

# Large-scale recoding of a bacterial genome by iterative recombineering of synthetic DNA

Yu Heng Lau<sup>1,2</sup>, Finn Stirling<sup>1,2</sup>, James Kuo<sup>1,2</sup>, Michiel A. P. Karrenbelt<sup>1,3</sup>, Yujia A. Chan<sup>1,2</sup>, Adam Riesselman<sup>4</sup>, Connor A. Horton<sup>1,2</sup>, Elena Schäfer<sup>1,2</sup>, David Lips<sup>1,2</sup>, Matthew T. Weinstock<sup>5</sup>, Daniel G. Gibson<sup>5,6</sup>, Jeffrey C. Way<sup>1,2</sup> and Pamela A. Silver<sup>1,2,\*</sup>

<sup>1</sup>Wyss Institute for Biologically Inspired Engineering, Harvard University, 3 Blackfan Circle, 5th Floor, Boston, MA 02115, USA, <sup>2</sup>Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Alpert 536, Boston, MA 02115, USA, <sup>3</sup>Systems and Synthetic Biology, Wageningen University, Dreijenplein 10, 6703 HB Wageningen, The Netherlands, <sup>4</sup>Program in Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA, <sup>5</sup>Synthetic Genomics, Inc., 11149 North Torrey Pines Road, La Jolla, CA 92037, USA and <sup>6</sup>Synthetic Biology and Bioenergy Group, J. Craig Venter Institute, 4120 Capricorn Lane, La Jolla, CA 92037, USA

Received March 15, 2017; Revised April 26, 2017; Editorial Decision April 28, 2017; Accepted May 02, 2017

## ABSTRACT

The ability to rewrite large stretches of genomic DNA enables the creation of new organisms with customized functions. However, few methods currently exist for accumulating such widespread genomic changes in a single organism. In this study, we demonstrate a rapid approach for rewriting bacterial genomes with modified synthetic DNA. We recode 200 kb of the *Salmonella typhimurium* LT2 genome through a process we term SIRCAS (stepwise integration of rolling circle amplified segments), towards constructing an attenuated and genetically isolated bacterial chassis. The SIRCAS process involves direct iterative recombineering of 10–25 kb synthetic DNA constructs which are assembled in yeast and amplified by rolling circle amplification. Using SIRCAS, we create a *Salmonella* with 1557 synonymous leucine codon replacements across 176 genes, the largest number of cumulative recoding changes in a single bacterial strain to date. We demonstrate reproducibility over sixteen two-day cycles of integration and parallelization for hierarchical construction of a synthetic genome by conjugation. The resulting recoded strain grows at a similar rate to the wild-type strain and does not exhibit any major growth defects. This work is the first instance of synthetic bacterial recoding beyond the *Escherichia coli* genome, and reveals that *Salmonella* is remarkably amenable to genome-scale modification.

## INTRODUCTION

The next widely anticipated breakthrough in genetic engineering is the ability to rapidly rewrite the genomes of industrially relevant microbes, plants, and animals. Rewriting entire genomes will deepen our understanding of the genetic code and dramatically transform human health, food and energy production, and our environment (1–6). A major challenge identified by the Genome Project-Write consortium is the efficiency of building and testing large modified genomes (1). In particular, genome recoding involves synonymous replacement of all instances of specific codons throughout an entire genome (2), requiring efficient assembly of large constructs containing thousands of designed base changes (3). New foundational technologies are therefore crucial for accelerating the pace of genome synthesis and modification.

New organisms based on recoded genomes can serve a variety of useful functions (1–3,7). While recoded genomes still encode for the same translation products, the replaced target codons can be repurposed for incorporating non-standard amino acids into proteins and peptides (8–10). Loss of the target codon tRNAs in a recoded organism can also impair its ability to express foreign DNA acquired by horizontal gene transfer or viral infection (2,11). This may enhance the genetic containment of an organism (12,13), a desirable feature when considering safety and stability of engineered organisms for applications in open environments (14).

Several recent examples of modifying bacterial genomes illustrate the challenges associated with the construction process. Church and co-workers reported a piecewise strategy for debugging their design of a 57-codon *Escherichia coli* genome (3). Testing 55 different strains containing independent ~50 kb segments of their recoded genome design

\*To whom correspondence should be addressed. Tel: +1 617 432 6401; Email: pamela.silver@hms.harvard.edu

on plasmids, 44 segments conferred viability. Although the remaining 11 segments contained lethal elements, the piecewise strategy helped to expedite the troubleshooting process by narrowing down the number of possible offending mutations, enabling a design flaw to be fixed in one of the segments. While cryptic ribosome binding sites and changes in mRNA folding have been identified as potentially important factors when recoding (15,16), underlying principles of genome redesign are generally still not well-defined and fixing design flaws remains challenging (17). The empirical nature of recoding is exemplified in recent work by Chin and co-workers, in which recoding of an essential 20 kb region of *E. coli* was attempted (7). Three of the eight different codon replacement schemes gave viable organisms and one lethal scheme could be rescued by fixing a single offending codon, while the remaining four schemes led to inviability. Furthermore, lessons from the *de novo* synthesis of a ‘minimal bacterial genome’ by Venter and co-workers have implications for genome recoding (5). While the authors halved the genome size of *Mycoplasma mycoides* by removing non-essential elements, this required iterative testing to address complexities such as quasi-essential genes and synthetic lethal gene pairs. Similar issues may arise when implementing genome-wide recoding.

Herein, we report a streamlined method for accumulating large-scale recoding in a single strain of *Salmonella enterica* serovar Typhimurium LT2 (*S. typhimurium*), bypassing the reliance of previous methods on site-specific enzymes. *S. typhimurium* is of interest in engineering the gut microbiome for diagnosis, vaccination and treatment of human disease (18–20). An attenuated and recoded *S. typhimurium* could serve as a stable chassis for housing various diagnostic and therapeutic circuits that can autonomously detect and act upon changes in the gut environment (21). In this work, we accumulate 1557 codon changes in 176 genes across 200 kb of the *S. typhimurium* genome (Figure 1), using an iterative process we term ‘stepwise integration of rolling circle amplified segments’ (SIRCAS).

## MATERIALS AND METHODS

### Assembly of synthetic DNA in yeast

Synthetic DNA fragments were either purchased from commercial vendors (Gen9, SGI-DNA and Integrated DNA Technologies), or obtained by automated enzymatic assembly of oligonucleotides (BioXp™ 3200 instrument, SGI-DNA). Full details of construct design are provided in the Supporting Information (Supplementary File SI 1 Section S3).

The DNA assembly method by transformation-associated recombination was adapted from previously reported protocols (22,23). To prepare yeast spheroplasts for transformation, *Saccharomyces cerevisiae* VL6-48 (ATCC® MYA3666™) was grown in 50 ml YPAD medium to an optical density OD<sub>600</sub> of 1. Cells were then suspended in 20 ml of 1 M sorbitol and kept at 4°C for 4 h. Spheroplasts were generated in 20 ml SPE buffer containing 20 µl beta-mercaptoethanol (Sigma-Aldrich) and 250 µg Zymolyase-20T (US Biological) at 30°C. When the OD<sub>600</sub> of cells diluted 1:10 in SPE buffer was 3–4 times the OD<sub>600</sub>

value of cells diluted in 2% SDS in SPE buffer, the cells were washed twice with 50 ml 1 M sorbitol. Cells were then resuspended in 2.5 ml STC buffer. After 20 min, 200 µl of spheroplasts were added to a 50 µl solution of DNA fragments (total ~1 µg DNA, equimolar fragments). After 10 min, 1 ml of freshly prepared 20% PEG 8000 solution was added. After a further 20 min, the spheroplasts were resuspended in 800 µl SOS and incubated at 30°C for 30 min. The recovered cells were then suspended in top agar kept at 55°C and poured onto synthetically defined minus Trp plates. Buffer and media formulations are detailed in the Supporting Information (Supplementary File SI 1 Section S1).

Successful assemblies were identified either by diagnostic digests of RCA products or multiplex PCR across assembly junctions (Supplementary File SI 1 Section S3.6).

### Rolling circle amplification of DNA from yeast artificial chromosomes

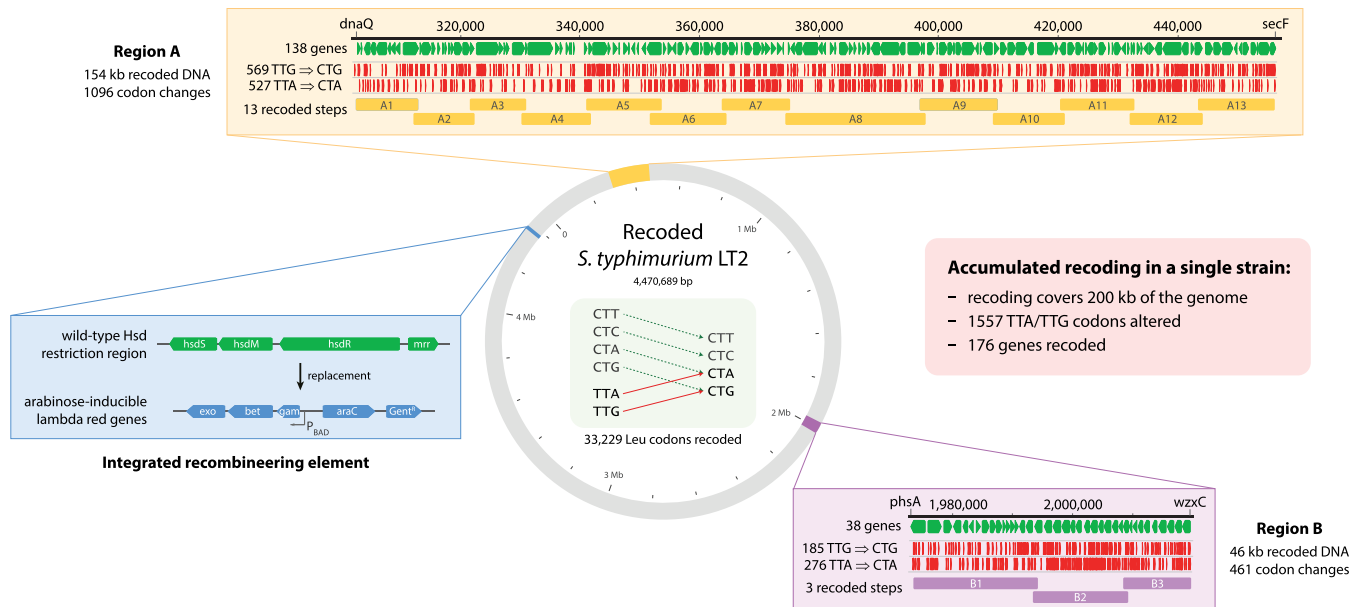
Adapting the protocol by Hutchison *et al.* (5), yeast centromeric plasmid DNA was isolated from *Saccharomyces cerevisiae* using a modified miniprep workflow. Overnight 5 mL cultures of yeast were suspended in 250 µl P1 buffer (Qiagen) containing 250 µg Zymolyase-20T (US Biological). After incubation at 37°C for 1 h, 250 µl P2 buffer and 250 µl P3 buffer (Qiagen) were added sequentially. The plasmid DNA in the supernatant was precipitated with 1 ml isopropanol, washed with 1 ml of 70% ethanol, and the resulting DNA pellet was air-dried. The plasmid DNA was then dissolved in 40 µl Tris–EDTA buffer pH 8.

Rolling circle amplification was conducted using reagents from the TempliPhi Large Construct Kit (GE Healthcare). 2 µl of plasmid DNA solution was added to 15 µl of sample buffer and heated at 95°C for 3 min. Upon cooling, 15 µl of reaction buffer, 1 µl enzyme and 2 µl of 10 mM dNTPs (New England Biolabs) were added, and the reaction was incubated for 24 h at 30°C.

To release the recoded DNA as a linear construct, the amplified DNA was diluted with 40 µl water, then 8 µl of 10× FastDigest buffer and 2 µl LguI restriction enzyme (Thermo Fisher) was added. After 1 h at 37°C, the linearized DNA was precipitated with ethanol and redissolved in 40 µl Tris–EDTA buffer pH 8.

### Integration into the *Salmonella* genome

Overnight cultures of *S. typhimurium* were diluted 1:100 into a 50 ml volume in LB containing 10 mM L-arabinose (Sigma-Aldrich). When an OD<sub>600</sub> of 0.5 was reached, cells were washed twice with 20 ml and once with 1 ml of ice-cold 10% glycerol, then resuspended in 250 µl of 10% glycerol. 50 µl of the competent cell suspension was electroporated with the appropriate DNA construct (100 ng per kb of construct, Bio-Rad MicroPulser using the Ec2 setting) in a 0.2 cm cuvette (USA Scientific). After 5 h outgrowth at 37°C in 1 ml SOC medium, 250 µl of cells was plated onto antibiotic selection plates. Typically, at least 50 colonies were patched onto double antibiotic selection plates (chloramphenicol and kanamycin) to determine which colonies contained the correct markers.



**Figure 1.** Accumulated recoding covering 200 kb of the *S. typhimurium* genome. An *in silico* design of the recoded genome replaced all 33 229 instances of TTA/TTG leucine codons with synonymous CTA/CTG codons respectively. From this design, regions A and B (yellow and purple) were independently recoded using SIRCAS, then combined into one strain by conjugative assembly. The ‘integrated recombinering element’ (inserted at the native *hsd* restriction system of *S. typhimurium*, blue) is an arabinose-inducible lambda red cassette that facilitates recoding by enhancing the homologous replacement of wild-type genomic DNA with synthetic fragments.

See the Supporting Information for further details about the SIRCAS protocol (Supplementary File SI 1, Section S4).

### Conjugative assembly of recoded genomic regions

Donor strains were prepared by integrating a cassette containing spectinomycin resistance and the origin of transfer from RK2 plasmid (24) upstream of recoded region A by recombinering, and transforming plasmid pTA-Mob (25) containing necessary parABCDE (‘partitioning’) genes. Recipient strains were prepared by integrating an ampicillin resistance cassette at the start of the recoding region to be transferred from the donor.

Conjugation experiments were based on a protocol previously described by Ma, Moonan and Isaacs (26). Donor and recipient strains were grown in separate 30 mL cultures, inoculated from overnight cultures by 1:100 dilution. The recipient strain lambda red system was induced with arabinose to enhance homologous recombination. Cultures were grown to  $OD_{600} \sim 0.5$ , centrifuged and resuspended in LB to  $OD_{600} \sim 15$ . Donor and recipient cells were mixed in 80–120  $\mu$ l aliquots and pipetted as 10–20  $\mu$ l spots on LB plates, incubated at 30°C for 1–2 h, then washed and plated as 250  $\mu$ l of a 1:100 dilution on selection plates. Candidates were patched out on multiple selection plates to check for correct phenotype and incubated at 37°C.

### Next-generation sequencing of recoded strains

*Salmonella* genomic DNA was prepared using a DNeasy Blood and Tissue kit (Qiagen) and quantified using a Qubit 2.0 fluorometer (Thermo Fisher). DNA libraries for next-generation sequencing were prepared using a Nextera XT

kit (Illumina) with size selection using SPRIselect beads (Beckman-Coulter) according to manufacturer’s protocols. Sequencing was conducted on a MiSeq platform using either a MiSeq or MiSeq Nano reagent kit v2 (Illumina), running 300 cycles with paired ends. Reads were trimmed and aligned to a reference genome using the Geneious software package (version 9.1.5) (27). Further details can be found in Supplementary File SI 1 Section S6.

### Measurement of recoded *Salmonella* growth rates

Cells were inoculated into 200  $\mu$ l LB in a 96-well clear flat-bottom plate (Corning) and incubated overnight at 37°C with continuous shaking. From these overnight stationary cultures, 1  $\mu$ l was added to 200  $\mu$ l fresh LB in a new 96-well plate, and incubated at 37°C with continuous shaking in a plate reader (BioTek Synergy HT).  $OD_{630}$  measurements were taken every 10 min for at least 16 h. Measurements were performed in technical triplicates, and at least three biological replicates were obtained for each sample. Errors are calculated as the standard error of the mean between the biological replicates.

Doubling times were determined in a similar manner to Lajoie *et al.* (2), based on linear regression of  $\ln(OD_{630})$  using five adjacent timepoints (40 min). The doubling time was calculated by  $t_d = \ln(2)/m$ , where  $m$  is the maximum gradient of  $\ln(OD_{630})$  as determined by the linear regression analysis. Additional growth data can be found in Supplementary File SI 1, Section S7.1.

For strains B2, B3 and A13-B3, cells were inoculated into 5 ml LB cultures overnight. From these overnight stationary cultures, 250  $\mu$ l was added to 25 ml fresh LB in a 250 ml unbaffled Erlenmeyer flask, and grown in a Multitron Standard (INFORS HT) shaking incubator at 37°C and 200

rpm. OD<sub>600</sub> measurements were taken every 15–25 min during early exponential phase on an Ultrospec 10 (Amersham Biosciences), removing an aliquot of 700 µl each time. Doubling times were calculated using the same procedure except using three adjacent timepoints which had OD values lying between 0.05 and 0.75. Colony forming unit (CFU) measurements were performed by counting colonies from appropriate dilutions on LB plates.

As a control, growth curves for wild-type LT2 were obtained in both plate reader and Erlenmeyer flask format, and the doubling times were found to be consistent with each other.

## RESULTS

### Computational design of a leucine-recoded *S. typhimurium* genome

A fully-recoded *S. typhimurium* genome was generated computationally (Figure 1 and Supplementary Figure SI 1 Section S2), replacing all 33 229 instances of leucine codons TTA and TTG in open reading frames (28) with synonymous CTA and CTG codons respectively. The decision to target leucine codons was due to their high frequency of occurrence throughout the genome (Supplementary File SI 1, Section S2.2), and similar to previous recoding efforts (3,7), the two specific codons were chosen because their corresponding tRNA anticodons are not involved in decoding the remaining four leucine codons (29). This recoding scheme also minimizes the number of base pair changes, requiring only a single T to C base change for each codon. In cases where target codons were present in overlapping reading frames, synonymous changes in both proteins were made where possible. In 49 instances where synonymous mutations were not possible, point mutations were chosen to minimize the impact on biological function as predicted by PROVEAN (30), an alignment-based tool for predicting the effect of amino acid substitutions and indels (Supplementary File SI 1 Table S2.3.1). There was one specific instance in which overlapping genes were split due to the presence of four target codons within the overlap (STM0521 and STM0522, see Supplementary File SI 1 Section S2.3). In addition to codon changes, ~400 kb of pseudogenes, mobile elements and pathogenicity islands were removed to reduce genetic instability, genome size and pathogenicity (28) (Supplementary File SI 1 Table S2.4.1). Finally, 754 instances of the recognition site for restriction enzyme LguI (5'-GCTCTTC/GAAGAGC-3') were removed to facilitate downstream cloning.

### Construction of an *S. typhimurium* recombineering strain

We used a recombineering approach based on the lambda red system for increasing integration efficiency (31–33). Recombineering is routinely used in *S. typhimurium* and other bacteria for small insertions and knockouts (34), but the efficiency decreases with increasing DNA insert size (35). Although the use of endonucleases in tandem with recombineering has been shown to improve integration efficiency (7 kb with I-SceI (35) and 100 kb insertions with Cas9 (7)), we postulated that these extra accessory proteins are not necessary to achieve practical rates of integration sufficient for

genome recoding, even when working with large DNA segments.

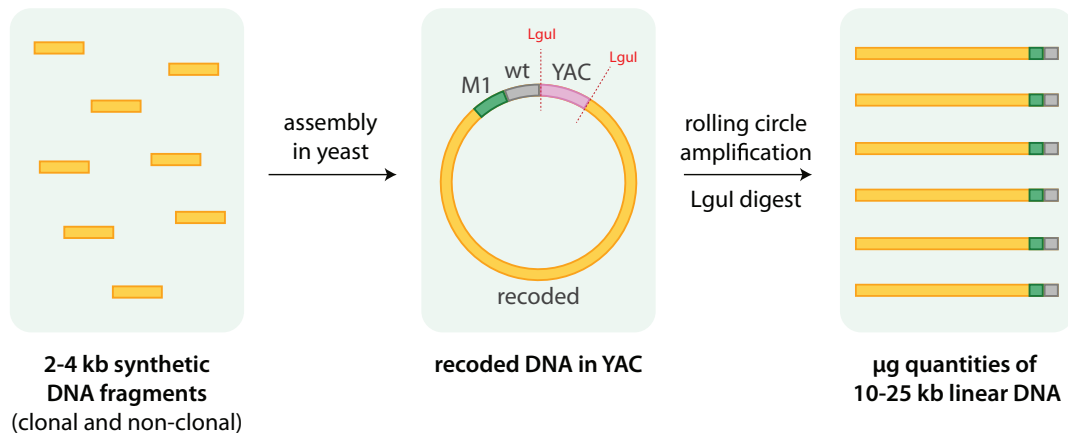
To generate a stable recombineering strain of *S. typhimurium*, we generated a construct containing the lambda red genes under arabinose-inducible control, a gentamicin resistance cassette, and homology arms for replacing the *hsd* region of the genome (Figure 1 and Supplementary File SI 1, Section S3.5). Using an *S. typhimurium* strain containing the recombineering plasmid pKD46 (32,36), the P<sub>ara</sub>-lambda red cassette was successfully integrated into the *hsd* locus, simultaneously removing the native *hsd* restriction system (37) which could otherwise impede transformation. Curing the pKD46 plasmid at elevated temperature resulted in the final restriction-deficient strain containing an integrated recombineering element (IRE).

### Preparation of large recoded DNA segments via rolling circle amplification

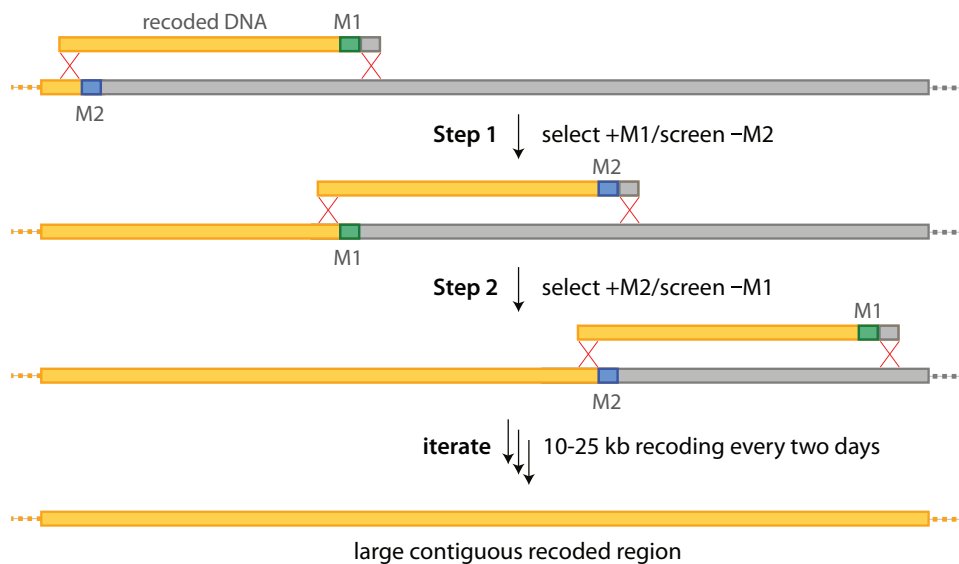
From the *in silico* genome design, sixteen recoded segments were constructed (A1–A13 and B1–B3, Figure 1 and Supplementary File SI 3) constituting two separate arbitrary regions of the recoded genome (Regions A and B). Each segment contained 10–25 kb of recoded DNA, a selection marker, and 1 kb flanking homology regions for integration. The 10–25 kb size range was chosen to simplify construction, decrease the likelihood of unwanted internal recombination events, and minimize the cost of fixing an error in any one segment. Each segment was assembled in yeast (23) from commercially synthesized 2–4 kb DNA fragments (Figure 2A). The cloned assemblies in yeast were checked by diagnostic digests, but were otherwise used immediately without further sequencing. To assess the utility of different commercial DNA sources, a mixture of 156 kb of clonal sequence-verified DNA and 44 kb of non-clonal DNA was used for constructing the segments (Supplementary File SI 2, Table S1). Clonal DNA refers to sequence-verified constructs propagated through a bacterial host, while non-clonal DNA constructs are synthesized completely *in vitro* in a faster and cheaper process to generate a heterogeneous pool containing errors derived from chemical synthesis. SIRCAS uses a marker swapping approach alternating between chloramphenicol and kanamycin selection (Figure 2B) for a simple phenotypic readout, similar to the strategy for building chromosomes in *S. cerevisiae* (4,6) and combining genomes in *B. subtilis* (38).

We bypassed the use of bacterial plasmids, site-specific enzymes (eg. site-specific integrases, Cas9), and associated cloning steps by amplifying each recoded segment directly from yeast using rolling circle amplification (39). RCA utilizes the bacteriophage ø29 DNA polymerase and random hexamers to selectively amplify circular DNA. Linearizing the resulting concatemer by LguI digestion gave microgram quantities of DNA ready for direct integration (Figure 2B). Carrying each segment on a bacterial plasmid would otherwise have required additional negative selection against the plasmid backbone to distinguish between the desired integration event versus the plasmid existing as an extrachromosomal replicative element (3,7). Propagation through a bacterial cloning host may also lead to toxicity issues arising from unintended expression of *Salmonella* genes.

## A - Assembly and amplification of recoded segments



## B - Direct iterative recombineering using SIRCAS

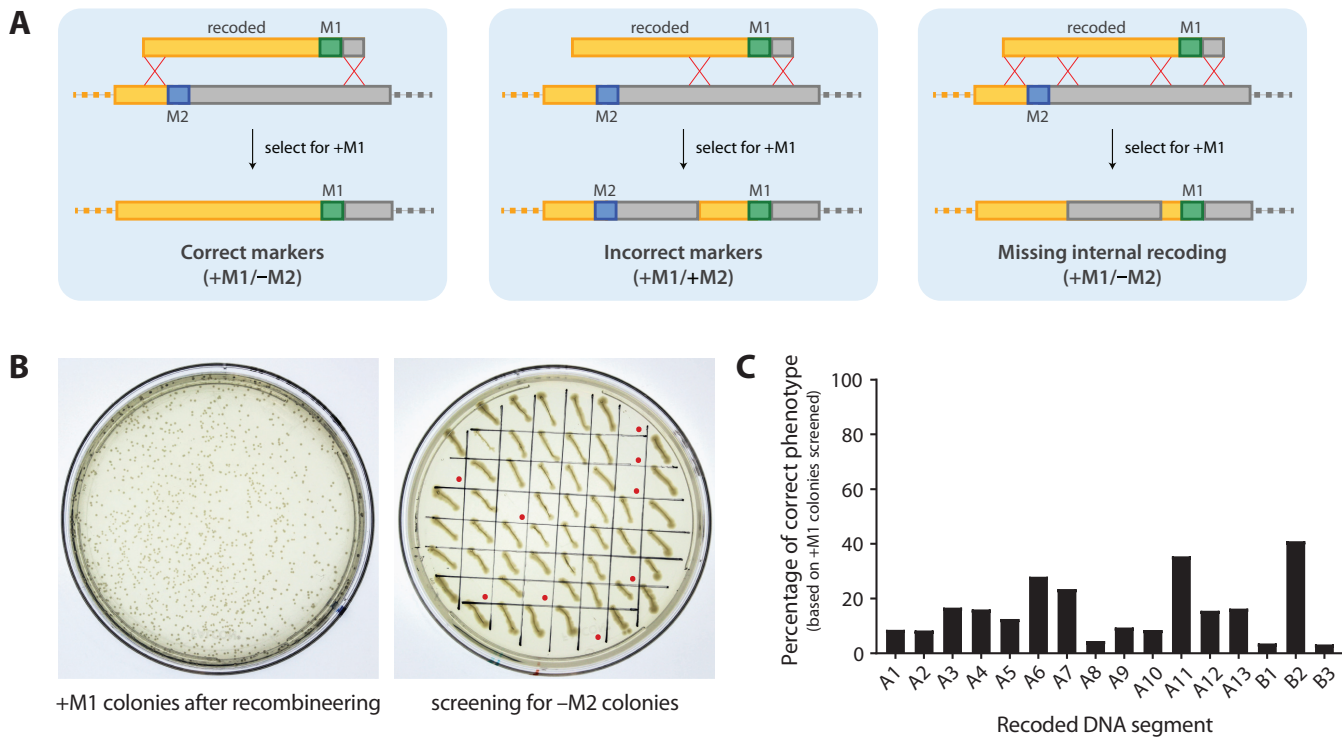


**Figure 2.** (A) Large 10–25 kb recoded DNA segments were created by assembling short synthetic DNA fragments into a YAC, followed by rolling circle amplification and linearization. Each recoded segment contains a selection marker (M1 or M2, typically kanamycin or chloramphenicol resistance cassettes) and flanking homology regions for integration. (B) Accumulated genome recoding by SIRCAS involves iterative recombination of recoded DNA segments. After each step, one selection marker is gained while the other is lost, providing a readout for successful recoding.

### Accumulated genome recoding by SIRCAS

We successfully performed SIRCAS on all sixteen recoded DNA segments to accumulate recoding in two independent *S. typhimurium* strains, one recoded from segment A1 to A13 and the other from B1 to B3. Each round of SIRCAS required only two days to complete. In a typical round of SIRCAS, hundreds of marker-positive colonies were obtained after recombineering (Figure 3). No colonies were obtained in negative controls in which arabinose or DNA was omitted. Upon screening for loss of the previous marker, a median of 14% of colonies was found to have correctly swapped markers, while the remaining colonies incorrectly contained both selection markers (Figure 3B and Supplementary File SI 1, Table S4.2.1). This frequency is

higher than naively predicted based on lengths of DNA in the distal and central segments, but is expected based on the results of Matic *et al.* (40), who found that mismatches in recombining DNA can significantly suppress the isolation of recombinants. The rate of obtaining the correct marker phenotype ranged from 3–41% (Figure 3C), presumably due to differences in marker integration locus (41), as well as the size and content of the incoming recoded DNA. To check whether recombinants with the correct markers were missing internal recoding due to multiple crossovers (Figure 3A), selected colonies with the correct phenotype were Sanger sequenced at one location in the middle of each newly recoded segment before progressing to the next round of SIRCAS. Correct recoding was observed in 83% of all sequenced colonies with the correct marker phenotype over



**Figure 3.** Screening for stepwise replacement of genomic segments. (A) Three possibilities after recombineering and selection for marker M1 are full integration of the entire segment leading to marker swapping, partial integration leading to the presence of both markers, or internal crossovers leading to missing recoding. (B) Hundreds of transformants were typically obtained on M1 selection plates (recoding of segment A13 shown here). On average, one in six colonies (16%) had the correct phenotype +M1/-M2, as identified by screening for no growth on M1+M2 plates. In this instance, 9 out of 60 colonies were correct (each marked with a red dot). (C) The rate of correct marker phenotype varied between different rounds of SIRCAS.

sixteen rounds of SIRCAS, indicating that multiple internal crossovers do occur but are relatively uncommon (Supplementary File SI 1 Table S4.3.1). This inference was supported by whole-genome sequencing of the final organism containing all 200 kb of recoded segments.

### Hierarchical assembly of SIRCAS-recoded regions by conjugation

After complete recoding of regions A and B by SIRCAS in two separate strains, the two regions of recoding were consolidated into one strain by conjugative assembly (42). An origin of transfer from the Gram-negative bacterial RK2 plasmid, which directs RK2-mediated conjugal transfer of DNA to a broad range of hosts (24), was integrated upstream of the recoded region of strain A13 by recombineering using an associated spectinomycin resistance cassette (Supplementary File SI 1 Figure S5.2.1). This donor strain was transformed with the 52.7 kb conjugation plasmid pTA-Mob (25), which contains the RK2 parABCDE ('partitioning') genes necessary for conjugation. The B3-based recipient strain was prepared by integrating an ampicillin resistance cassette at the start of region A such that upon conjugation, the resistance marker should be lost. Conjugation between these strains resulted in the transfer of the entire 154 kb recoded segment from the donor strain to the recipient, resulting in an exconjugant strain A13-B3 containing a total of 200 kb of recoded genomic DNA. Con-

jugation success rate as determined by ampicillin marker loss was 40% (17 out of 42 colonies screened).

### Characterization of accumulated recoding in *S. typhimurium*

In the final strain (designated A13-B3), recoding throughout both targeted regions spanning 200 kb of the genome was confirmed by next-generation sequencing, with a total of 1557 leucine codon changes successfully installed. An additional five point substitutions, one single point deletion and one single point insertion were found across 156 kb of the recoded regions constructed using clonal sequence-verified DNA, constituting an error rate of approximately 1 in 20 000. This number is consistent with the error rate of RCA (5) (approximately 1 in 50 000). In contrast, a total of 51 errors were found across the 44 kb of recoded regions constructed using non-clonal DNA, for an approximate error rate of 1 in 860 (Supplementary File SI 1 Section S6.3 and Supplementary File SI 2 Tab S7). The majority of these are single base deletions, consistent with expected errors in the synthesis of the chemical oligonucleotide precursors used to generate non-clonal DNA, which occur at a rate of better than 1 in 200 (43,44). Despite these errors (Supplementary File SI 2 Tab S8), the final strain A13-B3 was still viable and no other instances of incorrect leucine codon recoding were found.

Growth rates of all recoded strains were measured to analyze the impact of accumulated recoding on the viability of *S. typhimurium*. Despite the vast number of changes in-

roduced as part of the recoding process, no major cumulative growth defect was observed over the sixteen rounds of SIRCAS, demonstrating that bacterial genomes appear to be amenable towards large-scale rational re-engineering. No downward trend in fitness was observed across the entire 200 kb of recoding, with comparable doubling times measured across all strains at 37°C in LB (Figure 4A). Similar trends in growth behavior were also seen with growth at both 30°C and 42°C (Supplementary File SI 1 Figure S7.1.3). Sequencing also did not reveal any clear compensatory mutations in non-recoded regions of the genome which might suppress any defects (Supplementary File SI 2 Table S9).

Minor phenotypic changes were observed at one step of the recoding process. When grown in a plate reader, cells containing segment B2 (ie. strains B2, B3 and A13-B3) were observed to sediment unevenly (Figure 4B), resulting in anomalous growth curves. Growth experiments repeated in Erlenmeyer flasks showed no major differences in doubling time (Figure 4A). Although final optical density measurements differed between strains, growth as measured by colony forming units was consistent between wild-type and recoded strains B2, B3 and A13-B3 (Figure 4C). The sedimentation effect may be due to changes in the cell surface arising from non-clonal DNA errors in regions responsible for *O*-antigen biosynthesis (Supplementary File SI 1 Figure S6.3.1). No stepwise accumulation of growth defects was observed upon recoding all other segments.

## DISCUSSION

We have developed SIRCAS (stepwise integration of rolling circle amplified segments) as a rapid method for making genomic changes that bypasses the requirement for site-specific enzymes. Applying SIRCAS to recoding of the *S. typhimurium* genome, we installed 1557 TTA/TTG to CTA/CTG leucine codon changes across 200 kb of genomic DNA in a single strain. A key aspect of SIRCAS is the use of rolling circle amplification to produce DNA of sufficient quality and quantity appropriate for transformation. We found that recombineering with extended homology regions was effective on 10–25 kb linear DNA segments, without needing assistance from extra accessory proteins such as endonucleases. Sixteen cycles of integration into the genome were achieved efficiently, with each integration step requiring two days to complete. The recoded strains did not show any major fitness defects when compared to the wild-type strain, demonstrating the remarkable plasticity of bacterial genomes. Overall, this work paves the way for creating a completely recoded *Salmonella* as a genetically isolated chassis for downstream therapeutic applications.

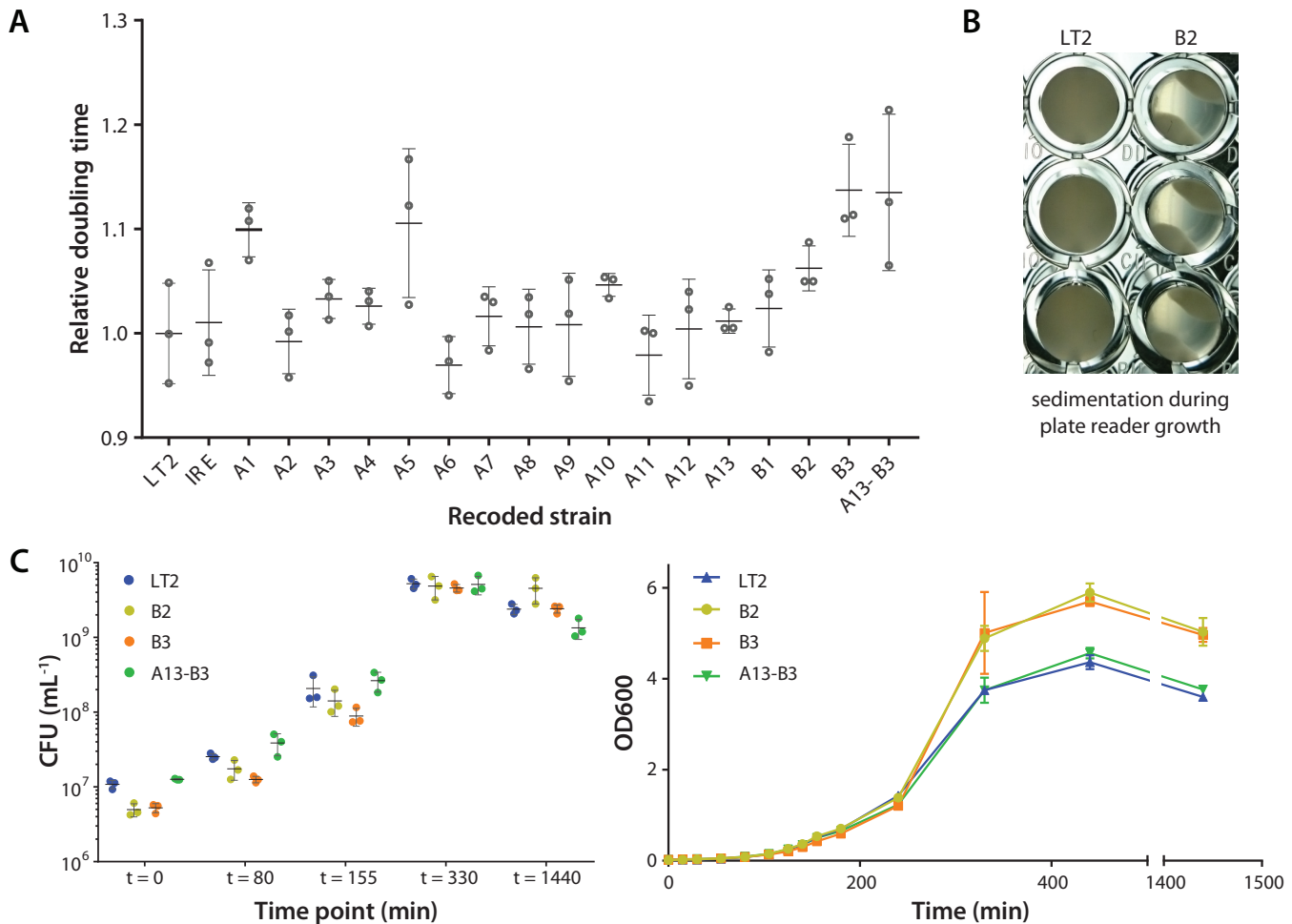
The SIRCAS approach can be carried out in parallel in many strains for more rapid construction. In this study, we maintained two different recoded strains and unified the recoded regions by conjugation. For completing the construction of a fully recoded *S. typhimurium* genome, we anticipate recoding 8–16 strains in parallel, followed by hierarchical assembly via conjugation. Recoding 20 kb every two days, it is theoretically possible to rewrite 300 kb of the genome per strain in a month.

The *Salmonella* genome tolerated a vast number of designed codon changes, suggesting that recoding the entire genome of *Salmonella* will be feasible. This result is consistent with similar recoding studies in *E. coli* (3,7). Design flaws may affect viability, and troubleshooting may be a significant challenge when constructing a complete recoded genome. As observed in the ‘minimal genome’ work by Venter *et al.* (5), synthetic lethal gene pairs may become apparent as more recoding is accumulated in a single strain. In these cases, the piecemeal nature of the SIRCAS strategy may prevent the process of redesigning problematic pieces from becoming a bottleneck. Further understanding of principles governing codon choice will also help improve our existing designs (15,16).

There are currently few alternative methods for large-scale genomic engineering. Multiplex automated genome engineering (45) and multiplexed CRISPR/Cas9 systems (46) are not sufficiently high throughput to install the large number of designed changes required for leucine recoding. Church and co-workers used site-specific integrases for testing a 57-codon *E. coli* genome design (3), integrating 10 out of 87 possible 50 kb recoded segments into independent strains. Site-specific integrases are an effective means for testing individual recoded segments, as there is less opportunity for internal crossover events when compared to homologous recombination. However, unification of recoded segments into a single strain is more challenging. Integrase-based methods require the installation and subsequent removal of landing pads for each integration step. The integration event itself also requires several steps, starting with introduction of the recoded segment on a plasmid, removal of the wild-type genomic segment, integration of the recoded segment, and finally removal of the plasmid. This multi-step approach compromises the efficiency of the construction process when compared to the single-step integration of SIRCAS.

The 20 kb size range may be an optimal compromise between the number of integration steps and ease of construction and troubleshooting. Chin and co-workers have combined Cas9 and recombineering to insert over 100 kb of DNA into the *E. coli* genome (7). The major advantage of using large 100 kb constructs is that fewer rounds of integration are required to achieve complete genome recoding when compared to the 10–25 kb segments we used for SIRCAS. Conversely, there is a greater likelihood of incomplete recoding due to multiple internal crossover events with increasing size. Larger constructs are also more difficult to assemble from small fragments (efficiency drops as more fragments are included in the assembly (23)) and more prone to shearing when handled. Furthermore, finding the causative mutations when recoding changes impact fitness becomes more challenging with increasing size. A noteworthy troubleshooting method was described by Chin and co-workers, using a sequencing approach to pinpoint where single lethal mutations arise. However, this method was only demonstrated for recoding a 20 kb region of the *E. coli* genome, and may be more problematic in segments with more than one lethal mutation.

The overall error rates of SIRCAS are competitive with that of other genome recoding methods. Within the 156 kb of recoded regions that was written with clonal sequence-



**Figure 4.** (A) Doubling times of various recoded *S. typhimurium* strains growing at 37°C in LB. Each data point is the average of three technical replicates, and the error bars represent the standard deviation of three biological replicates. No major fitness defect was observed throughout the recoding process. The strain nomenclature ‘A<sub>n</sub>’ refers to the strain containing all recoded segments from A1 up to A<sub>n</sub>. We note that doubling times for B2, B3 and A13-B3 were calculated from growth curves conducted in batch cultures grown in flasks due to sedimentation, while the remaining data was obtained on a plate reader. (B) Uneven sedimentation of cells is observed when strains B2, B3 and A13-B3 are grown in a plate reader (wells in right column). (C) Growth data for B2, B3 and A13-B3 was obtained by growth in culture flasks rather than in 96-well plates to avoid sedimentation artifacts. Wild-type LT2 was used as a control and had comparable doubling times in both plate and flask growth (see Supplementary File S1 1 Figure S7.1.4).

verified DNA, an overall error rate of 1 in 20 000 was observed (7 point mutations and one leucine codon reversion in 200 kb). In comparison, Church and co-workers (3) reported an error rate of 1 in 5000 (an average of 9.7 mutations and 0.6 codon reversions in 50 kb). For the 44 kb of recoded regions written using non-clonal DNA, an overall error rate of 1 in 860 was found (51 errors, primarily single point deletions). Use of clonal DNA for SIRCAS was therefore preferable, although issues associated with construct errors in non-clonal DNA may become less significant in the future as DNA synthesis and error-correction methods continue to improve.

Beyond genome recoding, SIRCAS is a powerful enabling technology for building and testing large *de novo* designs that span hundreds of genes. SIRCAS is not limited to *Salmonella*, and could be used in any recombinering competent host to integrate large naturally occurring or *de novo* designed gene clusters, facilitating the development of new industrial production strains. SIRCAS also enables interro-

gation of bacterial genetics on a scale that is not possible with traditional techniques. By making genome-scale construction more accessible, SIRCAS will help to empirically elucidate the underlying principles for completely *de novo* genome design.

#### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

#### ACKNOWLEDGEMENTS

We thank Eric Kofoed and Natalie Duleba from the laboratory of Prof. John Roth (Department of Microbiology, University of California Davis) for the kind gift of the recombinering *Salmonella* strain TT22971 containing plasmid pKD46. We also thank Prof. George Church (Department of Genetics, Harvard Medical School) and members of his laboratory for general discussions about genome recoding.



**Author contributions:** The overall study was conceived by P.A.S. and J.C.W. Experimental design was conducted by Y.H.L., F.S., J.K., M.A.P.K. and A.R. The computational design of the recoded genome and other computational analyses were performed by M.A.P.K., A.R. and Y.H.L. Assembly and RCA of DNA constructs was performed by Y.H.L., F.S., C.A.H., E.S. and D.L. The protocols for DNA assembly and RCA were established through guidance and troubleshooting from M.T.W., D.G.G. and Y.A.C. Genomic integration (SIRCAS) was conducted by Y.H.L. and C.A.H. Conjugation experiments were performed by J.K. Sequence analysis was conducted by Y.H.L. and F.S. Growth data for recoded strains was obtained by Y.H.L. All authors contributed to writing the manuscript.

## FUNDING

DARPA [BRICS HR0011-15-C-0094]; Wellcome Trust for the award of a Sir Henry Wellcome postdoctoral fellowship [107402/Z/15/Z to Y.H.L.]; acknowledges funding from a DOE CSGF fellowship [DE-FG02-97ER25308 A.R.]; NIH training grant [5T32GM007598 F.S.]. Funding for open access charge: BRICS [HR0011-15-C-0094].

**Conflict of interest statement.** D.G.G. is co-founder of SGI-DNA and Vice President of DNA Technologies at Synthetic Genomics, Inc., and M.T.W. is an employee of Synthetic Genomics, Inc.

## REFERENCES

- Boeke, J.D., Church, G., Hessel, A., Kelley, N.J., Arkin, A., Cai, Y., Carlson, R., Chakravarti, A., Cornish, V.W., Holt, L. *et al.* (2016) The genome project-write. *Science*, **353**, 126–127.
- Lajoie, M.J., Rovner, A.J., Goodman, D.B., Aerni, H.-R., Haimovich, A.D., Kuznetsov, G., Mercer, J.A., Wang, H.H., Carr, P.A., Mosberg, J.A. *et al.* (2013) Genomically recoded organisms expand biological functions. *Science*, **342**, 357–360.
- Ostrov, N., Landon, M., Guell, M., Kuznetsov, G., Teramoto, J., Cervantes, N., Zhou, M., Singh, K., Napolitano, M.G., Moosburner, M. *et al.* (2016) Design, synthesis, and testing toward a 57-codon genome. *Science*, **353**, 819–822.
- Annaluru, N., Muller, H., Mitchell, L.A., Ramalingam, S., Stracquadanio, G., Richardson, S.M., Dymond, J.S., Kuang, Z., Scheifele, L.Z., Cooper, E.M. *et al.* (2014) Total synthesis of a functional designer eukaryotic chromosome. *Science*, **344**, 55–58.
- Hutchison, C.A., Chuang, R.-Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H., Gill, J., Kannan, K., Karas, B.J., Ma, L. *et al.* (2016) Design and synthesis of a minimal bacterial genome. *Science*, **351**, aad6253.
- Richardson, S.M., Mitchell, L.A., Stracquadanio, G., Yang, K., Dymond, J.S., DiCarlo, J.E., Lee, D., Huang, C.L.V., Chandrasegaran, S., Cai, Y. *et al.* (2017) Design of a synthetic yeast genome. *Science*, **355**, 1040–1044.
- Wang, K., Fredens, J., Brunner, S.F., Kim, S.H., Chia, T. and Chin, J.W. (2016) Defining synonymous codon compression schemes by genome recoding. *Nature*, **539**, 59–64.
- Chin, J.W. (2012) Reprogramming the genetic code. *Science*, **336**, 428–429.
- Kipper, K., Lundius, E.G., Ćurić, V., Nikić, I., Wiessler, M., Lemke, E.A. and Elf, J. (2017) Application of noncanonical amino acids for protein labeling in a genomically recoded *Escherichia coli*. *ACS Synth. Biol.*, **6**, 233–255.
- Zheng, Y., Lajoie, M.J., Italia, J.S., Chin, M.A., Church, G.M. and Chatterjee, A. (2016) Performance of optimized noncanonical amino acid mutagenesis systems in the absence of release factor 1. *Mol. Biosyst.*, **12**, 1746–1749.
- Ma, N.J. and Isaacs, F.J. (2016) Genomic recoding broadly obstructs the propagation of horizontally transferred genetic elements. *Cell Syst.*, **3**, 199–207.
- Mandell, D.J., Lajoie, M.J., Mee, M.T., Takeuchi, R., Kuznetsov, G., Norville, J.E., Gregg, C.J., Stoddard, B.L. and Church, G.M. (2015) Biocontainment of genetically modified organisms by synthetic protein design. *Nature*, **518**, 55–60.
- Rovner, A.J., Haimovich, A.D., Katz, S.R., Li, Z., Grome, M.W., Gassaway, B.M., Amiram, M., Patel, J.R., Gallagher, R.R., Rinehart, J. *et al.* (2015) Recoded organisms engineered to depend on synthetic amino acids. *Nature*, **518**, 89–93.
- Moe-Behrens, G.H.G., Davis, R. and Haynes, K.A. (2013) Preparing synthetic biology for the world. *Front. Microbiol.*, **4**, 5.
- Napolitano, M.G., Landon, M., Gregg, C.J., Lajoie, M.J., Govindarajan, L., Mosberg, J.A., Kuznetsov, G., Goodman, D.B., Vargas-Rodriguez, O., Isaacs, F.J. *et al.* (2016) Emergent rules for codon choice elucidated by editing rare arginine codons in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E5588–E5597.
- Kelsic, E.D., Chung, H., Cohen, N., Park, J., Wang, H.H. and Kishony, R. (2016) RNA structural determinants of optimal codons revealed by MAGE-Seq. *Cell Syst.*, **3**, 563–571.
- Lajoie, M.J., Kosuri, S., Mosberg, J.A., Gregg, C.J., Zhang, D. and Church, G.M. (2013) Probing the limits of genetic recoding in essential genes. *Science*, **342**, 361–363.
- Kong, W., Brovold, M., Koenen, B.A., Clark-Curtiss, J. and Curtiss, R. (2012) Turning self-destructing *Salmonella* into a universal DNA vaccine delivery platform. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 19414–19419.
- Frahm, M., Felgner, S., Kocijancic, D., Rohde, M., Hensel, M., Curtiss, R., Erhardt, M. and Weiss, S. (2015) Efficiency of conditionally attenuated *Salmonella enterica* Serovar Typhimurium in bacterium-mediated tumor therapy. *MBio*, **6**, doi:10.1128/mBio.00254-15.
- Zheng, J.H., Nguyen, V.H., Jiang, S.-N., Park, S.-H., Tan, W., Hong, S.H., Shin, M.G., Chung, I.-J., Hong, Y., Bom, H.-S. *et al.* (2017) Two-step enhanced cancer immunotherapy with engineered *Salmonella typhimurium* secreting heterologous flagellin. *Sci. Transl. Med.*, **9**, eaak9537.
- Kotula, J.W., Kerns, S.J., Shaket, L.A., Siraj, L., Collins, J.J., Way, J.C. and Silver, P.A. (2014) Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 4838–4843.
- Kouprina, N. and Larionov, V. (2008) Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast *Saccharomyces cerevisiae*. *Nat. Protoc.*, **3**, 371–377.
- Gibson, D.G., Benders, G.A., Axelrod, K.C., Zaveri, J., Algire, M.A., Moodie, M., Montague, M.G., Venter, J.C., Smith, H.O. and Hutchison, C.A. (2008) One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 20404–20409.
- Guiney, D.G. and Jakobson, E. (1983) Location and nucleotide sequence of the transfer origin of the broad host range plasmid RK2. *Proc. Natl. Acad. Sci. U.S.A.*, **80**, 3595–3598.
- Strand, T.A., Lale, R., Degnes, K.F., Lando, M. and Valla, S. (2014) A New and Improved Host-Independent Plasmid System for RK2-Based Conjugal Transfer. *PLoS One*, **9**, e90372.
- Ma, N.J., Moonan, D.W. and Isaacs, F.J. (2014) Precise manipulation of bacterial chromosomes by conjugative assembly genome engineering. *Nat. Protoc.*, **9**, 2285–2300.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C. *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.
- McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F. *et al.* (2001) Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature*, **413**, 852–856.
- Giege, R., Sissler, M. and Florentz, C. (1998) Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.*, **26**, 5017–5035.

30. Choi, Y., Sims, G.E., Murphy, S., Miller, J.R. and Chan, A.P. (2012) Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS One*, **7**, e46688.
31. Sawitzke, J.A., Thomason, L.C., Costantino, N., Bubunenko, M., Datta, S. and Court, D.L. (2007) Recombineering: in vivo genetic engineering in *E. coli*, *S. enterica*, and beyond. *Methods Enzymol.*, **421**, 171–199.
32. Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 6640–6645.
33. Yu, D., Ellis, H.M., Lee, E.-C., Jenkins, N.A., Copeland, N.G. and Court, D.L. (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 5978–5983.
34. Karlinsey, J.E. (2007)  $\lambda$ -Red genetic engineering in *Salmonella enterica* serovar typhimurium. *Methods Enzymol.*, **421**, 199–209.
35. Kuhlman, T.E. and Cox, E.C. (2010) Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Res.*, **38**, e92.
36. Slechta, E.S., Bunny, K.L., Kugelberg, E., Kofoid, E., Andersson, D.I. and Roth, J.R. (2003) Adaptive mutation: general mutagenesis is not a programmed response to stress but results from rare coamplification of *dinB* with *lac*. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 12847–12852.
37. Colson, C. and Van Pel, A. (1974) DNA restriction and modification systems in *Salmonella*. *Mol. Gen. Genet.*, **129**, 325–337.
38. Itaya, M., Tsuge, K., Koizumi, M. and Fujita, K. (2005) Combining two genomes in one cell: Stable cloning of the *Synechocystis* PCC6803 genome in the *Bacillus subtilis* 168 genome. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 15971–15976.
39. Dean, F.B., Nelson, J.R., Giesler, T.L. and Lasken, R.S. (2001) Rapid Amplification of Plasmid and Phage DNA Using Phi29 DNA Polymerase and Multiply-Primed Rolling Circle Amplification. *Genome Res.*, **11**, 1095–1099.
40. Matic, I., Rayssiguier, C. and Radman, M. (1995) Interspecies gene exchange in bacteria: The role of SOS and mismatch repair systems in evolution of species. *Cell*, **80**, 507–515.
41. Englaender, J.A., Jones, J.A., Cress, B.F., Kuhlman, T.E., Linhardt, R.J. and Koffas, M.A.G. (2017) Effect of genomic integration location on heterologous protein expression and metabolic engineering in *E. coli*. *ACS Synth. Biol.*, doi:10.1021/acssynbio.6b00350.
42. Isaacs, F.J., Carr, P.A., Wang, H.H., Lajoie, M.J., Sterling, B., Kraal, L., Tolonen, A.C., Gianoulis, T.A., Goodman, D.B., Reppas, N.B. *et al.* (2011) Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science*, **333**, 348–353.
43. Kosuri, S. and Church, G.M. (2014) Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods*, **11**, 499–507.
44. Ma, S., Saaem, I. and Tian, J. (2012) Error correction in gene synthesis technology. *Trends Biotechnol.*, **30**, 147–154.
45. Wang, H.H., Isaacs, F.J., Carr, P.A., Sun, Z.Z., Xu, G., Forest, C.R. and Church, G.M. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, **460**, 894–898.
46. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.