Transient overexpression of DNA adenine methylase enables efficient and mobile genome engineering with reduced off-target effects

Lennen, Rebecca; Nilsson Wallin, Annika; Pedersen, Margit; Bonde, Mads; Luo, Hao; Herrgard, Markus; Sommer, Morten Otto Alexander

Published in:
Nucleic Acids Research

Link to article, DOI:
10.1093/nar/gkv1090

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Published online 22 October 2015

Nucleic Acids Research, 2016, Vol. 44, No. 4 ***e36***
doi: 10.1093/nar/gkv1090

Transient overexpression of DNA adenine methylase enables efficient and mobile genome engineering with reduced off-target effects

Rebecca M. Lennen†, Annika I. Nilsson Wallin†, Margit Pedersen†, Mads Bonde, Hao Luo, Markus J. Herrgård and Morten O. A. Sommer*

The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kogle Allé 6, 2970 Hørsholm, Denmark

Received July 31, 2015; Revised September 16, 2015; Accepted October 08, 2015

ABSTRACT

Homologous recombination of single-stranded oligonucleotides is a highly efficient process for introducing precise mutations into the genome of *E. coli* and other organisms when mismatch repair (MMR) is disabled. This can result in the rapid accumulation of off-target mutations that can mask desired phenotypes, especially when selections need to be employed following the generation of combinatorial libraries. While the use of inducible mutator phenotypes or other MMR evasion tactics have proven useful, reported methods either require non-mobile genetic modifications or costly oligonucleotides that also result in reduced efficiencies of replacement. Therefore a new system was developed, Transient Mutator Multiplex Automated Genome Engineering (TM-MAGE), that solves problems encountered in other methods for oligonucleotide-mediated recombination. TM-MAGE enables nearly equivalent efficiencies of allelic replacement to the use of strains with fully disabled MMR and with an approximately 12- to 33-fold lower off-target mutation rate. Furthermore, growth temperatures are not restricted and a version of the plasmid can be readily removed by sucrose counterselection. TM-MAGE was used to combinatorially reconstruct mutations found in evolved salt-tolerant strains, enabling the identification of causative mutations and isolation of strains with up to 75% increases in growth rate and greatly reduced lag times in 0.6 M NaCl.

INTRODUCTION

The ability to rapidly and precisely manipulate the genome of *E.coli* was first revolutionized by the development of mobile plasmids and prophages encoding inducible λ Red recombinase (1,2). The success of these systems was due to their mobility and curability as plasmids with temperature-sensitive origins, the enhanced rates of recombination afforded by expressing a viral recombinase, and the natural propensity of the γ subunit of λ Red to inhibit *E. coli* RecBCD exonuclease V. The latter feature allowed the transformation of linear double-stranded DNA (dsDNA) cassettes without degradation occurring too rapidly relative to recombination. It was additionally discovered that single-stranded DNA (ssDNA) oligonucleotides could be used to mediate homologous recombination without the drawbacks of degradative activity against dsDNA, with only the β subunit of λ Red recombinase being required for recombination (3). This discovery ultimately led to the development of more specific techniques for ssDNA oligonucleotide-mediated recombineering (4) and multiplex automated genome engineering (MAGE), where an electroporation mixture is cycled multiple times with multiplexed pools of oligonucleotides synthesized individually or on arrays to generate a combinatorial library of mutants (5,6).

Due to its ability to rapidly generate diversity at precisely controlled locations, MAGE has opened up a diverse array of applications in the fields of metabolic engineering, synthetic biology and evolutionary biology. In metabolic engineering, these include generation of precise mutant libraries of interest, such as ribosome binding site libraries targeting one or more genes, promoter libraries and simultaneous manipulation of expression of genes in multiple pathways (5–7). In synthetic biology, applications have thus far included genome-wide genetic reprogramming, such as replacement of all UAG stop codons with UAA (8) and rare codon replacement in a panel of essential genes (9). In evolutionary biology, MAGE has been used to combinatorially reconstruct mutations identified in adaptive evolution experiments to assess causality (10) and condition-specific selection of antibiotic resistance (11).

*To whom correspondence should be addressed. Tel: +45 21 51 83 40; Fax: +45 45 25 80 01; Email: msom@bio.dtu.dk
†These authors contributed equally to this work as the first authors.

In spite of the wide range of opportunities presented by MAGE, applications have lagged due to some limitations of the MAGE systems published to date. Most notably, the majority of previously reported systems achieved high efficiencies of replacement of single nucleotides through the use of strains defective in mismatch repair (MMR) (by deletion of *mutS*, encoding one subunit of the MutHLS methyl-directed mismatch repair system) (4,5). Strains with disabled MMR systems achieve multiple orders-of-magnitude increases in efficiency for replacement of targeted single nucleotides, however, they exhibit a mutator phenotype with approximately 1000-fold increased spontaneous mutation rates relative to strains with intact MMR (12,13). These off-target mutations can potentially obfuscate results obtained from libraries that undergo large numbers of MAGE cycles or in MAGE applications where the generated combinatorial library is subsequently subjected to selections to isolate mutants with improved phenotypes. Resulting isolated mutants also exhibit genomic instability unless MMR is restored, thus deterring efforts to perform high-throughput screening or subsequent strain manipulations. Additional minor limitations include the use of relatively non-mobile prophage-based systems, often with temperature-inducible recombinases (4,5). These restrict the ability to easily remove the system following MAGE and limit growth both during and after MAGE to temperatures of 32°C or lower, respectively.

The limitations in using MMR defective strains have previously been addressed through the use of non-MMR defective strains together with modifications that evade mismatch repair. This includes strategies such as generating mutations using particular mismatching base pairs (e.g. C:C or A:G) (4), forcing the mutation of more than five bases in a row (4), and the use of chemically modified bases that can functionally replace standard bases while evading mismatch repair (14). The former two strategies are undesirable as they significantly limit the scope of mutations possible to introduce by MAGE and preclude many applications such as reconstructing combinations of specific mutations observed in evolved strains and constructing fully diversified ribosome binding site libraries. The latter strategy introduces very significant cost increases for oligonucleotides and greatly reduced maximum measured allelic replacement efficiencies of up to 5-fold were observed for modified G (iso-deoxyguanosine) and A (2,6-diaminopurine) bases (14).

An alternative strategy for overcoming the high off-target mutation rate induced by permanent disabling of MMR is the use of conditionally active mismatch repair proteins (15). A MAGE host strain was constructed by Nyerges *et al.* by introducing previously discovered mutations in MutS and MutL that result in temperature-sensitive mismatch repair activity (16) plus a prophage containing a temperature-inducible λ Red recombinase. This strain exhibited comparable allelic replacement efficiencies to a *mutS*$^{-}$ strain of up to 25%, however, the spontaneous mutation rate remained low only at temperatures below 32°C. While a promising strategy, the system would be difficult to employ in alternative host strains and would require either maintaining resulting isolated mutants at or below 32°C, or undergoing a time-consuming process to remove the MAGE system and revert the MMR proteins back to wild-type alleles.

In this study, we report a new plasmid-based MAGE system, Transient Mutator Multiplex Automated Genome Engineering (TM-MAGE) (Figure 1A), that overcomes the limitations of previously reported methods that include high off-target mutation rates, lack of mobility and curability of the system, and the lack of ability to cultivate resulting mutant strains at 37°C. This is accomplished by transiently disabling MMR during MAGE cycling by inducible overexpression of *E. coli* Dam methylase, together with overexpression of the β subunit of λ Red in an artificial operon on a plasmid. The *dam* gene target was identified from an earlier screening study searching for native genes that increased the spontaneous mutation rate in *E. coli* when overexpressed (17). The efficiency and off-target mutation rates were measured for this system and compared to the EcNR2 strain (5), a plasmid-based system harboring only β in a *mutS*$^{+}$ host, and the same system in a *mutS*$^{-}$ host. To further demonstrate the utility of this system, MAGE was used to introduce several mutations that had previously been identified in strains evolved for growth in high NaCl concentrations (18). Subsequent selection was used to enrich the population in these mutants, and the off-target mutation rate was shown to be substantially lower using TM-MAGE than performing MAGE with a plasmid-based system in a *mutS*$^{-}$ strain. A version of the plasmid containing the *sacB* gene (encoding levansucrase from *Bacillus subtilis*) enables one-step curing of the system by sucrose counterselection.

MATERIALS AND METHODS

Chemicals, reagents and media

All chemicals and reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise stated. Tryptone and yeast extract used for media preparation were purchased from Sigma-Aldrich. LB medium (Lennox) used for cultivations during MAGE cycling contained 10 g/l tryptone, 5 g/l yeast extract and 5 g/l sodium chloride (Sigma). An additional 0.5 mM $MgSO_4$ was added to all LB media to correct an inherent magnesium limitation observed with these media components. Strains harboring plasmids were maintained with 100 μg/ml ampicillin except during growth for MAGE cycling and subsequent selections. MacConkey agar was purchased from Becton Dickinson (Franklin Lakes, NJ).

Oligonucleotides

All oligonucleotide sequences used in this study are shown in Supplementary Table S1. MAGE oligonucleotides used to introduce precise mutations that were previously identified for improved salt tolerance (18) (oligonucleotides **1** through **7**) were designed using the MODEST oligonucleotide design tool (19), with the length set to 70 bp, 15 bp minimum end homology and other options kept as default. Other MAGE control oligonucleotides (to introduce premature stop codons in *galK* and *xylA*, oligonucleotides **8** and **9**) were designed prior to the development of MODEST and feature mutated bases centered in the middle of

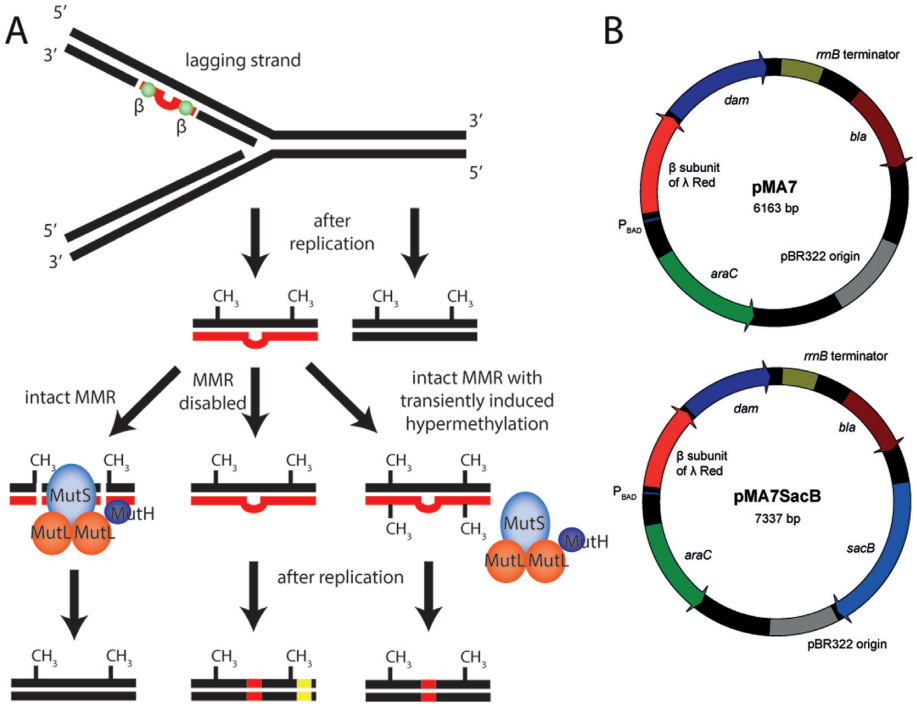

Figure 1. (**A**) Proposed mechanism of Transient Mutator Multiplex Automated Genome Engineering (TM-MAGE). Chromosomal mutations are introduced via single-stranded oligonucleotide recombination on the lagging strand of replicating DNA, mediated by the β subunit of λ Red recombinase. Under ordinary conditions (left), the newly-synthesized strand is hemimethylated, which enables the MMR system, composed of MutS, MutL and MutH, to bind to mismatched lesions and restore the wild-type allele. In standard MAGE (middle), the MMR system is eliminated, allowing incorporation of the mismatched base and its complement base into the genome of one daughter cell after a second round of replication (red). However, permanent disabling of mismatch repair also allows the accumulation of off-target mutations (yellow) during subsequent manipulations. In TM-MAGE (right), transient overexpression of DNA adenine methylase (Dam) is hypothesized to generate significant methylation of the newly-synthesized strand, disabling detection of incorporated mismatches by the MMR system and reducing the rate of off-target mutations that accumulate in subsequent manipulations as Dam expression returns to its baseline level. (**B**) Schematic of TM-MAGE system plasmids pMA7 (top) and pMA7SacB (bottom). Plasmid pMA7 contains an artificial operon of the gene encoding the β subunit of λ Red recombinase and *dam* controlled by an arabinose-inducible (P_{BAD}) promoter. Plasmid pMA7SacB additionally contains constitutively expressed *sacB* (levansucrase from *B. subtilis*) to allow for plasmid curing by sucrose counterselection.

the oligonucleotide. For demonstration of the maximum achievable MAGE efficiency for a single bp mismatch, a 90-bp oligonucleotide that results in the same *xylA* premature stop codon was designed with 4 phosphorothioated bases introduced at the 5′ and 3′ termini (oligonucleotide **10**). All oligonucleotides were synthesized by Integrated DNA Technologies (Leuven, Belgium) with standard desalting and without additional purification.

Strains

Relevant strains used in this study are shown in Table 1. K-12 MG1655 was kindly donated by Adam Feist (University of California, San Diego), which is equivalent to ATCC 47076/CGSC 6300 (20). EcNR1 and EcNR2 are K-12 MG1655 and K-12 MG1655 *mutS::cat* harboring a λ prophage at the *attB* site which contains a temperature inducible λ Red recombinase (5). K-12 MG1655 *mutS::cat* was constructed by P1*vir* transduction of *mutS::cat* from strain EcNR2 into donor strain K-12 MG1655. The *mutS* replacement was confirmed by colony PCR using oligonucleotides **11** and **12** as primers (Supplementary Table S1).

Plasmid construction

Plasmid pMA1 (GenBank accession number KT285943) was constructed by PCR amplifying the gene encoding β from strain EcNR1 (5), while also introducing a Shine–Dalgarno sequence immediately upstream of the start codon, using primers **13** and **14** and inserting the PCR product between the NheI and HindIII sites in pBAD24 (21). Plasmid pMA3 was constructed by amplifying *dam* from K-12 MG1655, while also introducing a Shine–Dalgarno sequence immediately upstream of the start codon, using primers **15** and **16** and inserting between the NheI and HindIII sites in pBAD18-Cm (21). Plasmid pMA7 (GenBank accession number KT285941) was constructed by standard USER cloning (22) of three fragments, with the first fragment amplified using primers **17** and **18** from template pMA1, the second fragment amplified using primers **19** and **20** from template pMA1, and the third fragment amplified using primers **21** and **22** from template pMA3. Plasmid pMA7SacB (GenBank accession number KT285942) was constructed using a Gibson Assembly Kit (New England Biolabs). Primers **23** and **24** were used to PCR amplify the parental pMA7 plasmid. The *sacB* gene was amplified from *E. coli* DA16543 (donated by Dan Andersson, Uppsala University) using primers **25** and **26**. Resulting DNA fragments were assembled according to the manufacturer's

Table 1. Strains and plasmids used in this study

Strain/plasmid	Relevant genotype/property[a]	Source/ Reference
Strains		
K-12 MG1655	$F^-\lambda^-$ *ilvG- rfb-50 rph-1*	A. Feist, UCSD
EcNR1	K-12 MG1655 λ-Red1 *bla bio*$^-$ Cm^R	5
EcNR2	K-12 MG1655 λ-Red1 *bla bio*$^-$ *mutS*$^-$ Cm^R	5
Plasmids		
pMA1	β subunit of λ Red under P_{BAD} promoter, Amp^R, ColE1 origin from pBAD24	This study
pMA3	*dam* under P_{BAD} promoter, Cm^R, ColE1 origin from pBAD18-Cm	This study
pMA7	β subunit of λ Red and *dam* under P_{BAD} promoter, Amp^R, ColE1 origin from pBAD24	This study
pMA9	β subunit of λ Red and *dam* under P_{BAD} promoter, Amp^R, RSF1030 origin	This study
pMA10	β subunit of λ Red and *dam* under P_{BAD} promoter, Amp^R, CloDF13 origin	This study
pMA7SacB	β subunit of λ Red and *dam* under P_{BAD} promoter, *sacB* from *B. subtilis* under constitutive promoter, Amp^R, ColE1 origin from pBAD24	This study

[a]Abbreviations: Amp, ampicillin; Cm, chloramphenicol; R, resistance.

instructions. All plasmids generated in this study are listed in Table 1.

MAGE cycling

Starter cultures of K-12 MG1655 harboring plasmids pMA7, pMA1 or pMA7SacB, or strain K-12 MG1655 *mutS::cat* harboring plasmid pMA1 were grown overnight in LB + 100 μg/ml ampicillin at 37°C with 250 rpm shaking. Strain EcNR2 was grown in LB overnight at 30°C with 250 rpm shaking. After ∼16 h, 15 ml of LB medium (containing 100 μg/ml ampicilling where appropriate) in 50 ml Falcon tubes was inoculated to an initial OD_{600} of 0.05, and incubated at either 37°C (K-12 MG1655 strains) or 30°C (EcNR2) until the OD_{600} reached 0.6–0.7. For strains harboring plasmids pMA7, pMA1 or pMA7SacB, the recombinase and *dam* functions were induced by addition of L-arabinose to a final concentration of 0.2% w/v, and incubation with shaking was continued for 10 min. For strain EcNR2, the recombinase function was induced by transferring the culture to a 42°C water bath with 200 rpm linear shaking for 15 min. Cultures were then moved to an ice water bath and chilled for 15 min. Cells were centrifuged in Falcon tubes at 4000 x *g* for 5 min at 4°C, the supernatant was decanted and cells were subsequently resuspended in 30 ml, and thereafter 1 ml of autoclaved MilliQ water. Cells were transferred to 2.0 ml microcentrifuge tubes and centrifuged at 4000 x *g* for 5 min at 4°C, the supernatant was removed by pipetting, and cells were resuspended in 0.2 ml of autoclaved MilliQ water. Fifty microliters of cells were then transferred to 1 mm gap electroporation cuvettes, 0.5 μl of one oligonucleotide (for oligonucleotides **8**, **9** or **10** resuspended to 100 nmol/ml in MilliQ water) or 3.5 μl of a mixture of 7 oligonucleotides (for oligonucleotides **1** through **7**, pre-mixed in a tube with a total concentration of 100 nmol/ml) was added, and mixed with a Pasteur pipette. Cells were immediately electroporated using an ECM 399 electroporation system (Harvard Apparatus, Holliston, MA) at 1.8 kV and 36 μF (resulting in a 5 ms pulse length) and resuspended in 950 μl LB. The entire contents of the electroporation cuvette were either transferred into 50 ml Falcon tubes containing 14 ml of LB when more cycles were to be performed in a single day (for 6 cycle experiments, 3 cycles were performed per day), or into a 14 ml culture tube containing 1 ml of LB when the final cycle of the day was completed. Cultures were grown overnight at 37°C (or 30°C for strain EcNR2) with 250 rpm shaking to allow for chromosome segregation.

For experiments performed during initial optimization of *dam* expression, optimization of the pMA7 system, and fluctuation tests for determination of mutation rates, an identical procedure was performed except that 15 ml of LB was inoculated 1:60 with starting culture, and centrifugation conditions were 6500 x *g* for 7 min at 4°C for the first two spins and 7000 x *g* for 30 s at 4°C for the final spin in a microcentrifuge. Induction of either arabinose-inducible or temperature-inducible β or λ Red was performed as previously described. For optimization of *dam* expression, cells were either non-induced, induced before chilling and centrifugation, induced immediately after electroporation, or induced both before chilling and centrifugation and after electroporation.

Plasmid curing

Following MAGE cycling and plating, individual colonies were struck as wedges on plates containing 10 g/l tryptone, 5 g/l yeast extract, 15 g/l agar, and 5% w/v sucrose (added from a 50% w/v stock after autoclaving the agar and media components) and grown at 37°C overnight. The next day, individual colonies were restruck on LB agar and LB agar containing 100 μg/ml ampicillin to verify plasmid removal by loss of ampicillin resistance.

Fluctuation tests

For the non-induced fluctuation tests, cultures were grown to saturation overnight in LB, or when pMA7 or pMA1 plasmids were present, LB + 100 μg/ml ampicillin. The following day, ten tubes containing 1.5 ml or 5 ml LB were inoculated with 5 μl of the saturated cultures. For non-induced state fluctuation tests, cultures were grown again to saturation overnight. For induced state fluctuation tests, cultures were grown to OD_{600} 0.4, induced with 0.2% L-arabinose, and electrocompetent cells were prepared as described above. Following electroporation with 5 pmol of oligonucleotide **45**, cells were grown to saturation in 5 ml of LB overnight. The following day, 10 μl (if a mutator strain background) or 200 μl (if a non-mutator strain background) was plated on LB agar containing 50 mg/ml rifampicin, and 100 μl (if a mutator strain background) or 1 ml (if a non-mutator strain background) was plated on LB agar containing 50 mg/ml nalidixic acid. Cells were also se-

rially diluted and plated on LB agar to determine total viable cell counts. Mutation frequencies were determined as the ratio of the number of colonies per ml culture appearing on rifampicin or nalidixic acid plates to the total number of viable colonies per ml culture.

Selections of combinatorial populations generated by MAGE

One milliliter of MAGE library cultures that were generated after 6 cycles using the 7 oligonucleotide mixture (**1** through **7**) were washed once to remove LB medium by centrifuging at 4000 x *g* for 1 min, removing the supernatant, and resuspending in 1 ml of M9 medium. The resulting cell suspensions were diluted 1:100 (25 μl inoculum in 2.5 ml medium) into either M9 medium or M9 containing an additional 0.5 M NaCl in triplicates in 24 well plates. The M9 resuspended libraries were also serially diluted and plated on M9 agar or M9 agar + 0.6 M NaCl. After approximately 24 h, liquid cultures were passaged to an initial OD_{600} of 0.05 in fresh M9 medium, or in M9 medium containing 0.6 M NaCl, and this was repeated daily for 5 days (Supplementary Figure S3). Each day, approximately 1.5–2.0 × 10^9 cells were harvested from each culture by centrifugation at 16 000 x *g* for 2 min, and cell pellets were stored at −20°C. Following the fifth day, the cultures passaged in M9 + 0.6 M NaCl were serially diluted and spread on plates containing M9 agar + 0.6 M NaCl, to isolate individual colonies. Individual colonies were selected and regrown on LB agar for subsequent screening and analysis.

Determination of allele replacement efficiencies

Oligonucleotides **8, 9** and **10** were designed to introduce premature stop codon mutations in *xylA* (encoding xylose isomerase) and *galK* (encoding galactokinase), thereby rendering cells incapable of catabolizing D-xylose or D-galactose. To assess replacement efficiencies, cultures electroporated with these control oligonucleotides were serially diluted in phosphate buffered saline and plated on MacConkey agar plates containing 1% w/v D-xylose (for oligonucleotides **8** and **10**) or 1% w/v D-galactose (for oligonucleotide **9**). Plates were incubated overnight at 37°C and colonies that were either stained pink (wild-type) or white (mutant) were counted.

Another measure of allele replacement efficiency, which also enabled determination of replacement efficiencies using the pooled NaCl oligonucleotides (**1** through **7**), was performed using amplicon sequencing (23). Briefly, cell pellets containing an estimated 5.0 × 10^8 to 1.5 × 10^9 cells (assuming OD_{600} 1.0 = 1 × 10^9 cells/ml) were harvested by centrifugation at 16 000 x *g* for 2 min, either during exponential growth following an intermediate daily cycle, or following overnight outgrowth for the only or last daily cycle performed. The supernatant was removed and cell pellets were stored at −20°C until further use. Selected amplicons spanning a 150–170 bp region around targeted mutations were amplified by colony PCR using primers containing 5′ overhangs complementary to priming sequences for Nextera XT indexed adapters (Supplementary Table S1, primers **27** through **44**). Each PCR reaction contained 5 μl of 2X Phusion HotStart II master mix (Thermo Scientific), 0.25 μl of forward primer (20 μM), 0.25 μl of reverse primer (20 μM), 3.5 μl water and 1 μl of resuspended cells (each frozen cell pellet was resuspended in 300 μl MilliQ water), and was run in a thermal cycler with the following program: (i) 98°C for 3 min, (ii) 98°C for 10 s, (iii) 56°C for 30 s, (iv) 72°C for 10 s, (v) Repeat steps 2–4 25 times, (vi) 72°C for 5 min and (vii) hold at 4°C. Different amplicons that could be sequenced under the same index were pooled together by combining 5 μl of each PCR product and adjusting the final volume to 50 μl with water. The pooled PCR reactions were cleaned up using Agencourt AmpPure XP beads (Beckman Coulter) per manufacturer directions, with the final elution in 52.5 μl of 10 mM Tris-HCl pH 8.5, of which 50 μl was transferred to a new PCR plate. Pooled, cleaned-up PCR products were amplified in a second PCR reaction with Nextera XT index 1 and 2 primers supplied with the Nextera XT 96 Index Kit (either 24 indices or 96 indices). Each PCR reaction contained 12.5 μl of 2X KAPA HiFi HotStart ReadyMix (KAPA Biosystems), 2.5 μl of a Nextera XT index 1 primer, 2.5 μl of a Nextera XT index 2 primer, 5 μl of water and 2.5 μl of pooled, cleaned-up PCR product. The resulting products were cleaned-up again with AmpPure XP beads and resuspended in 27.5 μl of 10 mM Tris-Cl pH 8.5, of which 25 μl was transferred to a new PCR plate. Concentrations of each pooled indexed product was determined using a Qubit dsDNA BR assay kit (Life Technologies, Carlsbad, CA), and PCR products were pooled to a final total concentration of 10 nM of DNA in 10 mM Tris-Cl pH 8.5 containing 0.1% v/v Tween 20. Manufacturer's instructions were followed for subsequent sequencing of libraries on a MiSeq sequencer (Illumina, San Diego, CA).

Analysis of amplicon sequencing data

A custom MATLAB (MathWorks, Inc., Natick, MA) script was used to count the frequencies of wild-type and mutant alleles by searching fastq files for a 25-bp sequence string flanking the mutation site (Supplementary Table S2), and counting the numbers of wild-type and mutant sequences. The resulting counts were used directly to calculate allelic replacement frequencies (ARFs) as (no. of mutant sequences) / (no. of wild-type sequences + no. of mutant sequences). Three biological replicates were analyzed for allele frequencies during selections of MAGE-generated populations, and *P*-values were determined by a two-sided t-test when averaged measurements were compared.

Growth screening

Biological triplicate colonies of selected isolates regrown on LB agar were inoculated into 300 μl M9 medium in 96-well deepwell plates and grown overnight at 37°C with 300 rpm shaking. For primary screening, clear-bottomed white 96-well half-deepwell plates (Enzyscreen B.V., Haarlem, Netherlands) containing 270 μl of either M9 medium or M9 medium + 0.667 M NaCl were inoculated with 30 μl of the overnight cultures diluted 1:10 in M9 medium, to bring to a final concentration of 0.6 M NaCl. Plates were clamped down with sandwich covers (Enzyscreen) and incubated at 37°C with 225 rpm shaking in a Growth Profiler (Enzyscreen), which optically scans the bottom of the

plates. Scans were recorded every 15 min, and Growth Profiler software was used to integrate green pixels in selected 1 mm diameter circular regions in each well. The green pixel values were converted to approximate OD_{600} values using a previously established calibration curve.

For secondary screening, 48-well FlowerPlates (m2plabs GmbH, Baesweiler, Germany) were inoculated to an initial OD_{600} of 0.03 (based on the pathlength for 200 μl of diluted cultures in a 96-well plate, as measured on a Synergy Mx plate reader (BioTek Instruments, Winooski, VT)) into M9 medium or M9 containing an additional 0.6 M final concentration of NaCl. The final volume in each well was 1.4 ml. Cultures were incubated in a BioLector microbioreactor system (m2p-labs) at 37°C with 1000 rpm shaking, with monitoring of the light backscatter intensity. Growth curves were baseline-adjusted by subtracting out the minimum light backscatter intensity in each well from all other values in that well.

Whole genome resequencing of MAGE isolates

Single colonies from each isolate regrown on LB plates were inoculated into LB medium and grown overnight at 37°C with 250 rpm shaking. Cell pellets were frozen at −20°C following centrifugation of 0.5 ml of culture at 16 000 x *g* for 2 min and removal of the supernatant. Genomic DNA was extracted from cell pellets using the PureLink Genomic DNA Mini Kit (Life Technologies) following the manufacturer's instructions, and was further concentrated and purified by ethanol precipitation as previously described (24). Genomic DNA concentrations used to calculate input DNA for library preparation (200 ng in 55 μl volume) were quantified using a Qubit broad-range dsDNA assay (Invitrogen). Libraries were prepared for sequencing using the TruSeq Nano HT Library Preparation Kit (Illumina).

Sequence variants were determined from sequencing fastq output files using breseq (25) version 0.24rc6 and NCBI Reference Sequence NC_000913.3 as the reference sequence, with a base quality cutoff of 20. All called mutations were present in >80% of the spanning reads and in regions with coverage of more than 20-fold, with the exception of a *nagA* A203E mutation which did not meet these criteria but was detected in strain mutS_pMA1_NaCl_20.

RESULTS

Feasibility of *dam* overexpression for MAGE

A requirement for high replacement efficiencies for the majority of single base pair mutations by λ Red-mediated recombination of single-stranded oligonucleotides is inactivation of the MMR system. This is typically achieved by deletion of *mutS* or other genes in the MMR system, however, permanent inactivation can lead to a greatly elevated spontaneous mutation rate both during and after MAGE. Therefore, we reasoned that a gene that generates a mutator phenotype when overexpressed would allow MAGE to be performed with transient inactivation of MMR. The *dam* gene was previously identified as the top hit in a screen for genes that increase mutation rates when overexpressed (17). Therefore *dam* was selected for assessment of whether overexpression could enable high-efficiency allelic replacement through λ Red mediated recombination. A single cycle of MAGE was performed using oligonucleotide **9** (to introduce a premature stop codon in the *galK* gene) and strains EcNR1 (*mutS*$^+$) and EcNR2 (*mutS*$^-$) harboring pMA3, a medium-copy plasmid containing the *dam* gene under control of an arabinose-inducible promoter. Four different *dam* induction procedures were tested: non-induced, induced immediately after electroporation, induced together with the λ Red genes before electrocompetent cell preparation, and induced both before electrocompetent cell preparation and after electroporation. Allelic replacement frequencies (ARFs) were assessed by counting white (positive) and red colonies appearing on MacConkey agar containing 1% galactose (Supplementary Figure S1). As expected, EcNR2 exhibited similar ARFs (6–10%) regardless of *dam* induction. In strain EcNR1, an elevated ARF in strain EcNR2 was achieved upon induction of *dam* together with λ Red, thereby demonstrating the feasibility of *dam* overexpression as an alternative to the use of permanent hypermutator strains(Figure 1A).

Development of a mobile MAGE system

In order to develop a mobile system that could be used for growth at 37°C, the β subunit of λ Red recombinase was cloned into pBAD24 (21) under control of an arabinose-inducible promoter to generate plasmid pMA1 (GenBank accession number KT285943). Next, *dam* was cloned in an artificial operon downstream of β in pMA1 to generate plasmid pMA7 (GenBank accession number KT285941) (Figure 1B). Finally, as it would be desirable to remove any plasmid-based MAGE system following its usage to generate mutations on the genome, the *sacB* gene was cloned into pMA7 to generate plasmid pMA7SacB (GenBank accession number KT285942), thus enabling plasmid curing by sucrose counterselection (Figure 1B).

Optimization of the pMA7 system for MAGE

To determine the optimal induction time for the artificial operon containing *dam* and β, single cycles of MAGE were performed with control oligonucleotide **9** (which introduces a premature stop codon in *galK*, preventing catabolism of galactose) following induction of cultures with 0.2% w/v L-arabinose for time periods between 5 and 30 min. The frequency of replacement was assessed by plating cells on MacConkey agar containing 1% galactose (Figure 2). A maximum number of recombinants were obtained after 10 min induction time, therefore, this induction time was chosen for subsequent recombineering experiments. In addition to the use of the pBAD24 backbone, systems were also constructed by USER cloning to introduce origins of replication from pRSFDuet-1 and pCDFDuet-1 (Novagen) to generate plasmids pMA9 and pMA10, respectively. These systems were tested in K-12 MG1655 using control oligonucleotide **45** to introduce a premature stop codon in the *malK* gene, rendering cells that receive the mutation unable to catabolize maltose. Based on counting red and white colonies on MacConkey agar plates containing 1% maltose, no statistically significant differences in the efficiency of replacement were observed between these systems and pMA7.

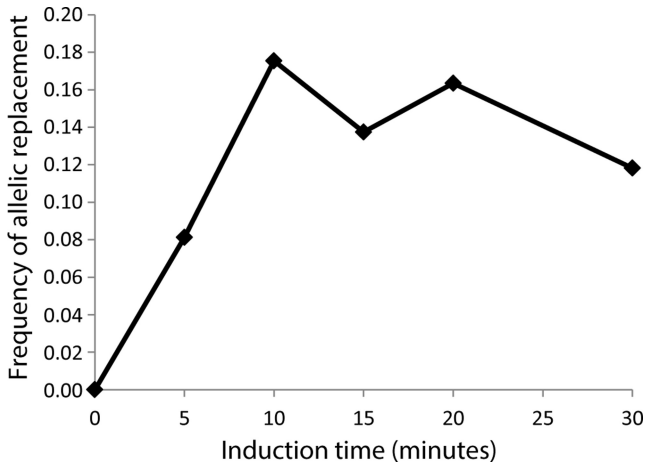

Figure 2. Allelic replacement efficiency (percent recombinants) with various lengths of Dam induction in K-12 MG1655/pMA7 prior to chilling cells for electrocompetent cell preparation. Ten minutes was found to result in optimal allelic replacement frequencies (ARFs) following a single cycle of MAGE with an oligonucleotide introducing a premature stop codon in *galK* (Y145*). ARFs were determined by counting red (wild-type allele) and white (mutant allele) colonies appearing following plating on MacConkey agar containing 1% galactose.

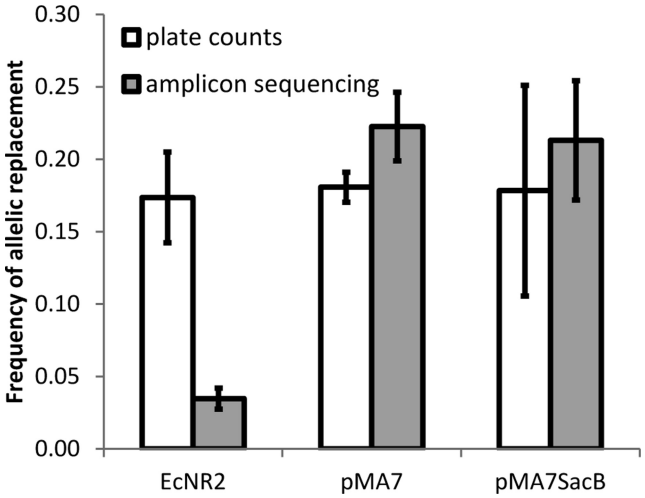

Figure 3. Transient expression of Dam enables efficient recombineering. ARFs for *xylA* Y13* in strains EcNR2, K-12 MG1655/pMA7 and K-12 MG1655/pMA7SacB following one cycle of MAGE with a 90 bp phosphorothioated oligonucleotide. ARFs were determined both by counting red (wild-type allele) and white (mutant allele) colonies appearing following plating on MacConkey agar containing 1% xylose, and by amplicon sequencing within the *xylA* locus. Error bars indicated standard deviations about the mean of three replicates (two for K-12 MG1655/pMA7SacB).

Therefore the pMA7 system was selected for further characterization.

Assessment of system performance

The replacement efficiencies obtained with K-12 MG1655 harboring the pMA7 and pMA7SacB systems (TM-MAGE) were compared to strain EcNR2 using control oligonucleotide **10** (which introduces a premature stop codon in *xylA*, preventing catabolism of D-xylose). This oligonucleotide had a length of 90 bp and incorporated four phosphorothioated bases from the 5′ terminus, as this length and 5′ phosphorothioation was determined to yield optimal ARFs when introducing a nonsense mutation in *lacZ* (5). Electroporated cells were outgrown overnight, and ARFs were determined both by plating on MacConkey agar containing 1% w/v D-xylose and by amplicon sequencing. The use of both pMA7 and pMA7SacB resulted in approximately 20% ARF of *xylA* with the *xylA* Y13* allele, as measured both by plate counts and amplicon sequencing (Figure 3). The use of EcNR2 resulted in a similar 17% ARF as measured by plate counts, but interestingly only a 3.5% ARF as measured by amplicon sequencing when searching for an exact sequence match to a 25 bp region centered on the mutant or wild-type base. This percentage increased to only 5.8 ± 0.7% when first searching for sequences containing the forward priming sequence for *xylA* amplicon sequencing (**43** in Supplementary Table S1) and then searching only for sequences containing the mutant or wild-type base and three preceding bases (TTCC or TTCA, respectively). Qualitative inspection of all sequences containing the forward priming sequence indicated an abundance of additional off-target mutations within the region complementary to the oligonucleotide. These mutations were likely responsible for alternative inactivations of *xylA* and could thus account for numerous false positives on plates. Far fewer off-target mutations were observed with the pMA7 and pMA7SacB systems, as evidenced by comparable ARFs determined by both plate counting and amplicon sequencing.

In a separate experiment, ARFs were compared over 6 MAGE cycles between K-12 MG1655/pMA7 (TM-MAGE), K-12 MG1655/pMA1 and K-12 MG1655 *mutS::cat*/pMA1 separately using oligonucleotides **8** and **10** (which introduce premature stop codons in *galK* and *xylA*, preventing catabolism of D-galactose and D-xylose, respectively) (Figure 4). These oligonucleotides have lengths of 70–72 bp and do not contain any phosphorothioated bases. Amplicon sequencing was performed using cells collected during exponential growth following each cycle or after overnight outgrowth (for the third and sixth cycles). Increasing ARFs with a saturating trend were observed using both control oligos, in general agreement with predictions of replacement frequency as a function of cycle number (5). The pMA7 system yielded similar but slightly reduced ARFs for most cycles compared to the hyper-mutator strain (*mutS*$^{-}$) harboring the pMA1 system. The pMA1 system had a maximum ARF of only 0.3% after 6 cycles, as expected due to active mismatch repair. Counting of white and red colonies on MacConkey agar plates following the sixth cycle yielded nearly identical ARF values to those measured by amplicon sequencing and counting exact matches to a 25 bp region centered on the wild-type or mutant base.

Curing of pMA7SacB

After 6 cycles of MAGE using either K-12 MG1655/pMA7SacB or pMA1, white colonies (*xyl*$^{-}$) growing on MacConkey agar plus 1% D-xylose were cured as described in methods. Colonies were obtained overnight

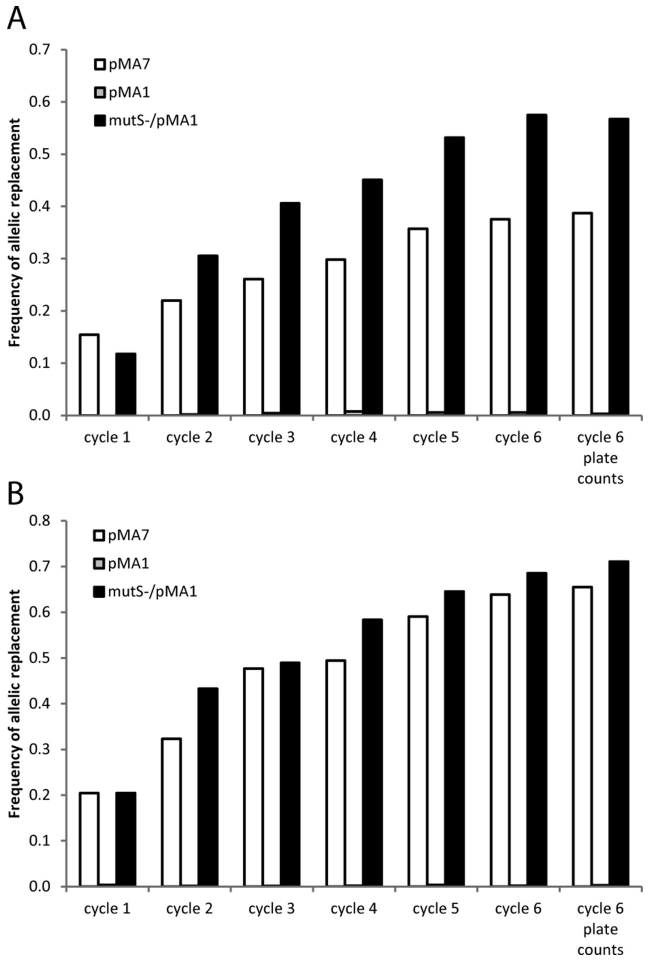

Figure 4. ARF from TM-MAGE with pMA7 are nearly equivalent to those obtained when mismatch repair is permanently disabled. ARFs for *galK* Y145* (top) and *xylA* Y13* (bottom) using 70-bp non-modified oligonucleotides were determined by amplicon sequencing using cell pellets obtained following each of 6 cycles of MAGE in strains K-12 MG1655/pMA1, K-12 MG1655/pMA7 and K-12 MG1655 *mutS::cat*/pMA1. ARFs were also determined from counting red (wild-type allele) and white (mutant allele) colonies appearing following plating after 6 cycles of MAGE on MacConkey agar containing galactose or xylose. Measurements were performed on single replicates.

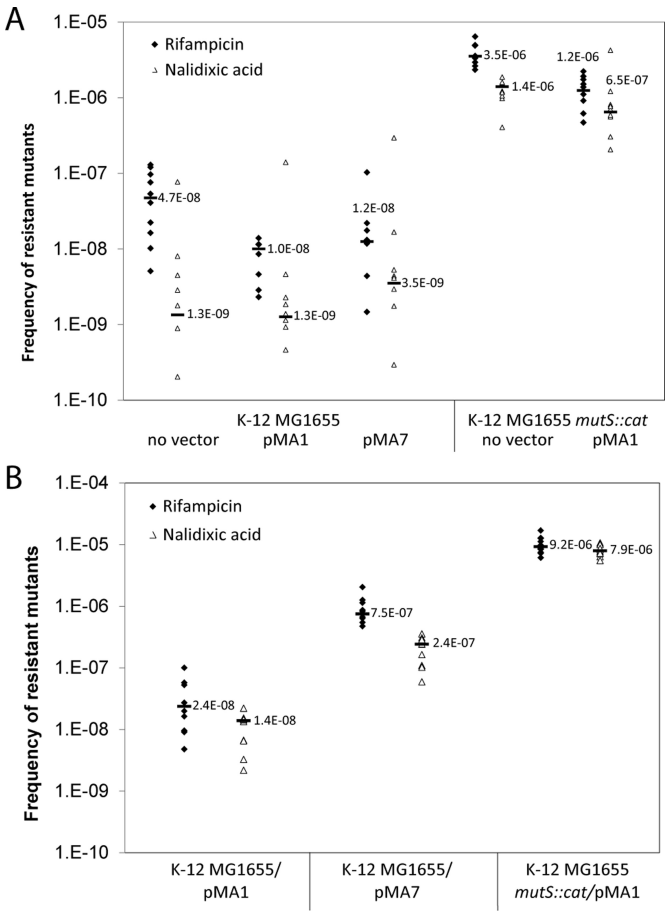

Figure 5. Spontaneous mutation rates are elevated only when Dam overexpression is induced, and are reduced compared to permanently disabled MMR. (**A**) Frequencies of rifampicin and nalidixic acid resistant mutants following plating of non-induced strains (K-12 MG1655 with and without pMA1, K-12 MG1655 *mutS::cat* with and without pMA1 and K-12 MG1655 with pMA7). (**B**) Frequencies of rifampicin and nalidixic acid resistant mutants arising following plating of induced strains that had undergone a procedure mimicking a single MAGE cycle. Individual replicate measurements are shown, with the median value denoted by a horizontal black line.

on sucrose and removal of pMA7SacB was verified by lack of growth on LB plates containing ampicillin (Supplementary Figure S2). Additionally, retainment of the *xylA* Y13* allele following curing was qualitatively verified by plating on MacConkey agar plus 1% D-xylose. Similar results with overnight curing of the plasmid were obtained with subsequent usage of pMA7SacB in our laboratory for the introduction of other point mutations, with routine removal of the plasmid by a single streaking and overnight growth on sucrose-containing plates.

Determination of mutation frequencies resulting from *dam* overexpression

While *dam* overexpression generates comparable ARFs as when *mutS*$^{-}$ background hosts are employed, the primary goal of finding an alternative inducible system that disables mismatch repair is to reduce the background mutation rate compared to that found in *mutS*$^{-}$ hosts. Therefore, fluctuation tests were performed to measure the relative mutation rates in K-12 MG1655 harboring pMA1, pMA7 and K-12 MG1655 *mutS::cat* harboring pMA1, both with and without induction of β and *dam* expression by addition of L-arabinose. Strains without plasmids were also tested as controls without induction. In non-induced tests, K-12 MG1655 harboring no vector, pMA1 and pMA7 exhibited similar median resistant mutation frequencies of between 1.0 to 4.7 $\times$ 10^{-8} for rifampicin, and 1.3 to 3.5 $\times$ 10^{-9} for nalidixic acid (Figure 5A). K-12 MG1655 *mutS::cat* harboring no vector and pMA1 exhibited mutation frequencies that were approximately 2 orders of magnitude higher for rifampicin and between 2 to 3 orders of magnitude higher for nalidixic acid. Thus, it is apparent that pMA7 does not increase mutation rates without or prior to arabinose induction, indicating effective repression of *dam* overexpression in the absence of inducing agent. With arabinose induction, median frequencies of resistant mutations for both ri-

fampicin and nalidixic acid remained similar (within about 1 order of magnitude) to those without induction for K-12 MG1655/pMA1 and K-12 MG1655 *mutS::cat*/pMA1, while those for K-12 MG1655/pMA7 increased by between 1 and 2 orders of magnitude, demonstrating an inducible mutator phenotype while maintaining a 12- to 33-fold reduced rate of mutation following induction as compared with a *mutS*$^{-}$ strain. This reduced mutation frequency is likely at least in part due to the transient nature of *dam* overexpression relative to the permanently disabled mismatch repair present in *mutS*$^{-}$ strains.

System performance with subsequent selections

One of the primary applications of MAGE is the generation of combinatorial libraries of a set of mutant alleles. In order to identify beneficial phenotypes from the resulting complex pool, performing subsequent high-throughput screening or selections is necessary. In the case when selections are performed, it was anticipated that low ARFs (e.g. the pMA1 system in a *mutS*$^{+}$ background) will lead to insufficient diversity in the starting pool of mutants and will hamper efforts to isolate strains with highly improved phenotypes derived from combinatorial mutations. It was also anticipated that in strains with permanently disabled MMR (e.g. *mutS*$^{-}$), accumulation of off-target mutations during the selection process would render phenotype to genotype relationships difficult to establish. TM-MAGE, which enables both high efficiency allelic replacement to achieve both a high diversity of combinatorial mutants and longer term genomic stability following MAGE cycling, was expected to lead to the isolation of strains with more improved phenotypes and with an improved ability to assign phenotype to genotype.

This process was tested by constructing combinatorial mutant libraries in strains K-12 MG1655/pMA7 (TM-MAGE), K-12 MG1655/pMA1 (no mismatch repair) and K-12 MG1655 *mutS::cat*/pMA1 (permanently disabled mismatch repair). A pool of 7 oligonucleotides (**1** through **7** in Supplementary Table S1) was transformed to introduce mutations that were previously identified in strains evolved for growth in high concentrations of NaCl (18). These oligonucleotides were designed to either directly introduce selected single nucleotide polymorphisms (SNPs) from the G6 and A3 strains (18), or to introduce premature stop codons at the nearest possible location to where insertion element sequences were located (as these mutations represent probable losses-of-function). The pooled oligonucleotides were transformed over 6 cycles of MAGE and ARFs were determined using amplicon sequencing of harvested cells after MAGE cycling. These frequencies are shown as the initial data points in Figure 6 for each mutant allele. Sixth cycle ARFs varied for different oligonucleotides and for the different systems, with K-12 MG1655 *mutS::cat*/pMA1 generally exhibiting the highest ARFs, TM-MAGE using K-12 MG1655/pMA7 exhibiting similar but slightly lower ARFs and K-12 MG1655/pMA1 exhibiting greatly reduced ARFs. The *mrdA* Q51L (T:A) and *proV* K12* (A:T) mutations, and to a lesser extent the *nagA* A203E (G:T) mutation, appeared to largely evade mismatch repair in K-12 MG1655/pMA1 by unknown mechanisms.

Following 6 MAGE cycles and overnight outgrowth, cells were plated on M9 agar with or without 0.6 M NaCl, and were also inoculated into liquid cultures of M9 medium with or without 0.5 M NaCl (this concentration was increased to 0.6 M NaCl with subsequent passages). Daily passaging was performed with the liquid cultures (Supplementary Figure S3), and after five passages and regrowth, cells were plated on M9 agar containing 0.6 M NaCl to select individual isolates. Additionally, after days 1, 3 and 5 of liquid culture passaging, cells were harvested for determination of the mutant allele frequency dynamics during selection using amplicon sequencing (Figure 6). Enrichment of the frequency of a mutation as a function of selection time in high NaCl concentrations is indicative of causality for salt tolerance, whether alone or in combination with other mutations. Only the *mrdA* Q51L, *msrB* C118F, *nagA* A203E and *rpsA* Q421K mutations were uniformly enriched during passaging in M9 + 0.6 M NaCl following MAGE in the *mutS*$^{-}$/pMA1 and pMA7 harboring strains. With the exception of the *nagA* A203E mutation in the population generated with pMA7, these trends were specific to selections in high salt and mutations were not observed to be enriched by passaging in M9 only. The *proV* K12* mutant was additionally enriched by passaging in high NaCl concentrations in *mutS*$^{-}$/pMA1 ($P = 0.002$) but not significantly for the pMA7-generated population. Following performing MAGE in K-12 MG1655/pMA1, no enrichment pattern could be discerned in high salt conditions, with the exception of *nagA* A203E. For many mutations, this was likely due to their extremely low starting frequencies as a result of active MMR during MAGE cycling. The *mrdA* Q51L mutation, which was present at a high starting frequency, exhibited a lack of enrichment in high salt. Because this mutation was enriched over 3–5 days using the other MAGE systems, a likely explanation is that this mutation is only beneficial in combination with one or more other mutations.

In summary, enrichment of particular introduced mutations following selection in high salt concentrations could only be successfully observed when high ARFs could be achieved by MAGE (i.e. with either K-12 MG1655 *mutS*$^{-}$/pMA1 or K-12 MG1655/pMA7). This is most likely due to the presence of strains in the starting population that possess higher numbers of combinatorial mutations (e.g. 3–5) that together impart tolerance to high salt concentrations. From our amplicon sequencing data of the selected populations, four causal mutations for NaCl tolerance out of the original seven introduced mutations could be putatively identified using TM-MAGE: *mrdA* Q51L, *msrB* C118F, *nagA* A203E and *rpsA* Q421K. The *proV* K12* mutation was only observed to be enriched in the *mutS*$^{-}$/pMA1 background strain, thus it was not possible to assign causality.

Screening and re-sequencing of selected isolates

For each of the three MAGE systems, a total of 20 isolates from M9 + NaCl plates (10 each from plates spread with cells immediately after MAGE cycling, and at the end of serial passaging) were screened for growth in liquid culture (M9 + 0.6 M NaCl) (Supplementary Figure S4). Primary screening results were used to select two or three isolates

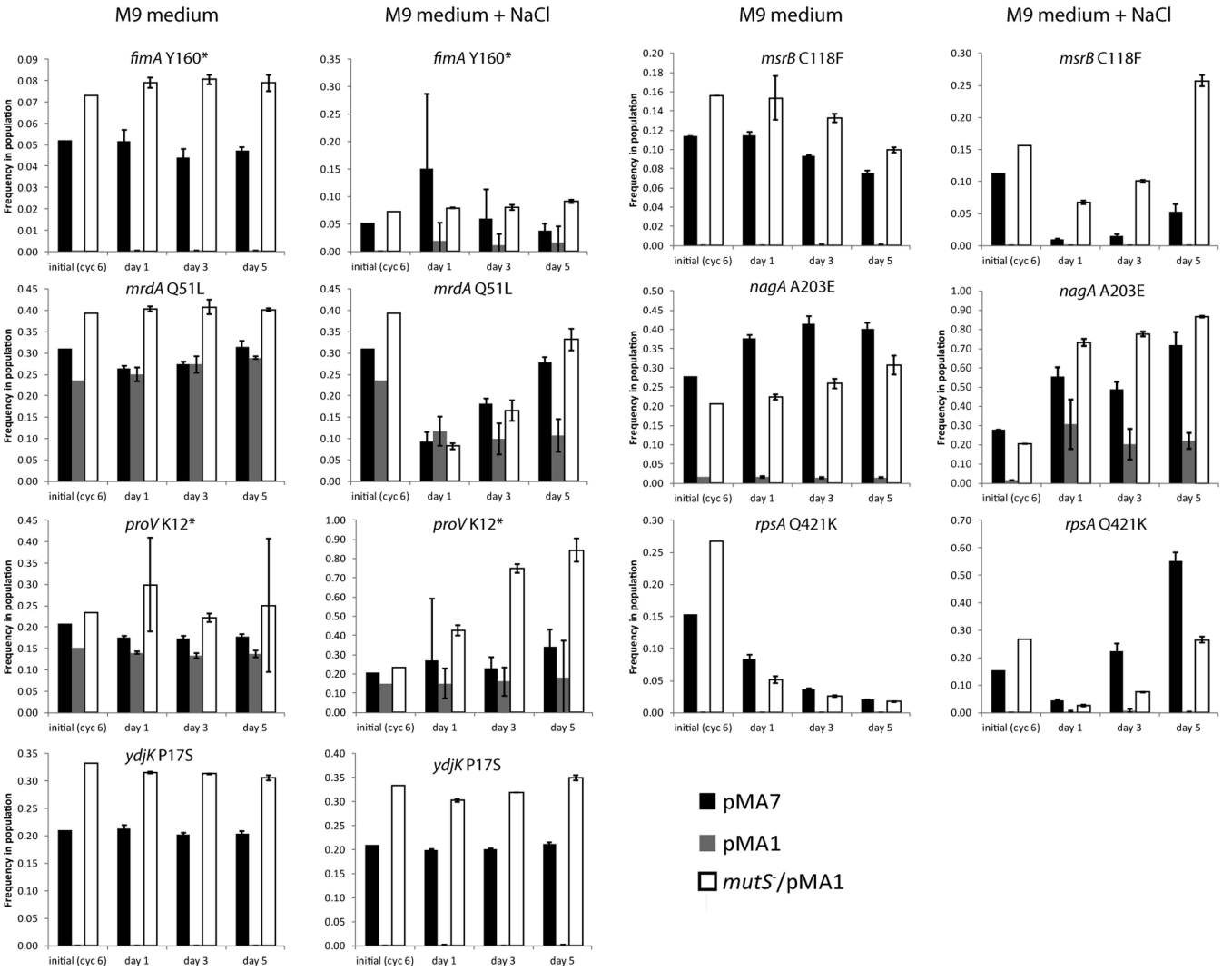

Figure 6. Population frequencies of mutant alleles (indicated at right) as determined by amplicon sequencing following 6 cycles of MAGE and after 1, 3 and 5 days of daily passaging in M9 medium or M9 medium plus NaCl (0.5 M for the first passage and 0.6 M for subsequent passages). Libraries were generated in K-12 MG1655/pMA7 (black bars), K-12 MG1655/pMA1 (gray bars) and K-12 MG1655 *mutS::cat*/pMA1 (white bars). Error bars indicate standard deviations about the mean values of three biological replicate cultures. Only a single replicate was sequenced following 6 cycles of MAGE, therefore, no error bars are shown.

per system, which underwent secondary screening in a Biolector microbioreactor system (Figure 7). Additionally for each system, three isolates from M9 plates that were spread with cells immediately after MAGE cycling and one *xyl*$^{-}$ isolate were screened under the same conditions.

A total of 20 isolates were whole-genome resequenced with the goal of assessing typical numbers of off-target mutations arising during multiple cycles of MAGE and subsequent selections, as well as validating the causative mutations inferred by amplicon sequencing (Table 2; full lists of called sequence variants are provided in Supplementary Data File 1). Although the mutation rate was shown to be elevated over 10-fold in K-12 MG1655/pMA7 following a typical single cycle of MAGE and plating on highly selective conditions compared to K-12 MG1655/pMA1, which harbors fully active mismatch repair (Figure 5), the duration of this period of elevated mutation rate following induction is not known largely due to the unknown turnover rate of the overexpressed Dam protein. Additional cycling would be expected to extend the duration of the elevated mutation rate as a result of Dam overexpression, however, that increase would only be expected to be linear with a high Dam turnover rate. After 6 cycles of MAGE and plating on MacConkey agar, *xyl*$^{-}$ isolates generated using both the pMA7 and pMA1 system in the *mutS*$^{-}$ background strain contained the target *xylA* Y13* mutation. The isolate generated by the pMA1 system contained two off-target mutations immediately adjacent to the target site (a 4-bp deletion beginning at the same genome coordinate as the Y13* mutation and a G→T mutation 5 bases away from that coordinate) that would also be expected to result in a loss-of-function of *xylA*. Aside from the off-target mutation in *xylA* using the pMA1 system, there was a single off-target mutation using pMA7, none using pMA1 and 16 using pMA1 in the *mutS*$^{-}$ strain (excluding mutations that were common

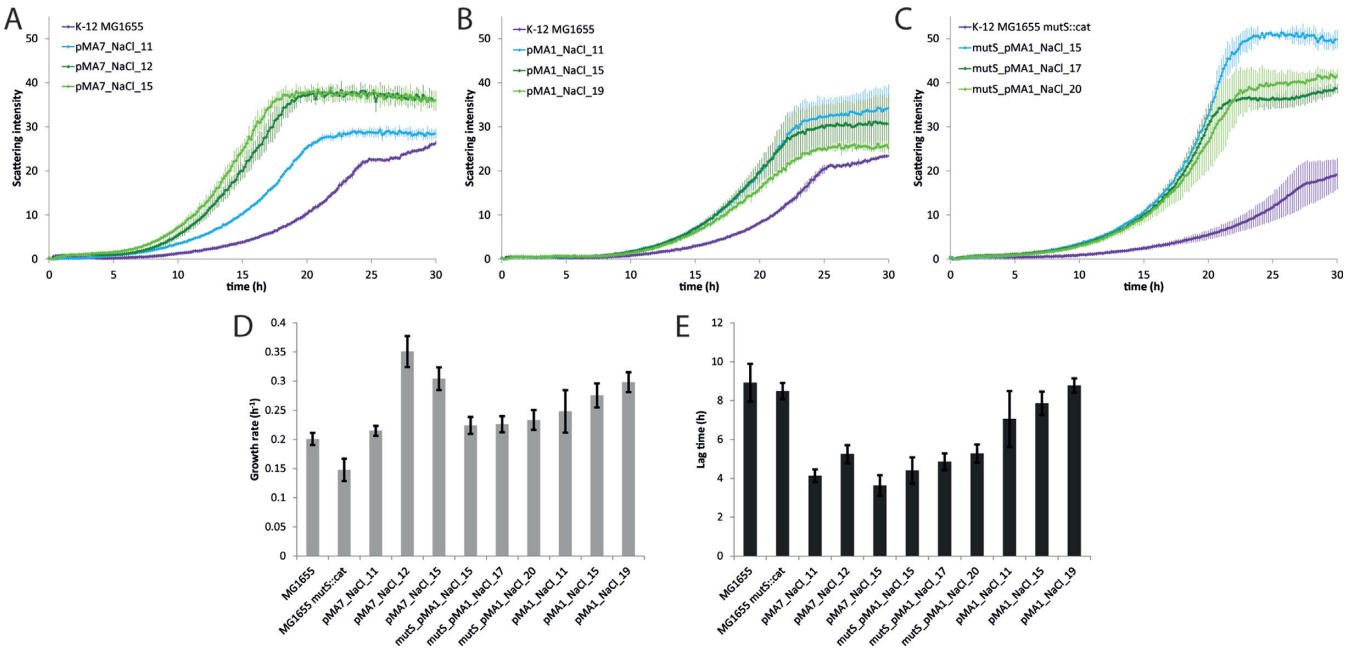

Figure 7. Strains with the highest growth rates and greatly reduced lag times were reliably isolated from NaCl-selected libraries generated using TM-MAGE, as compared to other MAGE systems. (**A**–**C**) Averaged growth curves across three biological replicate cultures of control strains and individual isolates (**A**: K-12 MG1655/pMA7, **B**: K-12 MG1655/pMA1, **C**: K-12 MG1655 *mutS::cat*/pMA1) following 5 days of passaging on M9 + NaCl. (**D**) Calculated growth rates and (**E**) lag times of starting strains and isolated strains following M9 + NaCl selections. Error bars represent standard deviations about the mean of three biological replicate measurements.

across all sequenced *mutS*$^{-}$ strains that can be assumed to be present in the original background strain).

Following selection on M9 plates or after liquid passaging in M9 + NaCl (Table 2), K-12 MG1655/pMA7 isolates harbored between 1 and 5 off-target mutations (in isolate pMA7-NaCl-15 with 5 off-target mutations, one was within the sequence of oligonucleotide **7**). By contrast, K-12 MG1655/pMA1 isolates harbored between 0 to 1 off-target mutations and strains using pMA1 in the *mutS*$^{-}$ background harbored 7–25 mutations (7–14 from selection in M9, and 9–25 from selection in M9 + NaCl). Thus, while the mutation rate using the pMA7 system is generally elevated compared to pMA1 in a *mutS*$^{+}$ background, there are fewer off-target mutations than in a *mutS*$^{-}$ background, and off-target mutations were not particularly more abundant following selection in the condition with a higher selective pressure (liquid passages followed by plating on M9 agar + NaCl) versus the condition with the lower selective pressure (plating on M9 agar). This is suggestive that most off-target mutations arose during MAGE cycling when Dam was periodically induced, rather than during subsequent selections.

Further investigating on-target mutations in the NaCl-selected mutants, strains isolated from the pMA7 population harbored between 3 and 5 on-target mutations, while those from the pMA1 population harbored 2 and those from the *mutS*$^{-}$/pMA1 population harbored 3 on-target mutations (Table 3). The strains isolated from the pMA1 population both contained the *mrdA* Q51L and *nagA* A203E mutations, which along with *proV* K12*, were the only mutations with elevated ARFs using this system, likely due to full or partial evasion of the mismatch repair system. These mutants exhibited only slightly improved growth in M9 + 0.6 M NaCl compared to the wild-type strain (Figure 7). By contrast, two out of three strains isolated from the pMA7 population contained the *mrdA* Q51L, *nagA* A203E, *rpsA* Q421K and *msrB* C118F mutations. The presence of these mutations is consistent with enrichments observed during passaging in M9 + NaCl by amplicon sequencing (Figure 6). Strains pMA7_NaCl_12 and pMA7_NaCl_15 exhibited a much greater improvement in growth in M9 + 0.6 M NaCl compared to the pMA1 isolates (Figure 7), with growth rates of 0.35 ± 0.03 and 0.30 ± 0.02 h^{-1}, respectively, compared to the K-12 MG1655 growth rate of 0.20 ± 0.01 h^{-1}. Lag times were also reduced in these isolates to 5.2 ± 0.5 and 3.6 ± 0.5 h, respectively, compared to 8.9 ± 1.0 h for K-12 MG1655. These strains in particular share the combination of all four of the above mentioned mutations. Strains isolated from the *mutS*$^{-}$ population harbored 3 mutations (all three contained *mrdA* Q51L and *rpsA* Q421K, with different third on-target mutations in each isolate) and also significantly improved growth phenotypes compared to K-12 MG1655 *mutS::cat* (0.22 to 0.23 h^{-1} compared with 0.15 h^{-1}). However, the presence of a large number of off-target mutations rendered it difficult to assign causation of the phenotype to the sets of on-target mutations that were identified.

DISCUSSION

The ability to use short (<100-bp), chemically synthesized, single-stranded oligonucleotides to precisely modify and edit genomes has opened up possibilities in a wide array of applications. To date, these have included the generation of combinatorial libraries targeting ribosome binding sites

Table 2. On-target and off-target mutations in re-sequenced isolates (20 total) following 6 cycles of MAGE to introduce the *xylA* Y13* mutation, in isolates following plating on M9 agar after MAGE cycling with a pool of 7 oligonucleotides designed to introduce mutations observed in evolved NaCl-tolerant strains, and in isolates cycled with the same oligonucleotide pool after 5 days of serially passaging cultures in M9 + NaCl and plating on M9 agar + NaCl

	6th cycle xylA*		M9		M9 + NaCl	
	on-target	off-target	on-target	off-target[a]	on-target	off-target[a]
pMA7 rep. 1	1	1	1	1	3	1
pMA7 rep. 2	-	-	3	4	4	4 + 1[b]
pMA7 rep. 3	-	-	0	2	5	4
pMA1 rep. 1	0	2[b]	1	0	-	-
pMA1 rep. 2	-	-	1	0	2	1
pMA1 rep. 3	-	-	0	0	2	1
mutS$^-$/pMA1 rep. 1	1	16	4	14	3	25[c]
mutS$^-$/pMA1 rep. 2	-	-	0	7	3	9
mutS$^-$/pMA1 rep. 3	-	-	2	10	3	15

[a]Not including pre-existing mutations occurring across all *mutS*- strains.
[b]Off-target mutations were within the oligonucleotide sequence.
[c]Not including a number of called missing coverage deletions (total of 43).

Table 3. On-target mutations identified in isolates following plating on M9 agar after MAGE cycling, and in isolates after 5 days of serially passaging cultures in M9 + NaCl and plating on M9 agar + NaCl

MAGE system	Isolate name	M9 on-target mutations	Isolate name	M9 + NaCl on-target mutations
K-12 MG1655/ pMA7	pMA7_M9_1	*mrdA* Q51L	pMA7_NaCl_11	*nagA* A203E, *rpsA* Q421K, *msrB* C118F
	pMA7_M9_2	*nagA* A203E, *ydjK* P17S, *proV* K12*	pMA7_NaCl_12	*mrdA* Q51L, *nagA* A203E, *rpsA* Q421K, *msrB* C118F
	pMA7_M9_3	none	pMA7_NaCl_15	*mrdA* Q51L, *nagA* A203E, *rpsA* Q421K, *msrB* C118F, *ydjK* P17S
K-12 MG1655/ pMA1	pMA1_M9_1	*mrdA* Q51L	pMA1_NaCl_11	-
	pMA1_M9_2	*proV* K12*	pMA1_NaCl_15	*mrdA* Q51L, *nagA* A203E
	pMA1_M9_3	none	pMA1_NaCl_19	*mrdA* Q51L, *nagA* A203E
K-12 MG1655 *mutS::cat*/ pMA1	mutS_pMA1_M9_1	*nagA* A203E, *ydjK* P17S, *msrB* C118F, *proV* K12*	mutS_pMA1_NaCl_15	*mrdA* Q51L, *rpsA* Q421K, *ydjK* P17S
	mutS_pMA1_M9_2	none	mutS_pMA1_NaCl_17	*mrdA* Q51L, *nagA* A203E, *rpsA* Q421K
	mutS_pMA1_M9_3	*msrB* C118F, *proV* K12*	mutS_pMA1_NaCl_20	*mrdA* Q51L, *rpsA* Q421K, *msrB* C118F

(RBS) of genes involved in biochemical production pathways (5), redesign of the entire genome via genome-wide stop codon replacement and replacement of rare codons (9), genome-wide promoter replacement (6), and combinatorial library generation followed by selection to identify causative combinations of mutations found in adaptively evolved strains (10). In all published applications besides the last example, which had utilized the pMA7 system in advance of this publication, strains deficient in mismatch repair (*mutS*$^-$) have been employed and the impact of off-target mutations on the phenotype of the final selected or screened strains has largely not been reported. An average of 113 off-target mutations were identified in strains with UAG stop codons converted to UAA (26), and 355 off-target mutations were reported in a strain with all UAG stop codons converted to UAA (8). These off-target mutations are likely to affect strain phenotype, particularly when strains are subjected to selective pressures to identify isolates with improved phenotypes. As an example of this, we have generated MAGE libraries containing random RBS insertions upstream of 70 transcription factors in a *mutS*$^-$ host and have conducted short-term selections (3 passages) in stress conditions to identify stress-tolerant mutants. Re-sequenced isolates from these selections exhibited greatly improved tolerance, however, only 4 out of 11 strains contained targeted RBS modifications, and of those that did, it was not possible to reconstitute tolerance by re-introducing only the identified RBS modifications into the background strains (Supplementary Figure S5).

In this study, a new plasmid-based system for performing MAGE, TM-MAGE, is reported that transiently inhibits methyl-directed mismatch repair through inducible overexpression of the Dam methylase while simultaneously expressing the β subunit of λ Red recombinase. Dam overexpression has long been known to result in an increased mutation rate due to increased methylation of both the template strand and newly synthesized strand during replication, due to hemimethylation being required for recognition of the newly synthesized strand by the MMR system (28). The system was shown to have a nearly equivalent frequency of allelic replacement as *mutS*$^-$ strains (either EcNR2 or K-12 MG1655 *mutS::cat*). Furthermore, while inducing a mutator phenotype, a much lower occurrence of off-target mutations was observed compared with permanent inactivation of MMR, with an over 5-fold reduction in the average number of off-target mutations in re-sequenced isolates following 6 cycles of MAGE and 5 days of selections under osmotic stress conditions compared to *mutS*$^-$ mutants. This is likely a result of transcription of plasmid-based *dam* only being induced during MAGE cycling, with natural protein turnover restoring ordinary DNA hemimethylation in subsequent generations. Nyerges *et al.* have also addressed the issue of off-target mutations in MMR-deficient hosts during MAGE through the introduction of temperature-sensitive MutS and MutL protein alleles into a host strain, resulting in an over 6-fold reduction in off-target mutations following 20 cycles of MAGE (15). However, utilization of this system necessitates that the MAGE background strain first possess the two temperature-sensitive alleles in addition to a system for expressing λ Red recombinase (in this case a prophage-based recombinase was employed). Accordingly, mutants constructed using this system must be cultivated at temperatures less than 32°C to ensure active mismatch repair unless the non-temperature-sensitive alleles are re-

stored. TM-MAGE addresses these limitations, as necessary functions are self-contained on a plasmid that can easily be transformed into any host strain, and can either be left in the host strain without cultivation temperature restrictions, or easily removed in a single day by sucrose counterselection when pMA7SacB is employed. These features enable robust generation of combinatorial libraries and precisely edited genomes with the ability to much more definitively assign phenotype to genotype due to the significantly reduced rate of accumulation of off-target mutations.

In the previous publication in which TM-MAGE (pMA7) was utilized (10), a detailed analysis of system performance with respect to ARFs and off-target mutation rates had not been performed. A total of nine cycles of MAGE were performed over three days (in comparison to the 6 cycles in 2 days performed in this study, thus lengthening the total time in which Dam was overexpressed), followed by serial passaging under the selection condition (42°C) during continuous exponential growth for 3 days. In 43 resequenced isolates (10), a median of 4 off-target mutations and 3 on-target mutations were observed. After removing off-target mutations that occurred within the oligonucleotide sequences, mutations that occurred spontaneously in the *pyrE*/*rph* intergenic region or *rph* gene, and larger detected areas of missing coverage in a few strains, the median number of off-target mutations dropped to two. These mutations were not considered due to possible oligonucleotide synthesis errors, mutations in *pyrE*/*rph* correcting a known defect in K-12 MG1655 growth in minimal medium (27) and appearing to be especially selected for at 42°C, and the likely presence of sequencing artifacts in the strains with many called missing coverage deletions (and if present, would unlikely be a result of disabled MMR), respectively. These results are comparable with the off-target mutation rates observed in this study and offer further support of the greatly reduced background mutation rate achievable by TM-MAGE relative to strains with permanently disabled MMR.

It should be noted that a potential limitation of TM-MAGE is the continued accumulation of off-target mutations during MAGE cycling. Similar numbers of off-target mutations were observed in re-sequenced isolates that had undergone 6 (this study) or 9 cycles of TM-MAGE (10). By constrast, a single cycle of MAGE was shown to generate an approximately one order-of-magnitude increase in the frequency of mutations providing resistance toward nalidixic acid or rifampicin (Figure 5B). This is suggestive of a nonlinear accumulation of off-target mutations as a function of MAGE cycle number, and is likely a result of elevated Dam protein levels persisting during and after cycling in at least a portion of the passaged population in subsequent cycles. Despite that significant increases in off-target mutations were not observed after 9 cycles, it is recommended that TM-MAGE cycles be limited in number in applications where off-target mutations would be particularly problematic. Experiments should be appropriately scaled with regards to both the pooled oligonucleotide diversity and the ability to screen or select desired phenotypes from the resulting populations. In the case of performing selections following any MAGE protocol, consideration must be given to the ability to enrich a particular mutant in a timeframe shorter than that required for cells to undergo adaptive laboratory evolution. Despite these potential limitations, the off-target mutation rate using TM-MAGE is significantly lower than when performing MAGE in strains with permanently disabled MMR and is thus better suited for experiments requiring larger numbers of cycles.

We demonstrated the power of using TM-MAGE to reconstruct causative mutations for tolerance to growth in high NaCl concentrations. These mutations were previously identified in the two best-performing isolates from adaptive laboratory evolutions of F^- and F^+ Hfr strains on high concentrations of NaCl (18). In that study, causative mutations were only partially explored through reconstruction of combinations of loss-of-function mutations, and through overexpression of the different mutated gene targets which only identified *ydjK* as improving osmotolerance. We here show that an improved combination of mutations with a relative fitness (growth rate of mutant / growth rate of wild-type − 1) in M9 + 0.6 M NaCl of 0.75 could be achieved using TM-MAGE, compared to reported fitnesses of 0.5 for the original evolved strains in M9 + 0.65 M NaCl and less than 0.5 for all overexpression and knockout mutants in M9 + 0.55 M NaCl (18). These TM-MAGE constructed strains harbored a combination of four mutations originating from both of the original evolved strains: *mrdA* Q51L, *nagA* A203E, *rpsA* Q421K and *msrB* C118F. Due to their extremely low initial frequency in the population, it was not possible to adequately select for mutants with this degree of improvement in growth phenotype from populations generated using pMA1 (which harbored an intact MMR system). And due to the high rate of off-target mutations likely occurring both before, during, and after MAGE in a strain with permanently disabled mismatch repair, it was also not possible to isolate mutants with significantly improved growth rates over wild-type K-12 MG1655, nor to assign causation from patterns in the growth behaviors of individual isolates. However, performing TM-MAGE with the pMA7 system, we were able to rapidly isolate improved osmotolerant mutants with combinations of mutations originally found in distinct evolved isolates.

TM-MAGE should find broad applicability as a replacement for other MAGE systems. New features that have previously hindered the use of MAGE include the ability to efficiently generate mutations and libraries of mutations in alternative host strains, and the rapid removal of the system for use of MAGE in metabolically engineered production hosts or any other application where a clean strain background is desired. Most importantly, we have demonstrated the ability to couple TM-MAGE robustly with subsequent selections, which will enable the full exploration of the generated mutational space for desired selectable phenotypes. This is particularly useful for reconstructing and identifying causal mutations in strains isolated by adaptive laboratory evolution. Use of TM-MAGE will also lead to the generation of more stable genotypes for subsequent screening applications, when compared to systems with permanently disabled MMR.

ACCESSION NUMBERS

Sequences for plasmids pMA1, pMA7, and pMA7SacB have been deposited as Genbank accession numbers KT285943, KT285941 and KT285942, respectively.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

The authors would like to acknowledge Anna Koza, Ida Bonde, and João Cardoso for their assistance with next-generation sequencing and bioinformatics analysis. M.O.A.S acknowledges funding from the European Research Council (grant: LimitMDR) and the Danish Free Research Councils.

FUNDING

The Novo Nordisk Foundation and the European Union Seventh Framework Programme (FP7-KBBE-2013-7-single-stage) [grant agreement no. 613745, PROMYS]; European Research Council [LimitMDR to M.O.A.S]; Danish Free Research Councils. Funding for open access charge: The Novo Nordisk Foundation.
Conflict of interest statement. None declared.

REFERENCES

1. Thomason,L., Court,D.L., Bubunenko,M., Constantino,N., Wilson,H., Datta,S. and Oppenheim,A. (2007) Recombineering: genetic engineering in bacteria using homologous recombination. *Curr. Protoc. Mol. Biol*. Unit1.16.
2. Datsenko,K.A. and Wanner,B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* using PCR products. *Proc. Natl. Acad. Sci U.S.A.*, **97**, 6640–6645.
3. 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 U.S.A.*, **98**, 6742–6746.
4. Sawitzke,J.A., Thomason,L.C., Bubunenko,M., Li,X., Constantino,N. and Court,D.L. (2013) Recombineering: highly efficient *in vivo* genetic engineering using single-strand oligos. *Methods Enzymol.*, **533**, 157–177.
5. Wang,H.H., Isaacs,F.J., Carr,P.A., Sun,Z.Z., Xu,G., Forest,C.R. and Church,G.M. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, **460**, 894–898.
6. Bonde,M.T., Kosuri,S., Genee,H.J., Sarup-Lytzen,K., Church,G.M., Sommer,M.O. and Wang,H.H. (2015) Direct mutagenesis of thousands of genomic targets using microarray-derived oligonucleotides. *ACS Synth. Biol.*, **4**, 17–22.
7. Ng,C.Y., Farasat,I., Maranas,C.D. and Salis,H.M. (2015) Rational design of a synthetic Entner-Doudoroff pathway for improved and controllable NADPH regeneration. *Metab. Eng.*, **29**, 86–96.
8. LaJoie,M.J., Rovner,A.J., Goodman,D.B., Aerni,H.R., Haimovich,A.D., Kuznetsov,G., Mercer,J.A., Wang,H.H., Carr,P.A., Mosberg,J.A. *et al.* (2013) Genomically recoded organisms expand biological functions. *Science*, **342**, 357–360.
9. LaJoie,M.J., Kosuri,S., Mosberg,J.A., Gregg,C.J., Zhang,D. and Church,G.M. (2013) Probing the limits of genetic recoding in essential genes. *Science*, **342**, 361–363.
10. Sandberg,T.E., Pedersen,M., LaCroix,R.A., Ebrahim,A., Bonde,M., Herrgard,M.J., Palsson,B.O., Sommer,M. and Feist,A.M. (2014) Evolution of *Escherichia coli* to 42°C and subsequent genetic engineering reveals adaptive mechanisms and novel mutations. *Mol. Biol. Evol.*, **31**, 2647–2662.
11. Munck,C., Gumpert,H.K., Wallin,A.I.N., Wang,H.H. and Sommer,M.O. (2014) Prediction of resistance development against drug combinations by collateral responses to component drugs. *Sci. Transl. Med.*, **6**, doi:10.1126/scitranslmed.3009940.
12. Horst,J.P., Wu,T.H. and Marinus,M.G. (1999) *Escherichia coli* mutator genes. *Trends Microbiol.*, **7**, 29–36.
13. Glickman,B.W. and Radman,M. (1980) *Escherichia coli* mutants deficient in methylation-instructed DNA mismatch correction. *Proc. Natl. Acad. Sci. U.S.A.*, **77**, 1063–1067.
14. 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.
15. Nyerges,Á., Csörgő,B., Nagy,I., Latinovics,D., Szamecz,B., Pósfai,Gy. and Pál,C. (2014) Conditional DNA repair mutants enable highly precise genome engineering. *Nucleic Acids Res.*, **42**, e62.
16. Hong,E.S., Yeung,A., Funchain,P., Slupska,M.M. and Miller,J.H. (2005) Mutants with temperature-sensitive defects in the *Escherichia coli* mismatch repair system: sensitivity to mispairs generated *in vivo*. *J. Bacteriol.*, **187**, 840–846.
17. Yang,H., Wolff,E., Kim,M., Diep,A. and Miller,J.H. (2004) Identification of mutator genes and mutational pathways in *Escherichia coli* using a multicopy cloning approach. *Mol. Microbiol.*, **53**, 283–295.
18. Winkler,J.D., Garcia,C., Olson,M., Callaway,E. and Kao,K.C. (2014) Evolved osmotolerant *Escherichia coli* mutants frequently exhibit defective *N*-acetylglucosamine metabolism and point mutations in cell shape-regulating protein MreB. *Appl. Environ. Microbiol.*, **80**, 3729–3740.
19. Bonde,M.T., Klausen,M.S., Anderson,M.V., Wallin,A.I., Wang,H.H. and Sommer,M.O. (2014) MODEST: a web-based design tool for oligonucleotide-mediated genome engineering and recombineering. *Nucleic Acids Res.*, **42**, W408–W415.
20. Freddolino,P.L., Amini,S. and Tavazoie,S. (2012) Newly identified genetic variations in common *Escherichia coli* MG1655 stock cultures. *J. Bacteriol.*, **194**, 303–306.
21. Guzman,L.M., Belin,D., Carson,M.J. and Beckwith,J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.*, **177**, 4121–4130.
22. Geu-Flores,F., Nour-Eldin,H.H., Nielsen,M.T. and Halkier,B.A. (2007) USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. *Nucleic Acids Res.*, **35**, e55.
23. Lee,J.S., Kallehauge,T.B., Pedersen,L.E. and Kildegaard,H.F. (2015) Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. *Sci. Rep.*, **5**, 8572.
24. Lennen,R.M. and Herrgård,M.J. (2014) Combinatorial strategies for improving multiple-stress resistance in industrially relevant *Escherichia coli* strains. *Appl. Environ. Microbiol.*, **80**, 6223–6242.
25. Deatherage,D.E. and Barrick,J.E. (2014) Identification of mutations in laboratory-evolved microbes from next-generations sequencing data using breseq. *Methods Mol. Biol.*, **1151**, 165–188.
26. Isaacs,F.J., Carr,P.A., Wang,H.H., Lajoie,M.J., Sterling,B., Kraal,L., Tolonen,A.C., 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.
27. Jensen,K.F. (1993) The *Escherichia coli* K-12 'wild-types' W3110 and MG1655 have an *rph* frameshift mutation that leads to pyrimidine starvation due to low *pyrE* expression levels. *J. Bacteriol.* **175**, 3401–3407.
28. Pukkila,P.J., Peterson,J., Herman,G., Modrich,P. and Meselson,M. (1983) Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in *Escherichia coli*. *Genetics*, **104**, 571–582.