colonies on 2xYT plates containing 50 μg/mL of kanamycin, 20 μg/mL chloramphenicol, 50 μg/mL apramycin and 2% glucose, we inoculated three single colonies into 2 mL of 2xYT containing 50 μg/mL apramycin and 0.2% arabinose. We grew induced cultures at 37°C with shaking at 220 rpm for ~6-8 hours to an OD_{600} of ~0.5-1.0 and subsequently harvested cells by centrifugation. Several dilutions were prepared (10^9 total cells, 10^{-0} to 10^{-7}) and spotted (~3 μL) and/or plated onto selective plates (2xYT agar, 50 μg/mL apramycin, 7.5% sucrose and 2.5 mM 4-chloro-phenylalanine). Following incubation for 18-24 hours 37°C, we phenotyped single colonies from the plate to confirm the loss of the upstream genomic landing site (e.g., chloramphenicol sensitivity for loss of *sacB-cat*) and gain of the downstream newly integrated landing site (e.g., kanamycin resistance for gain of *rpsL-kanR* from the BAC). Finally, we performed NGS as described in '*uREXER to integrate 100-kb recoded fragments*' to identify fully recoded clones.

Generation of NCS libraries on a BAC to fix non-recoded regions

To recode regions that were non-recodable with the vS33A7 recoding scheme, we constructed libraries of small BACs, termed NCS libraries (N-terminal coding sequences), where the 5′ ends of coding sequences of selected genes, such as essential or semi-essential genes, were varied. Libraries spanned the first 24-30 bp and preserved the amino acid sequence but varied the synonymous codons, either of recoding events only, or of all codons in that window, and reinstated canonical ATG start codons as needed. The libraries were assembled using degenerate primers (Merck) with which the selected genes were amplified. The entire synthetic insert sequence of the NCS-BAC was fused using overlap extension PCR. The BACs contained the same BAC/YAC backbone as the BACs constructed for uREXER and was amplified from a uREXER BAC with the correct marker setup for the landing site chosen. We assembled the NCS-BACs by standard Gibson assembly protocols, and then electroporated the assembled libraries into *E. coli* MDS42$^{rpsL(K43R)}$ or MDS42$^{rpsL(K43R)/\Delta recA}$ cells equipped with the corresponding landing site. We integrated the synthetic sequences into the genome via uREXER and sequenced for fully recoded clones either by Sanger sequencing or NGS.

pTarget protocol for fixing and for walking in 27B

To recode the 21-kb region within the 64 kb 27B fragment on the genome (figs. S13 and S15), we first divided this region into 4 pieces, each piece corresponding to the transcriptional unit from one of the 4 main promoters (fig. S15). We generated PCR products for each piece using four synthetic DNA templates (Twist Biosciences); these templates are the 21-kb region recoded with four 7-codon compression schemes (vS33A7, vS34A7, vS35A7, or vS36A7). These PCR products were generated with overlap extension PCR for all essential and semi-essential genes (all genes except *rpmJ* and *pioO*) to introduce NCS libraries (and internal promoter libraries for fragment 4) by using primers designed to include synonymisation of the first 8-10 codons and to restore the canonical ATG start codon as needed. Overall, the PCR products varied the recoding across the four schemes (vS33A7, vS34A7, vS35A7, vS36A7) and further varied the NCS and internal promoter sequences. We also split libraries 2 and 3, which contain a large number of genes, into smaller libraries (libraries 5-8) to more effectively cover NCS diversity within these genes (fig. S15B). The PCR products were used to replace the corresponding genomic sequence by pTarget-based recombination (CRISPR/Cas9 targeting cleavage of each wild-type genomic sequence and lambda-red recombination to integrate the PCR products into the genome (*53*)).

For pTarget-based recombination, we first transformed the pEcCas9 plasmid (*54*) into MDS42$^{rpsL(K43R)/\Delta recA}$ cells. We designed the pTarget plasmids by duplicating the sgRNA cassettes and replacing the two N20 sequences into new 20-bp sequences that target both ends of each fragment. To recode each fragment, we electroporated the PCR product (5,000-1,000 ng) and the corresponding pTarget plasmid (300-800 ng) simultaneously for targeted genome cutting and homologous recombination. After outgrowth for 3 hours, we plated cells under the selection of 50 μg/mL of kanamycin (pEcCas9) and 75 μg/mL of spectinomycin (pTarget). After incubation for 18 hours at 37°C, we then picked colonies for Sanger sequencing and NGS to confirm the recoding.

After confirming recoding solutions for all 4 fragments individually, we combined the solutions by introducing the sequences into the genome sequentially via pTarget-based recombination using the same protocol (fig. S15B). This led to genomic recoding of the 21-kb region, with a 73 bp wild-type sequence at the C terminus of *rplP*, which we subsequently replaced by lambda-red recombination with a synthetic sequence that implemented the original vS33A7 genome design.

Retron-mediated gene editing

For retron-mediated editing, we used a plasmid bearing the

retron sequence (pFR015) and a helper plasmid (pFR156) encoding CspRecT (*55*) and MutLE32K under the control of an arabinose-inducible promoter (*25*, *56*). The retron plasmid, which encodes the Ec86 retron scaffold and its cognate reverse transcriptase, was constructed by ligating annealed, phosphorylated ssDNA oligonucleotides into BsaI-digested pFR015. We designed the retrons to target the lagging strand of the genome or BAC to be edited as guided by short (30 to 35 base) homology arms and install the desired edit.

To perform the editing, the retron and helper plasmids were co-transformed into the strain harbouring the gene to be edited. An overnight culture of a co-transformant was then diluted 1:100 in fresh 2xYT medium supplemented with spectinomycin and gentamicin and incubated at 37C for 1 hour with shaking. Editing was induced via addition of 0.2% arabinose and the cultures were incubated for a further 24 hours. Cultures were then plated for single colonies and edited clones were identified via colony PCR and Sanger sequencing.

Multiple edits were performed to fix errors in BACs that arose during DNA synthesis or yeast assembly. A retron library was used to identify a solution for a non-recoded codon in 100k13.

Individual recoding landscapes and compiled recoding landscapes

We used an in-house Python script to generate recoding frequency landscapes. Briefly, each recoding event was annotated in the GenBank file using the format ["WT codon" to "Recoded codon"] (e.g., "TCA to AGT"). Next, NGS reads were pre-aligned and sorted by positional index using Bowtie 2 (*57*). Per sequenced clone, the sorted BAM file was then provided to the script, which scanned through all annotated recoding events in the GenBank file. For each event, the recoding frequency was calculated as the ratio of reads containing the recoded codon to the total number of reads at that position. The script produced individual CSV files for each clone, as well as a final summary CSV file compiling results from all sequenced clones. Resulting landscapes were then visualized in Prism.

Linkage mapping to determine growth-limiting recoded regions

In linkage mapping experiments for a given recoded strain, we conjugated genomic regions from a WT donor strain into a recoded recipient strain. The donor strain was prepared by integrating a *gentR-oriT* cassette facing toward the corresponding non-recoded section in the MDS42 genome through lambda-red recombineering (described above). The $RK24^{Lux}$ F-plasmid was introduced into the donor strain by electroporation or conjugation, and colonies were obtained on LB agar + 50 μg/mL apramycin and confirmed to contain the $RK24^{Lux}$ F-plasmid by luminescence (BioRad GelDoc, Chemiluminescence, 10 s). The recipient strain harboured the recoded region of interest flanked by a double-selection cassette as well as the pKW20_CDFSpec_pAra_RecA_Cas9_tracrRNA plasmid. Donor and recipient cells were grown overnight at 37°C in 2–15 mL of 2xYT selecting for all relevant genomic markers and the F-plasmid, and + 0.5% arabinose for inducing RecA expression in the *ΔrecA* strain. Cells were then harvested by centrifugation (5 min, 5000g) and washed three times with 1 mL of PBS. Each conjugation reaction contained a total of ~10^9 cells (estimated by OD_{600}) with a 1:1 ratio of donor and recipient cells; multiple replicates were set up in parallel. To initiate the conjugation the mixed cells were pelleted by centrifugation (5 min, 5000g), resuspended in 5–10 μL PBS and spotted on cut cellulose acetate filter membranes (Sartorius, 0.45 μm, Type 11106) placed on TYE/2xYT/LB agar plates (+ 0.5% arabinose for *ΔrecA* strains). The conjugation plates were incubated at 37°C for ~1.5 hours and cells were then recovered by vortexing the membranes in a 2 mL Eppendorf tube with 0.5 mL of SOB. Resuspended cells were then placed in an incubator shaker for 2 hours (37°C, 1,000 rpm). The cells were transferred into 10 mL LB or 2xYT supplemented with 50 μg/mL kanamycin ('even' ending sections) or 20 μg/mL chloramphenicol ('odd' ending sections) and 50 μg/mL spectinomycin, selecting for the recipient cells. The 10 mL cultures were passaged for several days by diluting them 1:100 into fresh media. All passaging intermediates were collected, and cells plated as single colonies (10^{-6}-10^{-7} dilution) on TYE/2xYT/LB agar supplemented with 50 μg/mL kanamycin ('even' ending sections) or 20 μg/mL chloramphenicol ('odd' ending sections). From these plates, phenotypically different colonies were picked from TYE plates from all timepoints and replicates into 50 μL of PBS in 1.2 mL 96-well plates (AB-1127). We then added 450 μL of LB containing the appropriate antibiotic to each well and incubated the plates for 24 hours (37°C, 800 rpm). We subjected all picked colonies to: i) growth curves to measure growth parameters, calculated as described in '*Growth rate measurement and analysis*'; ii) NGS analysis, as described in '*Preparation of whole-genome and BAC libraries for next-generation sequencing*'; and iii) recoding landscapes generation, as described in '*Individual recoding landscapes and compiled recoding landscapes*'.

We generated fitness-recoding linkage maps with the recoding and growth data listed above. To identify genomic regions associated with low fitness, we adopted the following strategy: a recoding event that is associated with poor fitness should exhibit a higher doubling time – the strain grows slower when it has not managed to cross out these recoding events across a distribution of different recoding landscapes. We first compiled together growth data and recoding landscapes, collected as described above, before filtering the following: i) samples which gained fitness through mutations

outside the recoded region of interest – assumed to have low doubling times/high max OD_{600} but maintain a fully recoded landscape; ii) samples which exhibit 'mixed traces' – assumed to derive from alignments to both a recoded and non-recoded copy of duplicated genome sections; for the purposes of fitness these should be considered non-recoded; and 3) samples with artefactual negative doubling times. Finally, for every recoding event, the minimum, maximum, mean and standard error of doubling times and max OD_{600} were identified, excluding doubling times >1,000 min (assumed non-growing). Linkage maps were plotted as minimum doubling times (black) and mean + stderr of doubling times (red) against corresponding recoding events as genomic coordinates, or maximum Max OD_{600} (black) and mean + standard error of Max OD_{600}.

RNA-seq mapping in growth limiting regions

We transformed BACs containing the 100-kb fragment of interest into MDS42$^{rpsL(K43R)/\Delta recA}$ cells and confirmed their integrity by NGS sequencing as described in *'Whole-genome and BAC library preparation for Illumina NGS'*. Cells containing the NGS-confirmed BAC were grown overnight in liquid culture in biological triplicates (2xYT + relevant BAC antibiotic); RNA was isolated from diluted overnight culture grown to OD_{600} = 0.2–0.3 (RNeasy Bacteria Mini Kit, Qiagen) and kept on ice during downstream processing. RNA integrity (RIN) scores were assessed on a Tapestation (Agilent) and ensured to be above 7. rRNA was depleted (NEBNext rRNA Depletion Kit (Bacteria), E7860L) to enrich mRNA and libraries were prepared (NEBNext Ultra II Directional RNA Library Prep Kit for Illumina, E7760S) for sequencing on an Illumina NextSeq2000 (P2 XLEAP-SBS (100 cycles), 60 bp PE reads).

Reference files for MDS42$^{rpsL(K43R)/\Delta recA}$ and each respective BAC were converted from GenBank to gff and fasta files using *emboss (58)*. For each BAC, the MDS42$^{rpsL(K43R)/\Delta recA}$ master gff/fa reference was concatenated with the respective BAC gff/fa. Illumina sequencing reads (FASTQ) were aligned to the 'MDS42 ΔrecA + BAC' reference (*Bowtie 2*, (*57*)) permitting multiple alignments, and bam files subsequently quantified using *HTSeq (59)*. Reads that aligned to more than one feature (i.e., reads which align to regions of the BAC with no recoding events) are counted for both the genome and the BAC and thus will not be considered differentially expressed.

Read count files were processed using R package *DESEQ2* (*60*) and transcripts corresponding to regions of the genome outside the 100 kb fragment of interest are excluded. Gene essentiality was annotated according to the Goodall database (*61*) and Log2FoldChange values plotted against adjusted p-values for each 100kbp fragment using R. Genes outside the range –1<Log2FoldChange <1 were classified as significantly differentially expressed.

Targeted adaptive evolution to improve recoded strain fitness

To improve the fitness of semi-synthetic strains we utilized oligo recombineering coupled with growth-based selections. Oligos were designed to form Okazaki fragments of the lagging strand during genome replication and were flanked by 30-35 bases of homology. Generally, we targeted N-terminal coding sequences by synonymising the first 8-10 codons and reinstating a canonical ATG start codon as needed, or we deleted refactoring insertions. Genomic regions were identified based on rational hypotheses, with essential genes prioritised, or as guided by uREXER recoding landscapes or fitness-recoding linkage maps (fig. S20). Strains to be evolved were transformed with plasmid pFR156 (*gentR*) or pFR157 (*hygR*) which encode for CspRecT and MutL-E32K (*62*, *63*) under the control of an arabinose-inducible promoter (*55*). An overnight culture of the strain was then diluted 50-fold in 100 or 250 mL 2xYT supplemented with the necessary antibiotics, grown at 37°C with shaking to OD_{600} = 0.2-0.3 and induced with 0.5% arabinose for 30 to 60 min. Electrocompetent cells were then prepared by washing the cells twice with 10% glycerol before suspending the cells in a final volume of 0.2 to 1.0 mL. Pools of synthetic oligos were then electroporated and the cells were recovered in 1 mL SOB for 2 hours at 37°C with shaking before transfer to a larger volume, typically between 50 to 250 mL, supplemented with the appropriate antibiotics. We serially passaged cultures for 2-7 days, in general, and cells were plated out for single colonies at regular intervals over this timeframe. Once heterogeneity in colony size emerged, we picked large colonies and measured individual growth curves in a 96-well plate format (see the '*Growth rate measurement and analysis'* section below). Clones exhibiting improved growth relative to the parental strain were then sequenced by NGS to identify whether on-target oligo-derived edits had occurred. The best performing clones were taken forward for further rounds of evolution or assembly steps.

Assembly of a synthetic genome from recoded sections by conjugation

Recoded sections of semi-synthetic strains were hierarchically combined to yield increasingly recoded strains via conjugative transfer of genomic DNA from a donor and subsequent *recBCD*-mediated recombination in the recipient (*48*). Donor strains were prepared by integrating an RK2 *oriT* cassette (*gentR-oriT*, *rpsL-kanR-oriT, sacB-nrsR-oriT*) upstream, and a double-selection cassette downstream, of the synthetic region to be transferred (fig. S23 and data S6). Additionally, the donor strains harboured an RK2 F-plasmid lacking an *oriT* which renders the F-plasmid incapable of self-transfer (*29*). The F-plasmid was introduced into the donor strain by electroporation or conjugation. Recipient

strains were prepared by integrating a compatible double-selection marker cassette, such as *pheS*-hygR*, downstream of a recoded stretch with homology to the region to be transferred from the donor strain. Regions of homology between donor and recipient were at least 3 kb in length and in some cases this homology was integrated simultaneously with the double-selection cassette integration. Longer homology regions generally led to increased assembly efficiency. Additionally, recipient strains harboured a plasmid encoding for recA under an arabinose-inducible promoter and a positive marker to be selected for after the conjugation, such as *specR*. The expression of recA from the plasmid was induced both in overnight cultures of recipient and in the agar plates on which the conjugation was run when genomic *recA* was lacking. For conjugation, overnight cultures of recipient and donor strains, supplemented with the necessary antibiotics, were washed thrice with PBS and mixed at donor:recipient ratios ranging from 4:1 to 1:1, generally at a total scale of 5 mL of OD_{600} = 2. These mixtures were then centrifuged, and the cell pellet was suspended in a minimal volume of residual PBS and spotted on cellulose acetate filter membranes (Sartorius) placed on a 2xYT, LB, or TYE agar plate. Conjugations were incubated at 37°C for 10 to 120 min after which cells were recovered from the membranes by vortexing the membrane in SOB. After shaking incubation at 37°C in SOB, for between 2 and 24 hours, the cells were plated on 2xYT or LB agar plates supplemented as needed to select for the marker plasmid in the recipient, the positive marker downstream of the synthetic region transferred from the donor, and against the negative marker downstream of the homology region in the recipient. Colonies that grew on these plates were then screened by phenotyping and confirmed by NGS before proceeding to the next assembly step.

Whole-genome and BAC library preparation for Illumina NGS

We isolated gDNA or BACs as described in '*uREXER to integrate 100-kb recoded fragments*' and diluted them at least 1:5 in Milli-Q filtered H_2O. BACs assembled in yeast were purified as described in '*BAC assembly and delivery*' We performed library preparation with assistance from an automated platform comprising a Biomek FXp (Beckman Coulter) with integrated thermocycler and fluorescence plate reader (Molecular Devices SpectraMAX I3), following protocols either described previously (*25*) – input DNA is quantified and diluted for the Nextera XT DNA Preparation Kit (Illumina) – or with the NEBNext UltraExpress FS DNA Library Prep Kit (New England Biolabs) without quantification following manufacturer's instructions at ~ 3X reduced volumes: 1) Fragmentation step: input DNA (3 μL), Milli-Q filtered H_2O (2.5 μL), FS enzyme mix (0.3 μL), FS reaction buffer (1.3 μL); 2) Adapter ligation: UltraExpress ligation master mix (8.5 μL), NEBNext Multiplex oligos for Illumina (Unique Dual Index UMI Adapters DNA Sets 1–4) at 1:10 dilution, (1.7 μL); 3) PCR: MTSC master mix (15.5 μL), NEBNext Primer Mix E7397A (1 μL), 5 cycles. Libraries were purified with AMPure XP (Beckman Coulter) according to manufacturer's instructions (23 μL) and eluted with 0.1X TE or Milli-Q filtered H_2O (13 μL). Libraries were subsequently quantified and pooled using Qubit 1X dsDNA HS assay kit (Thermo Fisher Scientific, Q33230) reagents adapted for a fluorescence plate reader.

Pooled libraries were paired-end sequenced on an Illumina NextSeq2000 according to manufacturer's instructions with either NextSeq2000 P1 (100/300 cycles), P2 (100/200 cycles) or P3 (100 cycles) reagent kits, utilizing either XLEAP-SBS or v3 SBS chemistry.

Illumina NGS data analysis

Reads were demultiplexed onboard (DRAGEN bcl2fastq 3.8.4 or 4.2.7) or from unindexed fastq.gz files with *demuxFQ* (CRUK CI genomics core). Reads were aligned with *Bowtie 2* (v2.5.1) and sorted and indexed with samtools (v1.17) for viewing in IGV (v2.13.2) or landscape generation.

Growth rate measurement and analysis

Bacterial clones were grown overnight at 37°C in 2xYT with the relevant antibiotic. Overnight cultures were diluted 1:100 and monitored for growth in a 200 μL volume in a 96-well plate. Measurements of OD_{600} were taken every 5 min for 18-24 hours on a Tecan Microplate reader.

To determine doubling times, we used the AMIGA fitness analysis software (*64*) to fit an exponential curve to a total of at least 3 independently grown biological replicates. For doubling time and max OD_{600} metrics, the mean and standard deviation from the mean were calculated for all n ≥ 3 replicates.

REFERENCES AND NOTES

1. F. H. Crick, The origin of the genetic code. *J. Mol. Biol.* **38**, 367–379 (1968). doi:10.1016/0022-2836(68)90392-6 Medline
2. D. Söll, Enter a new amino acid. *Nature* **331**, 662–663 (1988). doi:10.1038/331662a0 Medline
3. M. Kollmar, S. Mühlhausen, Nuclear codon reassignments in the genomics era and mechanisms behind their evolution. *BioEssays* **39**, 1600221 (2017). doi:10.1002/bies.201600221 Medline
4. Y. Shulgina, S. R. Eddy, A computational screen for alternative genetic codes in over 250,000 genomes. *eLife* **10**, e71402 (2021). doi:10.7554/eLife.71402 Medline
5. F. H. Crick, L. Barnett, S. Brenner, R. J. Watts-Tobin, General nature of the genetic code for proteins. *Nature* **192**, 1227–1232 (1961). doi:10.1038/1921227a0 Medline
6. G. Kudla, A. W. Murray, D. Tollervey, J. B. Plotkin, Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **324**, 255–258 (2009). doi:10.1126/science.1170160 Medline
7. B. K. Cho, K. Zengler, Y. Qiu, Y. S. Park, E. M. Knight, C. L. Barrett, Y. Gao, B. Ø. Palsson, The transcription unit architecture of the *Escherichia coli* genome. *Nat. Biotechnol.* **27**, 1043–1049 (2009). doi:10.1038/nbt.1582 Medline
8. G. W. Li, E. Oh, J. S. Weissman, The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* **484**, 538–541 (2012). doi:10.1038/nature10965 Medline

9. J. F. Curran, M. Yarus, Rates of aminoacyl-tRNA selection at 29 sense codons in vivo. *J. Mol. Biol.* **209**, 65–77 (1989). doi:10.1016/0022-2836(89)90170-8 Medline
10. Y. Jiang, S. S. Neti, I. Sitarik, P. Pradhan, P. To, Y. Xia, S. D. Fried, S. J. Booker, E. P. O'Brien, How synonymous mutations alter enzyme structure and function over long timescales. *Nat. Chem.* **15**, 308–318 (2023). doi:10.1038/s41557-022-01091-z Medline
11. C. Kimchi-Sarfaty, J. M. Oh, I.-W. Kim, Z. E. Sauna, A. M. Calcagno, S. V. Ambudkar, M. M. Gottesman, A "silent" polymorphism in the *MDR*1 gene changes substrate specificity. *Science* **315**, 525–528 (2007). doi:10.1126/science.1135308 Medline
12. G. Zhang, M. Hubalewska, Z. Ignatova, Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.* **16**, 274–280 (2009). doi:10.1038/nsmb.1554 Medline
13. J. B. Plotkin, G. Kudla, Synonymous but not the same: The causes and consequences of codon bias. *Nat. Rev. Genet.* **12**, 32–42 (2011). doi:10.1038/nrg2899 Medline
14. M. J. Lajoie, A. J. Rovner, D. B. Goodman, H.-R. Aerni, A. D. Haimovich, G. Kuznetsov, J. A. Mercer, H. H. Wang, P. A. Carr, J. A. Mosberg, N. Rohland, P. G. Schultz, J. M. Jacobson, J. Rinehart, G. M. Church, F. J. Isaacs, Genomically recoded organisms expand biological functions. *Science* **342**, 357–360 (2013). doi:10.1126/science.1241459 Medline
15. M. W. Grome, M. T. A. Nguyen, D. W. Moonan, K. Mohler, K. Gurara, S. Wang, C. Hemez, B. J. Stenton, Y. Cao, F. Radford, M. Kornaj, J. Patel, M. Prome, S. Rogulina, D. Sozanski, J. Tordoff, J. Rinehart, F. J. Isaacs, Engineering a genomically recoded organism with one stop codon. *Nature* **639**, 512–521 (2025). doi:10.1038/s41586-024-08501-x Medline
16. W. E. Robertson, L. F. H. Funke, D. de la Torre, J. Fredens, T. S. Elliott, M. Spinck, Y. Christova, D. Cervettini, F. L. Böge, K. C. Liu, S. Buse, S. Maslen, G. P. C. Salmond, J. W. Chin, Sense codon reassignment enables viral resistance and encoded polymer synthesis. *Science* **372**, 1057–1062 (2021). doi:10.1126/science.abg3029 Medline
17. N. J. Ma, F. J. Isaacs, Genomic recoding broadly obstructs the propagation of horizontally transferred genetic elements. *Cell Syst.* **3**, 199–207 (2016). doi:10.1016/j.cels.2016.06.009 Medline
18. A. Nyerges, S. Vinke, R. Flynn, S. V. Owen, E. A. Rand, B. Budnik, E. Keen, K. Narasimhan, J. A. Marchand, M. Baas-Thomas, M. Liu, K. Chen, A. Chiappino-Pepe, F. Hu, M. Baym, G. M. Church, A swapped genetic code prevents viral infections and gene transfer. *Nature* **615**, 720–727 (2023). doi:10.1038/s41586-023-05824-z Medline
19. J. F. Zürcher, W. E. Robertson, T. Kappes, G. Petris, T. S. Elliott, G. P. C. Salmond, J. W. Chin, Refactored genetic codes enable bidirectional genetic isolation. *Science* **378**, 516–523 (2022). doi:10.1126/science.add8943 Medline
20. M. Spinck, C. Piedrafita, W. E. Robertson, T. S. Elliott, D. Cervettini, D. de la Torre, J. W. Chin, Genetically programmed cell-based synthesis of non-natural peptide and depsipeptide macrocycles. *Nat. Chem.* **15**, 61–69 (2023). doi:10.1038/s41557-022-01082-0 Medline
21. F. J. Isaacs, P. A. Carr, H. H. Wang, M. J. Lajoie, B. Sterling, L. Kraal, A. C. Tolonen, T. A. Gianoulis, D. B. Goodman, N. B. Reppas, C. J. Emig, D. Bang, S. J. Hwang, M. C. Jewett, J. M. Jacobson, G. M. Church, Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333**, 348–353 (2011). doi:10.1126/science.1205822 Medline
22. F. R. Blattner, G. Plunkett III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, Y. Shao, The complete genome sequence of Escherichia coli K-12. *Science* **277**, 1453–1462 (1997). doi:10.1126/science.277.5331.1453 Medline
23. O. W. Jones Jr., M. W. Nirenberg, Degeneracy in the amino acid code. *Biochim. Biophys. Acta* **119**, 400–406 (1966). doi:10.1016/0005-2787(66)90198-5 Medline
24. K. Wang, J. Fredens, S. F. Brunner, S. H. Kim, T. Chia, J. W. Chin, Defining synonymous codon compression schemes by genome recoding. *Nature* **539**, 59–64 (2016). doi:10.1038/nature20124 Medline
25. J. F. Zürcher, A. A. Kleefeldt, L. F. H. Funke, J. Birnbaum, J. Fredens, S. Grazioli, K. C. Liu, M. Spinck, G. Petris, P. Murat, F. B. H. Rehm, J. E. Sale, J. W. Chin, Continuous synthesis of *E. coli* genome sections and Mb-scale human DNA assembly. *Nature* **619**, 555–562 (2023). doi:10.1038/s41586-023-06268-1 Medline
26. Y. H. Lau, F. Stirling, J. Kuo, M. A. P. Karrenbelt, Y. A. Chan, A. Riesselman, C. A. Horton, E. Schäfer, D. Lips, M. T. Weinstock, D. G. Gibson, J. C. Way, P. A. Silver, Large-scale recoding of a bacterial genome by iterative recombineering of synthetic DNA. *Nucleic Acids Res.* **45**, 6971–6980 (2017). doi:10.1093/nar/gkx415 Medline
27. L. Zhong, Q. Zhang, N. Lu, T. Wang, X. Xue, Z. Qin, The conjugation-associated linear-BAC iterative assembling (CALBIA) method for cloning 2.1-Mb human chromosomal DNAs in bacteria. *Cell Res.* **35**, 309–312 (2025). doi:10.1038/s41422-024-01063-7 Medline
28. D. G. Gibson, J. I. Glass, C. Lartigue, V. N. Noskov, R.-Y. Chuang, M. A. Algire, G. A. Benders, M. G. Montague, L. Ma, M. M. Moodie, C. Merryman, S. Vashee, R. Krishnakumar, N. Assad-Garcia, C. Andrews-Pfannkoch, E. A. Denisova, L. Young, Z.-Q. Qi, T. H. Segall-Shapiro, C. H. Calvey, P. P. Parmar, C. A. Hutchison III, H. O. Smith, J. C. Venter, Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52–56 (2010). doi:10.1126/science.1190719 Medline
29. J. Fredens, K. Wang, D. de la Torre, L. F. H. Funke, W. E. Robertson, Y. Christova, T. Chia, W. H. Schmied, D. L. Dunkelmann, V. Beránek, C. Uttamapinant, A. G. Llamazares, T. S. Elliott, J. W. Chin, Total synthesis of *Escherichia coli* with a recoded genome. *Nature* **569**, 514–518 (2019). doi:10.1038/s41586-019-1192-5 Medline
30. C. A. Hutchison III, R.-Y. Chuang, V. N. Noskov, N. Assad-Garcia, T. J. Deerinck, M. H. Ellisman, J. Gill, K. Kannan, B. J. Karas, L. Ma, J. F. Pelletier, Z.-Q. Qi, R. A. Richter, E. A. Strychalski, L. Sun, Y. Suzuki, B. Tsvetanova, K. S. Wise, H. O. Smith, J. I. Glass, C. Merryman, D. G. Gibson, J. C. Venter, Design and synthesis of a minimal bacterial genome. *Science* **351**, aad6253 (2016). doi:10.1126/science.aad6253 Medline
31. S. M. Richardson, L. A. Mitchell, G. Stracquadanio, K. Yang, J. S. Dymond, J. E. DiCarlo, D. Lee, C. L. V. Huang, S. Chandrasegaran, Y. Cai, J. D. Boeke, J. S. Bader, Design of a synthetic yeast genome. *Science* **355**, 1040–1044 (2017). doi:10.1126/science.aaf4557 Medline
32. J. S. Dymond, S. M. Richardson, C. E. Coombes, T. Babatz, H. Muller, N. Annaluru, W. J. Blake, J. W. Schwerzmann, J. Dai, D. L. Lindstrom, A. C. Boeke, D. E. Gottschling, S. Chandrasegaran, J. S. Bader, J. D. Boeke, Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471–476 (2011). doi:10.1038/nature10403 Medline
33. A. Nyerges, A. Chiappino-Pepe, B. Budnik, M. Baas-Thomas, R. Flynn, S. Yan, N. Ostrov, M. Liu, M. Wang, Q. Zheng, F. Hu, K. Chen, A. Rudolph, D. Chen, J. Ahn, O. Spencer, V. Ayalavarapu, A. Tarver, M. Harmon-Smith, M. Hamilton, I. Blaby, Y. Yoshikuni, B. Hajian, A. Jin, B. Kintses, M. Szamel, V. Seregi, Y. Shen, Z. Li, G. M. Church, Synthetic genomes unveil the effects of synonymous recoding. bioRxiv 2024.06.16.599206 [Preprint] (2024); https://doi.org/10.1101/2024.06.16.599206.
34. J. E. Venetz, L. Del Medico, A. Wölfle, P. Schächle, Y. Bucher, D. Appert, F. Tschan, C. E. Flores-Tinoco, M. van Kooten, R. Guennoun, S. Deutsch, M. Christen, B. Christen, Chemical synthesis rewriting of a bacterial genome to achieve design flexibility and biological functionality. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 8070–8079 (2019). doi:10.1073/pnas.1818259116 Medline
35. N. Ostrov, M. Landon, M. Guell, G. Kuznetsov, J. Teramoto, N. Cervantes, M. Zhou, K. Singh, M. G. Napolitano, M. Moosburner, E. Shrock, B. W. Pruitt, N. Conway, D. B. Goodman, C. L. Gardner, G. Tyree, A. Gonzales, B. L. Wanner, J. E. Norville, M. J. Lajoie, G. M. Church, Design, synthesis, and testing toward a 57-codon genome. *Science* **353**, 819–822 (2016). doi:10.1126/science.aaf3639 Medline
36. J. Koeppel, R. Ferreira, T. Vanderstichele, L. M. Riedmayr, E. M. Peets, G. Girling, J. Weller, P. Murat, F. G. Liberante, T. Ellis, G. M. D. Church, L. Parts, Randomizing the human genome by engineering recombination between repeat elements. *Science* **387**, eado3979 (2025). doi:10.1126/science.ado3979 Medline
37. K. Wang, D. de la Torre, W. E. Robertson, J. W. Chin, Programmed chromosome fission and fusion enable precise large-scale genome rearrangement and assembly. *Science* **365**, 922–926 (2019). doi:10.1126/science.aay0737 Medline
38. Y. Shao, N. Lu, Z. Wu, C. Cai, S. Wang, L.-L. Zhang, F. Zhou, S. Xiao, L. Liu, X. Zeng, H. Zheng, C. Yang, Z. Zhao, G. Zhao, J.-Q. Zhou, X. Xue, Z. Qin, Creating a functional single-chromosome yeast. *Nature* **560**, 331–335 (2018).

doi:10.1038/s41586-018-0382-x Medline
39. S. Pinglay, J.-B. Lalanne, R. M. Daza, S. Kottapalli, F. Quaisar, J. Koeppel, R. K. Garge, X. Li, D. S. Lee, J. Shendure, Multiplex generation and single-cell analysis of structural variants in mammalian genomes. *Science* **387**, eado5978 (2025). doi:10.1126/science.ado5978 Medline
40. W. E. Robertson, L. F. H. Funke, D. de la Torre, J. Fredens, K. Wang, J. W. Chin, Creating custom synthetic genomes in *Escherichia coli* with REXER and GENESIS. *Nat. Protoc.* **16**, 2345–2380 (2021). doi:10.1038/s41596-020-00464-3 Medline
41. P. D. Karp, S. Paley, R. Caspi, A. Kothari, M. Krummenacker, P. E. Midford, L. R. Moore, P. Subhraveti, S. Gama-Castro, V. H. Tierrafria, P. Lara, L. Muñiz-Rascado, C. Bonavides-Martinez, A. Santos-Zavaleta, A. Mackie, G. Sun, T. A. Ahn-Horst, H. Choi, M. W. Covert, J. Collado-Vides, I. Paulsen, The EcoCyc Database (2023). *Ecosal Plus* **11**, eesp00022023 (2023). doi:10.1128/ecosalplus.esp-0002-2023 Medline
42. G. Cambray, J. C. Guimaraes, A. P. Arkin, Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in *Escherichia coli*. *Nat. Biotechnol.* **36**, 1005–1015 (2018). doi:10.1038/nbt.4238 Medline
43. A. Espah Borujeni, D. Cetnar, I. Farasat, A. Smith, N. Lundgren, H. M. Salis, Precise quantification of translation inhibition by mRNA structures that overlap with the ribosomal footprint in N-terminal coding sequences. *Nucleic Acids Res.* **45**, 5437–5448 (2017). doi:10.1093/nar/gkx061 Medline
44. S. Bhattacharyya, W. M. Jacobs, B. V. Adkar, J. Yan, W. Zhang, E. I. Shakhnovich, Accessibility of the Shine-Dalgarno sequence dictates N-terminal codon bias in *E. coli*. *Mol. Cell* **70**, 894–905.e5 (2018). doi:10.1016/j.molcel.2018.05.008 Medline
45. R. Tian, Y. Liu, J. Chen, J. Li, L. Liu, G. Du, J. Chen, Synthetic N-terminal coding sequences for fine-tuning gene expression and metabolic engineering in *Bacillus subtilis*. *Metab. Eng.* **55**, 131–141 (2019). doi:10.1016/j.ymben.2019.07.001 Medline
46. J. Lederberg, Gene recombination and linked segregations in *Escherichia Coli*. *Genetics* **32**, 505–525 (1947). doi:10.1093/genetics/32.5.505 Medline
47. B. J. Bachmann, K. B. Low, A. L. Taylor, Recalibrated linkage map of *Escherichia coli* K-12. *Bacteriol. Rev.* **40**, 116–167 (1976). doi:10.1128/br.40.1.116-167.1976 Medline
48. J. Lederberg, E. L. Tatum, Gene recombination in *Escherichia coli*. *Nature* **158**, 558 (1946). doi:10.1038/158558a0 Medline
49. T. M. Wannier, A. M. Kunjapur, D. P. Rice, M. J. McDonald, M. M. Desai, G. M. Church, Adaptive evolution of genomically recoded *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 3090–3095 (2018). doi:10.1073/pnas.1715530115 Medline
50. A. Zaslaver, A. E. Mayo, R. Rosenberg, P. Bashkin, H. Sberro, M. Tsalyuk, M. G. Surette, U. Alon, Just-in-time transcription program in metabolic pathways. *Nat. Genet.* **36**, 486–491 (2004). doi:10.1038/ng1348 Medline
51. S. Pundir, M. J. Martin, C. O'Donovan, UniProt Protein Knowledgebase. *Methods Mol. Biol.* **1558**, 41–55 (2017). doi:10.1007/978-1-4939-6783-4_2 Medline
52. N. Kouprina, V. N. Noskov, M. Koriabine, S. H. Leem, V. Larionov, Exploring transformation-associated recombination cloning for selective isolation of genomic regions. *Methods Mol. Biol.* **255**, 69–89 (2004). doi:10.1385/1-59259-752-1:069 Medline
53. C. R. Reisch, K. L. Prather, The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in *Escherichia coli*. *Sci. Rep.* **5**, 15096 (2015). doi:10.1038/srep15096 Medline
54. Q. Li, B. Sun, J. Chen, Y. Zhang, Y. Jiang, S. Yang, A modified pCas/pTargetF system for CRISPR-Cas9-assisted genome editing in *Escherichia coli*. *Acta Biochim. Biophys. Sin. (Shanghai)* **53**, 620–627 (2021). doi:10.1093/abbs/gmab036 Medline
55. T. M. Wannier, A. Nyerges, H. M. Kuchwara, M. Czikkely, D. Balogh, G. T. Filsinger, N. C. Borders, C. J. Gregg, M. J. Lajoie, X. Rios, C. Pál, G. M. Church, Improved bacterial recombineering by parallelized protein discovery. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 13689–13698 (2020). doi:10.1073/pnas.2001588117 Medline
56. M. G. Schubert, D. B. Goodman, T. M. Wannier, D. Kaur, F. Farzadfard, T. K. Lu, S. L. Shipman, G. M. Church, High-throughput functional variant screens via in vivo production of single-stranded DNA. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2018181118 (2021). doi:10.1073/pnas.2018181118 Medline
57. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). doi:10.1038/nmeth.1923 Medline
58. P. Rice, I. Longden, A. Bleasby, EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000). doi:10.1016/S0168-9525(00)02024-2 Medline
59. S. Anders, P. T. Pyl, W. Huber, HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015). doi:10.1093/bioinformatics/btu638 Medline
60. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014). doi:10.1186/s13059-014-0550-8 Medline
61. E. C. A. Goodall, A. Robinson, I. G. Johnston, S. Jabbari, K. A. Turner, A. F. Cunningham, P. A. Lund, J. A. Cole, I. R. Henderson, The essential genome of *Escherichia coli* K-12. *mBio* **9**, e02096-17 (2018). doi:10.1128/mBio.02096-17 Medline
62. N. Costantino, D. L. Court, Enhanced levels of λ Red-mediated recombinants in mismatch repair mutants. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15748–15753 (2003). doi:10.1073/pnas.2434959100 Medline
63. A. Aronshtam, M. G. Marinus, Dominant negative mutator mutations in the mutL gene of *Escherichia coli*. *Nucleic Acids Res.* **24**, 2498–2504 (1996). doi:10.1093/nar/24.13.2498 Medline
64. F. S. Midani, J. Collins, R. A. Britton, AMiGA: Software for Automated Analysis of Microbial Growth Assays. *mSystems* **6**, e0050821 (2021). doi:10.1128/msystems.00508-21 Medline
65. K. C. Liu, Y. Gu, C. D. Day, J. W. Chin, Scripts for *Escherichia coli* with a 57-codon genetic code. Zenodo (2025): https://doi.org/10.5281/zenodo.15753696.
66. J. D. Heck, G. W. Hatfield, Valyl-tRNA synthetase gene of *Escherichia coli* K12. Molecular genetic characterization. *J. Biol. Chem.* **263**, 857–867 (1988). doi:10.1016/S0021-9258(19)35433-X Medline
67. H. Erhardt, S. Steimle, V. Muders, T. Pohl, J. Walter, T. Friedrich, Disruption of individual *nuo*-genes leads to the formation of partially assembled NADH:ubiquinone oxidoreductase (complex I) in *Escherichia coli*. *Biochim. Biophys. Acta* **1817**, 863–871 (2012). doi:10.1016/j.bbabio.2011.10.008 Medline
68. P. E. Lavery, S. C. Kowalczykowski, Biochemical basis of the constitutive repressor cleavage activity of recA730 protein. A comparison to recA441 and recA803 proteins. *J. Biol. Chem.* **267**, 20648–20658 (1992). doi:10.1016/S0021-9258(19)36735-3 Medline
69. A. V. Semerjian, D. C. Malloy, A. R. Poteete, Genetic structure of the bacteriophage P22 PL operon. *J. Mol. Biol.* **207**, 1–13 (1989). doi:10.1016/0022-2836(89)90437-3 Medline

ACKNOWLEDGMENTS

We thank Steven Wingett for assistance with Nextflow, Faye Rodgers for assistance with RNA-seq, and Simona Grazioli for assistance with sequencing. We thank Martyn Howard and Mark Cussens from the LMB Media Kitchen. **Funding:** This work was supported by the Medical Research Council (MRC), UK (MC_U105181009 and MC_UP_A024_1008) an ERC Advanced Grant SGCR, and a Wellcome Trust Investigator Award 220808/Z/20/Z, all to J.W.C. M.S. was funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) SP 1981/1-1 (project no. 493404643). F.B.H.R. was supported by a UK Research and Innovation (UKRI) Marie Skłodowska-Curie Actions (MSCA) guarantee fellowship (EP/Y014154/1) and an Investigator Grant (GNT2018461) from the National Health and Medical Research Council (NHMRC) Australia. Y.G. was supported by an EMBO fellowship (ALTF 93-2023). A.A.K. was supported by the Boehringer Ingelheim Fonds and the Cambridge Commonwealth, European and International Trust. R.L.S. was supported by the Harding Distinguished Postgraduate Scholars Programme. **Author contributions:** W.E.R. tested recoding schemes and generated the initial genome design. W.E.R., F.B.H.R., M.S., R.L.S., R.T., W.L., Y.G., A.A.K., K.C.L., J.F.Z., J.B., L.vB. assembled BACs and/or performed uREXER experiments. R.T. directed the use of N- terminal coding sequence libraries. W.E.R., F.B.H.R., M.S., R.L.S., R.T., W.L., Y.G. fixed non-recoded regions in individual fragments. M.S. developed and directed genomic integrations and refinement of defective sections, and linkage mapping for longer recoded sections. M.S., W.L., A.A.K performed mapping experiments. R.T. recoded the 27B fragment. W.E.R., R.L.S., C.F.D., K.C.L. coordinated RNA-seq experiments. F.B.H.R. developed and directed oligo-based evolution for fitness improvement. F.B.H.R., M.S., A.A.K., C.F.D. performed oligo-based

evolutions. W.E.R., F.B.H.R., M.S., R.L.S., R.T., C.F.D. performed conjugative assemblies of sections. R.L.S., Y.G., C.F.D., K.C.L., F.L.B. wrote and/or refined scripts for data analysis. F.B.H.R., M.S., R.L.S., A.A.K., Y.C. performed growth measurements. J.W.C. supervised the project. J.W.C. wrote the manuscript with input from W.E.R., F.B.H.R., M.S., R.L.S., R.T. **Competing interests:** J.W.C. and J.F.Z. have financial interests in Constructive Bio. J.F.Z. and K.C.L. consult for Constructive Bio. The MRC has filed a provisional patent application (GB2506688.7) related to this work on which W.E.R., F.B.H.R., M.S., R.L.S., R.T., and J.W.C are listed as inventors. **Data and materials availability:** All data are available in the main text or supplementary materials. The computer code for analyzing recoding landscapes, for generating linkage maps, and for RNA-seq analyses are available in Zenodo (*65*). The authors agree to provide any materials and strains used in this study upon request. This research was funded in whole or in part by UK Research and Innovation (MC_U105181009 and MC_UP_A024_1008) and the Wellcome Trust (220808/Z/20/Z), cOAlition S organizations. The author will make the Author Accepted Manuscript (AAM) version available under a CC BY public copyright license.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ady4368
Figs. S1 to S31
References (*66*–*69*)
MDAR Reproducibility Checklist
Data S1 to S10

Submitted 22 April 2025; accepted 18 July 2025
Published online 31 July 2025
10.1126/science.ady4368

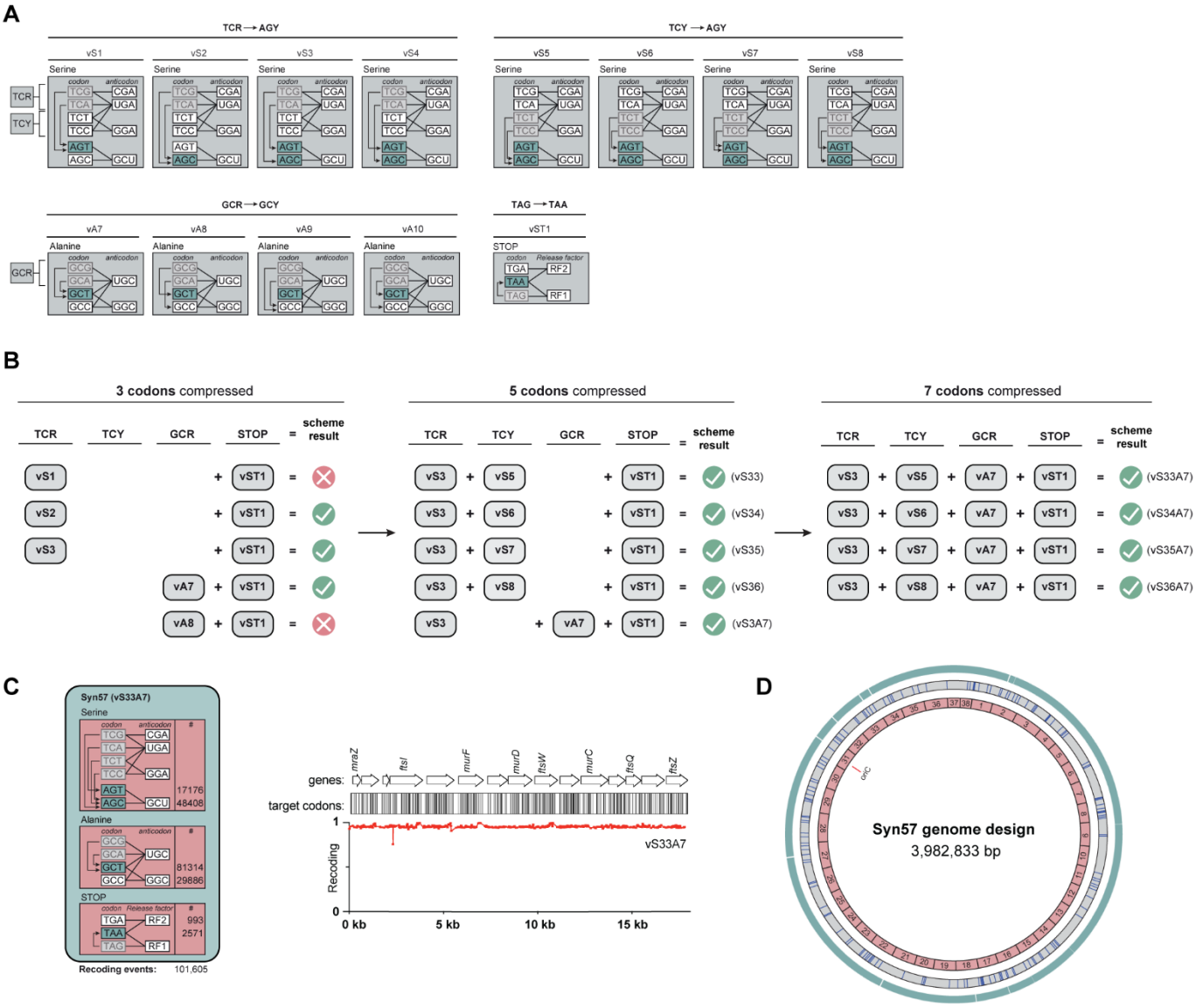

Fig. 1. Combining codon compression schemes to yield a 57-codon genome design. (**A**) Target codons to remove from the genetic code (grey boxes) were chosen from the serine and alanine sense codon boxes and the stop codon box. The recoding scheme determines which synonym (teal boxes) will replace the target codon. As input recoding schemes, we focused on the TCR and TCY codon boxes from serine, the GCR codon box from alanine, and the STOP codon box. We focused on building 3, 5 and 7 codon compression schemes by combining the indicated schemes. (**B**) The 3-codon compression schemes vS1, vS2, vS3, vA7, vA8 were previously tested using REXER (*24*), and the results are summarized here. To further compress the code by 5 codons, we combined 3-codon compression schemes targeting the TCR serine box with schemes targeting the TCY serine box (TCR + TCY) or the GCR alanine box (TCR + GCR) – this generated five different 5-codon compression schemes (vS33-vS36, and vS3A7). To remove 7 codons from the genetic code, we combined successful compression schemes from the TCR, TCY, GCR, and STOP codon boxes. Specifically, we combined the four successful TCR + TCY schemes with the GCR recoding scheme and the STOP recoding scheme – this generated four individual 7-codon compression schemes (vS33A7-vS36A7), all recoding 4 serine codons, 2 alanine codons, and 1 stop codon to different combinations of synonyms. We performed REXER with each of the indicated 3, 5 and 7 codon compression schemes (fig. S2). REXERs that led to recoding are indicated with a green tick, those that did not lead to recoding are indicated with a red X. (**C**) Implementing one of the 7-codon compression schemes, vS33A7, across the entire 4-megabase (Mb) genome sequence of *E. coli* MDS42 generates a recoded genome design with a 57-codon genetic code. The numbers adjacent to the codons indicate the total numbers of those codons in the designed, recoded genome. The recoding landscape following REXER with the vS33A7 scheme shows full recoding across the 20-kb *dcw* cluster. (**D**) Map of the synthetic 57-codon genome design with all TCG, TCA, TCC, TCT, GCA, GCG, and TAG codons removed. The outer ring (101,605 teal bars, outer wheel) indicates the position of all recoding events, while the blue bars (169) indicate refactoring events (grey middle wheel). The synthetic genome design was disconnected into 38 fragments of 100 kb each (pink inner wheel).

Fig. 2. Recoding every 100-kb fragment in the genome. (**A**) The genome design was split into 38 fragments, each fragment spans approximately 100 kb. The corresponding fragments on the parent genome (left) are shown in grey. Parallel uREXER experiments in wildtype cells were performed for each fragment with BACs containing synthetic DNA inserts that implement the corresponding genome design. 27 fragments (pink, middle) were fully recoded, while 11 fragments (grey) were either not recoded (100k27) by uREXER or only recoded across a part of their sequence. The ten partially recoded fragments were then further examined individually and the regions that were not replaced with the original synthetic genome design were identified, refined, and recoded according to one of four fixing strategies. This led to the recoding of all 38 fragments (right), with fragment 27 recoded across two sub-fragments (27A and 27B). (**B**) For each of the 11 100-kb fragments with refined regions, the sequence that was recoded with synthetic DNA is shown in pink, and the region that remained non-recoded is shown in grey. The total length of non-recoded (n.r.) sequence within each fragment is indicated. Non-recoded regions were recoded using the fixing strategies indicated by a black dot: N-terminal coding sequence libraries (NCS), application of an alternative recoding scheme (alt.), rational fixing of regulatory modules (reg.) (e.g., promoters), or adjusting the refactoring length for previously overlapping genes (ref.). 100k27 was split into the sub-fragments 27A (left pink bar) and 27B (right pink and grey bar).

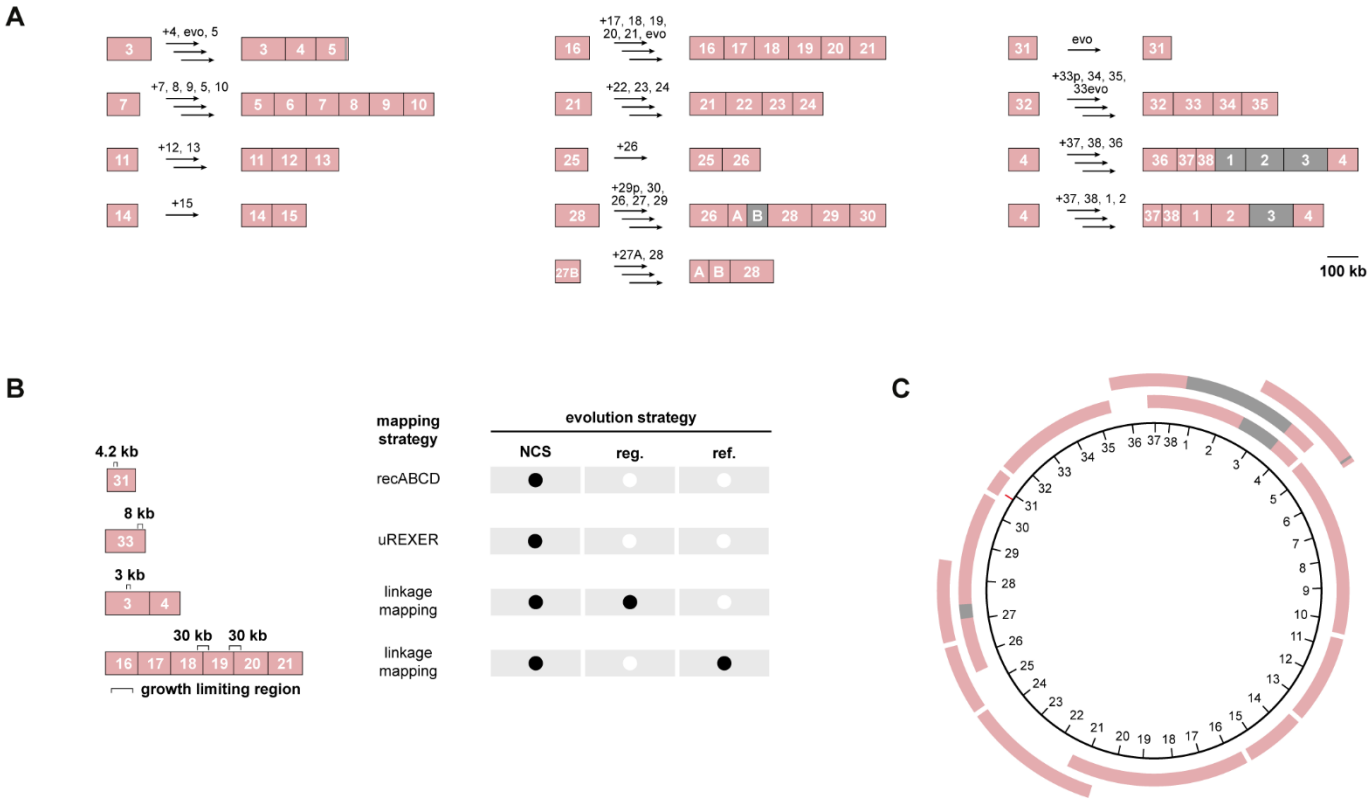

Fig. 3. Assembly and fitness improvement of larger recoded genomic sections. (**A**) The GENESIS (iterative uREXER) steps used to assemble larger recoded sections and the directed evolution steps used to improve fitness. The fragment number used in each step and directed evolution (evo) steps are indicated above the arrows, starting from individually recoded sections. Recoded sections are depicted in pink, non-recoded sections are depicted in grey, an apostrophe indicates evolved sections. The scalebar indicates 100 kb of genomic DNA. (**B**) We evolved four strains with recoded fragments or sections before their use in conjugative assembly. We utilized three different mapping strategies to define growth-limiting recoded regions as targets for evolution (figs. S10, S19, S21, and S22). The location and size (in kb) of the growth-limiting region is shown as a black bracket. The growth-limiting regions were targeted using the evolution strategies indicated by a black dot: N-terminal coding sequence libraries (NCS), rational fixing of regulatory modules (reg.) (e.g., promoters), or adjusting the refactoring length for previously overlapping genes (ref.) (figs. S18, S19, S21, and S22). (**C**) The recoded sections in the thirteen strains generated by GENESIS tiles the Syn57 genome as shown by the pink bars, tick marks show the genome fragment number.

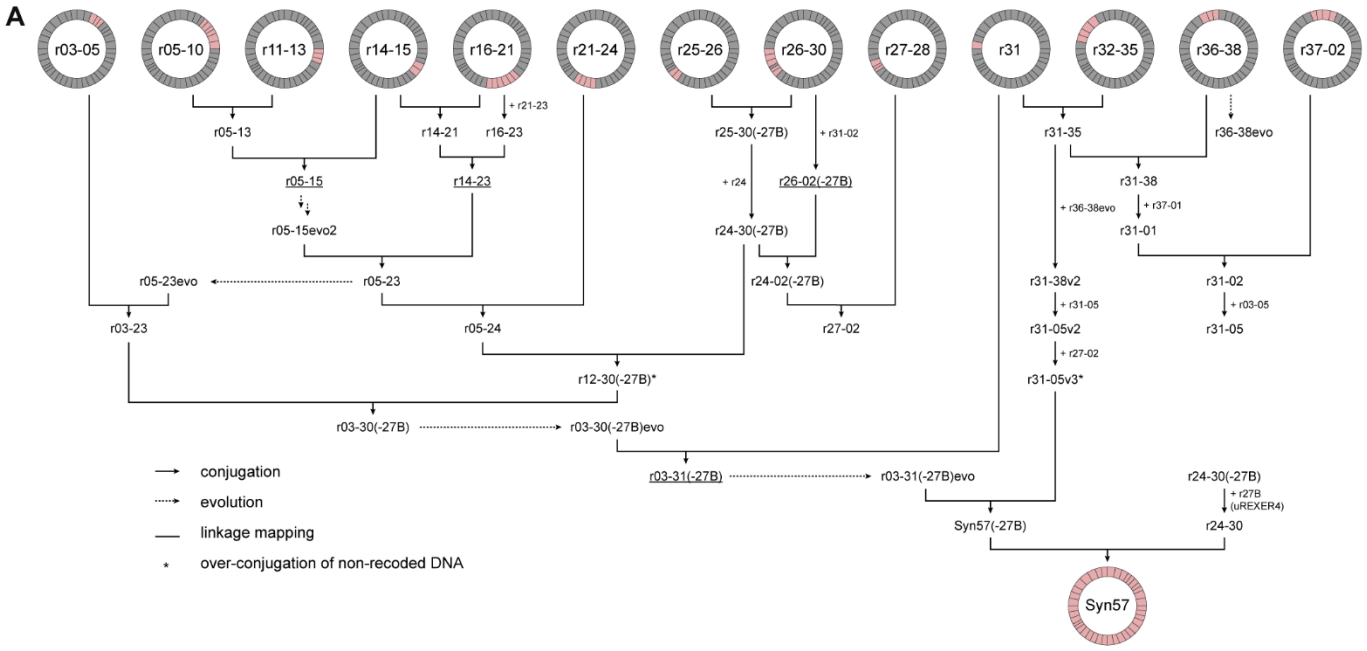

Fig. 4. Total genome assembly to create Syn57. (**A**) Starting with the set of 13 partially synthetic genome strains (shown in Fig. 3), we assembled a 57-codon genome into a single fully recoded strain using the indicated sequence of conjugations (black arrows). Donor and recipient strains contain unique recoded genomic sections (denoted in pink), with the donor stain harbouring a recoded genomic section downstream of the recoded genomic section in the recipient strain. Conjugation proceeded with interspersed steps of linkage mapping (underlined sections) and evolution (dotted arrows) to optimize strain fitness prior to downstream genome assembly steps, which ultimately yielded the final fully recoded strain Syn57. Changes present in the final genome relative to our initial design are depicted in fig. S30. A notable adaption occurred during the assembly of r27-02 - all recoded clones obtained had a duplication of a region containing *tufB* in r33. We hypothesized that this arose due the requirement to simultaneously recode the two gene copies encoding for EF-Tu, *tufA* (in r27B) and *tufB*. Over-conjugation during efforts to generate r27-05, via conjugation of a r31-05v2 donor on a r27-02 recipient, yielded r31-05v3 containing the duplication within r33. We then utilized r31-05v3 to generate Syn57(-27B) carrying the *tufB* duplication in preparation for the final conjugation step to transfer r27B. Details of all evolution edits and conjugative assembly steps are provided in data S5 and S6, respectively. The final sequence of the Syn57 genome is provided in data S8. (**B**) Overnight cultures of WT, r03-31(-27B)evo, r31-05v3, and Syn57 were normalized to OD_{600} 0.5 and spotted, in triplicate, on 2xYT solid medium and incubated at 37°C. Images of the plates were taken at the indicated timepoints. All plate images, and additional conditions, are shown in fig. S31A. (**C**) Doubling time (WT: 1.14 ± 0.03 hours; Syn57: 4.60 ± 0.31 hours) and maximum cell density (WT: 6.31 ± 0.028; Syn57: 1.03 ± 0.08) values were derived from the triplicate growth measurements of WT and Syn57 cells shown in fig. S31B, error bars indicate the standard deviation.