

## Probing the Limits of Genetic Recoding in Essential Genes M. J. Lajoie *et al. Science* **342**, 361 (2013); DOI: 10.1126/science.1241460

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

**Permission to republish or repurpose articles or portions of articles** can be obtained by following the guidelines here.

The following resources related to this article are available online at www.sciencemag.org (this information is current as of October 21, 2013):

**Updated information and services,** including high-resolution figures, can be found in the online version of this article at:

http://www.sciencemag.org/content/342/6156/361.full.html

Supporting Online Material can be found at: http://www.sciencemag.org/content/suppl/2013/10/16/342.6156.361.DC1.html

A list of selected additional articles on the Science Web sites **related to this article** can be found at: http://www.sciencemag.org/content/342/6156/361.full.html#related

This article **cites 35 articles**, 14 of which can be accessed free:

http://www.sciencemag.org/content/342/6156/361.full.html#ref-list-1

This article has been **cited by** 1 articles hosted by HighWire Press; see: http://www.sciencemag.org/content/342/6156/361.full.html#related-urls

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2013 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.

# **Probing the Limits of Genetic Recoding in Essential Genes**

M. J. Lajoie,<sup>1,2</sup>\* S. Kosuri,<sup>3</sup>\* J. A. Mosberg,<sup>1,2</sup> C. J. Gregg,<sup>1</sup> D. Zhang,<sup>3</sup> G. M. Church<sup>1,3</sup>†

Engineering radically altered genetic codes will allow for genomically recoded organisms that have expanded chemical capabilities and are isolated from nature. We have previously reassigned the translation function of the UAG stop codon; however, reassigning sense codons poses a greater challenge because such codons are more prevalent, and their usage regulates gene expression in ways that are difficult to predict. To assess the feasibility of radically altering the genetic code, we selected a panel of 42 highly expressed essential genes for modification. Across 80 *Escherichia coli* strains, we removed all instances of 13 rare codons from these genes and attempted to shuffle all remaining codons. Our results suggest that the genome-wide removal of 13 codons is feasible; however, several genome design constraints were apparent, underscoring the importance of a strategy that rapidly prototypes and tests many designs in small pieces.

The canonical genetic code is nearly universal (1), allowing natural organisms to share beneficial traits via horizontal gene transfer. Genetically modified organisms also share this code, rendering them susceptible to viruses and capable of releasing recombinant genetic material [e.g., resistance genes (2)] into

## Fig. 1. Codon reassignment across 42 essential genes. (A) *E. coli* MG1655 codon usage heat map; brightness increases as codon usage decreases. Black numbers are total codon usage based on

NC\_000913.2 (National Center for Biotechnology Information, 1 September 2011). The anticodon specificities (29, 30) are illustrated as dashed brackets; white indicates anticodons that were targeted for eventual removal. Amino acids are indicated in the yellow side bars. White boxes denote the 13 forbidden codons, and white numbers report how many instances of each codon were in the panel of 42 targeted essential genes. All 405 instances of these forbidden codons were successfully recoded across 80 E. coli strains. Additionally, all possible codons were swapped the environment. By redefining the genetic code, we hope to produce genomically recoded organisms (GROs) that are safe and useful.

In separate work, we have completely reassigned the UAG codon in *Escherichia coli* MG1655 (3). UAG was chosen for its rarity and simplicity of function, but our results (3) reinforce

that sense codons must also be reassigned to achieve robust genetic isolation, broad virus resistance, and expanded chemical versatility (4). However, sense codon reassignment poses a considerable challenge given that codon usage can strongly affect gene regulation (5), ribosome spacing (6, 7), translation efficiency (7, 8), translation levels (9), translation accuracy (10), and protein folding (11, 12). Furthermore, DNA/RNA motifs can provide additional noncoding functions such as regulating translation initiation via 5' mRNA secondary structure (13), sharing sequence with overlapping small RNAs (14), pausing the ribosome at internal Shine-Dalgarno sequences (15), and regulating mRNA localization (16). Therefore, it is difficult to predict the effects of a given codon change, and these factors may substantially constrain the malleability of the genome.

<sup>1</sup>Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. <sup>2</sup>Program in Chemical Biology, Harvard University, Cambridge, MA 02138, USA. <sup>3</sup>Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA.

\*These authors contributed equally to this work.

†Corresponding author. E-mail: gchurch@genetics.med.harvard.edu



to synonymous codons, and gene overlaps were removed by duplication (bottom). (**B**) Strategy for recoding essential genes. Recoded genes (blue rectangles) were synthesized from Agilent Oligonucleotide Library Synthesis arrays (24), then transcriptionally fused to  $kan^R$  (purple rectangles) by isothermal assembly (25). These cassettes were recombined into EcNR2 {*E. coli* MG1655  $\Delta$ (*ybhB-bioAB*)::[ $\lambda$ cl857 N(*cro-ea59*)::*tetR-bla*]  $\Delta$ *mutS*::*cat*} using  $\lambda$  Red (26), and recombinants were selected on kanamycin. Putative recombinants were screened with three sets of primers: Wild-type primers (gray) hybridize specifically to the natural gene sequence, mutant primers (blue) hybridize specifically to the recoded gene sequence, and boundary primers (black) hybridize to the

surrounding genomic DNA. Desired recombinants were detected by polymerase chain reaction and then verified by Sanger sequencing. We found that *kan<sup>R</sup>* (*"kan<sup>R</sup>* only") could be inserted downstream of all genes except for *rplO* without causing major deleterious effects. We attempted to replace all 42 natural genes with radically recoded versions ("Fully recoded"; blue rectangles and triangles are recoded sequence). To coarsely map problematic design elements in the failed cassettes, we prepared cassettes that preserved natural sequence at the N terminus ("Partially recoded"; gray rectangles and triangles are natural sequence). Finally, all remaining forbidden codons were recoded with CoS-MAGE (green triangles) and confirmed with Sanger sequencing.

## REPORTS

However, despite the myriad mechanisms by which swapping synonymous codons could be deleterious, efforts to express a codon-randomized *Klebsiella* nitrogenase gene cluster in *E. coli* have been successful, albeit with reduced activity compared with wild type (17).

Although such information is critical for reassigning the genetic code, genome-wide codon essentiality has largely been unexplored, perhaps due to the substantial degree of genetic modification necessary for addressing such questions. For example, the complete removal of 13 codons corresponding to the least frequently used anticodons (Fig. 1A and supplementary text) will require 155,224 changes in E. coli MG1655, several of which may not be tolerated. Although it has never been attempted, de novo genome design, synthesis, and transplantation (17) seem unlikely to produce a viable genome bearing this unprecedented number of potentially deleterious changes. Indeed, lethal genetic elements have been difficult to identify and eliminate using de novo genome transplantation (16). Therefore, we have developed in vivo multiplex genome editing technologies (18, 19) to rapidly prototype and manufacture genomes. Our approach exploits diversity and natural selection and is highly amenable to our goal of testing the flexibility of synonymous codon choices as they pertain to reassigning the genetic code.

To address this question, we attempted to individually recode 42 essential genes, including all 41 essential ribosomal protein-coding genes (20) and *prfB*, which relies on a programmed frameshift for proper translation (21). Because expression level correlates strongly with codon usage bias (9), the highly expressed and tightly regulated (22, 23) ribosomal genes should be among the most difficult to change. To study codon essentiality in each of these genes, we attempted to remove all instances of the aforementioned 13 codons (hereafter referred to as "forbidden" codons). In addition, we gauged tolerance for largescale DNA sequence alterations by shuffling all possible codons to synonymous alternatives. Replacement codons were chosen randomly from a weighted distribution, based on their frequencies in all E. coli genes (AUG and UGG codons were unchanged because they uniquely encode Met and Trp, respectively). Finally, we changed 1 non-AUG start codon to AUG, separated six gene overlaps, removed one frameshift, and avoided the use of six restriction sites used in gene assembly (supplementary text). Thus, whereas the protein sequence was 100% identical in our designs, the nucleotide sequence was on average only 65.4% identical, and the codon identity was only 4.44% (corresponding to the unchanged AUG and UGG codons) (table S1). Based on these radical design parameters, we did not expect all design elements to be tolerated. Therefore, individually recoding each gene was the most biologically relevant scale on which to assess the effects of recoding without sacrificing the ability to rapidly map design flaws.

We synthesized recoded genes from DNA microchips (24), transcriptionally fused each to a kanamycin resistance gene  $(kan^R)$  by isothermal assembly (25), and replaced the corresponding natural gene (one gene per strain) in vivo using  $\lambda$ Red recombination (26) (Fig. 1B). We also introduced  $kan^R$  downstream of the natural genes and found that 41 of 42 (table S2) allowed insertion with an average growth defect of 15% in LB-Lennox medium (12% in Teknova Hi-Def Azure medium). Insertion downstream of rplO was unsuccessful, indicating that disrupting operon structure-and, by extension, refactoring overlapping genes-is a potential failure mode for redesigning genomes. For the recoded genes, we found that 26 of 42 (table S2) were successful, with an average growth defect of 20% in LB-Lennox (14% in Azure) compared with kan<sup>K</sup> insertion controls (Fig. 2). In the recoded prfB strain, removing the frameshift and recoding an upstream AGG codon that may be involved in pausing translation and enhancing frameshifting (15) did not significantly affect fitness (t test, P =0.86). Finally, to test the independence of the growth defects, we inserted a recoded rplM or rpsI gene transcriptionally fused to spectinomycin resistance into three recoded strains with varying fitness (rpmC syn1, rplE syn1, and rplP syn1). All double-mutant strains exhibited better fitness than predicted assuming that the fitness defects were independent, although this does not rule out potential cumulative effects from combining multiple deleterious designs (fig. S1).

The 16 unsuccessfully recoded genes provided an opportunity to identify failure modes for recoding. We coarsely mapped deleterious alleles in the remaining genes by recoding only the C-terminal half of each gene (successful for 9 of 16 genes, table S2). Of these 9 genes, 7 were also amenable to recoding all but the first 30 codons (Fig. 1B).

Although not conclusive based on the limited sample size, these remaining failed replacements may be caused by the disruption of endogenous control mechanisms upstream of the gene (23) or by codon bias affecting expression (7, 12). Using the above synthetic complementation approaches, we recoded a total of 294 of 405 forbidden codons in 35 of 42 targeted essential genes across 35 strains (one recoded gene per strain) (tables S2 and S3). This generated 4375 out of 6496 total desired nucleotide changes and introduced 29 synthesis errors and/or spontaneous mutations (1 error per 436 base pairs) (Fig. 3 and table S4). Although synthesis errors sometimes introduced de novo forbidden codons, additional screening invariably found alternative clones lacking forbidden codons. We hypothesize that the remaining genes (7 of 16) failed due to perturbations in gene expression arising from separating overlapping genes and/or nonviable changes introduced while shuffling codons that were not forbidden.

To determine whether any remaining instances of the forbidden codons were essential, we used coselection multiplex automated genome engineering (CoS-MAGE) (27) to remove all remaining forbidden codons in small groups across a population of cells (111 desired mutations in 45 clones) (Fig. 3 and table S5). The CoS-MAGE recombinants exhibited robust fitness (Fig. 2), indicating that none of the forbidden codons provide a systematic barrier to removal. Furthermore, this suggests that unsuccessful gene replacements using fully recoded cassettes were not due to the removal of forbidden codons. Our initial designs vielded all desired mutations except for one (rplO U162G). Unexpectedly, when we attempted to replace this CUU (Leu) codon using a pool of oligos encoding all Leu, Ile, Val, and Ala codons (table S6), only CUG (Leu), UUG (Leu), and GUG (Val) were not observed (table S7). Therefore,



**Fig. 2. Recoded strain doubling times.** Recoded strain doubling times in (**A**) LB-Lennox media and (**B**) Teknova Hi-Def Azure media. Each data point represents the average doubling time of a given strain with a portion of a ribosomal gene recoded (n = 3). Error bars for each group represent mean  $\pm$  SD. Under assay conditions, the parental strain {*E. coli* MG1655  $\Delta$ (*ybhB-bioAB*)::[ $\lambda$ cl857 N(*cro-ea59*)::*tetR-bla*]  $\Delta$ *mut5::cat*, "EcNR2"} exhibited a 49  $\pm$  4 min doubling time in LB-Lennox and a 84  $\pm$  5 min doubling time in Teknova Hi-Def Azure. Strain genotypes and doubling times are summarized in tables 54 and 55. *Kan<sup>R</sup>* insertion into natural sequences (with no recoding) seldom impaired fitness. Still, we could not introduce *kan<sup>R</sup>* downstream of *rplO* after three attempts. Fully or partially recoded gene recombinants exhibited the broadest range of fitness defects. For successful recombinants, position of the recoded gene in its operon did not appear to correlate strongly with fitness. The CoS-MAGE recombinants exhibited robust fitness, indicating that all tested forbidden codons are readily dispensable in small groups.

CUU is not essential, but 3 out of 12 tested replacement codons (all ending in UG) were either deleterious or recalcitrant to  $\lambda$  Red-mediated allele replacement in a way that was not anticipated. We note that the native gene sequence at this locus (ACT CTT GCC) contains most of the CTWGG Vsr recognition motif [Vsr is a mismatch repair endonuclease that is somewhat MutS-independent (28)] but that the position (nucleotide 3 instead of nucleotide 2) and identity (T:C instead of T:G) of the oligo-mediated mismatch are noncanonical. We mutated codons 23 and 24 of vsr to in-frame stop codons but were still unable to isolate rplQ recombinants with CUG, UUG, or GUG codons at position 162, thus suggesting that Vsr is not the cause of these failed replacements. It is likely that further recoding will uncover additional cryptic

design flaws; nevertheless, our strategy is well suited to rapidly identify alternative solutions that are viable.

Our results provide three important insights for designing recoded genomes. First, when tested individually or in groups, all 405 instances of the forbidden codons were nonessential, suggesting that they are amenable to genome-wide removal. Second, our inability to replace CUU with CUG, UUG, and GUG at position 162 in *rplQ* demonstrates that synonymous codons can be nonequivalent in unpredictable ways. Nevertheless, our ability to successfully remove all instances of 13 codons from a panel of highly expressed essential genes indicates that radical genome recoding is feasible. Finally, most of the recoded genes displayed reduced fitness, and combining the cur-





rent designs into a single genome could lead to unacceptable fitness impairment. In contrast, we did not observe significantly altered growth rates for the CoS-MAGE strains in which only forbidden codons were changed (table S5). Therefore, our future strategies for genome-wide codon reassignment will only change codons of interest while selecting for variants with normal growth. This approach leverages diversity and evolution to overcome such uncharacterized genome design constraints, allowing researchers to focus on creating genomes possessing new and useful functions.

#### **References and Notes**

- 1. A. Ambrogelly, S. Palioura, D. Söll, *Nat. Chem. Biol.* 3, 29–35 (2007).
- 2. M. G. Schafer et al., PLoS ONE 6, e25736 (2011).
- 3. M. J. Lajoie et al., Science 342, 357-360 (2013).
- C. C. Liu, P. G. Schultz, Annu. Rev. Biochem. 79, 413–444 (2010).
- 5. M. Frenkel-Morgenstern *et al., Mol. Syst. Biol.* **8**, 572 (2012).
- N. T. Ingolia, S. Ghaemmaghami, J. R. S. Newman, J. S. Weissman, *Science* 324, 218–223 (2009).
- 7. T. Tuller et al., Cell 141, 344-354 (2010).
- H. Gingold, Y. Pilpel, *Mol. Syst. Biol.* 7, 481 (2011).
  J. B. Plotkin, G. Kudla, *Nat. Rev. Genet.* 12, 32–42
- (2011).
- 10. H. Akashi, Genetics 136, 927-935 (1994).
- 11. E. Angov, Biotechnol. J. 6, 650-659 (2011).
- S. Pechmann, J. Frydman, Nat. Struct. Mol. Biol. 20, 237–243 (2013).
- 13. M. Bulmer, Genetics 129, 897-907 (1991).
- 14. A. Shinhara et al., BMC Genomics 12, 428 (2011).
- G.-W. Li, E. Oh, J. S. Weissman, *Nature* 484, 538–541 (2012).
- K. Nevo-Dinur, A. Nussbaum-Shochat, S. Ben-Yehuda, O. Amster-Choder, *Science* 331, 1081–1084 (2011).
- 17. D. G. Gibson et al., Science 329, 52-56 (2010).
- 18. F. J. Isaacs et al., Science 333, 348-353 (2011).
- 19. H. H. Wang et al., Nature 460, 894-898 (2009).
- Y. Yamazaki, H. Niki, J. Kato, *Methods Mol. Biol.* 416, 385–389 (2008).
- 21. K. Higashi et al., J. Biol. Chem. 281, 9527-9537 (2006).
- M. Kaczanowska, M. Rydén-Aulin, *Microbiol. Mol. Biol. Rev.* 71, 477–494 (2007).
- M. T. Sykes, E. Sperling, S. S. Chen, J. R. Williamson, Anal. Chem. 82, 5038–5045 (2010).
- 24. S. Kosuri et al., Nat. Biotechnol. 28, 1295-1299 (2010).
- 25. D. G. Gibson et al., Nat. Methods 6, 343-345 (2009).
- D. G. Yu et al., Proc. Natl. Acad. Sci. U.S.A. 97, 5978–5983 (2000).
- 27. P. A. Carr et al., Nucleic Acids Res. 40, e132 (2012).
- 28. R. Zell, H. J. Fritz, EMBO J. 6, 1809–1815 (1987).
- 29. M. A. Sørensen et al., J. Mol. Biol. 354, 16-24 (2005).
- 30. A. C. Forster, G. M. Church, Mol. Syst. Biol. 2, 45 (2006).

Acknowledgments: We thank S. Vassallo and J. Ho for technical assistance and U. Laserson, D. Goodman, N. Eroshenko, D. Mandell, D. Söll, L. Ling, and F. Isaacs for helpful comments. Funding was from the U.S. Department of Energy (DE-FG02-02ER63445), NSF (SA5283-11210), the Defense Advanced Research Projects Agency (N66001-12-C-4040), the U.S. Office of Naval Research (N000141010144), Agilent Technologies, Wyss Institute, and Department of Defense National Defense Science and Engineering Graduate Fellowship (M.J.L.).

### Supplementary Materials

www.sciencemag.org/content/342/6156/361/suppl/DC1 Materials and Methods Supplementary Text Fig. S1 Tables S1 to S11 References (*31–36*)

10 April 2013; accepted 19 September 2013 10.1126/science.1241460