Genome-scale promoter engineering by coselection MAGE

Harris H Wang[1,2,6], Hwangbeom Kim[3,6], Le Cong[1,4,5], Jaehwan Jeong[3], Duhee Bang[3] & George M Church[1,4]

Multiplex automated genome engineering (MAGE) uses short oligonucleotides to scarlessly modify genomes; however, insertions >10 bases are still inefficient but can be improved substantially by selection of highly modified chromosomes. Here we describe 'coselection' MAGE (CoS-MAGE) to optimize biosynthesis of aromatic amino acid derivatives by combinatorially inserting multiple T7 promoters simultaneously into 12 genomic operons. Promoter libraries can be quickly generated to study gain-of-function epistatic interactions in gene networks.

The ability to directly manipulate and add new genetic regulatory elements to the chromosome is highly desirable as it bypasses the need to resort to plasmid-based systems in which copy-number variation can substantially affect the dynamics and robustness of synthetic networks[1]. Previously, we had developed the MAGE method to combinatorially introduce sequence changes to many sites in the *Escherichia coli* genome using libraries of 90-base synthetic oligonucleotides[2,3]. Although the efficiency for making small modifications is high (>30% for <4 base pairs (bp))[2,4], insertion efficiency per MAGE cycle drops substantially for larger inserts (<2% for >20 bp), thus limiting its utility for integration of regulatory elements or larger sequence motifs. However, we had unexpectedly found that a small fraction of isolates contained more than one mutated site in each cycle of MAGE. We hypothesized that isolation of clones containing a mutation at one site may increase the likelihood of finding other mutated sites because a certain cell subpopulation may be more electrocompetent or more available to perform oligo-based allelic replacement. We call the process of population enrichment for highly mutated genomes CoS-MAGE, which we demonstrate here as an enhanced scarless genome engineering approach to efficiently introduce larger regulatory elements such as promoters into targeted regions on the chromosome.

We chose the 20-bp T7 promoter (5′-TAATACGACTCACTATAGGG-3′) for insertion upstream of 12 genes or operons in the *E. coli* genome that are associated with the biosynthesis of aromatic amino acids (**Fig. 1** and **Supplementary Fig. 1**). In the presence of T7 RNA polymerase, transcription from the T7 promoter is strong and orthogonal to that at endogenous promoters. In *E. coli*, the industrially relevant blue dye, indigo, and a compound used in leukemia treatment[5], indirubin, can be produced from the tryptophan-biosynthesis intermediate, indole, by heterologous expression of a *Methylophaga* sp. flavin-containing monooxygenase, encoded by *bfmo*[6]. Using the MAGE-competent EcNR2 strain[2], we first transformed the bfmo-encoding pJ401 plasmid to generate EcHW7. We then removed known feedback regulation and allosteric inhibition associated with tryptophan biosynthesis[7–10], which involved nonsense inactivation of *trpR* and introducing mutations in *aroF* (to encode a P148L mutant protein), *trpE* (to encode a M293T mutant protein) and in *aroG* (to encode a D146N mutant protein). We then inactivated chromosomal *galK*, *malK*, *cat* and *bla* by introducing a revertible premature stop codon in each gene with high efficiency by MAGE using 90-bp oligos to generate the strain EcHW47 (Online Methods). We transformed EcHW7 and EcHW47 with pN249 to produce strains E7N and E47N, respectively.

To perform CoS-MAGE, we spiked the mixed oligo pool that targets all 12 insertion sites at a molar ratio of 50:1 with a small amount of an oligo (called CoS oligo) that reverted the function of an inactivated genomic selectable marker (that is, *galK*$^-$, *malK*$^-$, *cat*$^-$ or *bla*$^-$). Restoration to *galK*$^+$ or *malK*$^+$ enabled growth on M9 minimal medium supplemented with galactose or maltose, respectively, as the sole carbon source. Restoration to *cat*$^+$ or *bla*$^+$ conferred antibiotic resistance to chloramphenicol or carbenicillin, respectively.

In the first coselection stage, we performed four MAGE cycles targeting 12 sites and *cat*$^-$. At the end of the coselection stage, we isolated 96 colonies plated with and without chloramphenicol and genotyped them by multiplex allele-specific colony PCR[3] to query for T7 promoter insertions at the 12 target sites (Online Methods). Among the isolated clones, target sites near the coselection marker were highly enriched for T7 promoter insertions only when the chloramphenicol coselection was applied (**Fig. 1b**). At these sites, the frequency of T7 promoter integration was as high as 25% with coselection in comparison to 2–3% without it. We observed this increased frequency of integration after only two MAGE cycles (**Fig. 1b**). We hypothesize that oligo-mediated genomic modifications near the coselection marker were enhanced by capturing cells whose replication forks were transiently in the open state and available for allelic replacement (P. Carr and H.H.W., unpublished data).

By selecting against cells that do not generate oligo-mediated allelic replacements, we can substantially enrich the population for individual clones with multiple T7 promoter insertions. This coselection effect can be enhanced through multiple MAGE cycles.

[1]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts, USA. [2]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA. [3]Department of Chemistry, Yonsei University, Seoul, Korea. [4]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. [5]Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, Massachusetts, USA. [6]These authors contributed equally to this work. Correspondence should be addressed to H.H.W. (harris.wang@wyss.harvard.edu) or D.B. (duheebang@yonsei.ac.kr).

RECEIVED 3 OCTOBER 2011; ACCEPTED 22 FEBRUARY 2012; PUBLISHED ONLINE 8 APRIL 2012; DOI:10.1038/NMETH.1971

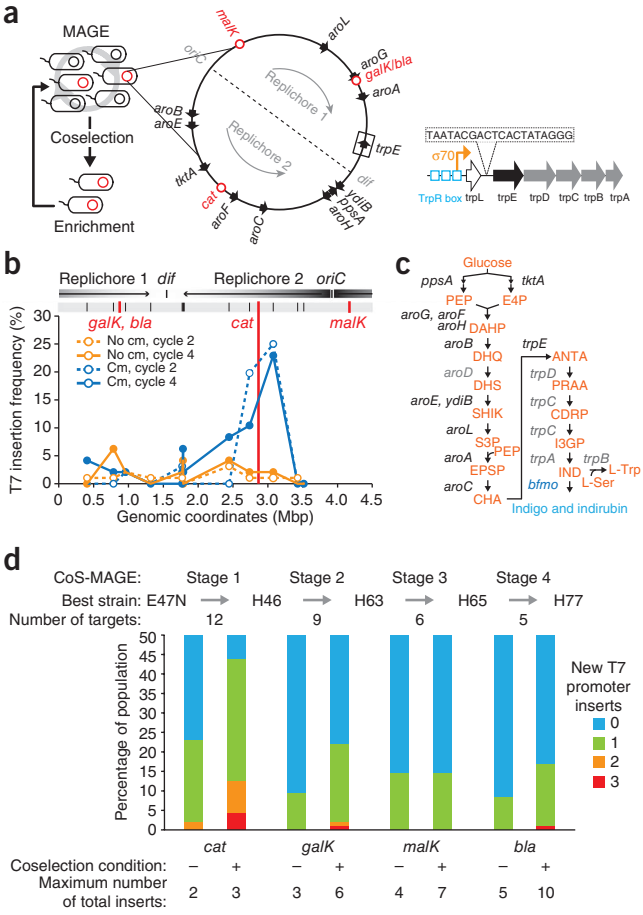


Figure 1 | CoS-MAGE strategy for enriching highly modified genomes. (**a**) Cells undergoing cycles of MAGE were enriched by a coselection stage through phenotypic selection of a revertible genomic locus (*malK*, *galK*, *cat* or *bla*) and used for subsequent iterative cycles of MAGE (left). Twelve genomic operons were targeted for insertion of the T7 promoter sequence upstream of the first relevant open reading frame. For example, T7 promoter sequence (bottom right) targeting *trpE* was inserted upstream of the first relevant open reading frame. TrpR box, tryptophan repressor binding sequence; σ70, endogenous promoter. (**b**) T7 promoter insertion frequency near the *cat* locus with and without coselection by chloramphenicol (cm) after two and four MAGE cycles as determined by multiplex allele-specific colony PCR (Online Methods). Genomic coordinates based on the coordinate designations in ancestral strain MG1655 (GenBank U00096.2). (**c**) Schematic of indigo and indirubin biosynthesis. First gene in each relevant operon shown in black; targeted genes that are not first in the operon are shown in gray. (**d**) Percentage of colonies with 0–3 new T7 promoter insertions for the indicated conditions. The most enriched strains at the end of each coselection stage are labeled.

Without coselection, 10–25% of the population contained cells with one T7 promoter after four MAGE cycles. With coselection, more than 40% of cells had at least one T7 promoter insertion and 5% had three insertions (**Fig. 1d**). Using clone H46 that contained the most T7 promoter insertions (3 of 12 sites), we performed a second stage of CoS-MAGE by targeting the remaining nine sites using another coselection marker, *galK*. This stage produced clones with up to three more T7 promoter insertions. Two more stages of CoS-MAGE generated a clone with ten T7 promoter insertions (**Fig. 1d**). In comparison, we could only generate up to five insertions in a single clone in the absence of coselection. Coselection with *malK* during the third stage yielded mutants with only one additional insertion, likely because *malK* is far from any target site (>500 kilobases; **Supplementary Table 1**), providing support for the replication-dependency hypothesis in which coselection affects neighboring chromosomal regions. These results demonstrate that CoS-MAGE can substantially enhance the generation of genomic insertions (**Supplementary Note**). We inserted T7 promoter into the two remaining sites by CoS-MAGE to produce a clone that contains all 12 T7 promoters to orthogonally drive gene expression.

One important feature of CoS-MAGE is the capacity to produce combinatorial libraries of variants. In the process of generating the 12-target strain, we isolated intermediate strains with fewer T7 promoter insertions but in different combinations (**Fig. 2**). We recovered 80 unique variants (named H1–H80) with 1–12 T7 promoter insertions including all 12 single inserts and two complete variants of 11 double inserts (*aroG* or *trpE* with all others). Using combinatorial T7 promoter expression, we probed the tunable parameters in the biosynthesis network. We indirectly measured the amount of tryptophan produced through expression of *bfmo* from each strain in the library when driven by a plasmid-encoded T7 polymerase (pN249). We readily extracted indigo and indirubin pigments, the major products of the bfmo monooxygenase[6], by dimethyl sulfoxide (DMSO) treatment (Online Methods).

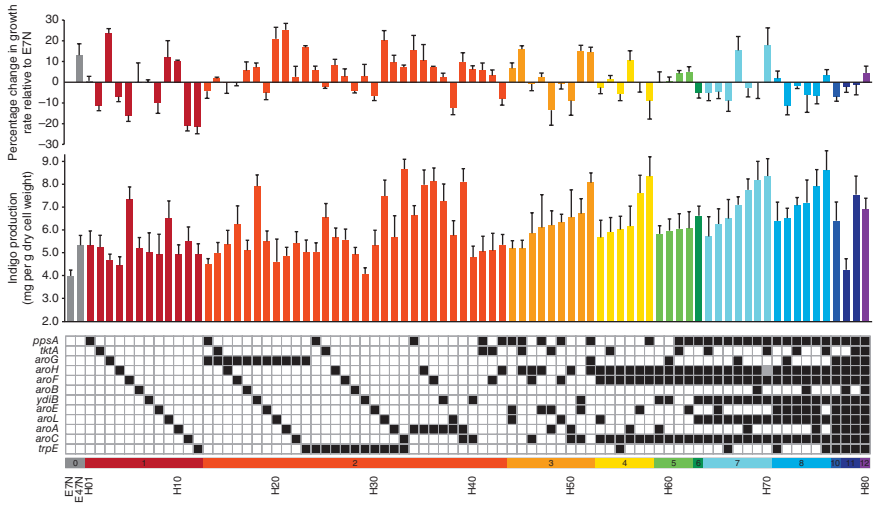


Figure 2 | Indigo production from synthetic T7 promoter–regulated operons. Relative change in growth rate for E47N (E7N derivative with inactivated *trpR*, and mutated *aroF*, *trpE* and *aroG*) and H1–H80 (derivatives of E47N with T7 promoter insertions) was normalized to growth rate of E7N (E2N-derivative with *bfmo*-containing pJ401 plasmid) as $(y - x)/y$, where x is the doubling time of each strain and y is doubling time of the E7N wild type (top). Indigo production based on measured indirubin amounts (middle; Online Methods). All error bars, s.e.m. ($n = 3$). Color coding reflects grouping by the number of T7 insertions as indicated on the bottom. Genotypes for each clone are marked by black boxes and ordered by their relative location along the biosynthesis pathway (*ppsA* is proximal and *trpE* is distal). The gray box of strain H70 corresponds to an insert in *aroH* with a mutated T7 sequence (5′-TAATACACTCAGTATGGG-3′).

In our library, we found and visually detected strains with more than fourfold improvement in indigo production over the E2N ancestor strain (expressing T7 RNA polymerase; **Fig. 2** and **Supplementary Fig. 2**). Based on calibration with commercially available indigo, we estimate that our best variants (H33 and H76) produced >8.6 mg g^{-1} dry cell weight of indigo. We observed synergistic epistasis between certain T7 promoter insert combinations. For example, T7 promoter insertions in *aroC* (variant H11) and *trpE* (variant H12) did not significantly improve indigo production individually relative to the E47N ancestral strain ($P = 0.29$ and 0.11, respectively; Student's *t*-test here and below), but in combination (variant H33) produced 62% more than the E47N strain ($P = 8.4 \times 10^{-10}$). Many other pairs also showed synergetic interactions, such as *aroG* with *ydiB* (variant H18), *aroA* with *aroH* (H35), *aroA* with *aroF* (H36), *aroA* with *ydiB* (H37) and *aroA* with *aroC* (H39).

Analysis of higher-order mutants revealed more unexpected results. Strain H77, with 10 of 12 inserts, showed a 20% improvement in indigo production over the E47N strain ($P = 2.5 \times 10^{-3}$), but the addition of the T7 promoter insert in *aroB* in the H78 strain significantly reduced the yield to 20% below that in the E47N strain ($P = 4.2 \times 10^{-3}$). Addition of the final T7 promoter insert in *tktA* in the 12-promoter H80 strain returned yield to 29% above that in the E47N strain ($P = 2.5 \times 10^{-3}$). Such large and opposite effects of *aroB* insertion and subsequent *tktA* insertion suggest that interactions in this gene network are unexpectedly complex, nonlinear and epistatic. Growth rate measurements of each strain under *bmfo*-induced conditions also revealed changes ranging from −30% to +30% relative to the ancestral strain (**Fig. 2**). However, we extracted no notable correlation between growth rate and indigo amount (**Supplementary Fig. 3**). We hypothesize that such complex epistatic interactions are more prevalent in endogenous networks than previously expected, especially in the context of large-scale modulation of gene expression using synthetic regulatory modules.

Using CoS-MAGE, we can probe epistasis in endogenous and engineered genetic pathways. With a fully synthetic regulon such as the 12 promoter–containing H80 strain, a reverse optimization can be done by CoS-MAGE using degenerate oligos to fine-tune expression of each T7 promoter in the pathway with orthogonal T7 polymerases (K. Temme, R. Hill, T.H. Segall-Shapiro and C.A. Voigt, unpublished data). Switchable coselection markers (antibiotic-resistance genes, fluorescent proteins or metabolic genes) can be placed anywhere in the genome and can be easily activated or inactivated with individual oligos. We mapped 21 useful genomic markers with their relevant switching oligos and selection schemes (**Supplementary Fig. 4** and **Supplementary Table 2**). Longer oligos can be used in MAGE, with increased homology arm length correlating with increased integration efficiency (**Supplementary Fig. 5**). We speculate that the use of longer oligonucleotides (100–200 bases) may enable the integration of even larger regulatory sequences for efficient CoS-MAGE. The efficiency improvements of CoS-MAGE over other MAGE implementations (**Supplementary Fig. 6** and **Supplementary Table 3**) make it more feasible to use the method without automated instrumentation. The coselection strategy can be applied to improve engineering of other organisms in which oligo-mediated recombineering may be less efficient[11]. Plasmid engineering can benefit from coselection to combinatorially generate vector libraries of synthetic circuits[12]. Multiplexed modification of endogenous genes and regulatory elements will enable larger genome-scale engineering efforts to push the limits of engineered biological systems[13].

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank J. Keasling, A. Juminaga, P. Carr, S. Kosuri, J. Aach and T. Gianoulis for helpful discussions and data interpretation, H. Salis (Pennsylvania State University) for providing pJ401, and C. Voigt (Massachusetts Institute of Technology) for providing pN249. This work was funded by multiple programs from the US National Science Foundation (SynBERC, SA5283-11210), the Department of Energy (Genome to Life Center, DE-FG02-03ER6344) and the Wyss Institute for Biologically Inspired Engineering. Additional funding was from the Next-Generation BioGreen 21 Program (SSAC, PJ008109), Rural Development Administration and the Intelligent Synthetic Biology Center of Global Frontier Project funded by the Ministry of Education, Science and Technology (2011-0031956), Korea. H.H.W. is supported by the Wyss Institute Technology Development Fellowship and the US National Institutes of Health Director's Early Independence Award (grant DP5OD009172). J.J. is supported by the National Junior research fellowship from the National Research Foundation of Korea (2012-0000391).

AUTHOR CONTRIBUTIONS

H.H.W., H.K., D.B. and G.M.C. designed the study. H.H.W., H.K., L.C. and J.J. performed the experiments. H.H.W., H.K. and D.B. analyzed the data and prepared the initial manuscript. All authors edited and revised the final manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

Published online at http://www.nature.com/doifinder/10.1038/nmeth.1971. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Eldar, A. & Elowitz, M.B. *Nature* **467**, 167–173 (2010).
2. Wang, H.H. *et al. Nature* **460**, 894–898 (2009).
3. Wang, H.H. & Church, G.M. *Methods Enzymol.* **498**, 409–426 (2011).
4. Costantino, N. & Court, D.L. *Proc. Natl. Acad. Sci. USA* **100**, 15748–15753 (2003).
5. Hoessel, R. *et al. Nat. Cell Biol.* **1**, 60–67 (1999).
6. Choi, H.S. *et al. Biochem. Biophys. Res. Commun.* **306**, 930–936 (2003).
7. Caligiuri, M.G. & Bauerle, R. *J. Biol. Chem.* **266**, 8328–8335 (1991).
8. Kikuchi, Y., Tsujimoto, K. & Kurahashi, O. *Appl. Environ. Microbiol.* **63**, 761–762 (1997).
9. Weaver, L.M. & Herrmann, K.M. *J. Bacteriol.* **172**, 6581–6584 (1990).
10. Bongaerts, J., Kramer, M., Muller, U., Raeven, L. & Wubbolts, M. *Metab. Eng.* **3**, 289–300 (2001).
11. Swingle, B. *et al. Mol. Microbiol.* **75**, 138–148 (2010).
12. Yokobayashi, Y., Weiss, R. & Arnold, F.H. *Proc. Natl. Acad. Sci. USA* **99**, 16587–16591 (2002).
13. Isaacs, F.J. *et al. Science* **333**, 348–353 (2011).

ONLINE METHODS

Media, chemicals and reagents. All strains were grown in rich LB-Lennox media containing tryptone (10 g l^{-1}), yeast extract (5 g l^{-1}), and NaCl (5 g l^{-1}), and buffered to pH 7.45 with NaOH. Chloramphenicol at 20 µg ml^{-1}, kanamycin at 30 µg ml^{-1}, carbenicillin at 50 µg ml^{-1} or spectinomycin at 100 µg ml^{-1} was supplemented to liquid LB-min medium or LB-min agar plates (LB-min with 15 g l^{-1} agar) for selection. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was used at 0.25 mM to induce the *lac* promoter. MacConkey-galactose (Mac-Gal) or MacConkey-maltose (Mac-Mal) agar plates were made by supplementing Difco MacConkey agar base (40 g l^{-1}), which contains peptone (17.0 g l^{-1}), proteose peptone (3.0 g l^{-1}), bile salts no. 3 (1.5 g l^{-1}), NaCl (5.0 g l^{-1}), agar (13.5 g l^{-1}), neutral red (30 mg l^{-1}) and crystal violet (1 mg l^{-1}), with D-(+)-galactose or D-(+)-maltose monohydrate at 10 g l^{-1}, respectively. M9 minimal medium was made by adding 5× M9 base (64 g l^{-1} $Na_2HPO_4{\cdot}7H_2O$ or 30 g l^{-1} Na_2HPO_4, 15 g l^{-1} KH_2PO_4, 2.5 g l^{-1} NaCl and 5.0 g l^{-1} NH_4Cl) with $MgSO_4{\cdot}7dH_2O$ (1 mM), vitamin B1 (0.05 g l^{-1}), D-biotin (0.2 µg l^{-1}) and 0.2% D-(+)-galactose or D-(+)-maltose monohydrate. Multiplex PCR kits were purchased from Qiagen. Standard 96-well format agarose gel electrophoresis system (Bio-Rad Sub-Cell Model 96 Cell) and reagents were used.

Oligonucleotides. All primers and oligonucleotides were obtained from Integrated DNA Technologies with standard purification. MAGE oligos were designed to insert T7 promoters 35 bp upstream of the start codon of the protein coding sequence of 12 genes or operons (**Supplementary Fig. 1**). These oligos had minimized secondary structure (>−12 kcal mol^{-1}) and targeted the replicating lagging strand. Additionally, each oligo contained four phosphorothioate linkages at the 5′ end. Sequences of all MAGE oligos and primers used for sequencing and multiplex allele-specific colony (MASC)-PCR are listed in **Supplementary Table 4**.

CoS-MAGE cycling process. MAGE cycling was performed as previously described[3]. In brief, cells were grown in 3 ml volume with needed antibiotics in a rotating incubator at 30 °C. At OD_{600} of 0.6, cells were heat-shocked in a shaking water bath at 42 °C for 15 min. One milliliter of cells was made electrocompetent per reaction at 4 °C by several buffer exchanges with sterile distilled water through centrifugation and resuspension. Oligos were added to concentrated cells for each 50 µl electroporation reaction. For multiplex reactions, targeting oligos were mixed in equal molar to reach a final total oligo concentration of ~12 µM. CoS oligos (for example, cat_restore_oligo, galK_restore_oligo, malK_restore_oligo and bla_restore_oligo), which would restore the inactivated nonsense mutation to functional wild-type sequence, were added to the oligo pool at a molar ratio of 1:50 (1 CoS oligo per 50 targeting oligos). Electroporation was carried out in 1-mm gap cuvettes with the following conditions: 1.8 kV, 200 Ω and 25 µF; cells were recovered into fresh LB medium for subsequent MAGE cycle. Each stage of coselection was done at the end of four cycles of MAGE by plating on the appropriate selective medium (LB with chloramphenicol for *cat*$^+$, LB with carbenicillin for *bla*$^+$, M9-Gal for *galK*$^+$, Mac-Gal for red *galK*$^+$, M9-Mal for *malK*$^+$ or Mac-Mal for red *malK*$^+$). Isolating colonies and MASC-PCR genotyping 48–96 colonies resulted in many clones with insertions that were subsequently verified by sequencing. A streamlined CoS-MAGE protocol is described here. Each cycle of MAGE takes ~2 h, and four cycles can be easily done in a single day, ending with plating on the selection medium plates (for the coselection step). Overnight selection on plates will yield colonies in the morning, which can be picked directly into MASC-PCRs to identify the best clone (2 h turnaround). At the same time, each colony can be grown in 96-well format in parallel to the PCR-genotyping step. The best clone (taken from the already 2-h–grown 96-well plate) can then be used for the next four cycles of MAGE in the same day (day 2) and iterated in subsequent days in the same manner. For selection that required a minimal medium growth step (M9-Gal and M9-Mal), an additional day was needed to allow the colonies to form for each coselection step, thus doubling the total time needed. We note that four cycles of MAGE is not needed each day before the coselection enrichment is applied (two or even one cycle may be sufficient). We performed four cycles per day because it was practically as easy as two cycles with low sample numbers, and some minor benefit can be gained from additional rounds of standard MAGE. This coselection method is also a practical way of implementing MAGE in the absence of an automated instrument to perform MAGE, while still achieving a high frequency of site-directed mutagenesis.

Multiplex allele-specific colony PCR. MASC-PCR was done as previously described[3] as a fast, reproducible and cost-effective way to query the genotype of clones at up to 12 loci per reaction. In brief, two sets of primers (wild-type forward and reverse, designated as f/r, and mutant forward and reverse, designated as fm/r) were used to query each genomic locus with f/r corresponding to the wild-type allele and fm/r corresponding to the mutant allele. The forward primers (f and fm) were identical except at the 3′ end, which corresponded to either the wild-type (f) or mutant (fm) allele. The reverse primer was the same for both alleles. If a clone has the wild-type allele, only primer set f/r should produce a product and not primer set fm/r. Conversely, if a clone has the mutant allele, only primer set fm/r should produce a PCR product and not f/r. This allele-specific PCR can be multiplexed to amplify amplicons of different lengths (100, 150, 200, 250, 300, 400, 500, 600, 700, 850, 1,000 and 1,200 bp) corresponding to each of the 12 loci in the MASC-PCR. Primers were generally designed for a target melting temperature (T_m) of 62 °C, but a gradient PCR was done to experimentally determine the optimal T_m to improve specificity of distinguishing wild-type and mutant alleles. The Multiplex PCR kit from Qiagen was used in 20-µl reactions with 1 µl of a 1 in 100 water dilution of a picked colony and final primer concentration of 0.2 µM. PCR cycles were as follows: step 1, 15 min at 95 °C; step 2, 30 s at 94 °C; step 3, 30 s at optimal T_m; step 4, 80 s at 72 °C; step 5, repeat steps 2–4 26 times; and step 6, 5 min at 72 °C. Electrophoresis was done on a 1.5% ethidium bromide agarose gel and visualized under UV light.

Base strain construction. The MAGE-competent Δ*mutS*::*cat* base strain EcNR2 has been previously described[2]. EcNR2 contained a heat shock–inducible genomic λ-Red system that required growth at 30–32 °C, with the *bla* gene replacing the *bioA* locus. Transformation of EcNR2 with pJ401 (courtesy of H. Salis, Pennsylvania State University), which expressed *bfmo* by 0.25 mM IPTG induction and selection on LB with kanamycin plates, yielded EcHW7. The tryptophan repressor (TrpR) was

doi:10.1038/nmeth.1971

inactivated in EcHW7 by MAGE using trpR_KO_Oligo. MASC-PCR screening of 48 colonies (using primer set trpR_KO_f/fm/r) identified several clones, one of which was designated EcHW44. To remove tryptophan-associated allosteric feedback, single nucleotide mutations were introduced to *aroF trpE*, and *aroG* of EcHW44 using aroF(P148L)_oligo, trpE(M293T)_oligo, aroG(D146N)_oligo by applying four cycles of MAGE along with oligo galK_mut45_oligo. After the four MAGE cycles, cells were plated on Mac-Gal with kanamycin plates and 48 *galK*$^-$ white colonies were picked after overnight growth, pooled and regrown together for another two cycles of MAGE with aroF(P148L)_oligo, trpE(M293T)_oligo, aroG(D146N)_oligo and galK_mut45_oligo (total oligo concentration of 10 μM). After the two MAGE cycles, cells were plated on Mac-Gal with kanamycin plates, and 32 *malK*$^-$ white colonies were picked after overnight growth and genotyped for the *aroF*, *trpE* and *aroG* mutations by MASC-PCR using primer sets aroF(P148L)_f/fm/r, trpE(M293T)_f/fm/r and aroG(D146N)_f/fm/r, which identified clone EcHW45 that had all the mutations. EcHW45 was used to generate the inactivated *cat*$^-$ genotype using cat_mut45_oligo and subsequently made *bla*$^-$ with bla_mut45_oligo to generate EcHW47. Screening of *cat*$^-$ and *bla*$^-$ was easily done by replica plating from LB plates onto LB with chloramphenicol and carbenicillin or more quickly done by picking ~30 colonies into LB and LB chloramphenicol and carbenicillin to identify chloramphenicol- or carbenicillin-sensitive strains. EcHW47 contained *aroF*(P148L), *trpE*(M293T), *aroG*(D146N) and revertible nonsense mutations in the inactivated *galK*$^-$, *malK*$^-$, *cat*$^-$ and *bla*$^-$ loci. Transformation of the IPTG-inducible T7 RNA polymerase gene on pN249 (courtesy of C. Voigt, Massachusetts Institute of Technology) into EcNR2, EcHW7 and EcHW47 and selection on LB with spectinomycin yielded E2N, E7N and E47N, respectively.

T7 promoter insertions by CoS-MAGE. The E47N strain was used to insert T7 promoters into 12 genomic loci 35 bp upstream of *aroL*, *aroG*, *aroA*, *trpE*, *ydiB*, *ppsA*, *aroH*, *aroC*, *aroF*, *tktA*, *aroE* and *aroB*. For the first stage of CoS-MAGE, 12 oligos were pooled together in equal molar ratio for a final concentration of 1 μM per oligo. The cat_restore_oligo (CoS oligo at final concentration of 0.02 μM) was added to this pool. The E47N strain was cycled four times by MAGE using the oligo pool over the course of 8 h (a typical workday). In general, the number of *cat*$^+$ cells in the population was ~1 in 2,000, so dilutions were made accordingly when plating. The cell population was diluted and spread on LB with chloramphenicol (for coselection for *cat*) or on LB (without coselection) overnight. We isolated and genotyped 96 colonies from each plate by MASC-PCR for presence of the T7 promoter sequences at each of the 12 target sites. Of the coselected colonies assayed, clone H46 had the most T7 promoter inserts (*aroH*, *aroC* and *aroF*), so it was used for the second stage of CoS-MAGE. In the second stage, 9 oligos corresponding to the remaining targets were pooled together in equal molar ratio as well as the galK_restore_oligo (CoS oligo at final per oligo concentration of 1 μM). We increased the CoS oligo to target oligo ratio because at the original (1:50 ratio) we expected to find 1 red *galK*$^+$ colony in ~2,000 colonies on MAC-Gal, which would be infeasible for clonal isolation. Boosting the CoS oligo concentration can increase the number of *galK*$^+$ colonies to levels screenable by plating. Alternatively, plating on M9-Gal for *galK*$^+$ colonies was also feasible and would not need higher *galK* CoS oligo concentration, but at a longer delay owing to slower colony growth on minimal medium (~2.5× of wild type). The 9 target plus *galK* CoS oligo pool was used in the second coselection stage through four cycles of MAGE on H46 strain, and the cell population was plated on Mac-Gal. Isolation and MASC-PCR genotyping of 96 red *galK*$^+$ colonies grown overnight resulted many additional T7 inserts. The best clone (H63) had six inserts in total (*aroF*, *aroC*, *aroH*, *ppsA*, *ydiB* and *aroL*). The third coselection stage was done on the remaining six targets similarly as described above using malK_restore_oligo as the CoS oligo and plating on Mac-Mal to isolate red *malK*$^+$ colonies. The best clone (H65) containing seven inserts (*aroE, aroF*, *aroC*, *aroH*, *ppsA*, *ydiB* and *aroL*) was isolated. The fourth coselection stage was done on the remaining five targets using bla_restore_oligo. Selection on LB with carbenicillin and isolation of 96 colonies resulted in the best clone, H77, that contained ten inserts (without *tktA* and *aroB*). We reiterated CoS-MAGE on the remaining targets to generate the H80 clone, which contained all 12 desired T7 promoter insertions as well as other clones containing unique insertions through MASC-PCR every two MAGE cycles. A summary of genomic distance of each locus to the relevant coselectable marker is shown in **Supplementary Table 1**. In total, 80 genotypically unique strains (H1-H80) were isolated containing 1–12 T7 promoter insertions. We transformed the T7 polymerase–encoding pN249 plasmid into each strain and selected on LB with spectinomycin for the presence of the plasmid.

Sequencing verification. We sequenced all 12 target sites for the expected T7 promoter sequence in each of the 80 strains (H1-H80). The PCR primers amplified ~150–180 bp around each T7 insertion locus by a standard protocol. All sequenced sites were verified to contain the correct T7 sequence (TAATACGACTCACTATAGGG) except site *aroH* of strain H70, in which the T7 sequence was TAATACACTCAGTATGGG.

Pigment extraction and quantification. To quantify the amount of upregulation in the modified biosynthesis pathway, we measured indigo and indirubin production in strains H1–H80 and ancestral strains E2N, E7N and E47N following previously described methods[3]. Each strain was grown in 10 ml of LB with kanamycin, spectinomycin and IPTG for 24 h at 30 °C before extraction. Strain E2N was grown in LB with spectinomycin and IPTG only. For pigment extraction, 1 ml of each strain was concentrated by centrifugation in a deep-well 96-well plate. Excess medium was discarded, and each cell pellet was resuspended in 1 ml of DMSO. The samples were sonicated for 10 s and centrifuged again to concentrate the cell debris, which appeared white. The blue supernatant was removed and stored at room temperature in the dark to allow accurate measurement of indigo and indirubin levels. After 3 d, 200 μl of each supernatant was transferred to a 96-well plate, and absorbance at 550 nm was measured on a Molecular Devices spectrophotometer to quantify relative indirubin amounts. Indigo at 620 nm was also measured. However, the absorbance values decayed over time (**Supplementary Fig. 2**), likely owing to indigo instability. Thus, to determine absolute indigo production levels, we took the relative indirubin values against the amounts in strain H33 (best producer) and normalized them against the measured indigo production level (at Abs_{620}) of an immediately extracted H33 strain sample. Indigo amounts were determined by mapping the Abs_{620} to a calibration curve generated using dilutions

of commercially obtained indigo powder. Five milliliters of cells were dried and weighed to derive the dry cell weight. Pigment extractions were repeated four separate times on different days for each strain to ensure statistical confidence.

Statistical analysis. *P* values described in the paper were determined by performing the Student's *t* test between the two sample groups to derive the *t* score, which is then used to compute the *P* value.

doi:10.1038/nmeth.1971

Nature Methods

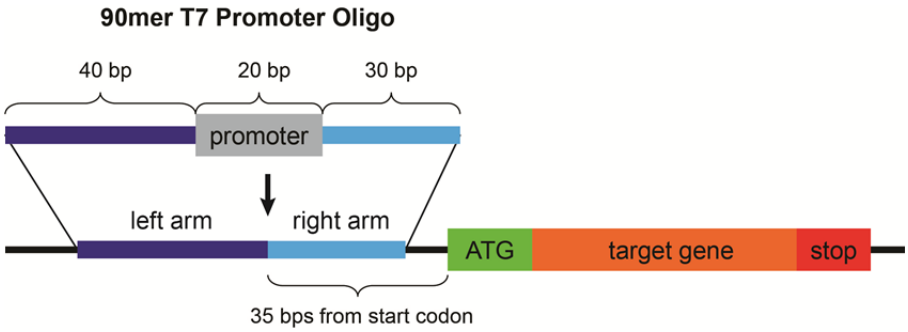
Genome-scale promoter engineering by coselection MAGE

Harris H Wang, Hwangbeom Kim, Le Cong, Jaehwan Jeong, Duhee Bang & George M Church

Supplementary Figure 1	Insertion design used to introduce T7 promoters upstream of genes/operons.
Supplementary Figure 2	Pigment quantification.
Supplementary Figure 3	Correlation of indigo production level and growth rate
Supplementary Figure 4	Genomic map and list of 21 available Co-Selection markers that can be used to enhance MAGE efficiency.
Supplementary Figure 5	Characterization of the insertion efficiency of various lengths
Supplementary Figure 6	Comparison of the length of time in experimental hours needed for different MAGE strategies to generate clones with 20bp T7 promoter insertions.
Supplementary Table 1	Summary of stages of CoS-MAGE to generate the H77 strain.
Supplementary Table 3	Side-by-side comparison of various genome engineering strategies to make a strain that contains 20bp T7 insertion mutation across 10 genomic loci.
Supplementary Note	Co-selection mechanism and enhancement analysis

Note: Supplementary Tables 2 and 4 are available on the Nature Methods website.

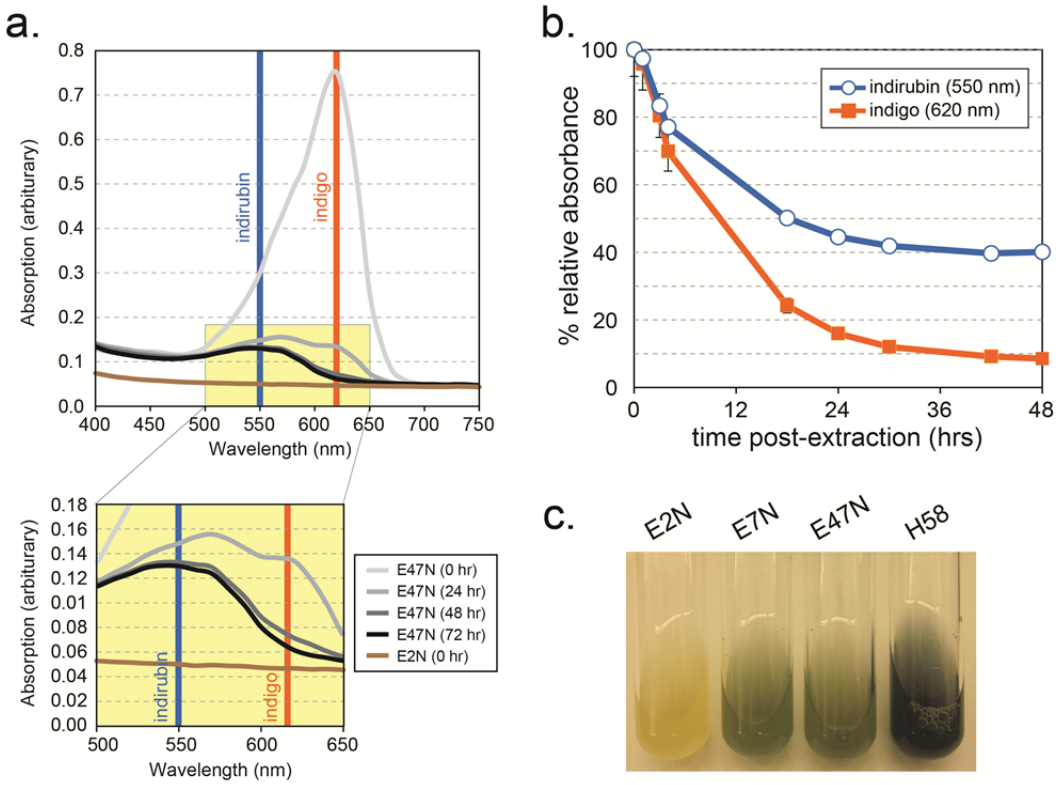
Nature Methods: doi:10.1038/nmeth.1971



Supplementary Fig. 1

Insertion design used to introduce T7 promoters upstream of genes/operons. The T7 promoter was inserted 35 bp upstream of the open reading frame to avoid disrupting the ribosomal binding site.

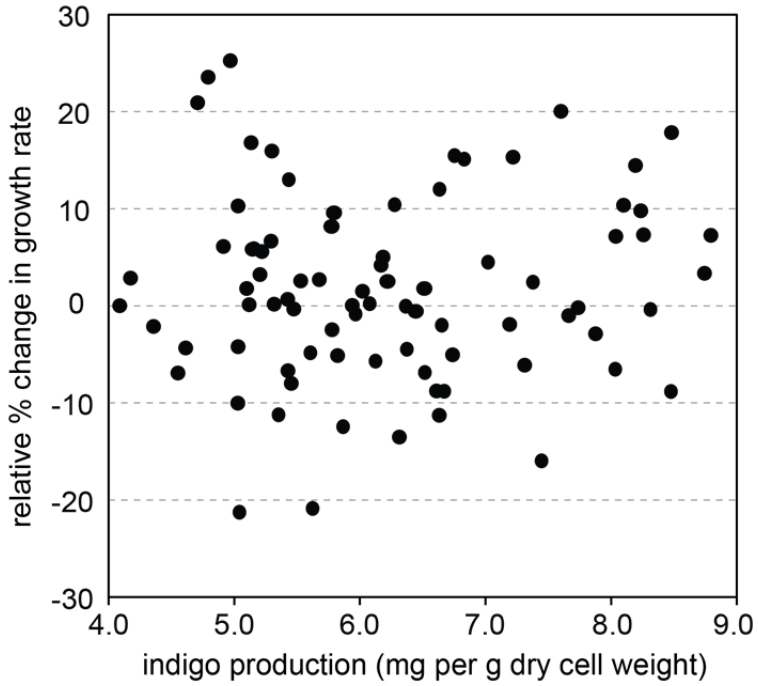
Nature Methods: doi:10.1038/nmeth.1971



Supplementary Fig. 2

a, Absorbance spectrum of extracted pigment from E47N measured at 400 to 750 nm in DMSO after incubation at room temperature for 0, 24, 48 and 72 hours (gray to black lines). The insert highlights the relevant wavelengths 550 nm and 620 nm corresponding to max absorbance peaks for indirubin and indigo respectively. Indirubin levels stabilized after 48 hrs. **b,** Calculation of relative absorbance change over time for 550 nm peak (indirubin) and 620 nm peak (indigo). Absorbance at 620 nm dropped substantially after 2 days likely due to indigo instability. Absorbance at 550 nm stabilized to ~40% of initial levels, likely after interference of from indigo at 620 nm has disappeared. Indirubin levels (Ab_{550}) was thus used to quantify pigment production. **c,** Picture of 3 mL cultures of E2N ancestral control, E7N and E47N controls and high pigment producer clone H58 grown in LB-min media for 24 hrs. The blue pigmented cultures were the result of indigo and indirubin production.

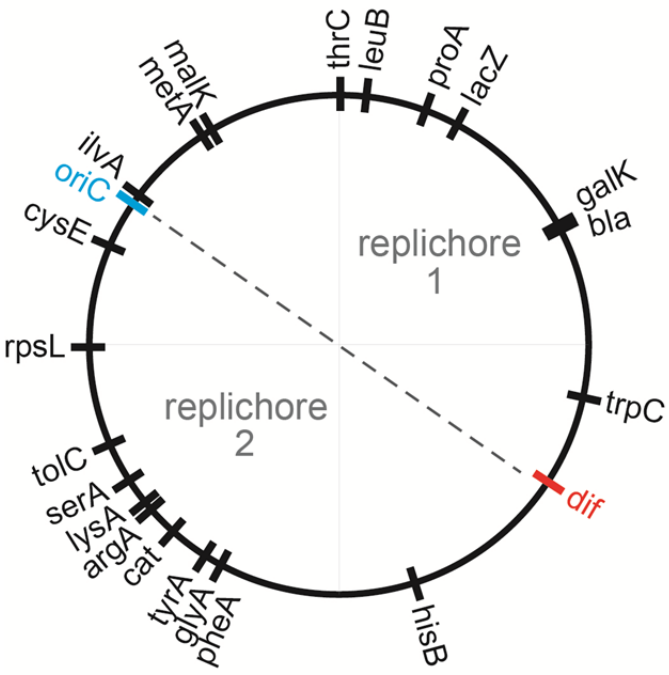
Nature Methods: doi:10.1038/nmeth.1971



Supplementary Fig. 3

Indigo production level and relative % change in growth rate does not appear to have any notable correlation for strains H1-H80, E2N, E7N, and E47N.

Nature Methods: doi:10.1038/nmeth.1971

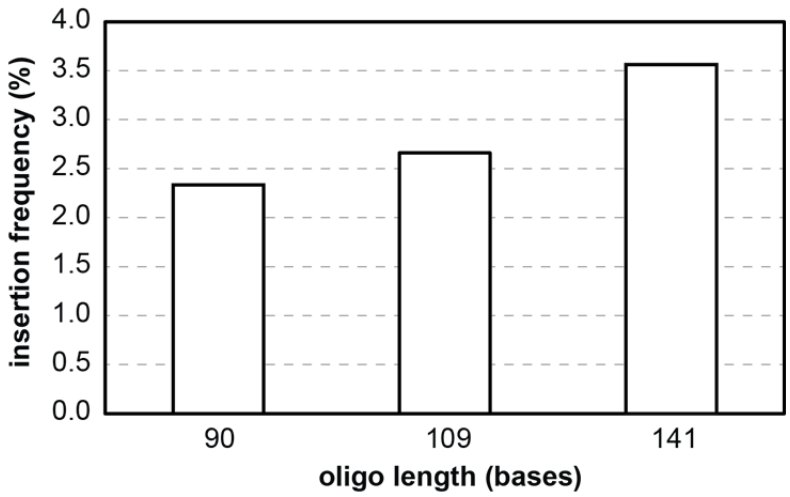


REPLIC	MARKER	LOCATION
1	thrC	3783
1	leuB	81795
1	proA	260812
1	lacZ	365478
1	galK	788768
1	ΔbioA::bla	808383
	dif	1588788
1	trpC	1317764
2	hisB	2091541
2	pheA	2735846
2	glyA	2683484
2	tyrA	2738026
2	ΔmutS::cat	2855406
2	argA	2947499
2	lysA	2976834
2	serA	3056383
2	tolC	3176216
2	rpsL	3472446
2	cysE	3780470
	oriC	3923883
1	ilvA	3953577
1	metA	4212352
1	malK	4245059

Supplementary Fig. 4

Genomic map and list of 21 available Co-Selection markers that can be used to enhance MAGE efficiency. The genomic coordinate of each CoS marker is given. CoS markers most efficiently enhance nearby target sites within ~500 kb on the same replichore due to the nature of the bidirectional θ-replication of the *E. coli* genome. CoS oligos that inactivate or restore each marker are provided in Supplementary Table 2.

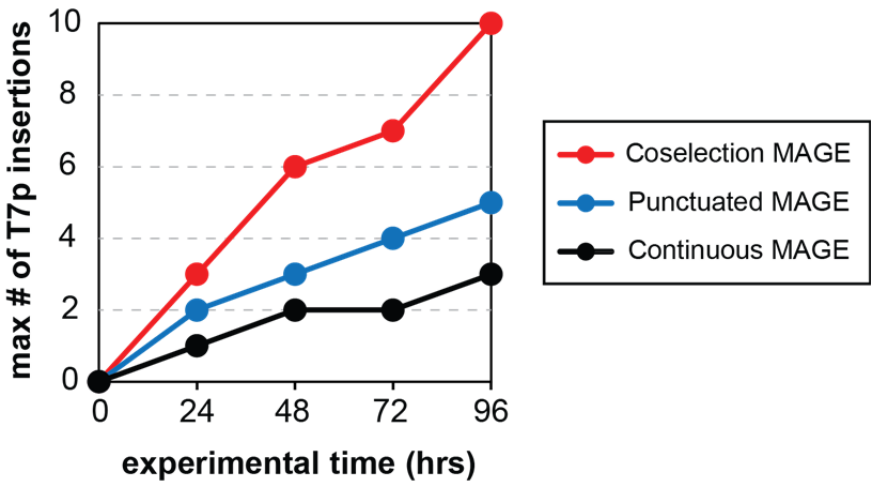
Nature Methods: doi:10.1038/nmeth.1971



Supplementary Fig. 5

Characterization of the insertion efficiency of a 19 bp sequence (ATGATGATGATGATGATGA) versus length (and homology arm) of the oligo. Insertion of the 19 bp sequence inactivated the *lacZ* gene by generating a premature stop codon. The 90-base oligo (GATTCACTGGCCGTCGTTTTACAACGTCGTGACTGATGATGATGATGATGATGAGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGC) has 35-36 bases of homology to the target sequence on each arm. The 109-base oligo (CATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGATGATGATGATGATGATGAGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCC) has 45 bases of homology to the target sequence on each arm. The 141-base oligo (AGGAAACAGCTATGACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGATGATGATGATGATGATGAGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGG) has 61 bases of homology to the target sequence on each arm. Oligos were obtained from Integrated DNA Technology's Ultramer Custom Oligo service. Increased insertion efficiency was observed with increased homology in the absence of inhibitory oligo secondary structures. All oligos had minimal secondary structures (ΔG of -7.7 to -11.4 kcal/mol).

Nature Methods: doi:10.1038/nmeth.1971



Supplementary Fig. 6

Comparison of the length of time in experimental hours needed for different MAGE strategies to generate clones with 20bp T7 promoter insertions. Continuous MAGE is the standard MAGE process ~10 cycles/day with no clone isolation until the end of the cycling process. Punctuated MAGE is the MAGE process with ~4 cycles/day with plating, clone isolation, and enrichment for the most mutated clone every 24 hrs. Coselection MAGE is our enhanced MAGE process using CoS markers to generate clones with up to 3 insertions per day.

Nature Methods: doi:10.1038/nmeth.1971

Supplementary Table 1

Summary of Stages of CoS-MAGE to generate the H77 strain.

Stage	Best Strains	CoS marker	Genes with T7 insertions	Distance from CoS marker (bp)
0	E47N			
1	H46	cat	aroH	1068923
			arof	117280
			aroC	410972
2	H63	galK	ppsA	1003911
			ydiB	992966
			aroL	382425
3	H65	malK	aroE	816765
4	H77	bla	aroG	32881
			aroA	149505
			trpE	510878

*Distance values are calculated from start codon of CoS marker genes to start codon of target genes.

Nature Methods: doi:10.1038/nmeth.1971

Supplementary Table 3

Side-by-side comparison of various genome engineering strategies to make a strain that contains 20bp T7 insertion mutation across 10 genomic loci.

	dsDNA Recombineering	continuous MAGE	punctuated MAGE	coselection MAGE
Multiplexibility	no	yes	yes	yes
Insertion efficiency	n/a	0.02	0.02	>0.2 near CoS
# of loci	10	10	10	10
Per locus insertion eff	n/a	0.002	0.002	>0.02 near CoS
# of cycles needed	n/a	610	40	16
Days of experiments	20	61	10	4
Max # of mutations made per day	0.5	0.1	1.0	2.5

Nature Methods: doi:10.1038/nmeth.1971

Supplementary Note

Co-Selection Mechanism and Enhancement Analysis

Here, we speculate on some mechanistic rationale for the observed improvements in insertion efficiency by CoS MAGE. Our findings that nearby sites are found to be incorporated at higher frequency in these enriched cells suggest a strong temporal component of this process, which can be explained by the replicative process. We believe that these consecutive oligo integration events are in fact happening along the same lagging strand of the replicating DNA. Additionally, this mechanism seems to fall in line with the observation that CoS markers do not confer much benefit to other far-off sites or cross-replichore sites, thus further pointing to a local replication fork accessibility mechanism. For any loci, it is accessible to lagging strand oligo integration at max once every 25 min with each 1-3 kb stretch of DNA existing in the open state for only ~1 second every cell doubling (assuming 25 min doublings). The CoS marker may perhaps be viewed as artificially synchronizing DNA replication in the cell population by selecting for replication forks that are open and have undergone an oligo-mediated integration event at a given site.

The crucial feature of CoS-MAGE is the increased capability to enrich for cells with higher number of mutations and longer base insertions, thus extending the multiplexing capabilities of MAGE. With CoS-MAGE, ~5% of cells had 3 insertions. With standard MAGE, no 3-insert cells were found when the same numbers of colonies were screened. Standard MAGE generates mutations based on a binomial distribution. Assuming that the efficiency of each insertion is 2% for a set of 12 targets, then the frequency of generating 1 or more inserts is 21.5% within the population as calculated by $\sum_{n=1}^{12}\binom{12}{n}(0.02)^{n}(1-0.02)^{12-n}$. Isolation of a 1-mutation clone at a 95% likelihood would need to satisfy the condition $(1\text{-}0.215)^{x} < 0.05$, which gives an *x* of 12; meaning that 12 clones need to be screened to find one that has at least 1 mutation at 95% confidence. To find a clone with 3 or more mutations in this population

Nature Methods: doi:10.1038/nmeth.1971

(frequency of 0.2%), one would need to screen 1948 colonies. Thus, there is a 2 order of magnitude difference in the screening requirements for finding a 1 vs 3 mutation clone. Given our CoS-MAGE experimental results of ~5% abundance of cells with 3-inserts or more, we can back calculate to find that our individual insertion efficiency would be 8% (vs 2% with standard MAGE). Thus with co-selection, the effect on multiplexing is exponential such that we are able to reduce the screening requirement for finding highly modified cells by substantial enrichment (>2 log). Further comparisons of CoS MAGE with other modalities of MAGE and dsDNA recombineering techniques are outlined in **Supplementary Table 3**.

Nature Methods: doi:10.1038/nmeth.1971