Design, Synthesis and Testing of a 57-Codon Genome

Matthieu Landon[1,2,3]*†(matthieulandon@fas.harvard.edu), Nili Ostrov[1]†, Marc Guell[1,9]†, Gleb Kuznetsov[1,5]†, Jun Teramoto[1,4], Julie E. Norville[1], Natalie Cervantes[1], Minerva Zhou[7], Michael G. Napolitano[6] , Mark Moosburner[1], Ben Pruitt[9], Nick Conway[9], Dan Goodman[1], Kerry Singh[7], Cameron L. Gardner[1], Barry L. Wanner[1,4,8], Marc Lajoie[1], **George M. Church[1]**

[1]Department of Genetics, Harvard Medical School, Boston, MA; [2]Program in Systems Biology, Harvard University, Cambridge, MA; [3]Ecole des Mines de Paris, Mines Paristech, 75272 Paris, France; [4]Department of Biological Sciences, Purdue University, West Lafayette, IN; [5]Program in Biophysics, Harvard University, Cambridge, MA; [6]Program in Biological and Biomedical Sciences, Harvard University, Cambridge, MA; [7]Department of Biological Engineering, MIT, Cambridge, MA; [8]Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA; [9]Wyss Institute for Biologically Inspired Engineering, Boston, MA; †Equal contribution

http://arep.med.harvard.edu (Church lab)

Project Goals: Recoding, replacing codons genome-wide in an organism, is a unique and transformative way to explore currently unknown biological landscapes. By changing all instances of a target codon to synonymous codons and disabling the target codon function, invading DNA from natural sources cannot be translated, making the GRO resistant to infection and unreceptive to horizontal transfer. Unused codons can also be assigned to nonstandard amino acids (NSAAs) to generate proteins with novel functions[2], or to make the GRO biocontainable by making provision of NSAAs essential to growth[1]. We previously created a GRO in which all instances of the very rare UAG codon in *E. coli* were changed to UAAs (321 changes)[3]. To create GROs with larger scales of change, we have been developing radical recoding methods by which the genome is resynthesized with all target codons reassigned to other values. Here, our goal has been to develop these methods and generate an *E. coli* that lacks 7 codons. This work underscores the feasibility of extensively rewriting genomes, and establishes a framework for large-scale design, assembly, troubleshooting and phenotypic analysis of synthetic organisms.

We report the computational design, synthesis and partial testing of segments of a recoded *Escherichia coli* genome in which all 62,214 instances of seven codons were replaced with synonymous alternatives across all protein coding genes. Replacement values for target codons AGA (Arg), AGG (Arg), AGC (Ser), AGU (Ser), UUG (Leu), UUA (Leu) and UAG (Stop) were chosen by in-house designed software that attempts to retain key mRNA features such as secondary structure and ribosome binding sites, refactors overlapping genes, and avoids sequences that are difficult to synthesize. The entire computationally redesigned 3.98Mb genome was divided into 1254 fragments and synthesized *de novo* by a variety of vendors. As even a single lethal design error can lead to an inviable strain, we developed a pipeline for assembling the fragments into ~50 kb segments, testing these segments as plasmids in separate *E. coli* strains, deleting the corresponding non-recoded chromosomal sequence to demonstrate that the

recoded sequence complements the original, and integrating the recoded sequence directly into the chromosome. To debug failures at the various steps of this process, we also developed a troubleshooting pipeline to diagnose the causative gene, codon, or other feature, and to determine corrections. At this time, we have assembled the entire recoded genome in 87 segments and validated 54% of genes, including 50% of all essential genes, by individually testing 47 segments up to the chromosomal integration step. We found that 99.3% of 1918 recoded genes retained functionality with only a modest effect on strain fitness, and investigated 13 design flaws using our troubleshooting pipeline. As we push towards generating viable, chromosomally integrated strains for all 87 recoded segments, we are simultaneously developing and optimizing segment assembly strategies for creating a strain with a fully recoded genome. Once complete, a 57-codon organism could broaden synthetic functionality in living cells.

References

1. D. J. Mandell *et al.*, Biocontainment of genetically modified organisms by synthetic protein design. *Nature* **518**, 55 (2015).
2. C. C. Liu, P. G. Schultz, Adding new chemistries to the genetic code. *Annu Rev Biochem* **79**, 413 (2010).
3. AM. J. Lajoie *et al.*, Genomically recoded organisms expand biological functions. *Science* **342**, 357 (2013).

This project has been funded by DOE grant DE-FG02-02ER63445 and DARPA grant BAA-12-64.