ACS SyntheticBiology

Letter

pubs.acs.org/synthbio

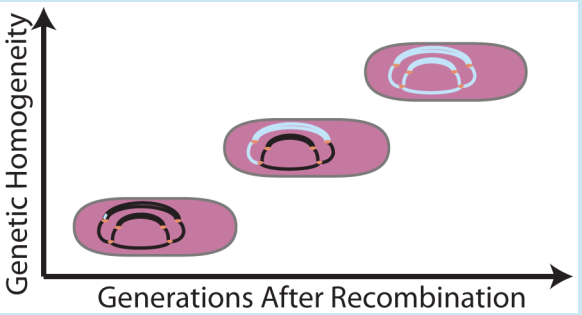
Quantifying Impact of Chromosome Copy Number on Recombination in *Escherichia coli*

T. Steele Reynolds and Ryan T. Gill*

Department of Chemical and Biological Engineering, 596 UCB, JSCBB, University of Colorado Boulder, Boulder, Colorado 80309, United States

Supporting Information

ABSTRACT: The ability to precisely and efficiently recombineer synthetic DNA into organisms of interest in a quantitative manner is a key requirement in genome engineering. Even though considerable effort has gone into the characterization of recombination in *Escherichia coli*, there is still substantial variability in reported recombination efficiencies. We hypothesized that this observed variability could, in part, be explained by the variability in chromosome copy number as well as the location of the replication forks relative to the recombination site. During rapid growth, *E. coli* cells may contain several pairs of open replication forks. While recombineered forks are resolving and segregating within the population, changes in apparent recombineering efficiency should be observed. In the case of dominant phenotypes, we predicted and then experimentally confirmed that the apparent recombination efficiency declined during recovery until complete segregation of recombineered and wild-type genomes had occurred. We observed the reverse trend for recessive phenotypes. The observed changes in apparent recombination efficiency were found to be in agreement with mathematical calculations based on our proposed mechanism. We also provide a model that can be used to estimate the total segregated recombination efficiency based on an initial efficiency and growth rate. These results emphasize the importance of employing quantitative strategies in the design of genome-scale engineering efforts.



KEYWORDS: *genome engineering, recombination efficiency*

The continued development of the lambda red recombination system (recombineering) and multiplex single-strand DNA (ssDNA) synthesis technologies has enabled increasingly complex genome-scale design and engineering efforts.[1−7] As this complexity increases, so does the need to reliably calculate genome editing efficiencies used to estimate progress throughout construction efforts.[5,7−9] Reported efficiencies for ssDNA recombination range from less than 1%[10] to as high as 70%.[11] Efforts within our own lab often result in recombination efficiencies that vary by an order of magnitude. In such cases, the amount of time required to construct a specific genome design or create a genome-scale library by recombination techniques, such as MAGE,[5] will similarly vary, resulting in a variety of subsequent challenges (e.g., incomplete library coverage, an inability to estimate design completion dates, etc.) to the overall engineering platform.

Recombination efficiency is traditionally measured by using single- or double-strand DNA (dsDNA) oligomers (<200 nucleotides) to restore the function of a previously altered gene that encodes an easily scored phenotype.[11] While not often considered, the mechanism of recombination can affect the results of such efficiency assays. During periods of rapid growth, when recombination is typically performed, *Escherichia coli* cells contain multiple pairs of open replications forks on the same genome, which enables cell doubling times less than the time needed for chromosomal replication. It has been reported that *E. coli* can contain as many as 16 partial copies of its chromosome depending on growth conditions.[12] Multiple replication forks result in the presence of multiple partially complete chromosomes within the same cell, which provide multiple targets for oligo recombination.[13] Modifying only a subset of these target sites leads to genetic polymorphism, potentially confounding the results of a recombination assay.[11,14−18] Herein, we describe a mechanism explaining how genetic polymorphism would impact recombination efficiency when measured using a screenable phenotype.

Two recombination scenarios are presented in Figure 1 for a dominant recombinant phenotype coded on the lagging strand. In Figure 1A, a recombination event in a cell with two open replication forks (one pair) is shown. During recombination, the oligonucleotide anneals to the exposed lagging strand template before synthesis of an Okazaki fragment can occur, resulting in a mismatch with that template. After one generation of recovery following recombination, a replication fork uses the recombineered DNA as the template for leading-strand synthesis and the wild-type DNA as the template for lagging-strand synthesis. Cell division results in one mutant and

Received: October 15, 2014
Published: March 12, 2015

DOI: 10.1021/sb500338g
ACS Synth. Biol. 2015, 4, 776−780

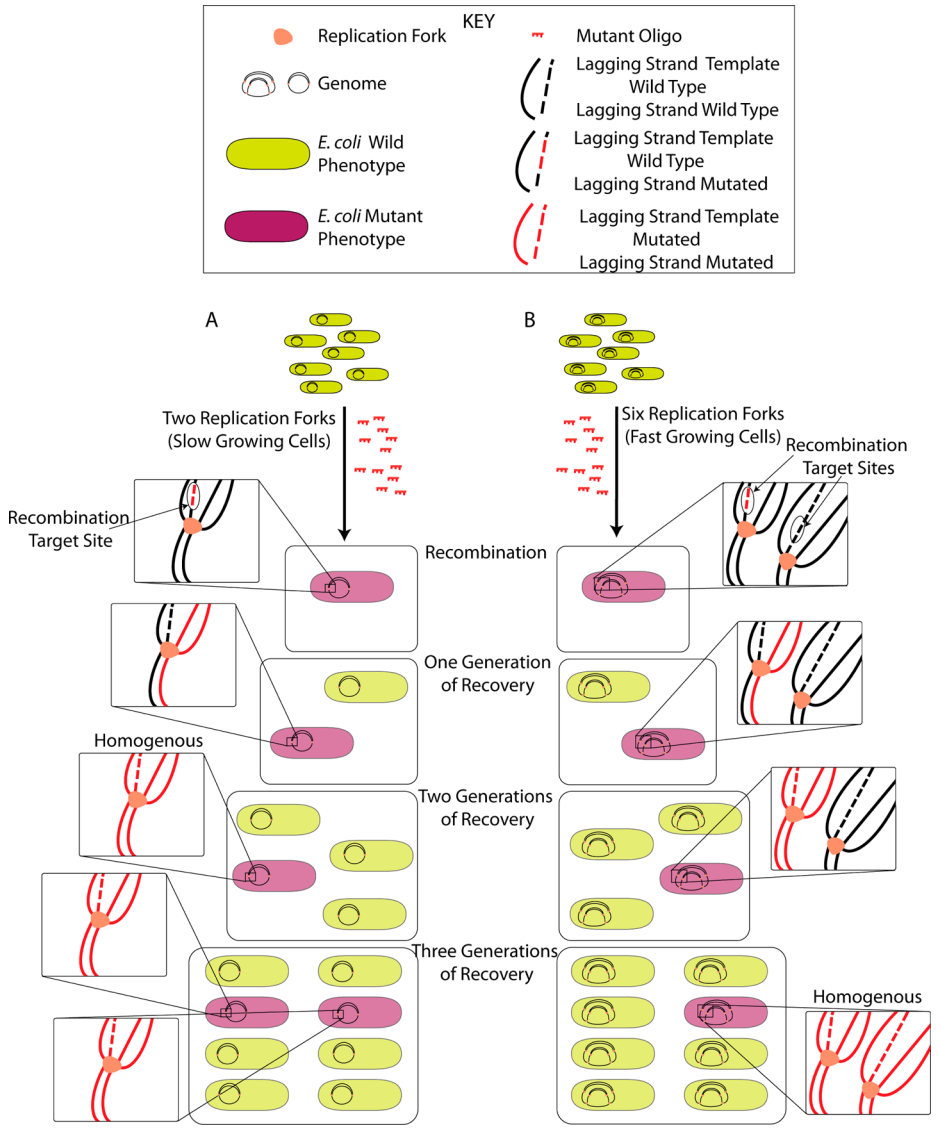


Figure 1. Impact of replication fork number on final recombination efficiency. The freshly synthesized lagging strand is indicated by a dashed line. (A) Recombination in a cell with two open replication forks is shown. (B) Recombination in a cell with six open replication forks is shown. Because more forks must be resolved in scenario B to reach genotypic homogeneity within the recombineered cell, the final recombination efficiency is lower than that in the two replication fork scenario.

one wild-type cell. After a second generation of recovery, another pass of the replication fork at the recombination site uses the recombineered DNA as the template for both leading- and lagging-strand synthesis, resulting in a homogeneous mutant cell. Cell division yields one mutant and three wild-type cells. The total segregated recombination efficiency is one-fourth the apparent efficiency immediately following recombination.

Figure 1B depicts recombination in a cell containing six replication forks (three pairs). In this scenario, the cells are growing faster than the time required for complete chromosome synthesis. In order to maintain steady-state growth, after each cell division a new pair of replication forks is opened on all available *oriC* sites.[19] This results in multiple replication forks proceeding simultaneously on the same chromosome or partially complete chromosome. With two open forks inside the cell at the recombination target site, there are now two loci where recombination can occur. As a result, one additional generation of recovery is required to obtain homogeneous recombineered cells compared to the two replication forks scenario. This results in a total segregated recombination efficiency one-eighth of the initial apparent recombination efficiency.

The simple mechanism described above can be modeled using inputs of growth rate, genomic position of target mutation, initial apparent recombination efficiency, and rate of genome replication. By estimating the average number of replication forks in front of the target mutation (which depends on the distance from the origin and growth rate), the total segregated (or final) recombination efficiency can be computed based on the same basic mathematics laid out in Figure 1. Importantly, growth rate, initial recombination efficiency, and genomic position[20] data are relatively simple to obtain. For this study, genome replication rate was estimated assuming a C period of 41 min and a D period of 22 min.[21] This basic model, along with a sensitivity analysis, is included as Supporting Information. The estimates provided by the model are compared to experimental data in Table 1.

DOI: 10.1021/sb500338g
ACS Synth. Biol. 2015, 4, 776−780

Table 1. Measured and Predicted Segregated Recombination Efficiencies

activated gene	initial apparent recombination efficiency	total segregated efficiency ± SD	predicted segregated efficiency	specific growth rate (h^{-1})
galK	30%	4.1 ± 0.5%	3.8%	2.0
mtlA	30%	2.7 ± 0.6%	2.6%	2.5
lacZ	11%	1.9 ± 0.2%	1.7%	1.7

To validate this model, we inserted loss of function mutations (stop codons) into the genes *mtlA* and *galK* in *E. coli* strain EcNB14 using ssDNA lambda phage recombineering.[3] We then used recombination to simultaneously restore function of both genes. Following recombination, cells were recovered in LB and plated on MacConkey agar containing mannitol (for *mtlA* screening) or galactose (for *galK* screening) at approximately 30 min intervals. After 24 h of growth on the MacConkey agar plates, recombination efficiency was assessed by counting pink and white colonies (Figure 2). The percentage of colonies metabolizing mannitol (functional *mtlA)* after 30 min was found to be 30%, and the percentage of colonies metabolizing galactose (functional *galK*) after 30 min of recovery was also found to be 30%. As predicted, the percentage of colonies metabolizing mannitol and galactose decreased with recovery time. After 5 to 6 h of recovery time, the apparent recombination efficiency dropped to 2.7 and 4.0% for *mtlA* and *galK*, respectively, after which the recombination efficiency was observed to be stable.

To further validate the model proposed above, individual *galK* "on" colonies (galactose metabolizing) were picked from successive recovery plates and restreaked onto a second MacConkey agar plate (Supporting Information Figure 1). The restreak of the colony from the earliest time point is mostly white, indicating that *galK* is inactive in many of the daughter cells from the picked colony, whereas streaks from successively later time points are increasingly pink, i.e., *galK* active. These results further support the hypothesis that cells engineered by ssDNA recombineering become increasingly genetically homogeneous as the recovery time increases.

We confirmed the above results were not due to the use of the MacConkey agar assay, the particular *E. coli* strain employed, or the *galK* and *mtlA* genes evaluated. Specifically, using *E. coli* EcNR2, we first disrupted *lacZ* function and then restored function in the same manner as above. *lacZ* function was scored using blue/white screening on LB agar plates containing Xgal. The recombination efficiency was lower for *lacZ* than that for *mtlA* and *galK*. However, as shown in Figure 3, the percentage of colonies metabolizing lactose similarly

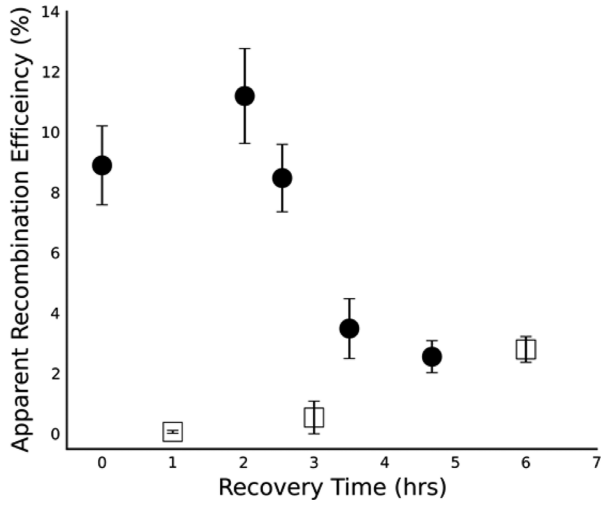


Figure 3. Comparison of dominate and recessive phenotypes. Recombination efficiency of the *lacZ* on oligo is shown (filled circles). The resulting *lacZ* on cells were then deactivated using the *lacZ* off oligo (empty squares). Error bars represent standard deviation.

decreases as recovery proceeds, with apparent recombination efficiency declining from an observed maximum of 11% to 2.5% after 4.5 h of recovery (Figure 3). After approximately 24 h of recovery, 1.9% of cells were found to have *lacZ* function.

The results of these three experiments were compared to predictions from the model described above using a binomial distribution. It was found that all three total segregated recombination efficiency measurements were within 1 standard deviation of their predicted value.

To further test the mechanism presented in Figure 1, we next measured apparent recombination efficiency as a function of recovery time for a recessive phenotype. Genetic homogeneity is necessary for a recessive phenotype to be observed. Therefore, we predicted apparent recombination efficiency for a recessive phenotype should increase during recovery time, as genetic homogeneity increases, rather than decline, as seen in a dominate phenotype. To test this prediction, *lacZ* function was

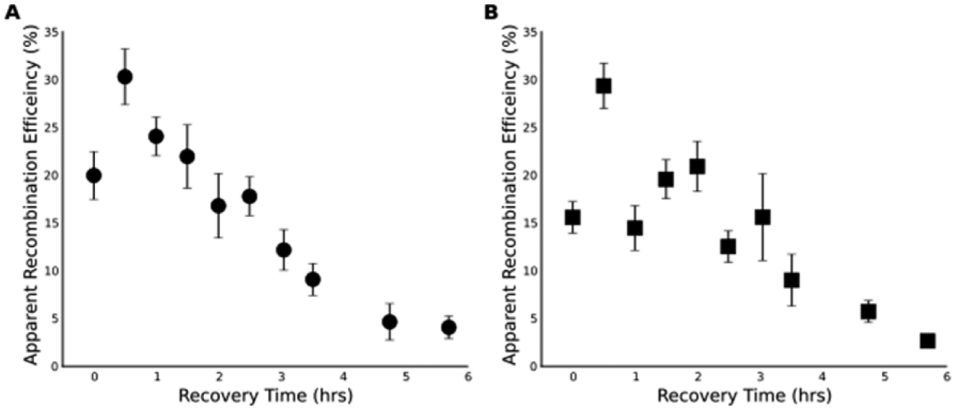


Figure 2. Recombination at different genomic positions. (A) Measured values for the recombination efficiency of the *galk* on oligo. (B) Recombination efficiency for the *mtlA* on oligo. Error bars represent standard deviation.

DOI: 10.1021/sb500338g
ACS Synth. Biol. 2015, 4, 776−780

terminated by the same ssDNA recombineering approaches used above in one of the EcNR2 colonies with functional *lacZ*. Initial apparent recombination efficiency was near 0% (Figure 3), efficiency increased during recovery, and eventually 2.5% of colonies showed no *lacZ* function. It is interesting to note the total segregated recombination efficiencies of the *lacZ* on and off oligos were similar (2.5 vs 1.9%), suggesting that difficulties in detecting recessive gene mutations may be due to the use of insufficient recovery times.

The importance of using total segregated recombination efficiency instead of the initial apparent recombination efficiency is best illustrated by considering a population of cells subjected to successive rounds of recombination. Suppose a genome engineering project required recombination at 10 sites in the genome of a single cell. If, mistakenly, the initial apparent recombination efficiency is used as measured for *galK* above (30%), then it would be estimated[22] that eight rounds of recombination would be required for 50% of the cell population to contain all 10 desired mutations. However, if the total segregated recombination efficiency was used (4.1%), approximately 65 rounds of recombination would be required for 50% of the cell population to have all desired mutations. Accurate estimations of the time and effort required for genome engineering projects are crucial to equipment scheduling and experimental efficiency.

The scale and complexity of synthetic biology and genome engineering efforts continue to expand. Understanding the impact of chromosome replication on recombination efficiency is important to developing robust strategies for multiplex genome engineering. For this reason, we advocate that recombination efficiencies be reported as total segregated recombination efficiencies rather than apparent efficiencies. Furthermore, we provide a tool to help predict total segregated efficiency, which should aid in planning genome engineering projects. Additionally, obtaining genetic homogeneity, via increased recombination recovery time, in mutant cells should ensure long-term genetic stability of mutations. The tool and associated understanding presented here can aid in the development of more robust genome editing and engineering efforts in a broad range of applications.

METHODS

Strains and Recombination. *lacZ* experiments were performed in EcNR2 (specific growth rate: 2.3 h^{-1}). EcNR2 is a derivative of *E. coli* MG1655 with ΔbioA::*λ*-prophage and ΔmutS::cmR.[5] EcNB14 is derived from EcNR2 with mtlA::-STOP and galk::STOP. The oligos used in this study are listed in Table 2.

Recombination. Recombination methods have been described previously.[23] Briefly, *E. coli* cells were grown overnight in lysogeny broth (LB). Cells were then transferred to shake flasks and grown to an optical density of approximately 0.6. One milliliter of cells was spun at 14 000 rpm for 1 min, and the pellet was resuspended with 1 mL of chilled ultrapure water. This process was repeated two additional times, and the water was decanted after each spin. One microliter of 100 μM oligo was added to the pellet and resuspended in 50 μL of ultrapure water. Cells were then electroporated, transferred to 3 mL of LB, and allowed to recover at 30 °C.

Recovery Time Assay. At 30 min time intervals, the transformed cells were diluted and spread on MacConkey agar plates for *mtlA* and *galK* or LB Agar plates with X-gal for *lacZ* in triplicate. Plates were incubated at 30 °C. After approximately

Table 2. Oligos Used in the Study

oligo	sequence
lacZ off	G★G★A★A★ACAGCTATGACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCG
lacZ on	G★G★A★A★ACAGCTATGACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCG
mtlA on	C★T★C★G★GTTCTATGATTGCAGGTCCGCTGGGCGGCTGGTGCATTAAGCACTTCGACCGCTGGGTAGACGGTAAGATCAAATCCGGTTTTGAG
galK on	G★C★T★T★CACTGGAAGTCGCGGTCGGAACCGTATTGCAGCAGCTTTATCATCTGCCGCTGGACGGCGCACAAATCGCGCTTAACGGTCAGGAA

DOI: 10.1021/sb500338g
ACS Synth. Biol. 2015, 4, 776−780

36 h, the colonies on the plates were counted. For each data point, an average of 993 colonies were counted. Apparent recombination efficiency at each time point was determined by dividing the number of recombinant colonies (pink for *mtlA and galK* on, blue for *lacZ* on, and white for *lacZ* off) by the total number of colonies on the respective plate.

■ ASSOCIATED CONTENT

Supporting Information

Figure S1: Colonies expressing *galK* following *galk* on oligo recombination. Spreadsheet of the model described in this work. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: (303) 492-2627; E-mail: rtg@colorado.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank OPX Biotechnologies for their financial support of this work as part of a sponsored research agreement with the University of Colorado. We also thank N. Boyle (Colorado School of Mines) for providing the strains and oligos used in this study and D. Clough (University of Colorado) for assistance with statistical techniques.

■ REFERENCES

(1) Murphy, K. C. (1998) Use of bacteriophage λ recombination functions to promote gene replacement in *Escherichia coli*. *J. Bacteriol. 180*, 2063−2071.

(2) Zhang, Y., Buchholz, F., Muyrers, J. P., and Stewart, A. F. (1998) A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat. Genet. 20*, 123−128.

(3) Yu, D., Ellis, H. M., Lee, E.-C., Jenkins, N. A., and Copeland, N. G. (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A. 97*, 5978−5983.

(4) Alper, H., Miyaoku, K., and Stephanopoulos, G. (2005) Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets. *Nat. Biotechnol. 23*, 612−616.

(5) 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.

(6) 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., and Mosberg, J. A. (2013) Genomically recoded organisms expand biological functions. *Science 342*, 357−360.

(7) Sandoval, N. R., Kim, J. Y., Glebes, T. Y., Reeder, P. J., Aucoin, H. R., Warner, J. R., and Gill, R. T. (2012) Strategy for directing combinatorial genome engineering in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A. 109*, 10540−10545.

(8) Carr, P. A., Wang, H. H., Sterling, B., Isaacs, F. J., Lajoie, M. J., Xu, G., Church, G. M., and Jacobson, J. M. (2012) Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection. *Nucleic Acids Res. 40*, e132−e132.

(9) Swingle, B., Markel, E., Costantino, N., Bubunenko, M. G., Cartinhour, S., and Court, D. L. (2010) Oligonucleotide recombination in Gram-negative bacteria. *Mol. Microbiol. 75*, 138−148.

(10) Li, X. t., Thomason, L. C., Sawitzke, J. A., Costantino, N., and Court, D. L. (2013) Bacterial DNA polymerases participate in oligonucleotide recombination. *Mol. Microbiol. 88*, 906−920.

(11) Sawitzke, J. A., Costantino, N., Li, X.-t., Thomason, L. C., Bubunenko, M., Court, C., and Court, D. L. (2011) Probing cellular processes with oligo-mediated recombination and using the knowledge gained to optimize recombineering. *J. Mol. Biol. 407*, 45−59.

(12) Sergueev, K., Court, D., Reaves, L., and Austin, S. (2002) *E. coli* cell-cycle regulation by bacteriophage lambda. *J. Mol. Biol. 324*, 297−307.

(13) Boyle, N. R., Reynolds, T. S., Evans, R., Lynch, M., and Gill, R. T. (2013) Recombineering to homogeneity: extension of multiplex recombineering to large-scale genome editing. *Biotechnol. J. 8*, 515−522.

(14) Maresca, M., Erler, A., Fu, J., Friedrich, A., Zhang, Y., and Stewart, A. F. (2010) Single-stranded heteroduplex intermediates in λ Red homologous recombination. *BMC Mol. Biol. 11*, 54.

(15) Skerker, J. M., Leon, D., Price, M. N., Mar, J. S., Tarjan, D. R., Wetmore, K. M., Deutschbauer, A. M., Baumohl, J. K., Bauer, S., and Ibáñez, A. B. (2013) Dissecting a complex chemical stress: chemogenomic profiling of plant hydrolysates. *Mol. Syst. Biol. 9*, 674.

(16) Witkin, E. M. (1951) Nuclear segregation and the delayed appearance of induced mutants in *Escherichia coli*. *Cold Spring Harbor Symp. Quant. Biol. 16*, 357−372.

(17) Poteete, A. R., Fenton, A. C., and Nadkarni, A. (2004) Chromosomal duplications and cointegrates generated by the bacteriophage lamdba Red system in *Escherichia coli* K-12. *BMC Mol. Biol. 5*, 22.

(18) Morris, R. T., and Drouin, G. (2004) Ectopic gene conversions in four *Escherichia coli* genomes: increased recombination in pathogenic strains. *J. Mol. Evol. 58*, 596−605.

(19) Skarstad, K., Steen, H. B., and Boye, E. (1985) *Escherichia coli* DNA distributions measured by flow cytometry and compared with theoretical computer simulations. *J. Bacteriol. 163*, 661−668.

(20) Keseler, I. M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martínez, C., Fulcher, C., Huerta, A. M., Kothari, A., and Krummenacker, M. (2013) EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res. 41*, D605−D612.

(21) Hill, N. S., Kadoya, R., Chattoraj, D. K., and Levin, P. A. (2012) Cell size and the initiation of DNA replication in bacteria. *PLoS Genet. 8*, e1002549.

(22) Gallagher, R. R., Li, Z., Lewis, A. O., and Isaacs, F. J. (2014) Rapid editing and evolution of bacterial genomes using libraries of synthetic DNA. *Nat. Protoc. 9*, 2301−2316.

(23) Sharan, S. K., Thomason, L. C., Kuznetsov, S. G., and Court, D. L. (2009) Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protoc. 4*, 206−223.

DOI: 10.1021/sb500338g
ACS Synth. Biol. 2015, 4, 776−780