

Improved bacterial recombineering by parallelized protein discovery

Timothy M. Wannier^{a,1} , Akos Nyerges^{a,b} , Helene M. Kuchwara^a, Márton Czikkely^b, Dávid Balogh^b, Gabriel T. Filsinger^a , Nathaniel C. Borders^a, Christopher J. Gregg^a, Marc J. Lajoie^a , Xavier Rios^a, Csaba Pál^b, and George M. Church^{a,1} 

^aDepartment of Genetics, Harvard Medical School, Boston, MA 02115; and ^bSynthetic and Systems Biology Unit, Institute of Biochemistry, Biological Research Centre, Szeged HU-6726, Hungary

Edited by Susan Gottesman, National Institutes of Health, Bethesda, MD, and approved April 17, 2020 (received for review January 30, 2020)

Exploiting bacteriophage-derived homologous recombination processes has enabled precise, multiplex editing of microbial genomes and the construction of billions of customized genetic variants in a single day. The techniques that enable this, multiplex automated genome engineering (MAGE) and directed evolution with random genomic mutations (DIVERGE), are however, currently limited to a handful of microorganisms for which single-stranded DNA-annealing proteins (SSAPs) that promote efficient recombineering have been identified. Thus, to enable genome-scale engineering in new hosts, efficient SSAPs must first be found. Here we introduce a high-throughput method for SSAP discovery that we call “serial enrichment for efficient recombineering” (SEER). By performing SEER in *Escherichia coli* to screen hundreds of putative SSAPs, we identify highly active variants PapRecT and CspRecT. CspRecT increases the efficiency of single-locus editing to as high as 50% and improves multiplex editing by 5- to 10-fold in *E. coli*, while PapRecT enables efficient recombineering in *Pseudomonas aeruginosa*, a concerning human pathogen. CspRecT and PapRecT are also active in other, clinically and biotechnologically relevant enterobacteria. We envision that the deployment of SEER in new species will pave the way toward pooled interrogation of genotype-to-phenotype relationships in previously intractable bacteria.

recombineering | genome engineering | synthetic biology | RecT | MAGE

The increasing accessibility of whole-genome sequencing to microbiologists has created a gap between researchers’ ability to read vs. their ability to write/edit genetic information. As the availability of sequencing data increases, so too does its hypothesis-generating capacity, which motivates the development of genome-editing tools that can be employed to create user-defined genetic variants in a massively parallelizable manner. Currently, most techniques for editing microbial genomes that meet these criteria exploit bacteriophage-derived homologous recombination processes (1–3). Molecular tools derived from phages enable recombineering (recombination-mediated genetic engineering), which uses short homology arms to efficiently direct the integration of double-stranded DNA cassettes and single-stranded DNA (ssDNA) into bacterial genomes (4–6). Improvements to recombineering in *Escherichia coli* enabled multiplex, genome-scale editing and the construction of billions of genetic variants in a single experiment (7–9). Two of these recombineering-based methods, multiplex automated genome engineering (MAGE) and directed evolution with random genomic mutations (DIVERGE), have been used for a variety of high-value applications (3, 10–14), but at their core they offer the ability to generate populations of bacteria that can contain billions of precisely targeted mutations. However, while these techniques function well in *E. coli* and some closely related enterobacteria, efforts to reproduce these results in other bacterial species have been sporadic and stymied by low efficiencies (SI Appendix, Table S1) (15–31).

The incorporation of genomic modifications via oligonucleotide annealing at the replication fork, called oligo-mediated

recombineering, is the molecular mechanism that drives MAGE and DIVERGE (4, 32, 33). This method was first described in *E. coli*, and is most commonly promoted by the expression of *bet* (here referred to as Red β) from the Red operon of *Escherichia* phage λ (5, 6, 34). Red β is an ssDNA-annealing protein (SSAP) whose role in recombineering is to anneal ssDNA to complementary genomic DNA at the replication fork. Although improvements to recombineering efficiency have been made (7–9), the core protein machinery has remained constant, with Red β representing the state-of-the-art in *E. coli* and enterobacterial recombineering. Red β additionally does not adapt well to use outside of *E. coli*, displaying host tropism, presumably toward hosts that are targets of infection for *Escherichia* phage λ . To enable recombineering in organisms in which Red β does not work efficiently, most often Red β homologs from prophages of genetically similar bacteria are screened (15, 16, 21, 23, 30), but high levels of recombineering efficiency, as seen in *E. coli*, remain elusive.

Significance

Bacterial recombineering allows researchers to interrogate microbes by modifying their genomic DNA. Improvements to the efficiency of recombineering have allowed many simultaneous edits to be made at once. Here we describe “serial enrichment for efficient recombineering” (SEER), a method for identifying efficient single-stranded DNA-annealing proteins (SSAPs) in a microbe of interest. We use SEER to identify two SSAPs: 1) CspRecT doubles editing efficiency over Red β , the state-of-the-art in *Escherichia coli* recombineering; and 2) PapRecT achieves high efficiency in *Pseudomonas aeruginosa*, a widely studied human pathogen. We show that these SSAPs work effectively across a broad range of Gammaproteobacteria, demonstrate vastly improved performance in multiplex applications, and provide broad host-range plasmid resources.

Author contributions: T.M.W., A.N., G.T.F., C.J.G., M.J.L., X.R., C.P., and G.M.C. designed research; T.M.W., A.N., H.M.K., M.C., D.B., and N.C.B. performed research; T.M.W., A.N., H.M.K., M.C., and D.B. analyzed data; and T.M.W. and A.N. wrote the paper.

Competing interest statement: G.M.C. has related financial interests in EnEvolv, GRO Biosciences, and 64-x. G.M.C., C.J.G., M.J.L., and X.R. have submitted a patent application relating to pieces of this work (WO2017184227A2). T.M.W., G.T.F., and G.M.C. have submitted a patent application related to the improved single-stranded DNA-annealing proteins variants referenced here. A.N. and C.P. have submitted a patent application related to directed evolution with random genomic mutations (DIVERGE) (PCT/EP2017/082574 [WO2018108987] Mutagenizing Intracellular Nucleic Acids).

This article is a PNAS Direct Submission.

Published under the PNAS license.

Database deposition: All next-generation sequencing data are available in SI Appendix and Dataset S1. Plasmids are available with Addgene, and all code and supporting data has been made available on GitHub at <https://github.com/churchlab/SEER>.

¹To whom correspondence may be addressed. Email: timothy_wannier@hms.harvard.edu or gchurch@genetics.med.harvard.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2001588117/-DCSupplemental>.

First published May 28, 2020.