Parallelized *in vivo* Construction of a Synthetic 57-Codon *E. coli* Genome

Nili Ostrov,[1] Maximilien Baas-Thomas[2]* (maximilien@g.harvard.edu), Alexandra Rudolph,[2] Matthieu Landon,[3] and **George Church**[1,4]

[1]Department of Genetics, Harvard Medical School, Boston, MA; [2]Program in Biological and Biomedical Sciences, Harvard University, Cambridge, MA; [3]Program in Systems Biology, Harvard University, Cambridge, MA; and [4]Wyss Institute for Biologically Inspired Engineering, Boston, MA

http://arep.med.harvard.edu

Project Goals: We are building a fully recoded, 3.97 Mb *Escherichia coli* genome, in which seven codons are replaced with synonymous alternatives in all protein coding genes. For this aim, the recoded genome was *de-novo* synthesized and assembled in vivo into 87 segments. In the final steps of genome construction, we combine all segments to a single fully recoded genome.

We present progress towards a fully recoded, 57-codon *Escherichia coli* genome, in which seven codons are replaced with synonymous alternatives in all protein coding genes. Synthetic recoded DNA was assembled into 50 kb episomal segments, which were individually tested and combined *in vivo* into larger recoded "clusters" via Cas9-assisted recombineering. Each segment replaces its corresponding wild-type sequence. This integration efficiency for 50 kb segments was enhanced from 30% to 80% by increasing homology size and optimization of gRNA sequences. Cycle time was further decreased by streamlining of screening and sequence analysis. Currently, we are simultaneously constructing eight recoded clusters, aiming for 500 kb of recoded DNA in each. Thus far, no significant decrease in growth rate has been observed in intermediate strains carrying up to 200 kb of genomically recoded DNA (~200 recoded genes). In parallel, we are actively pursuing de-risking of the final assembly protocol by conjugative assembly (CAGE) of recoded clusters. As we approach the final assembly steps of a virus-resistant *E.coli* genome, intermediate clusters are also used to test expression of genes which depend on non-standard amino acids, to enable strict biocontainment of the final strain. Our work expands the toolkit available for large scale engineering in living cells and opens new avenues for rewriting genomes.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, under Award Number DE-FG02-02ER63445 and Defense Advanced Research Projects Agency grant BAA-12-64. MML and GMC are inventors on patent application #62350468 submitted by President and Fellows of Harvard College that covers Methods for Rule-based Genome Design. Dr. Church is a founder of the following companies in which he has related financial interests: ReadCoor; EnEvolv; and 64-x. For a complete list of Dr. Church's financial interests, see also http://arep.med.harvard.edu/gmc/tech.html.