

Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection

Peter A. Carr^{1,2,*}, Harris H. Wang^{3,4,*}, Bram Sterling^{1,2}, Farren J. Isaacs⁵,
Marc J. Lajoie^{6,7}, George Xu^{6,8,9}, George M. Church^{4,6,*} and Joseph M. Jacobson^{1,2}

¹The Center for Bits and Atoms, Massachusetts Institute of Technology, Cambridge, MA 02139, ²Media Lab, Massachusetts Institute of Technology, Cambridge, MA 02139, ³Department of Systems Biology, Harvard Medical School, Boston, MA 02115, ⁴Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, ⁵Department of Molecular, Cellular and Developmental Biology, Systems Biology Institute, Yale University, New Haven, CT 06520, USA, ⁶Department of Genetics, Harvard Medical School, Boston, MA 02115, ⁷Program in Chemical Biology, Harvard University, Cambridge, MA 02138, ⁸Program in Medical Engineering Medical Physics, Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA 02139 and ⁹Program in Biophysics, Harvard University, Cambridge, MA 02138, USA

Received December 24, 2011; Revised March 27, 2012; Accepted April 30, 2012

ABSTRACT

Genome-scale engineering of living organisms requires precise and economical methods to efficiently modify many loci within chromosomes. One such example is the directed integration of chemically synthesized single-stranded deoxyribonucleic acid (oligonucleotides) into the chromosome of *Escherichia coli* during replication. Herein, we present a general co-selection strategy in multiplex genome engineering that yields highly modified cells. We demonstrate that disparate sites throughout the genome can be easily modified simultaneously by leveraging selectable markers within 500 kb of the target sites. We apply this technique to the modification of 80 sites in the *E. coli* genome.

INTRODUCTION

New capabilities that enhance the engineering of organisms at the whole-genome scale provide avenues to construct biological systems with new properties. Such engineering can produce minimized (1) or mosaic (2) genomes, or ones that may contain new genes, pathways, metabolisms and fundamentally different regulatory structure (3). However, these projects require significant time and resources to accomplish by traditional genetic engineering or lab-scale

evolution approaches. Prior work demonstrated that targeted chromosomal modifications could be efficiently introduced in *Escherichia coli* using synthetic oligonucleotides (oligos) complementary to the lagging strand of the replicating chromosome (4–6), which we refer to as oligo-mediated λ -Red allelic replacement (AR). This type of method has been used in other prokaryotic and eukaryotic systems (7–9). Such an approach provides several advantages: AR is an extremely general mechanism; user-defined oligos can target anywhere in the chromosome without a need for site-specific nucleases; no antibiotic or other functional selection is necessary and the mutagenesis process leaves no sequence-based ‘scars’ in the genome. Furthermore, oligo-mediated genome engineering can also be multiplexed and automated (10).

We recently described multiplex automated genome engineering (MAGE), using AR to combinatorially modify 24 targeted sites throughout the genome, to rapidly increase the output of a metabolic pathway (10). We have also applied MAGE to the modification of hundreds of genome sites of *E. coli* MG1655 in pursuit of a re-engineered genetic code (11). Herein, we present a general co-selection (CoS) strategy based on MAGE to isolate highly modified cells with many chromosomal modifications. We demonstrate that one or more selectable genetic markers (within ~500 kb of the targeted sites) can be used to obtain as many as eight targeted modifications in a single MAGE cycle. We further iterate these

*To whom correspondence should be addressed. Tel: +1 781 981 7488; Fax: +1 781 981 7848; Email: carr@ll.mit.edu

Correspondence may also be addressed to Harris H. Wang. Tel: +1 617 955 9575; Fax: +1 617 432 6513; Email: harris.wang@wyss.harvard.edu

Correspondence may also be addressed to George M. Church. Tel: +1 617 432 7562; Fax: +1 617 432 6513;

Email: gchurch@genetics.med.harvard.edu

Present addresses:

Peter A. Carr, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA 02420, USA.

Bram Sterling, Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA.

The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

cycles to accumulate many more modifications over a 1.1 Mb span of the *E. coli* chromosome. This type of CoS strategy can also be applied to incorporate multiple new regulatory elements into the chromosome (12).

MATERIALS AND METHODS

Multiplex automated genome engineering

MAGE uses single-stranded oligos to modify the deoxyribonucleic acid (DNA) sequence of many different chromosomal sites *in vivo*, frequently applying multiple oligos over several iterations (10). A singleplex, single cycle of MAGE is comparable with the oligo recombineering technique of Court and co-workers (4,6,13).

Allele switching and selection

MAGE was used to switch gene functions off and on *in vivo*, using CoS oligos that either introduce (and later remove) a stop codon or remove (and later re-introduce) the translation start site. In multiplex experiments, dilute quantities of these CoS oligos were used, typically at 1% (each) of the total oligo concentration. For example, we find that a 10-plex cycle of MAGE is maximized when using 0.5 μM of each oligo (5 μM total) and 0.05 μM of each CoS oligo. The selectable genes used in this study are *kan*, *bla*, *cat* and *tolC*. The *lacZ* gene was also used as a means to quantify and screen for modified cells.

Media, chemicals and reagents

Liquid cultures of all strains were grown in LB-Lennox media (referred to hereafter as LB) containing tryptone (10 g/l), yeast extract (5 g/l) and NaCl (5 g/l) and buffered to pH 7.45 with NaOH. Chloramphenicol (*cat*), kanamycin (*kan*) or carbenicillin (*carb*) were added to LB cultures or LB-agar plates (LB with 15 g agar/L) at concentrations of 20 $\mu\text{g/ml}$, 30 $\mu\text{g/ml}$ or 50 $\mu\text{g/ml}$, respectively. X-Gal (40 $\mu\text{g/ml}$) and isopropyl β -D-1-thiogalactopyranoside (0.1 mM) were used on LB agar plates for functional assay of β -galactosidase activity. Multiplex polymerase chain reaction (PCR) kits were purchased from Qiagen (Cat no. 206143). Standard agarose gel electrophoresis reagents were used. Colicin E1 was expressed in strain JC411 and purified as described by Schwartz and Helinski (14).

Primers and other oligos

All oligos were obtained from Integrated DNA Technologies with no additional purification. Oligos for AR contained either two phosphorothioate linkages at the 3' and 5' terminus or four phosphorothioate linkages at the 5' terminus unless designated otherwise. We have found that protection of the oligo with at minimum two phosphorothioate linkages at the 5' terminus improves MAGE efficiency by a factor of 2 (10).

Escherichia coli strains

Construction of EcNR1, EcNR2 and EcFI5 strains was previously documented (10). In brief, our λ -Red construct

(including the ampicillin resistance gene *bla*) derived from a defective temperature-inducible λ was introduced by P1 transduction to *E. coli* MG1655 to produce EcNR1 (Δ *bioA:: λ -Red-*bla**). EcNR2 is an EcNR1 derivative with Δ *mutS::cat*. EcFI5 is an EcNR2 derivative with Δ *galK::kan* and made kanamycin and chloramphenicol sensitive by inactivating the *kan* and *cat* gene using two oligos *kan_off* and *cat_off* that introduced a nonsense mutation in each gene to create the genotype Δ *galK::kan(-)* and Δ *mutS::cat(-)*. EcBS2 is derived from EcFI5 using oligo *bla_off* to deactivate the *bla* gene. EcBS3 is derived from EcFI5 using oligo *tolC_off* to deactivate the *tolC* gene. EcBS5 was derived from EcNR1 by (i) switching the *mutS* gene into an inactive state using MAGE and oligo *mutS_off* (Supplementary Table S4); (ii) deleting the endogenous *tolC* gene and (iii) inserting the *tolC* gene within the region to be modified, between the *yohC* and *yohD* genes. (Experiments using other strains than EcBS5 used the endogenous *tolC* marker.) Colicin E1 expression strain JC411 was obtained from Roberto Kolter.

Oligo-mediated AR

Allele-replacement-competent cells were generated as described previously (10). In brief, individual colonies from a freshly streaked overnight plate were inoculated into 3 ml LB aliquots and grown in a rotator drum at 300 rpm at 32°C. On reaching OD₆₀₀ of 0.7, the glass tubes were moved to a 42°C shaking water bath for 15 min to induce the expression of λ -Red proteins. Cells are then immediately chilled on ice for at least 5 min and subsequently made electrocompetent in 1 ml aliquots by twice pelleting and resuspending in cold-sterile dH₂O. Cells are finally concentrated 20-fold into 50 μl dH₂O containing oligos. This 50 μl volume is electroporated with a BioRad GenePulser (set to 1800 V, 25 μF , 200 Ω) using a 1-mm gap cuvette. Electroporated reactions are immediately added to 3 ml of warm LB media and recovered for at least 3 h (to allow segregation of modified alleles and division into clonal daughter cells) before plating on LB-agar. A typical multiplex MAGE recombination would use a total oligo concentration of 5 μM (i.e. 250 picomoles in the 50 μl electroporation volume). Thus when using 10 non-selectable oligos, these would be present at concentrations of 0.5 μM each (25 picomoles each in the 50 μl) and when using CoS, 0.05 μM of each CoS oligo was used (2.5 picomoles each in the 50 μl , i.e. approximately 1% of the total oligo concentration).

Multiplex allele-specific colony PCR

Multiplex allele-specific colony PCR (MASC-PCR) as described previously (11,15) was used to genotype sets of clones to estimate allele replacement (AR) frequencies and assess the distribution of modifications among groups of clones. In short, two sets of primers were synthesized for each genomic locus, one corresponding to the mutant allele and one corresponding to the wild-type (WT) allele. The forward primers were identical except at the 3' terminal nucleotide that corresponded to the specific sequence of either the mutant or WT allele. The reverse

primer was the same for both alleles. Primers were designed for a target T_m of 62°C. To query the genotype in a clone, the mutant and WT allele primer sets were run in separate colony PCR reactions. A clone containing the mutant allele will generate PCR products only using the mutant allele primers and not the WT primers and vice versa for a clone with the WT allele. Non-specific amplification of both mutant and WT primers were observed when a suboptimal annealing temperature was used. A gradient PCR was thus done to experimentally determine the optimal annealing temperature, which tended to vary from 62°C to 67°C depending on the primer sequence. Multiple loci were queried in a single PCR reaction using the multiplex PCR kit from Qiagen and pooled primer sets that produced amplicons of length 100, 150, 200, 250, 300, 400, 500, 600, 700 and 850 bp, corresponding to up to 10 different genomic loci. In each 20 μ l PCR reaction, 1 μ l of a 1 in 100 dilution in water of a saturated clonal culture (i.e. produced from a colony one wishes to assay) generated the best MASC-PCR specificity. PCR cycles were heat activation and cell lysis for 15 min at 95°C, denaturing for 30 s at 94°C, annealing for 30 s at the optimized annealing temperature, extension for 80 s at 72°C, repeated cycling for 26 times and final extension for 5 min at 72°C. Gel electrophoresis on a 1.5% agarose gel produced the best separation for a 10-plex MASC-PCR reaction.

Multiplex allele-specific colony quantitative PCR

To complement the MASC-PCR analyses, we also used a highly multiplexed quantitative PCR (qPCR) screen to rapidly identify clones that contained the highest degree of modification. This technique is reported in full detail elsewhere (11). Two qPCR reactions were compared for each clone evaluated, one with 10 or more pairs of primers matched to the unmodified TAG genes and the other with the same number of primer pairs matched to the intended TAA modifications. The TAG reactions were expected to proceed most efficiently with a WT template and the TAA reactions most efficiently with a fully modified template. Intermediate values between these extremes also provided an effective, though non-linear gauge of the extent of modification for each clone.

Each colony was used as template for a pair of qPCR reactions comparing the amplification efficiency when matched to primers terminating in WT or targeted mutant sequence. The experimental measurement for a given clone is then compared with the equivalent values measured for the unmodified starting (negative control) strain. This reference value is subtracted from each ΔC_q to yield a $\Delta\Delta C_q$, with unmodified clones scoring close to zero (as with the negative control colonies). The largest $\Delta\Delta C_q$ values were expected to indicate the most modified clones, which we confirmed by genotyping clones with varying $\Delta\Delta C_q$ values. Large numbers of clones could be quickly assessed with this approach (up to 191 per 384-well plate, plus a negative control). A typical assessment of MAGE-cycled clones partitioned a 384-well plate into 4 groups of 96 wells. Each group of 96 wells was thus used to assay 48 colonies (44–46 queried colonies plus 2–4

control colonies) at 10 loci. After identification of the most promising clones, site-specific qPCR genotyping was used to identify which specific sites had been modified, selecting the best clones for further modification.

Individual bacterial colonies were picked from LB-agar plates by touching a 20 μ l pipette tip to a colony and suspending this small amount of cells in 0.5 ml sterile distilled deionized water; 5 μ l of this suspension was used as template in 20 μ l qPCR reactions containing 1 \times NovaTaq buffer, 0.5 U NovaTaq Hotstart DNA Polymerase (EMD Biosciences), 250 μ M each deoxynucleotide triphosphate (dNTP), 0.5 \times SYBR Green I (Invitrogen) and 5% dimethyl sulfoxide (DMSO). Primer concentrations were 50 nM for each of 10 primer pairs, i.e. 500 nM total forward primers and 500 nM total reverse primers. A typical qPCR program included a 10 min hot start at 95°C, followed by 40 cycles (95°C for 30 s, 60°C for 30 s and 72°C for 30 s) finishing with a melt curve analysis. All reactions were performed in a 7900 HT system (Applied Biosystems). PCR primers for all sites were designed to have a melting temperature estimated at 62°C. Reverse primers were chosen to yield amplicons in the size range of 200–225 bp. No optimization was needed for qPCR primer sequences or for multiplex/singleplex reaction conditions.

Allele frequency qPCR to determine AR frequencies

Allele frequency qPCR (AF-qPCR) was used to measure AR frequencies throughout larger cell populations produced by MAGE. This technique queries the ensemble population of a cell culture instead of individual clones. The method of Germer *et al.* (16) was modified to more accurately quantify extreme high (>90%) and low (<10%) frequencies (P.A.C., B.S. and J.M.J, in preparation). As with MASC-PCR earlier, two pairs of primers are used, matching either WT or mutant sequences to discriminate between alleles.

AF-qPCR templates were either homogeneous cultures grown in LB (positive and negative controls) or heterogeneous mixtures variable for two alleles, most frequently the TAG WT and TAA mutant stop codons described elsewhere (11). Each culture was used as template for a pair of PCR reactions comparing the amplification frequency when matched to primers terminating in WT or targeted mutant sequence. The difference in these two frequencies (in amplification threshold cycle, ΔC_q) for a control reaction defines the lower and upper (0% and 100%) limits of the measurement. The experimental measurement for a mixed culture is then compared with these reference values to calculate a percentage representation for each allele in the pool. The AF method of Germer *et al.* (16) was used as a starting point, with refinements to the calculation to more accurately determine high frequencies (90–99%) and low frequencies (1–10%); 5 μ l of cell culture (typically diluted 1:100 or 1:1000 into dH₂O) was used as template in 20 μ l qPCR reactions containing 1 \times NovaTaq buffer, 0.5 U NovaTaq Hotstart DNA Polymerase (EMD Biosciences), 250 μ M each dNTP, 0.5 \times SYBR Green I (Invitrogen) and 5% DMSO. Primer concentrations were 500 nM each. A typical qPCR program included a 10 min hot start at 95°C, followed by 40 cycles (95°C for 30 s, 60°C

for 30 s and 72°C for 30 s) finishing with a melt curve analysis. All reactions were performed in a 7900 HT system (Applied Biosystems). PCR primers for all sites were designed to have a melting temperature estimated at 62°C. Reverse primers were chosen to yield amplicons in the size range of 200–225 bp.

CoS singleplex experiments

Escherichia coli strain EcBS2 was subjected to a single MAGE cycle to modify genes at 20 locations dispersed throughout the genome. Each separate electroporation was performed as a singleplex experiment, i.e. dominated by the presence of one oligo with the intent to modify one site. Trace amounts of three other oligos were included for restoring function at selectable loci *cat*⁻, *kan*⁻ and *bla*⁻. Resultant cultures were grown under selective and non-selective conditions, and the extent of modification at each target site was assessed by AF-qPCR. The impact of CoS for these functions was calculated as a CoS factor $F = E(\text{selection})/E(\text{no selection})$.

Observing CoS synergies

AR frequency was measured for one MAGE cycle at a single chromosomal site *pphB* flanked by a pair of deactivated selectable marker genes, *tolC*⁻ and *cat*⁻. Contexts ranged from singleplex (1 oligo, mole fraction = 1) to highly multiplexed (24 different oligos, i.e. each at mole fraction = 0.04). The groupings of sites and their targeted mutations are indicated in Supplementary Table S3. Trace amounts of CoS oligos (1% of total) were included for restoring selectable function. Resulting cultures were grown under conditions selecting for re-activated proximal marker (*cat*, <5 kbp), distal marker (*tolC*, >300 kbp), both and neither.

Multicycle CoS MAGE with alternating selections

Performing multiple cycles of CoS MAGE in series requires either many markers that can each be selected once or a single marker that can alternately be selected in the on and off states. For addressing a set of 80 sites in the *E. coli* chromosome, we used many cycles of CoS MAGE using the dually selectable *tolC* gene. When the CoS oligo used

was *tolC*_{on}, the post-electroporation culture was allowed to recover at 30°C with agitation for 1 h before plating 80 µl on LB agar with carbenicillin and 0.005% sodium dodecyl sulphate (SDS) to isolate *tolC*⁺ colonies. When the CoS oligo used was *tolC*_{off}, the post-electroporation culture was allowed to recover and grow at 30°C with agitation for no less than 5 h. This culture was then allowed to grow to mid-log and OD600 of 0.4–0.6. At the same time, a known *tolC*⁺ culture was brought to the same state of growth to serve as a negative control. Each of these was used to inoculate a tube of 2 ml LB and 20 µl colicin E1 preparation with 20 µl cell growth. These cultures were allowed to grow for 8–12 h. Each was then plated (typically 100 µl of a 10⁻⁵ dilution of mid-log culture or a 10⁻⁶ dilution of confluent culture) onto LB agar with carbenicillin to isolate colonies for screening. Strain EcBS5 was used as the starting strain for this experiment. This strain has the endogenous *tolC* gene seamlessly deleted and re-inserted within the region to be modified (40 targets on either side in the chromosome).

RESULTS

To efficiently modify genomes, we introduce several distinct oligos into the cells simultaneously, which integrate into actively replicating chromosomes at high frequency (10). For any given site, the specific oligo anneals to the lagging strand of the DNA replication fork, resolving into one of the daughter cells (Figure 1) (17). Although each site is modified by incorporating a unique oligo, we hypothesized that oligos targeting multiple sites in close proximity should integrate into the same newly synthesized strand of the chromosome. When a daughter cell containing one such modification is isolated by selection, we expect that this cell would be highly enriched for other modifications at nearby sites in a co-operative manner, a process we refer to as CoS (Figure 1). With this strategy, selectable genes can be used as CoS markers in various combinations across different regions of the genome to enhance MAGE through CoS. These markers can be pre-existing in the genome or inserted into the chromosome

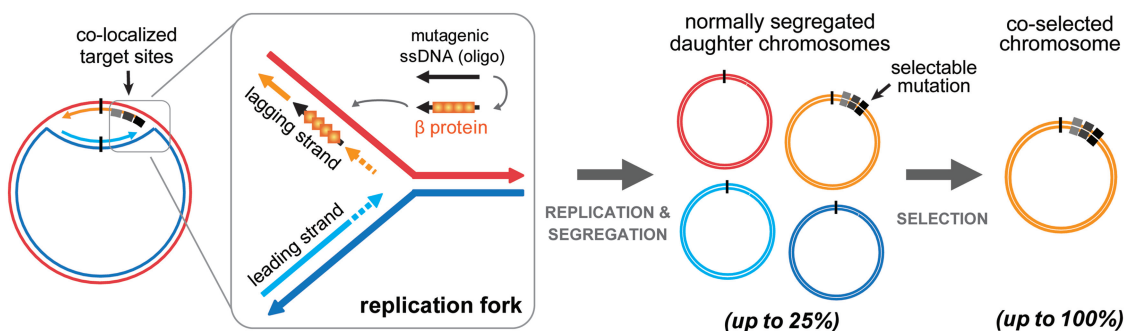


Figure 1. Mechanism for MAGE AR with CoS. The dividing chromosome is schematized, with integration of a mutagenic oligo into the genome at a replication fork [adapted from Costantino and Court (4)]. An oligo electroporated into the cell is bound by multiple copies of the λ -bacteriophage β protein and anneals to the lagging strand during DNA replication. When multiple oligos are incorporated into nearby sites (black and gray rectangles), they are predicted to co-segregate at high frequency, often inherited by the same daughter cell. Thus, a permissive replication fork seems to be a limiting factor in MAGE. Using one of these modifications to change the function of a selectable gene allows selection to remove unmodified cells.

by double-stranded DNA recombineering (8) and switched on or off by oligo-mediated AR.

To characterize the effect of CoS on *E. coli*, we first measured single AR frequencies at 20 sites spaced around the 4.6 Mb chromosome (Figure 2A). Each of the oligos used generated a single-basepair silent mutation (stop codon TAG to TAA). These sites were drawn from a larger group of 314 such targeted codon replacements reported previously (11). In addition, three CoS marker sites were chosen (two in close proximity to each other and one on the opposite replichore). These markers are inactivated *bla* (ampicillin/carbenicillin), *kan* (kanamycin) and *cat* (chloramphenicol) resistance genes on the chromosome, each encoding a reversible nonsense mutation. We used CoS oligos (*bla_on*, *kan_on* and *cat_on*) that can restore the selectable phenotype to these markers. For each of the 20 sites, a ‘singleplex’ MAGE experiment targeted the modification of one TAG site, while also including more dilute amounts of CoS oligos for CoS. Thus, each electroporation introduced one specific targeting oligo (5 μ M concentration) plus the three CoS oligos (0.05 μ M each). We then measured the resulting AR frequencies with and without CoS.

CoS yielded notable enhancement of AR frequencies, quantified as a CoS factor—the ratio of AR frequencies with and without selection (Figure 2A). Consistent with our hypothesis, we observed that CoS enhanced AR frequencies especially at sites within \sim 500 kb of the selected CoS markers on the same replichore. (We note that the rightmost data point in Figure 2A appears to show a strong *cat* CoS enhancement at a much greater distance. However, one of the two unselected AR frequencies was quite high at this locus, a potential outlier giving rise to the large error bar.) Moreover, greater enhancements were observed at sites in phase with the direction of the replication fork, i.e. ‘downstream’ of the CoS markers (farther from the origin of replication).

To further investigate these proximity effects, we assessed the effect of CoS with multiplex MAGE experiments using 24 target sites flanked by two CoS markers. These sites spanned a 320 kb region in the same replichore of the chromosome, flanked by inactivated CoS markers (*cat*⁻ and *tolC*⁻). We characterized the AR frequencies at multiple sites, with and without CoS. Figure 2B shows AR frequencies for one representative site (*pphB*) within 5 kb of the *cat*⁻ marker and more than 300 kb from the *tolC*⁻ marker. CoS for conversion of the (*cat*⁻) marker proximal to *pphB* produced a greater enhancement than CoS for conversion of the (*tolC*⁻) marker distal to *pphB*. CoS for both flanking markers produced the greatest enhancement. These effects are observed for both singleplex conditions (one site modified, i.e. using one oligo plus CoS oligos) but are greatest under highly multiplexed conditions (up to 24 distinct oligos plus CoS oligos, see also Supplementary Figure S1 and Supplementary Table S3).

To explore the extent to which CoS can enhance multiplexed genome engineering in greater detail, we quantified the AR frequencies for 37 sites throughout the chromosome. These sites were divided into four subsets (A–D, Supplementary Table S2) in varying positions relative to two CoS markers, *cat*⁻ and *kan*⁻, located on opposite replichores (Figure 3A). In Replichore 1, Group A sites are clustered in close proximity to the *kan*⁻ marker, and Group B sites are more dispersed. In Replichore 2, Group C sites are clustered in close proximity to the *cat*⁻ marker, and Group D sites are more dispersed. AR frequencies for these targets were evaluated in up to 10-plex MASC-PCR reactions (Supplementary Methods). This allowed us to both measure AR frequencies at individual sites (Figure 3B) and to assess the distributions of modifications in the resulting clones (Figure 3C). The frequencies of AR across the 37 target sites were measured by screening up to 48 isolated clones under each CoS condition (none, kanamycin, chloramphenicol or both). The

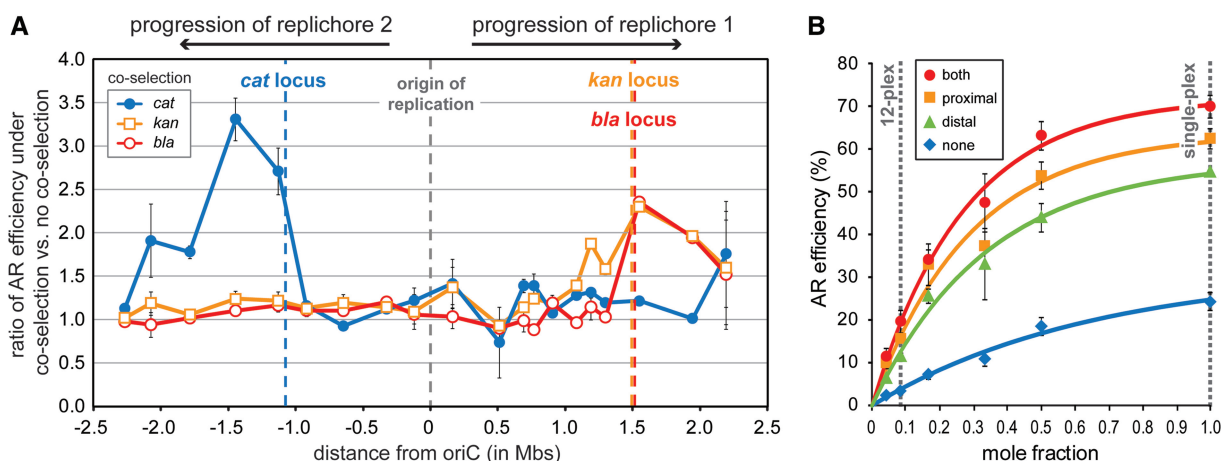


Figure 2. (A) 20 sites were separately modified throughout the *E. coli* genome in singleplex MAGE experiments in the presence of CoS oligos that re-activate the genes *cat*⁻, *kan*⁻ and *bla*⁻. Resultant cultures were grown under selective and non-selective conditions, and the AR frequencies at each site were assessed by AF-qPCR. A CoS factor (the ratio of AR frequencies with and without CoS) is shown—factors greater than one indicate increased frequencies due to CoS. (B) AR frequency was measured for one MAGE cycle at a single chromosomal site (*pphB*) flanked by a pair of deactivated CoS markers. Contexts ranged from singleplex (one oligo, 100%) to highly multiplexed (24 oligos, each \sim 4% of the pool) not including the CoS oligos. Resulting cultures were grown under conditions selecting for either re-activated proximal marker (*cat*⁻, <5 kb), distal marker (*tolC*⁻, >300 kb), both and neither.

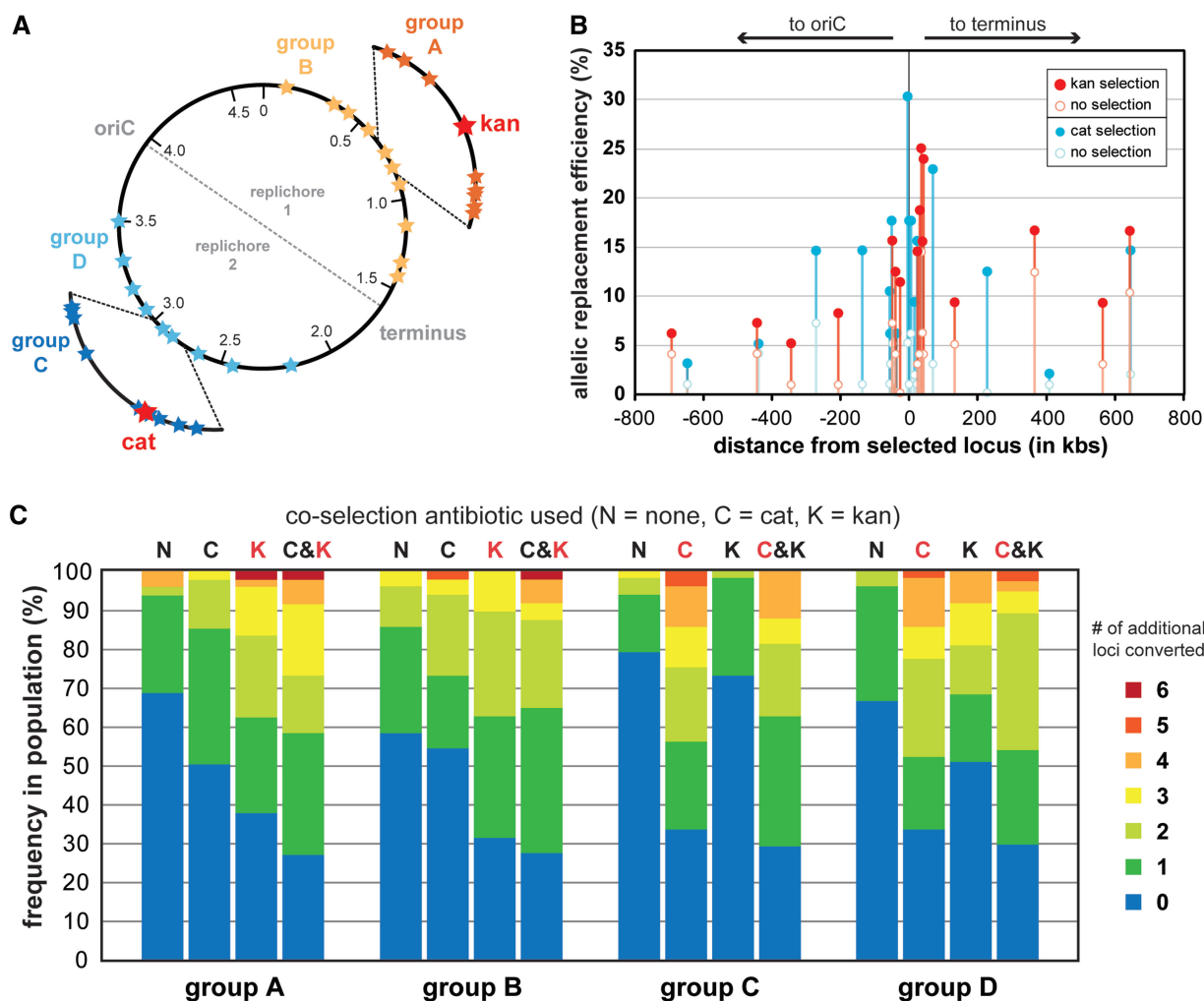


Figure 3. Co-selection enhances multiplex AR. (A) Diagram of four groups of targets A–D based on proximity to selectable loci *kan* (A, B) and *cat* (C, D). Each group contained 8–10 targets, not counting CoS markers (Supplementary Table 2). Groups A (eight targets) and C (9 targets) are within 56 kb from *kan* and *cat*, respectively, whereas Groups B and D (10 targets each) are within 694 kb from *kan* and *cat*, respectively. (B) AR frequency of Groups A–D mapped to distance to nearest selected locus, *kan* or *cat*, and the increased frequency for each site when CoS was applied. (C) Frequency distribution of clones in the population with different numbers of additional conversions in each of four Groups A–D, under no selection, *cat*, *kan* or *cat/kan* CoS. Same-replichore CoS is indicated in red lettering.

average AR frequencies for these experiments are given in Supplementary Table S1. When cells were not co-selected for restoration of either *kan*⁻ or *cat*⁻ marker, the AR frequency in the multiplexed reaction was low, averaging 3.7% per site. Targets in close proximity to the co-selected markers on the same replichore (Group A/*kan* and Group C/*cat*, <56 kb) showed the greatest frequency improvements under CoS—giving average CoS factors from 3.3- to 5.5-fold. In contrast, CoS at the opposite replichore (Group A/*cat* and Group C/*kan*, >1.4 Mb) yielded a modest improvement—CoS factors of 1.3- to 1.6-fold. When frequencies are plotted against the distance to the nearest CoS markers on the same replichore (Figure 3B), we find the greatest improvements clustering near these markers.

To evaluate the synergistic effects among sites under CoS, we further analyzed the distribution of conversions accumulated for individual clones in Groups A–D (Figure 3C, Supplementary Figure S4). Without CoS, most of the population (~70%) remained unmodified.

Cross-replichore CoS showed marginal increases in the frequency at which mutants were found. In contrast, same-replichore CoS dramatically increased the frequency of mutants with large numbers of modifications. Double CoS of one marker on each replichore further increased multiplexed AR frequencies, giving rise to conversions in 70% of the cell population. These populations contained individuals with as many as 8 of 10 targeted conversions (Group A: six co-selected sites verified by MASC-PCR, plus two marker sites by *kan/cat* double selection). These results showed that a single cycle of MAGE can operate in a single cell at 8 or more spatially distinct loci.

We posit that CoS in general isolates cells that have taken up more oligos, giving rise to the modest increases in AR frequencies during cross-replichore CoS. Moreover, proximity-based CoS (within ~500 kb) especially increases the likelihood of isolating cells which had chromosomes at an optimal stage of replication for obtaining correlated AR events. This effect is notable in our ability to easily

isolate highly modified cells. Without CoS, the average AR frequency across all sites of Groups A–D was $3.7(\pm 3.4)\%$ per target. With double CoS of the *kan* and *cat* markers, the average was $15.6(\pm 9.4)\%$, a 4-fold improvement (Supplementary Figure S4). If these co-selected frequencies were independent of each other, the population of modified clones would be described by a binomial distribution. For AR frequencies of 3.7%, only one colony in 1.5×10^7 would contain six or more modifications of eight (excluding the CoS markers), and we would need to screen at least 4.5×10^7 clones for a 95% likelihood of obtaining one (Figure 4; see later for detailed calculations). Frequencies of 15.6%, would require 10^4 colonies to meet the same goal, a >4000-fold decrease in the scale of colony screening. However, in the earlier experiment, the 6-conversion mutant was found by screening only 48 colonies with CoS.

We have consistently achieved this level of performance in CoS experiments using from 6 to 24 different oligos, yielding clones with 5–8 modifications (Figure 5) per round of CoS-MAGE. We attribute our increased ability to enrich for highly modified cells to co-operative effects of CoS, isolating groups of sites that are converted together. Using simple MASC-PCR or multiplex allele-specific colony qPCR (MASC-qPCR) methods, we can readily screen 100 or more colonies for conversion at up to 12 target sites simultaneously with a turnaround

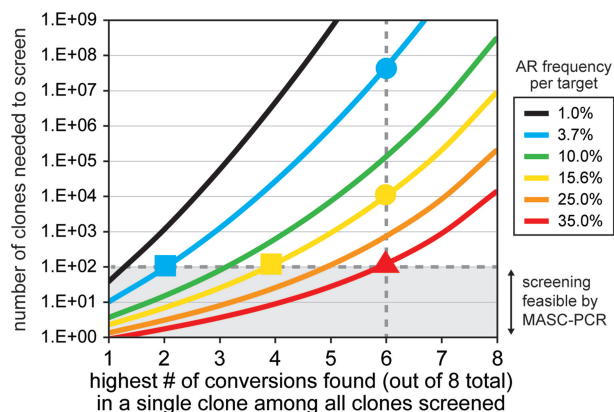


Figure 4. Colony screening requirement to isolate clones with the highest number of conversions based on a binomial model. AR process is evaluated as a binomial process with each site being independently converted. The solid lines correspond to relationship between the number of clones needed to be screened from the population after AR with oligos targeting eight sites and the highest number of conversions that are expected to be found in one clone among the clones screened. Each line corresponds to a different average per site AR frequency. AR frequency of 3.7% and 15.6% are empirically determined values for 37 sites with and without CoS, respectively. At a 95% confidence interval, the number of colonies needed to find a 6-conversion clone (indicated by the vertical gray dotted line) is shown as solid circles. If 100 colonies are screened, one expects to find 2- and 4-conversion clones for 3.7% and 15.4% AR efficiency, respectively (solid squares and horizontal gray-dotted line). To find a 6-conversion clone by screening 100 colonies, we expect to need an average AR frequency per site of 35% (solid triangle) assuming a binomial process. In our experience, CoS yields a population of highly modified clones that consistently outperforms expectations based on independent AR frequencies.

time of 3–4 h. Without CoS, one either needs to screen or genotype an impractical number of clones, or expect to perform many more cycles of MAGE (11,15).

Only a low concentration of the CoS oligos (relative to total oligo concentration) is necessary to achieve CoS enhancement (Supplementary Figure S2). A low CoS oligo-to-total oligo ratio minimizes competition for entry to the cell between the CoS oligo and the rest of the oligo pool. However, a lower fraction of CoS oligo leads to a smaller population of surviving cells as a result of selection (i.e. fewer cells recombine a molecule of the CoS oligo, so fewer cells survive selection). This bottleneck reduces the size of the population and diminishes the diversity accessed using MAGE. Thus, for more extreme dilutions of CoS oligos (0.01% of total or less, or applying CoS at multiple CoS markers simultaneously), we have observed co-selected populations dominated by small numbers of genotypes. A smaller surviving population also produces a longer delay, as a selected culture grows back to the required cell density for the next MAGE cycle. We found that diluting the fraction of CoS oligo to 0.1–1% of the total oligo pool led most consistently to the greatest CoS enhancement (Supplementary Figures S2, S3 and S4), without overly restricting the cell population.

To illustrate how one might perform cycles of CoS-MAGE in series, we used this strategy to recode 80 sites across one-fourth of the *E. coli* genome, spanning 1.1 megabase pairs (Figure 5). A single dually selectable *tolC* marker (11,18), inserted into the center of this region, was repeatedly re-used for this purpose in strain EcBS5. Odd cycles of co-selected MAGE switched the *tolC* gene off by removing the start codon, whereas even cycles reversed this change to switch *tolC* on. At each cycle, up to 191 colonies were screened by MASC-qPCR to identify highly modified clones. Groups of 10 sites as identified previously (11) were targeted for modification with each pair of cycles. Odd cycles included oligos to modify all 10 sites within a group, whereas even cycles were used to finish off any unmodified sites (from 2 to 6) of that group. Any sites still left unmodified were then carried over, i.e. included as targets for the following odd cycle. Thus for some cycles, 11 or 12 conversions were attempted simultaneously (not including the CoS marker conversion). A total of 18 cycles of CoS-MAGE were used (Figure 5). All 80 sites were modified over the course of this experiment, although one site (*yegV*, modified at cycle 2) was modified back to WT inadvertently, being overwritten by an overlapping oligo acting on another nearby site (*yegW*, at cycle 8). Re-conversion of *yegV* to a TAA stop codon was attempted again once (Figure 5, open square at cycle 18) but was unsuccessful.

Cycles attempting the most modifications also yielded the most modifications (Figure 5). For example, when 10–12 modifications were attempted, the results ranged from 5 to 8 sites modified (with a mode of seven sites). This outcome suggests that the most efficient approach to modifying a large group of sites would be to maximize the number of sites addressed at every cycle of the process. However, in experiments addressing as

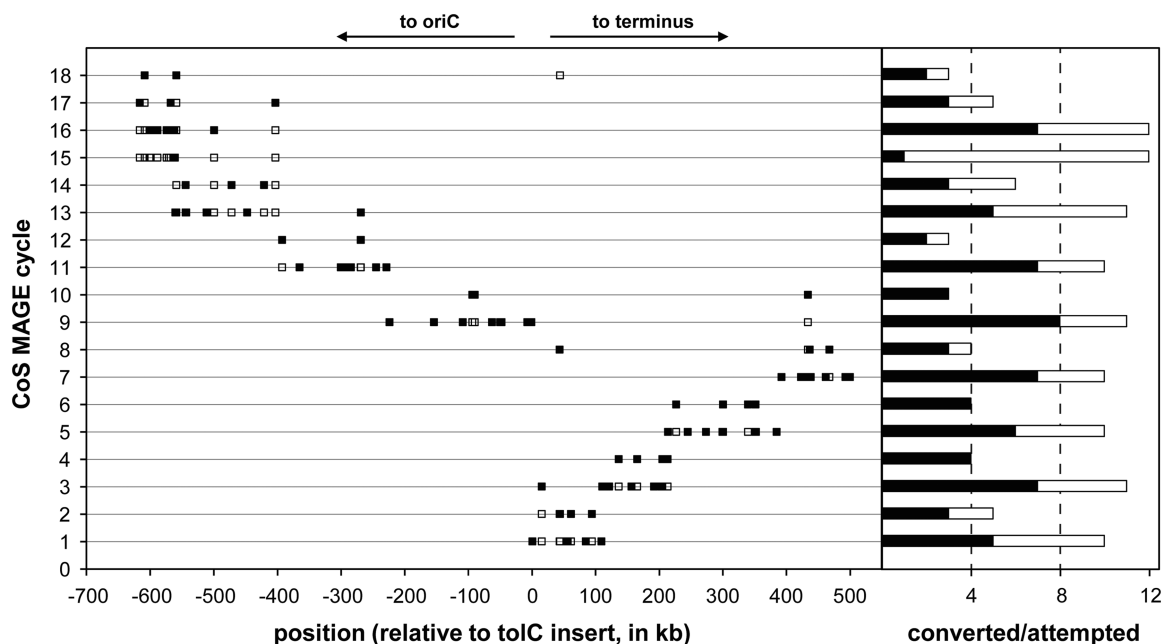


Figure 5. Serial CoS-MAGE to address 80 genome sites. Repeated cycles of MAGE used CoS at a dually selectable *tolC* locus to reformat all amber (TAG) stop codons spanning one-fourth of the *E. coli* MG1655 genome. Odd-numbered cycles switched *tolC* off and selected for the absence of the gene product using colicin E1. Even-numbered cycles switched the *tolC* to an active state and selected for the presence of the protein with SDS. After each cycle colonies were screened by MASC-qPCR to identify clones containing the largest number of successful edits. One cycle (15) performed poorly (1 conversion of 12 attempted) due to arcing during electroporation. **Left:** Progressive editing of the genome is shown for each cycle, with sites successfully modified shown with solid squares and those unmodified shown with open squares. Site locations are indicated relative to the inserted *tolC* CoS marker. **Right:** Numbers of conversions at each cycle. The greatest number of conversions was obtained for cycles attempting 10 or more (versus 6 or fewer). These yielded 5–8 conversions per cycle, with a median (and mode) of 7, not counting the conversion of the CoS marker.

many as 20 sites (data not shown), we have not yet observed more conversions than when addressing 12. Strategies that allow the screening of many more clones at many more sites are also likely to yield more highly modified clones. For screening by MASC-PCR or MASC-qPCR, multiplexing detection up to 12 sites per PCR reaction well has been implemented. Performing these techniques more intensively would permit screening a few thousand clones per day (e.g. 2-hour qPCR runs, 192 clones/run). If screening limits were abolished, we anticipate even more modified clones would be obtained. Other factors that likely affect the reach of CoS-MAGE are the number of oligo molecules entering the cell, the kinetics of oligo survival in the cell and the rate of oligo incorporation into the chromosome. Experiments to address these factors are beginning to provide further improvement, such as by limiting the action of endogenous nucleases (J.A. Mosberg, C.J. Gregg, *et al.*, in preparation).

In the earlier experiment, significant time and effort was required to produce and screen colonies at each cycle. Although a cycle of MAGE normally can be performed in 3 h or less as part of a larger automation process (10), including CoS requires at least a similar amount of additional time for growth under selective conditions. Including screening at every cycle in the earlier example then required plating cells to grow colonies, followed by

colony picking and PCR-based colony screening, extending cycle times to at least 2 days. The 80-site experiment was paused at multiple intervals as convenient—thus ~40-day process was executed in approximately twice that amount of time. The choice of frequent screening had the benefit of yielding the most highly modified clones to pass to the next cycle. However, other applications of CoS-MAGE will likely find it most expedient to forgo frequent screening and simply take advantage of the strong enhancement provided by CoS, which can be 4-fold at each site (Supplementary Figures S2 and S3). The latter approach would be especially preferred in applications where one wishes to maximize the degree of diversity generated in a modified population, such as for optimizing a metabolic pathway (10) or tuning a genetic circuit (12).

We developed a mathematical model to anticipate the screening requirement when using MAGE to isolate highly modified clones, both with and without CoS. Modeling the multiplex AR as a simple binomial distribution with conversion events at each target site occurring independently, we first assume an average AR frequency per site of p for a group of k sites. The abundance of clones with n mutations in the population after a single MAGE cycle is $f = \binom{k}{n} p^n (1-p)^{k-n}$. To isolate this n -mutation clone at a 95% likelihood needs to satisfy the condition $(1-f)^s < 0.05$. When solving for s , we get

$s > \log(0.05)/\log(1 - f)$, which is the number of colonies to be isolated and screened to find at least one clone with n or more mutations at 95% likelihood (Figure 4). For $k = 8$ and $n = 6$ at $p = 0.037$ (the AR frequency of 3.7% without CoS observed above, Figure 3 and Supplementary Table S1), we calculate $s = 44965770$. At $p = 0.156$ (AR frequency of 15.6% observed with CoS in the same experiment), we find that $s = 10420$. This calculation indicates that CoS-MAGE could reduce screening requirements by >4000-fold under such conditions. Furthermore, with CoS-MAGE, we observe such highly modified clones at much higher frequencies than this (e.g. 1 in 50 instead of 1 in 10420 predicted). Consistent with our physical model (Figure 1), these results indicate that oligo incorporations into the same region of the chromosome can be highly correlated and not completely independent.

Examined another way, as we find a 6-conversion clone in our multiplex CoS experiment of eight target sites at an abundance of 2% ($f = 0.02$), this would only be predicted by the above model if $p = 0.35$ or 35% (twice the observed value). At this AR frequency, we need to screen 117 clones to confidently find a clone with six conversions or more, which is on par with our experimental findings. This analysis underscores two important points. First, a 4-fold improvement in AR frequency per site translates to a dramatically decreased screening need. Second, as the binomial model predicts AR frequencies of 35% required to yield these 6-conversion clones (in sharp contrast to the 15.6% frequency observed), the highly co-operative CoS process does not seem to follow a simple binomial distribution of independent events.

Another feature of CoS-MAGE is that it minimizes the number of growth cycles cells must spend in a mutator state. Performing MAGE at high efficiency typically has relied on cells deficient in mismatch repair, e.g. *mutS*⁻ (4). (Otherwise, the cell's mismatch repair pathway attempts to 'repair' the genome edits that are being attempted.) However, performing MAGE this way also increases the rate of accumulation of background mutations in the genome (11). One strategy to avoid this limitation is to leave the repair pathway intact and instead use oligo sequences that create mismatches poorly recognized by mismatch repair, such as CC mismatches (4), multiple mismatches (17,19) and mismatches produced with chemically modified bases (19). However, these approaches place some sequence limitations on which genome edits can be made.

When a genome engineering application requires shutting off mismatch repair, the amount of cell growth in the mutator state should be minimized. CoS-MAGE provides a benefit in this regard, requiring far fewer cycles (and cell divisions) to reach a given objective. In addition, we explored the possibility of turning mismatch repair off temporarily at the beginning of MAGE cycling, so that it could be turned back on when finished (Supplementary Figure S5A). A *mutS*_{off} oligo was designed to edit the ATG start codon of the *mutS* gene to ATC by creating a CC mismatch poorly recognized by the *mutS* protein, turning the gene off but not deleting it as in previous

studies (4). The *mutS*_{off} oligo was applied during the first cycle of a MAGE experiment where *tolC* on/off switching was also used as described earlier. The population fraction for the ATC (*mutS*_{off}) allele was measured for four cycles of co-selected MAGE. We anticipated that the *mutS*_{off} population would increase with each cycle, even though the *mutS*_{off} oligo was only applied in the first cycle: CoS-MAGE selects at each cycle for cells that have successfully taken up oligos and incorporated them into their genomes, and only *mutS*_{off} cells should do this (and survive) at high efficiency. We observed that the *mutS*_{off} cells became dominant in the population after only a few cycles of CoS-MAGE (Supplementary Figure S5B).

We have observed the effects of CoS acting far from the site of a given CoS marker and on both sides of the marker (Figures 2A, 3B and 5). Nevertheless, an asymmetry was observed, most prominently in Figure 2A (a singleplex experiment) with CoS effectiveness dropping off sharply for sites closer to the origin than the marker. Figures 3B and 5 (multiplex experiments) may also indicate a measure of this asymmetry, but if so, these effects are more modest. The reason for this asymmetry—and why it might display most prominently for singleplex experiments—is unclear. Part of the explanation may lie in the nature of the replicating arms of the chromosome, as the copy number of genes (and thus numbers of targets for oligo annealing) upstream of the marker will generally be higher.

DISCUSSION

Currently, genomes can be engineered by different complementary approaches including complete *de novo* synthesis (20) and editing techniques such as MAGE (10). *De novo* synthesis offers the ability to create new genomes without a physical template but is limited by the difficulty and cost of *in vitro* DNA assembly, by the technical challenges of 'booting' a synthetic genome and by the biological challenges of designing a highly modified genome that will still support life. In contrast, MAGE relies on the manipulation of an existing genomic template *in vivo* to produce newly engineered variants without the need for total re-synthesis. Such template-mediated genome engineering is especially attractive in cases where the new constructs share strong sequence similarity (>90%) with existing constructs. As our approach modifies an existing genome by living intermediates, MAGE facilitates efficient incorporation of specified mutations and real-time viability testing to identify and avoid any lethal mutations. In contrast, *de novo* synthesis uses an all-or-nothing synthesis and boot approach that does not lend itself to easy troubleshooting. Furthermore, template-based engineering can benefit from natural selection processes as new genomes progress by directed steps from existing functional genomes.

We have previously reported the development of hardware for efficient automation of MAGE processes (10). The selection steps of CoS MAGE can be easily

incorporated into the cycles performed by this system for obtaining the optimum combinations of modification, growth and selection. The CoS enhancement increases the ability of MAGE to make very large numbers of changes to a genome, especially when combined with other tools such as conjugative assembly genome engineering (CAGE) (11). As we previously reported using MAGE and CAGE together to make 314 genome edits in *E. coli*, CoS MAGE (with CAGE) provides the possibility of extending this reach to thousands of sites. For projects on a smaller scale, CoS can be used for a more modest number of cycles performed manually (without automation hardware) to modify dozens of sites (12).

CoS strategies dramatically increase our genome engineering capabilities. We have demonstrated in several experiments that CoS MAGE yields higher AR frequencies, often improving by a factor of 4. This enhancement is especially useful when making many modifications to a genome. For example, our recent report altering stop codons in *E. coli* (11) required 18 cycles of MAGE to produce an average of 8 modifications of 10 targeted sites. In contrast, with CoS-MAGE we now easily isolate cells that incorporate an average of seven modifications (plus one or two modified CoS marker genes) after only a single cycle. Including CoS marker conversions, we have demonstrated at least nine simultaneous modifications to the genome are possible. With further tuning or screening more clones, a greater number seems plausible. These increased frequencies are obtained using easily switchable genetic markers to co-select for several correlated AR events, targeting multiple chromosomal sites spanning as much as a megabase pair of a genome (up to 500 kb from the selection gene in either direction). Cells containing many such chromosomal modifications can be isolated efficiently by screening and can then be used for subsequent CoS-MAGE cycles. Markers with both positive and negative selection options (e.g. *tolC*, *galK* and *thyA*) are readily available for this purpose. Deploying these markers throughout the genome will generate programmable zones that are hyper-responsive to genome engineering. With the effects of CoS-MAGE spanning up to 1 megabase pair per marker, only a modest number of markers may be needed to fully address microbial genomes such as *E. coli* MG1655 (4.6 Mb).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–4, Supplementary Figures 1–5 and Supplementary Methods.

ACKNOWLEDGEMENTS

The authors thank Shuguang Zhang for helpful discussions and extensive use of laboratory resources. They also thank Joshua Mosberg and John Aach for discussions and manuscript comments.

FUNDING

Multiple programs from the National Science Foundation (SynBERC [0540879]; Center for Bits and Atoms [0122419]; Genes and Genomes Systems Cluster [0719344]); Department of Energy (Genome to Life Center [DE-FG02-03ER6344]); Technology Development Fellowship from the Wyss Institute for Biologically Inspired Engineering and the Director's Early Independence Award from the National Institutes of Health [1DP5OD009172-01 to H.H.W.]; a US DOD NDSEG fellowship (to M.J.L.). Funding for open access charge: The Center for Bits and Atoms, Massachusetts Institute of Technology, Cambridge, MA, USA.

Conflict of interest statement. None declared.

REFERENCES

- Posfai,G., Plunkett,G. 3rd, Feher,T., Frisch,D., Keil,G.M., Umenhoffer,K., Kolisnychenko,V., Stahl,B., Sharma,S.S., de Arruda,M. *et al.* (2006) Emergent properties of reduced-genome *Escherichia coli*. *Science*, **312**, 1044–1046.
- Itaya,M., Tsuge,K., Koizumi,M. and Fujita,K. (2005) Combining two genomes in one cell: stable cloning of the *Synechocystis* PCC6803 genome in the *Bacillus subtilis* 168 genome. *Proc. Natl Acad. Sci. USA*, **102**, 15971–15976.
- Isalan,M., Lemerle,C., Michalodimitrakis,K., Horn,C., Beltrao,P., Raineri,E., Garriga-Canut,M. and Serrano,L. (2008) Evolvability and hierarchy in rewired bacterial gene networks. *Nature*, **452**, 840–845.
- Costantino,N. and Court,D.L. (2003) Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc. Natl Acad. Sci. USA*, **100**, 15748–15753.
- Ellis,H.M., Yu,D., DiTizio,T. and Court,D.L. (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl Acad. Sci. USA*, **98**, 6742–6746.
- Li,X.T., Costantino,N., Lu,L.Y., Liu,D.P., Watt,R.M., Cheah,K.S., Court,D.L. and Huang,J.D. (2003) Identification of factors influencing strand bias in oligonucleotide-mediated recombination in *Escherichia coli*. *Nucleic Acids Res.*, **31**, 6674–6687.
- Wang,Z., Zhou,Z.J., Liu,D.P. and Huang,J.D. (2006) Single-stranded oligonucleotide-mediated gene repair in mammalian cells has a mechanism distinct from homologous recombination repair. *Biochem. Biophys. Res. Commun.*, **350**, 568–573.
- 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.
- 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.
- 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.
- Isaacs,F.J., Carr,P.A., Wang,H.H., Lajoie,M.J., Sterling,B., Kraal,L., Tolonen,A., Gianoulis,T.A., Goodman,D.B., Reppas,N.B. *et al.* (2011) Precise manipulation of chromosomes *in vivo* enables genome-wide codon replacement. *Science*, **333**, 348–353.
- Wang,H., Kim,H., Cong,L., Jeong,J., Bang,D. and Church,G. (2012) Genome-scale promoter engineering by co-selection MAGE. *Nat. Methods*, April 8 (doi: 10.1038/nmeth.1971; epub ahead of print).
- Swaminathan,S., Ellis,H.M., Waters,L.S., Yu,D., Lee,E.C., Court,D.L. and Sharan,S.K. (2001) Rapid engineering of

- bacterial artificial chromosomes using oligonucleotides. *Genesis*, **29**, 14–21.
14. Schwartz,S.A. and Helinski,D.R. (1971) Purification and characterization of colicin E1. *J. Biol. Chem.*, **246**, 6318–6327.
 15. Wang,H.H. and Church,G.M. (2011) Multiplexed genome engineering and genotyping methods applications for synthetic biology and metabolic engineering. *Methods Enzymol.*, **498**, 409–426.
 16. Germer,S., Holland,M.J. and Higuchi,R. (2000) High-throughput SNP allele-frequency determination in pooled DNA samples by kinetic PCR. *Genome Res.*, **10**, 258–266.
 17. 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.
 18. DeVito,J.A. (2008) Recombineering with tolC as a selectable/counter-selectable marker: remodeling the rRNA operons of *Escherichia coli*. *Nucleic Acids Res.*, **36**, e4.
 19. Wang,H.H., Xu,G., Vonner,A.J. and Church,G. (2011) Modified bases enable high-efficiency oligonucleotide-mediated allelic replacement via mismatch repair evasion. *Nucleic Acids Res.*, **39**, 7336–7347.
 20. Gibson,D.G., Benders,G.A., Andrews-Pfannkoch,C., Denisova,E.A., Baden-Tillson,H., Zaveri,J., Stockwell,T.B., Brownley,A., Thomas,D.W., Algire,M.A. *et al.* (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science*, **319**, 1215–1220.