# Exploration of Methods for Many-Site Genome Editing with MAGE

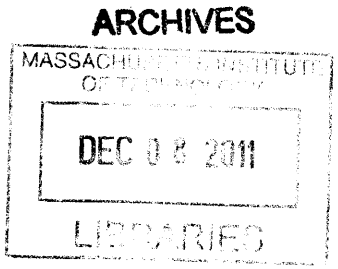
Bram Sterling

Bachelor of Science in Mathematics and Bachelor of Science in Biology, Massachusetts Institute of Technology, 2008

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning in partial fulfillment of the requirements for the degree of

Master of Science
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2011




© Massachusetts Institute of Technology 2011. All rights reserved.

Author:_____
Program in Media Arts and Sciences
August 5 2011

Certified by:_____
Joseph Jacobson
Associate Professor
Program in Media Arts and Sciences
Thesis Supervisor

Accepted by:_____
Mitchel Resnick
LEGO Professor of Learning Research
Academic Head
Program in Media Arts and Sciences

# Exploration of Methods for Many-Site Genome Editing with MAGE

Bram Sterling

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning on August 5 2011 in partial fulfillment of the requirements for the degree of Master of Science

## Abstract

In order to modify bacterial genetic codes, modifications must be made throughout the genome wherever the affected codon is used. Making such numerous and widespread genomic changes requires specialized techniques. MAGE is well-suited to this task, being highly amenable to multiplexing and having low time and resource costs per site. MAGE has been used as a first stage in recoding efforts, converting small clusters of sites in separate strains to be combined by other means, but improvements in MAGE technique suggest the possibility of using it to produce fully-recoded strains directly. To this end, I compare strategies based on co-selected MAGE and apply the best by performing 80 site conversions spread over 1/4 of the *E. coli* genome.

Thesis Supervisor: Joseph Jacobson
Title: Associate Professor

**Exploration of Methods for Many-Site Genome Editing with MAGE**

Thesis Committee:

Advisor: _____

Joseph Jacobson
Associate Professor
Program in Media Arts and Sciences

Reader: _____

Ed Boyden
Associate Professor
Program in Media Arts and Sciences

Reader: _____

Farren Isaacs
Assistant Professor
Molecular, Cellular & Developmental Biology
Systems Biology Institute
Yale University

# Table of Contents

## Background

The near universality, with some notable exceptions [1], of the genetic code which specifies the translation of nucleic acid sequences to polypeptides in natural lifeforms is a key fact of modern biology. The genetic code's portability across species is the foundation for the use of recombinant DNA, one of the primary components of modern biotechnology. The universal compatibility of genetic code is also exploited in nature, with horizontal transfer of genes between species.

Useful as the standardized genetic code is, there are many potential benefits to an experimental organism with a different code. Some of the advantages spring from the simple consequence of genetic incompatibility. An organism which follows an incompatible code could be immune to natural viruses, due to mistranslation of virus genes. Bacteriophage contamination of cultures is an industrially significant problem [2]. The same mutual unintelligibility could prevent spontaneous horizontal gene transfer into or out of the organism. Preventing gene transfer out of the experimental organism addresses some safety concerns for genetic engineering by ensuring that novel engineered genetics are not able to spread into wild microbial populations [3]. The incompatible organism would also be unable to make use of genes cloned directly from organisms using the original genetic code, but the widespread availability of gene synthesis makes recoded synthetic genes a viable substitute.

In addition to the inherent advantages of a sufficiently modified genetic code, certain specific changes have interesting possibilities. There is significant interest in incorporating amino acids outside the normal 20 into proteins. But the natural genetic code completely fills the coding space. Since every possible sequence can be interpreted under the existing code, it is impossible

to unambiguously code for a novel amino acid. Current methods of incorporating non-natural amino acids by translation are functional, but less than ideal. An engineered code can leave some codons coding for nothing, as openings which could be used to encode non-natural amino acids with greater efficiency [4] As well, the natural genetic code incorporates a mutation tolerance mechanism by arranging for codons to be clustered in mutation space with codons for the same or functionally similar amino acids. In nature, this improves mutation survival, but in an engineered organism survivability is secondary to genetic stability. A sparse genetic code where most mutations either result in meaningless codons which cause translation faults or code for radically different chemistry might result in fewer viable mutants, and higher effective genomic stability.

The Molecular Machines group led by Joseph Jacobson, part of the Center for Bits and Atoms at the MIT Media Lab, in collaboration with George Church's group at Harvard Medical Center, is working to produce useful working strains with recoded genomes supporting altered genetic codes [5]. The planned rE. coli series [6] of *E. coli* strains includes both compression of the genetic code to produce open codons and exchanging the meanings of codon groups to increase genetic incompatibility.

## Problem

The chief difficulty of re-engineering the genetic code of an organism is that each codon is used at many locations distributed across the entire genome. In order to cleanly re-assign a codon to a new role, the genome must be rewritten so that existing uses of the codon to be re-assigned are replaced by another appropriate codon. Even the least frequently used codon in the *E. coli* genome, the stop codon TAG, appears 314 times scattered across the 4.6 megabase

genome. This represents too many genomic changes to be feasible by conventional methods. Two approaches to making the changes appear possible. Complete resynthesis of the genome with desired changes, in an extension of recent work by the Venter group [7], or the multiplexed MAGE [8] the Molecular Machines group has chosen for the rE. coli project, of which the elimination of the TAG codon is the first stage.

Full genome re-synthesis promises the ability to implement any combination of desired genomic changes in a single step, and at a flat price. However, synthesis of a full genome is an expensive project, compared to a limited genome rewrite by other means. There is also the problem of genome boot up, or getting the synthetic genome active in a live cell, which may be difficult or impossible if the new genome has major incompatibilities with the host strain. If not otherwise resolved, this might require building multiple intermediate genomes in order to perform radical genetic code alterations, which would seriously undermine the advantages of the approach. Among genome editing methods, MAGE, derived from earlier oligo-based Lambda Red recombineering methods [9] stands out for codon reassignment. Using inexpensive oligos rather than larger constructs, which limits costs, MAGE works best at replacing small patches of genome, such as single codons or individual bases. MAGE has an extremely high baseline site replacement rate and can accomplish much more by multiplexing, and still more when taking advantage of co-selection [10], especially using doubly-selectable sites.

The rE. coli project spread the initial codon replacement across many strains in parallel, each receiving 10 consecutive target codon changes. These are to be combined into a single fully recoded strain by CAGE [5], a hierarchal conjugation-based strategy. At the time, converting the majority of a 10 site group required required days of continual MAGE cycling. As well as being

time consuming, such extended cycling is mutagenic and there was concern that it would result in the cells developing resistance to MAGE, as had been observed in previous experiments. Developments in MAGE since, particularly including a better understanding of linkage and performance enhancement by co-selection mean that today a single MAGE cycle followed by selection and screening can produce a number of site conversions comparable to the 18-cycle runs of unselected MAGE previously [5] used. This capability calls for a re-evaluation of how large numbers of site conversions are best performed. Increasing the rate of site conversion by MAGE directly increases the scale of feasible MAGE-based recoding projects, which is necessary if more ambitious recoding projects such as the planned rE. coli 2.0 and 3.0 [6] are to be carried out using MAGE genome editing.

## Related Work

There is no work in large-scale purely MAGE-based genome editing approaching the scale of this project. Our past work has included extensive experience with converting 10-site blocks, and some investigation of larger groups of up to 20 sites. This project investigates the potential of higher levels of multiplexing, and converts more total sites by MAGE alone than ever before in a single strain.

CAGE offers a hierarchic collection scheme that thus far has been used to accumulate small groups of changes initially made using basic MAGE. It has proven effective in producing genomes with the same set of modifications as this project targets, and others of equivalent complexity.

In the larger field of genome editing techniques, there have been a number of efforts which have made similar or greater numbers of discrete changes, over equal or larger genomic spans.

Such major genome engineering exercises are reviewed in Carr and Church [3]. The MGF-01 minimized-genome *E. coli* strain [11], with 53 sections of its genome removed, is a particularly interesting case. Cassette-based recombineering for individual deletions and transduction-based consolidation were used to produce a thriving strain with an appreciably reduced genome and favorable growth compared to the ancestral strain.

Still farther afield, the Venter group's fully synthetic *M. mycoides* genome [7] is the forerunner of a powerful alternative to genome editing. While in fact only a handful of moderate divergences from the natural sequence were made, for the sake of demonstration, the full re-synthesis approach in principle should be capable of arbitrarily extensive and detailed genome modification provided the altered genome is viable.

## Methods

The 80 sites to be converted comprise 8 consecutive 10-site blocks out of the 32 that make up the full list of TAG codons to be altered for rE coli 1.0 [5], specifically blocks 17 through 24, spanning over 1.1 megabases. Each site is one of the 314 TAG stop codons of the *E. coli* MG1655 genome, which are to be converted to TAA stop codons. Each block was to be converted by two rounds of multiplexed co-selection enhanced MAGE. After each round, colonies would be screened using two stages of qPCR to find a highly converted strain to proceed with. The second MAGE round would then target only those sites not successfully converted by the first round. Any sites still not converted after the second round would then be rolled over to be converted in the next pair of MAGE rounds along with the next 10-site block.

In principle, it should be possible to perform two rounds of co-selected MAGE in 5 days. One day to perform a MAGE round with tolC stop and begin selection, one day to plate for

colonies, one day to screen tolC- colonies, one day to perform a MAGE round with tolC restore and plate for colonies, and one day to screen tolC+ colonies.

As exploratory options MAGE was attempted with 20 and 40 target oligos, spanning blocks 19-20 and 17-20, respectively. The MASC-qPCR assay is not effective on such large numbers of sites. Addressing 20 or more targets in a single reaction works poorly due to cross-talk, and splitting them among multiple 10-site reactions produces separate rankings which are of limited

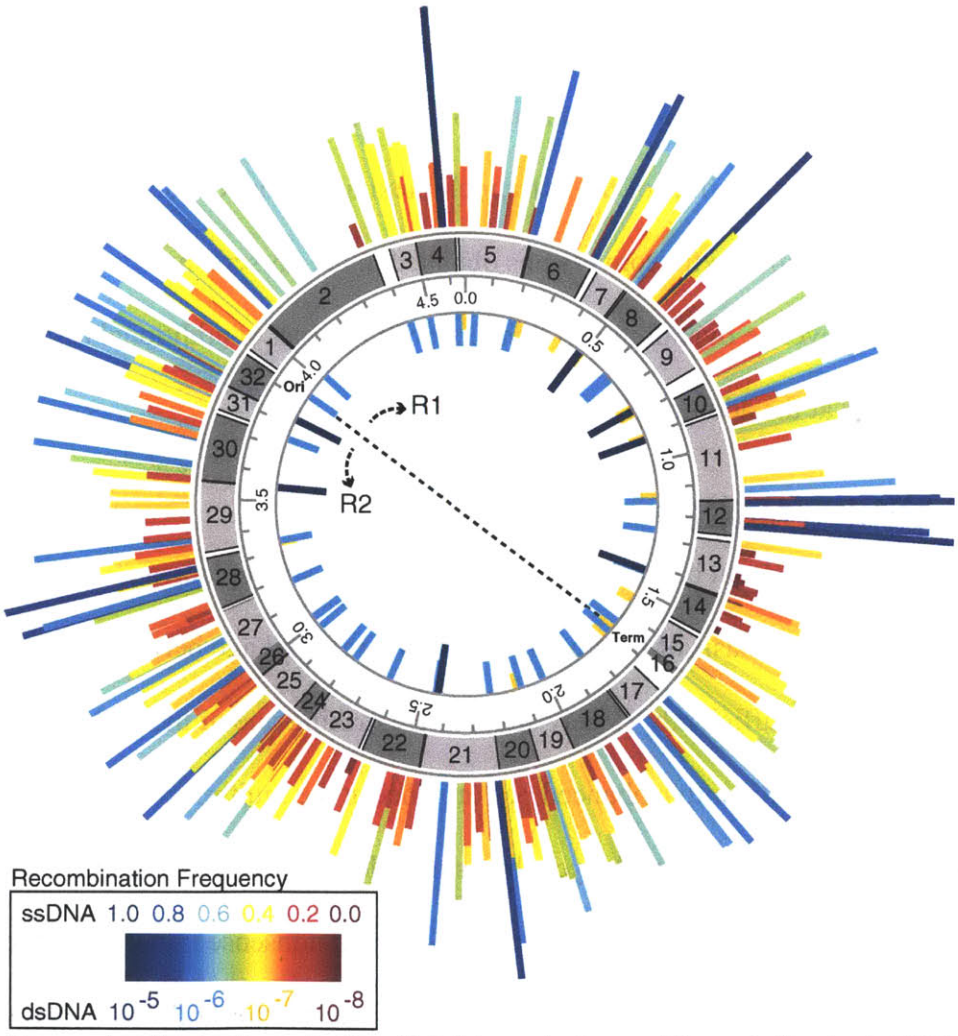



*Figure 1: Illustration of rE. coli 1.0 targets. Inward lines indicate dsDNA recombineering frequency at selected sites. Outward lines indicate conversion frequency of target sites after 18 rounds of standard MAGE. From Isaacs et. al. [5]*

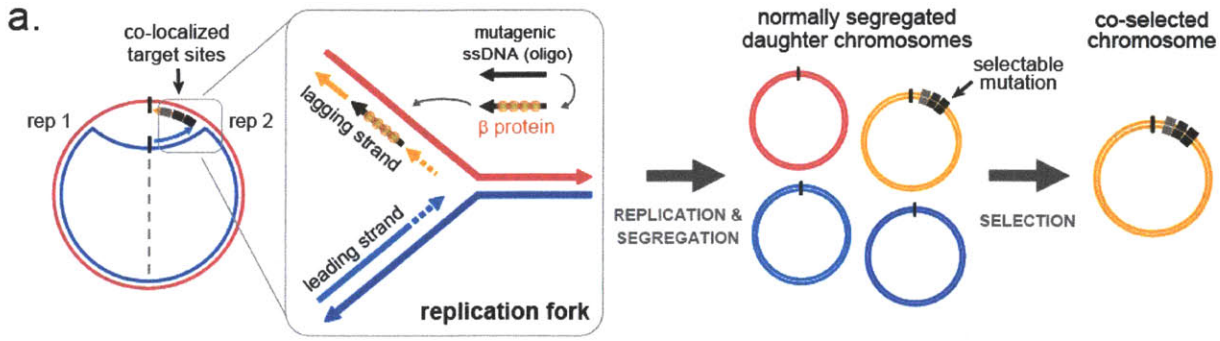



*Figure 2: Illustration of proposed model for MAGE co-selection, from Carr et. al. [10] Integration of oligos at the replication fork results in multiple changes made in one round to co-segregate, allowing selection for one target to enrich the population for conversion at all targets.*

use in picking out strains with many total conversions. So the resultant strains were screened on only 10 targets. If the detrimental effect of the extra multiplexing on conversion of those 10 assayed sites was sufficiently low, extra conversions obtained in the un-screened targets might accrue fast enough to improve the overall rate of conversion at the cost of additional assays to assess intermediate strains at the additional sites.

## Multiplex co-selected MAGE:

The basic MAGE method [8] can be made to yield much higher site replacement frequencies by use of co-selection [10]. The basis of this method is that there is linkage between MAGE replacement events which are sufficiently closely spaced. By making one of the replacements selectable, one may obtain significantly enhanced performance at the multiplexed sites.

Linked selection can be provided on each cycle by using a doubly-selectable marker. For the 80-site replacement, the chosen method is creating and reverting a nonsense mutation in the tolC gene, which codes for a transmembrane channel. Cells can be selected for both tolC+ and tolC- genotypes [12] by using SDS and Colicin E1, respectively.

The endogenous tolC gene is located at base pair 3176137, over 300kb toward the origin of

replication from the nearest of the 80 target sites. To position the selective marker more conveniently, the ancestral strain has had endogenous tolC seamlessly deleted and re-inserted at 2223738, in between blocks 20 and 21. This position was anticipated to provide strong co-selection for all sites in blocks 17-20, down-replicon from the marker, and an uncertain level of co-selection to blocks 21-24. Fall-off of co-selection in the more distant up-replicon blocks was a potential problem, as co-selection has observed to lose effect over similar up-replicon distances.

## qPCR strain screening:

The qPCR-based screen for highly converted strains uses two stages to examine a large number of clones while keeping to a manageable number of reactions. The first stage multiplexes over all target sites to allow tested clones to be roughly ordered by number of conversions. The second stage tests a handful of the most favorable clones for conversion at each target site.

The first stage uses multiplex allele-specific quantitative colony PCR (MASC-qPCR) [5] across the current target sites. 191 colonies are tested, and the 15 highest rated are passed on to the next stage.

The second stage uses the allele-specific colony qPCR to test each of the 15 top candidate strains at each of the current targets individually, providing a selection of full genotypes from which a strain can be chosen to begin the next cycle. The primary factor in the choice is total number of sites converted, but choosing between strains with equally numerous but different sets of conversions is necessary.

## *Materials:*

### Strain

The ancestral strain is EcNR1 [5], which is MG1655 with λ-prophage::bioA/bioB. MutS was deactivated by introducing a nonsense mutation by oligo recombineering. This strain was further seamlessly deleted for endogenous tolC by oligo recombineering to prepare for reintroduction of tolC by cassette recombineering. The resultant strain, EcNR1 mutS- ΔtolC (-> 20.21) responds well to MAGE and has the tolC gene located within hopeful co-selection range of all desired target sites. This strain is the ancestor of all other strains used in the experiment.

### Oligos

The MAGE 90-mers and MASC-qPCR primers used are exactly those used in Isaacs et al. [5] to convert and assess sites in regions 17 through 24. TolC function is toggled using the 90-mers tolC_on: tgggttcagttcgttgagccaggccgagaacctgatgcaagtttaTcagcaagcacgccttagtaacccggaat-tgcgtaagtctgccgc and tolC_off: tgggttcagttcgttgagccaggccgagaacctgatgcaagtttaAcagcaagcacgcct-tagtaacccggaattgcgtaagtctgccgc.

## *Techniques:*

### Co-selected MAGE

In each round of MAGE, a clonal liquid culture was used to inoculate 20 mL low salt (5g/L) Luria Broth (LSLB) in a baffle flask. This culture was grown at 30°C with agitation at 250 RPM to mid-logarithmic growth, determined by OD600 of 0.3-0.6. To induce the λ Red genes the flask was transferred to a shaking water bath at 42°C for 15 minutes, after which it was immediately placed on ice and moved to a 4° environment to be prepared for electroporation.

After the culture has chilled, two 1 mL samples were withdrawn. Each was pelleted by

centrifugation at 16100 g for 1 minute, and the supernatant fully removed, with repeat centrifugation as necessary. The pellet was then resuspended in 1 mL pure water. The pelleting and resuspension was then repeated, followed by a further pelleting and removal of supernatant. The pellets were then resuspended together in 90 μL pure water.

In duplicate, 45 μL of the electrocompetent cells was then added to 5 μL of MAGE 90-mer mix, for a final concentration of 990 nM total target site TAA -> TAG 90-mers, evenly divided among sites, and 10 nM tolC stop or tolC restore 90-mer for co-selection. This was transferred cold to a chilled electroporation cuvette with 1mm gap, and electroporated using a Bio-Rad Gene Pulser Xcell at 1800 V, 25 μF, 200Ω. The electroporated cells were immediately resuspended in 1 mL LSLB.

When the co-selection oligo used was tolC restore, the full 1 mL was taken from the cuvette and added to another 1 mL LSLB in a culture tube. This was then allowed to recover at 30°C with agitation for 1 hour before plating 80 μL on LSLB agar with carbenicilin and 0.005% SDS to isolate tolC+ colonies.

When the co-selection oligo used was tolC stop, 50 μL was taken from the cuvette and inoculated into 2 mL LSLB. This was allowed to recover and grow at 30°C with agitation for no less than 5 hours. After diluting if necessary, it was then allowed to grow to mid-log and OD600 of 0.4 to 0.6. At the same time, a known tolC+ culture was brought to the same state of growth to serve as a negative. Each of these was used to inoculate a tube of 2 mL LSLB and 20 μL colicin E1 preparation with 20 μL cell growth. These cultures were allowed to grow for 8-12 hours, until the electroporation-derived cultures become confluent. Each was then plated at $10^4$ and $10^5$ dilutions onto LSLB agar with carbenicilin.

### qPCR site assay

MASC-qPCR is performed as described in Isaacs et al. [5], with the exception that the combined primer concentration of 1 μM is divided evenly among whatever set of primers is required by the current reaction, covering up to 12 sites. Attempting larger numbers of sites results in increased crosstalk between the primer pairs, reducing the usefulness of the result. Prospective colonies are ranked in descending order of Δ-Ct. A single wild-type control is included for comparison.

Allele specific colony qPCR uses the same procedure as MASC-qPCR, but includes only a single primer pair in each reaction. A wild-type control colony is required to ensure that the unconverted genotype is represented among the samples. The ΔΔCt indicative of positive conversion is highly site-dependent, but a set of colonies including both TAG and TAA genotypes offers distinctively bimodal ΔCt values which can be evaluated by examination. In some cases intermediary ΔCts may be observed. Samples exhibiting such results should be avoided, as it may indicate a non-clonal cell sample or an undesired mutation close to the target site.

## Implementation and Results

Results of co-selected MAGE rounds to produce the 80-site converted strain are detailed in table 1. A number of rounds featured modifications to the standard plan or showed unexpected effects.

Round 1: Only 64 colonies in MASC-qPCR screen, and only 7 passed to the second-stage assay.

Round 3: Test-plated leading strains for SDS sensitivity. All tested positive.

Post-Round 6: Allele-specific qPCR tested current strain at all sites in blocks 18-20. All sites

were found to be converted to TAA except 20-8, contrary to the result after round 2.

Round 7: Test-plated winner to confirm SDS sensitivity.

Round 8: Based on later results, 20-7 was restored to wild type in the process of converting 20-8 in this round. The 20-7 and 20-8 90-mers each overlap the other's target site and frequently overwrite it to wild-type.

Round 10: Expanded second-stage screen to 31 colonies.

Round 12: Plating at 1 hour growth yielded insufficient colonies. Replating after further outgrowth was required.

Round 13: Test-plated leading strains for SDS sensitivity. 14 of 15 tested positive.

Round 14: Cells recovered 90 minutes before plating, yielded insufficient colonies. Only 95 colonies screened.

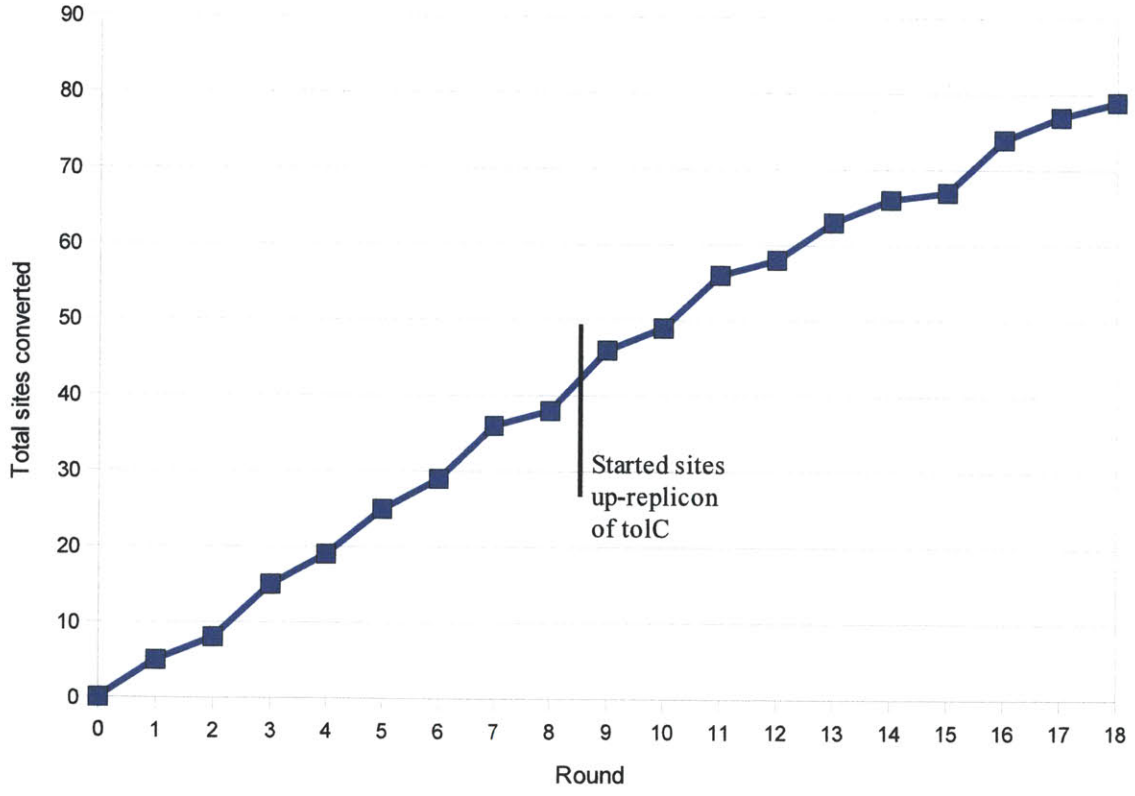



*Figure 3: Plot of conversion progress*

Round 15: Both electroporations arced. After colicin selection, estimated 1/4th of colonies are tolC-. Of 15 chosen for second-stage screen, 7 proved to be SDS sensitive on test-plating. One of these was chosen for continuation.

Round 16: Allowed recovery for 5 hours 40 minutes, plated at 1 and 10 fold dilution.

Round 17: Plating produced limited colonies, so only 143 colonies were available for MASC-qPCR screen.

Post-Round 17: In preparation for the final round, the current strain was assessed at all 80 sites by allele-specific qPCR. All sites were found to be converted except for 23-9, 24-9, and 20-7.

Round 18: Recovery extended to overgrowth, then plated at $10^3$, $10^4$, and $10^5$ fold dilution. Site screen included 20-8 as well as the target sites in an effort to recover a colony in which the 20-7 90-mer did not overwrite 20-8. The second stage of qPCR screening was expanded to examine 46 colonies, of which 8 were contaminated by pipetting error.

From electroporation for round 1 to second stage screen for round 18, the process ran 85 days. That is nearly double the 45 days that in theory should suffice to run this number of rounds.

| Round | Selective 90mer | Targets | Converted |
|---|---|---|---|
| 1 | tolC stop | All block 20 sites | 20-1, 20-3, 20-5, 20-6, 20-10 |
| 2 | tolC restore | 20-2, 20-4, 20-7, 20-8, 20-9 | 20-2, 20-4, 20-7, 20-8* |
| 3 | tolC stop | 20-9, all block 19 sites | 20-9, 19-3, 19-4, 19-5, 19-7, 19-9, 19-10 |
| 4 | tolC restore | 19-1, 19-2, 19-6, 19-8 | 19-1, 19-2, 19-6, 19-8 |
| 5 | tolC stop | All block 18 sites | 18-1, 18-2, 18-6, 18-7, 18-8, 18-10 |
| 6 | tolC restore | 18-3, 18-4, 18-5, 18-9 | 18-3, 18-4, 18-5, 18-9 |
| 7 | tolC stop | All block 17 sites | 17-1, 17-2, 17-4, 17-5, 17-8, 17-9, 17-10 |
| 8 | tolC restore | 17-3, 17-6, 17-7, 20-8 | 17-3, 17-6, 20-8 |
| 9 | tolC stop | 17-7, all block 21 sites | 21-1, 21-2, 21-3, 21-4, 21-5, 21-8, 21-9, 21-10 |
| 10 | tolC restore | 21-6, 21-7, 17-7 | 21-6, 21-7, 17-7 |
| 11 | tolC stop | All block 22 sites | 22-1, 22-2, 22-5, 22-6, 22-7, 22-8, 22-9 |
| 12 | tolC restore | 22-3, 22-4, 22-10 | 22-4, 22-10 |
| 13 | tolC stop | 22-3 and all block 23 sites | 22-3, 23-3, 23-6, 23-7, 23-10 |
| 14 | tolC restore | 23-1, 23-2, 23-4, 23-5, 23-8, 23-9 | 23-2, 23-4, 23-8 |
| 15 | tolC stop | 23-1, 23-5, and all block 24 sites | 24-1 |
| 16 | tolC restore | 23-1, 23-5, 23-9, and 24-2 to 24-10 | 23-5, 24-2, 24-4, 24-5, 24-6, 24-7, 24-8 |
| 17 | tolC stop | 23-1, 23-9, 24-3, 24-9, 24-10 | 23-1, 24-3, 24-10 |
| 18 | tolC restore | 20-7, 23-9, 24-9 | 23-9, 24-9 |

*Table 1: Production MAGE rounds. *20-8 was detected as converted in round 2 but as unconverted after round 6. It is accounted as having been converted in round 8.*

## Highly Multiplexed Co-selected MAGE Test

The trial of higher multiplexed MAGE used 64 clones in the first stage and 4 in the second stage screen for each of the two tested scales. In both cases, few conversions were detected in block 20. With the 20-target 19-20 MAGE pool, only 20-2 and 20-5 were found converted. With the 40-target 17-20 pool, 20-3, 20-4, and 20-10 were obtained. On assaying the top strains at the unscreened sites, the 19-20 winner was found to be converted at 19-9, for a total of 3 conversions, while the 17-20 winner was found to be converted at 17-1 and 17-10, bringing its total to 5. Considering that the latter result was obtained with only 4 second stage clones rather

than 7, it could be considered stronger than the 5 obtained from the conventional 10-site pool of block 20 only. However, the prospect of an expected 4 rounds to fully convert block 20, with little likelihood that blocks 19, 18, and 17 would all follow in a single round each, was considered unfavorable compared to 10-site pools which were strongly expected to convert a block every 2 rounds.

# Discussion

## *Potential Improvements*

Although the tested possibilities for improving performance were rejected, an opportunity for increased performance is evident from examining the results of the cycles. Cycles targeting 10 or more sites (rounds 1, 3, 5, 7, 9, 11, 13, 15, and 16) produced a median of 7 changes. The other cycles targeted at most 6 sites, and converted no more than 4. (See figure 4 or table 2). What this indicates is that to get many conversions, it is advisable to target many sites on every cycle. This could easily be implemented by adding new targets to the current set every cycle such that each cycle covers 10 to 12 sites, rather than introducing new targets 10 at a time on a rigid schedule. Simply adding one new target for each target converted should suffice.

Doing so might yield a conversion rate of roughly 7 sites per cycle until fewer than 10 unconverted targets remain. However, the process might be stalled by an accumulation of refractory sites as present targets. Some conversions are known to be significantly less likely to appear than others [5], and it is likely the case that some conversions make a weaker contribution to the MASC-qPCR signal and are thus less likely to be isolated after screening. Such conversions, when in a target pool with less-refractory targets, may be at risk of persistently failing to appear in the strain chosen for propagation. Rounds in which there are fewer targets

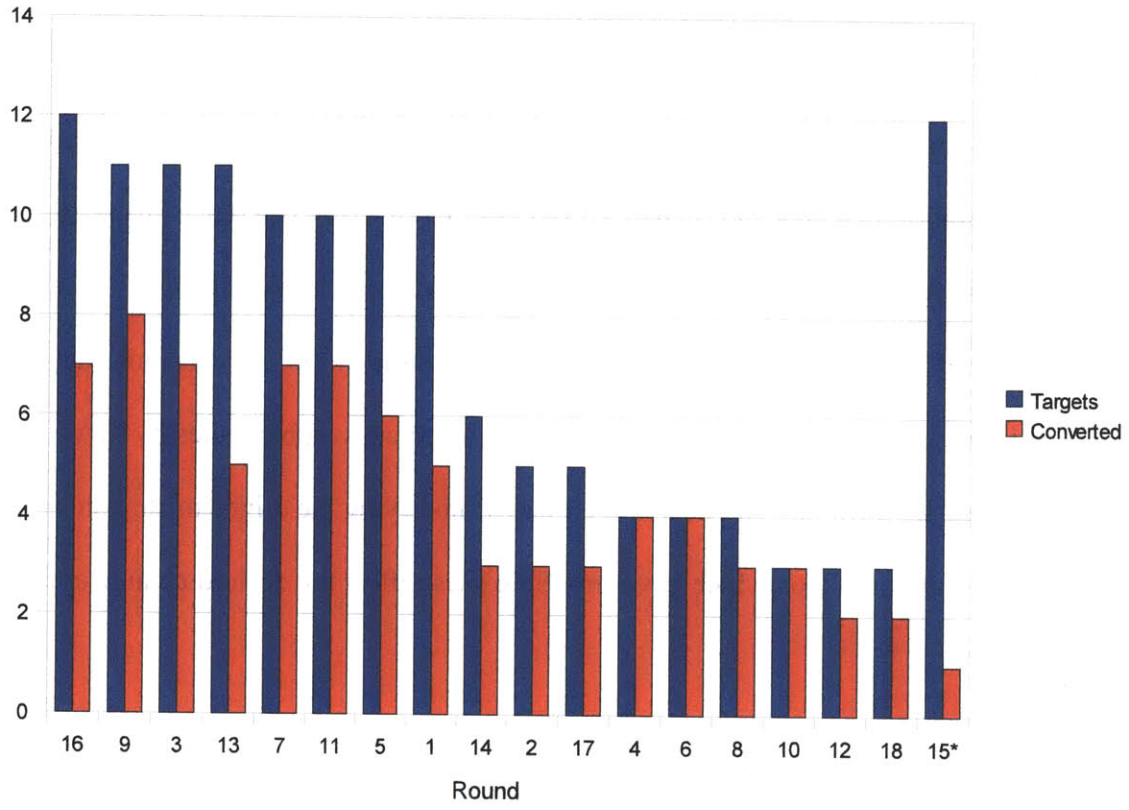



*Figure 4: Comparison of number of targets and number of conversions made across the 18 rounds of co-selected MAGE. *Due to arcing, round 15 is not comparable to other rounds, but is included for completeness.*

increase the 90-mer fraction dedicated to each and decrease competing signals in the MASC-qPCR, both of which would improve the chances of obtaining the problematic conversion.

This more fluid approach might better harness the potential of using larger multiplexed pools as well. With a rolling set of assayed sites rather than a fixed block by block frame, fully converting blocks ceases to be a hurdle. The sole factor becomes how many site conversions can be obtained per cycle, which this study has not firmly established. The necessary addition of an assay step to determine which conversions outside the screened set would be a minor cost. Based on the limited test, 40-site multiplexing with screening at 10 sites might yield more total conversions than the standard 10 site fully screened pool, but further testing is needed. In light

of the success of up-replicon co-selection as far as block 24, 80-site pools should be considered as well as the smaller pools tested for this project.

Based on the extremely poor performance of cycle 15, it would be advisable to discard any MAGE electroporation that arcs and prepare fresh cells. Even if an extra day of work is required, less time would be lost than in completing such an unproductive cycle.

Avoiding MAGE oligos which overlap and revert neighboring sites, such as those targeting sites 20-7 and 20-8, is also strongly advised. It is possible to convert both sites by MAGE in such a case[5], but the difficulty can be avoid either by modifying the oligos to stop short of the next target, or by making oligos which cover multiple sites reflect the desired change at all of them.

## *Performance and potential*

Co-selected MAGE has proven to be a viable method for genome-editing on the 80-site scale. Over the course of 18 cycles of MAGE, 79 of the 80 desired edits were obtained with no discernible breakdown of the genome modification process. It is expected that a number of off-target mutations will have accumulated in the genome due to long propagation in a mutS- state and repeated λ-red induction, but MAGE performance does not appear to have suffered the fall-off sometimes seen after many cycles of non-selected MAGE [5], nor has tolC selectability failed. Notably there does appear to be a possible decrease in rate of conversions in the later rounds, starting at round 13. The downturn is slight enough to be attributed to chance, but might also be indicative of weaker co-selection far up-replicon of tolC, or even a slight reduction in the strain's MAGE performance. It might also indicate an increased frequency of recalcitrant targets. Although there is no accumulation of problem sites from earlier blocks, block 23 may be

intrinsically challenging. In light of the high performance of round 16 any general drop in MAGE performance cannot be severe, but followup assessment the final strain's performance is called for.

This stability is likely a result of the complex selection pressure which is regularly applied throughout the process. To be propagated from one round to the next a strain need not only pass selection on tolC, but must be picked out as highly converted by the two-stage screen, and can at the same time be tested against the possibility of having subverted the primary selection by attaining a phenotype resistant to both SDS and colicin E1. In light of this success over 80 targets, this suggests that this method of MAGE with co-selection is stable and robust, and capable of scaling to larger tasks.

As a method capable of large-scale site editing, the co-selected MAGE method's nearest comparison is to the MAGE-CAGE hierarchical method. Recently, a strain converted at the same 80 sites was obtained by MAGE-CAGE [5]. Reportedly [Lajoie, MJ, private correspondence] strain construction with CAGE took approximately 5 weeks. Even accounting for time to generate the 10-site converted strains that are precursors to CAGE assembly, that is substantially less time than was required by the MAGE-only method, and somewhat faster than the theoretical potential speed of 18 rounds of the co-selected MASC-qPCR screened MAGE method. Nonetheless, co-selected MAGE offers some potential advantages.

The co-selected MAGE process scales linearly, converting new sites at a steady pace by simple repetition. While this leads to much slower progress on larger tasks than logarithmic scaling methods like hierarchical CAGE, it also means that the process is scale-invariant. Extending the current work from an 80-site quarter genome to a 160 site half genome is simply a

matter of repositioning the tolC gene and performing another series of cycles. While time consuming, this process could potentially convert the full 314 targets for rE. coli 1.0 in less than a year. Hierarchical CAGE might achieve the same far faster, but calls for larger conjugative genomic transfers to do so, potentially facing novel challenges.

Scaling down instead of up, co-selected MAGE remains well-suited to making a smaller number of changes. A single cycle of screened co-selected MAGE achieves slightly fewer changes than 18 cycles of unselected MAGE [5], but takes less time and avoids risk of phenotypic drift seen in the unselected process. Unselected MAGE can, ideally, be performed at a rate of one cycle per 2-2.5 hours, and screening 47 clones after 18 cycles with 10 90-mers can be expected to produce a strain with 8 conversions [5]. Based on this, and supposing that no problems arise from iterating that process, a plausible model for open-ended MAGE processing is 3 days of cycling at 6 cycles per day, followed by one day screening clones, to produce 8 site conversions every 4 days. This is much quicker than the co-selected method as performed, which yielded slightly less than one conversion per day on average. However the necessary steps for a co-selected MAGE round can be performed in 2.5 days, on average. With a rolling set of 11 targets, the expected yield would be 7 conversions per 2.5 days, 40% higher than the non-selected projection. In addition to now-proven stability, use of co-selection does offer the possibility of moderately more rapid MAGE genome editing.

The requirement for a selectable marker adds some overhead to co-selected MAGE, but is likely worthwhile for 20+ edits. On a smaller scale still, the overhead required by co-selection will likely overwhelm any time benefit for 4 or less sites. However, on 4 sites or less, screened co-selected MAGE has been observed to obtain full conversion in a single electroporation,

avoiding any intermediate genotypes. This might be valuable for some manipulations, such as introducing a mutation in one gene and a compensatory mutation in another together, or reconfiguring genetic controls for a synthetic pathway.

## Conclusion

Co-selected MAGE is capable of genome editing on the 80 site quarter genome scale, though not as fast as MAGE-CAGE. While its linear method makes it likely to scale to larger tasks, such as the 314 targets of rE. coli 1.0 [6], with minimal technical problems, linear scaling of project time is very unfavorable compared to the logistic time scaling of hierarchical methods. Because of this co-selected MAGE is unsuitable as the sole method for creating strains with very large numbers of changes, such as the thousands or tens of thousands required for larger genome recoding projects. However, for smaller scale projects of 40 or less changes, it has potential to be faster than either MAGE-CAGE or non-selected MAGE cycling. This could include initial production of changes to be accumulated by MAGE-CAGE.

Co-selection greatly increases the number of changes made per MAGE cycle and has potential to substantially improve the rate of changes over time, provided that an appropriately positioned doubly-selectable marker is available. Combined with fast, scar-less recombineering to manipulate markers, this makes it an appealing method for making moderate numbers of changes over a megabase-scale area of the genome.

## References

[1]: A. Ambrogelly, S. Palioura, D. Söll. Natural expansion of the genetic code. Nat Chem Biol. 2007 Jan;3(1):29-35.

[2]: Łos M, Czyz A, Sell E, Wegrzyn A, Neubauer P, Wegrzyn G. Bacteriophage contamination: is there a simple method to reduce its deleterious effects in laboratory cultures and biotechnological factories? J Appl Genet. 2004;45(1):111-20.

[3]: Carr PA, Church GM. Genome engineering. Nat Biotechnol. 2009 Dec;27(12):1151-62.

[4]: Liu CC, Schultz PG. Adding new chemistries to the genetic code. Annu Rev Biochem. 2010;79:413-44.

[5]: Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, Kraal L, Tolonen AC, Gianoulis TA, Goodman DB, Reppas NB, Emig CJ, Bang D, Hwang SJ, Jewett MC, Jacobson JM, Church GM. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. Science. 2011 Jul 15;333(6040):348-53.

[6]: Carr, P.A., and Isaacs, F. (2006). rE.coli: whole genome engineering. Presented at the Second International Conference on Synthetic Biology (Synthetic Biology 2.0), May 20–22, Berkeley, CA.

[7]: Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA 3rd, Smith HO, Venter JC. Creation of a bacterial cell controlled by a chemically synthesized genome. Science. 2010 Jul 2;329(5987):52-6. Epub 2010 May 20.

[8]: Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM. Programming cells by multiplex genome engineering and accelerated evolution. Nature. 2009 Aug 13;460(7257):894-8. Epub 2009 Jul 26.

[9]: Ellis HM, Yu D, DiTizio T, Court DL. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. Proc Natl Acad Sci U S A. 2001 Jun 5;98(12):6742-6. Epub 2001 May 29.

[10]: Carr PA, Wang HH, Sterling B, Isaacs FJ, Xu G, Church GM, Jacobson JM. Enhanced Multiplex Genome Engineering through Cooperative Oligonucleotide Co-selection. Manuscript in preparation.

[11]: Mizoguchi, H., Sawano, Y., Kato, J. & Mori, H. Superpositioning of deletions promotes growth of Escherichia coli with a reduced genome. DNA Res. 2008 Oct;15(5):277-84. Epub 2008 Aug 27.

[12]: DeVito JA. Recombineering with tolC as a selectable/counter-selectable marker: remodeling the rRNA operons of Escherichia coli. Nucleic Acids Res. 2008 Jan;36(1):e4. Epub 2007 Dec 15.

## Supplemental Data

| Round | Total Conversions | Round Targets | Conversions | First stage screen | Second stage screen |
|---|---|---|---|---|---|
| 1 | 5 | 10 | 5 | 64 | 7 |
| 2 | 8 | 5 | 3 | 191 | 15 |
| 3 | 15 | 11 | 7 | 189 | 15 |
| 4 | 19 | 4 | 4 | 191 | 15 |
| 5 | 25 | 10 | 6 | 191 | 15 |
| 6 | 29 | 4 | 4 | 191 | 15 |
| 7 | 36 | 10 | 7 | 191 | 15 |
| 8 | 38 | 4 | 3 | 191 | 15 |
| 9 | 46 | 11 | 8 | 191 | 15 |
| 10 | 49 | 3 | 3 | 191 | 31 |
| 11 | 56 | 10 | 7 | 191 | 15 |
| 12 | 58 | 3 | 2 | 191 | 15 |
| 13 | 63 | 11 | 5 | 191 | 15 |
| 14 | 66 | 6 | 3 | 95 | 15 |
| 15 | 67 | 12 | 1 | 191 | 15 |
| 16 | 74 | 12 | 7 | 191 | 15 |
| 17 | 77 | 3 | 3 | 143 | 15 |
| 18 | 79 | 2 | 2 | 191 | 38 |

*Table 2: Numerical summary of conversions and tabulation of number of clones screened. Colonies discarded due to contamination are not included.*

| Site | Gene | 18-cycle MAGE rate | Rounds to convert |
|---|---|---|---|
| 17-1 | ydhM | 60.0 | 1 |
| 17-2 | lhr | 73.3 | 1 |
| 17-3 | ynhG | --- | 2 |
| 17-4 | sufA | 47.7 | 1 |
| 17-5 | ydiA | 37.8 | 1 |

| | | | |
|---|---|---|---|
| 17-6 | ydiE | 43.9 | 2 |
| 17-7 | nlpC | 13.3 | 4 |
| 17-8 | btuC | 22.0 | 1 |
| 17-9 | arpB_1 | 69.8 | 1 |
| 17-10 | ydjX | 64.9 | 1 |
| 18-1 | nudG | 64.3 | 1 |
| 18-2 | b1788 | 35.6 | 1 |
| 18-3 | yeaL | 15.6 | 2 |
| 18-4 | yeaX | 17.8 | 2 |
| 18-5 | yobB | 31.1 | 2 |
| 18-6 | exoX | 14.7 | 1 |
| 18-7 | yecN | 56.5 | 1 |
| 18-8 | otsA | 28.9 | 1 |
| 18-9 | dcyD | 10.9 | 2 |
| 18-10 | yedM | 32.6 | 1 |
| 19-1 | fliE | 34.8 | 2 |
| 19-2 | fliN | 25.5 | 2 |
| 19-3 | fliP | 47.8 | 1 |
| 19-4 | fliQ | 34.1 | 1 |
| 19-5 | yedS_1 | 45.5 | 1 |
| 19-6 | cbl | 31.9 | 2 |
| 19-7 | b1996 | 32.6 | 1 |
| 19-8 | hisL | 42.2 | 2 |
| 19-9 | wbbJ | 62.8 | 1 |
| 19-10 | wcaM | 47.4 | 1 |
| 20-1 | wcaL | 17.0 | 1 |
| 20-2 | wcaC | 45.7 | 2 |
| 20-3 | asmA | 20.0 | 1 |
| 20-4 | baeR | 21.7 | 2 |
| 20-5 | gatR_1 | 36.2 | 1 |
| 20-6 | tra5_4 | 17.8 | 1 |
| 20-7 | yegV | 28.3 | 2 |

| | | | |
|---|---|---|---|
| 20-8 | yegW | 17.4 | 3* |
| 20-9 | yehQ | 38.3 | 3 |
| 20-10 | yohC | 70.2 | 1 |
| 21-1 | yohF | 33.3 | 1 |
| 21-2 | sanA | 85.2 | 1 |
| 21-3 | yejA | 25.0 | 1 |
| 21-4 | yejE | 33.3 | 1 |
| 21-5 | b2191 | 21.1 | 1 |
| 21-6 | rcsD | 34.9 | 2 |
| 21-7 | rcsC | 26.8 | 2 |
| 21-8 | yfaT | 50.0 | 1 |
| 21-9 | menF | 70.0 | 1 |
| 21-10 | yfcO | 2.9 | 1 |
| 22-1 | yfcU | 21.7 | 1 |
| 22-2 | tfaS | 29.8 | 1 |
| 22-3 | ypdI | 23.4 | 3 |
| 22-4 | yfdY | 21.7 | 2 |
| 22-5 | yfeO | 48.9 | 1 |
| 22-6 | mntH | 14.9 | 1 |
| 22-7 | xapR | 8.7 | 1 |
| 22-8 | yfeR | 23.4 | 1 |
| 22-9 | yffB | 41.3 | 1 |
| 22-10 | hda | 23.4 | 2 |
| 23-1 | yfgG | 6.8 | 4 |
| 23-2 | pbpC | --- | 2 |
| 23-3 | yphA | 17.0 | 1 |
| 23-4 | yfhB | 27.7 | 2 |
| 23-5 | kgtP | 38.3 | 3 |
| 23-6 | yfiA | 22.7 | 1 |
| 23-7 | yfjQ | 38.6 | 1 |
| 23-8 | yfjR | 17.0 | 2 |
| 23-9 | ypjC | 26.5 | 5 |

| | | | |
|---|---|---|---|
| 23-10 | ygaQ | 40.9 | 1 |
| 24-1 | ygaR | 57.1 | 1 |
| 24-2 | yqaC | 40.0 | 1 |
| 24-3 | gabT | 22.0 | 2 |
| 24-4 | ygaU | 37.2 | 1 |
| 24-5 | ygaM | 20.9 | 1 |
| 24-6 | luxS | 30.2 | 1 |
| 24-7 | mltB | 36.4 | 1 |
| 24-8 | srlE | 43.9 | 1 |
| 24-9 | norW | 14.6 | 3 |
| 24-10 | ascB | --- | 2 |

*Table 3: Conversion rate and rounds to conversion by site. 18-cycle MAGE rate is the frequency of conversion after 18 unselected MAGE rounds, as reported in Isaacs et. al. [5]. Rounds to convert is the number of rounds of co-selected MAGE the site was targeted in before being converted. Round 15 is not counted, except for 24-1, due to arcing. *20-8 is problematic due to its 90-mer overlapping 20-7, and due to its false positive in round 2.*

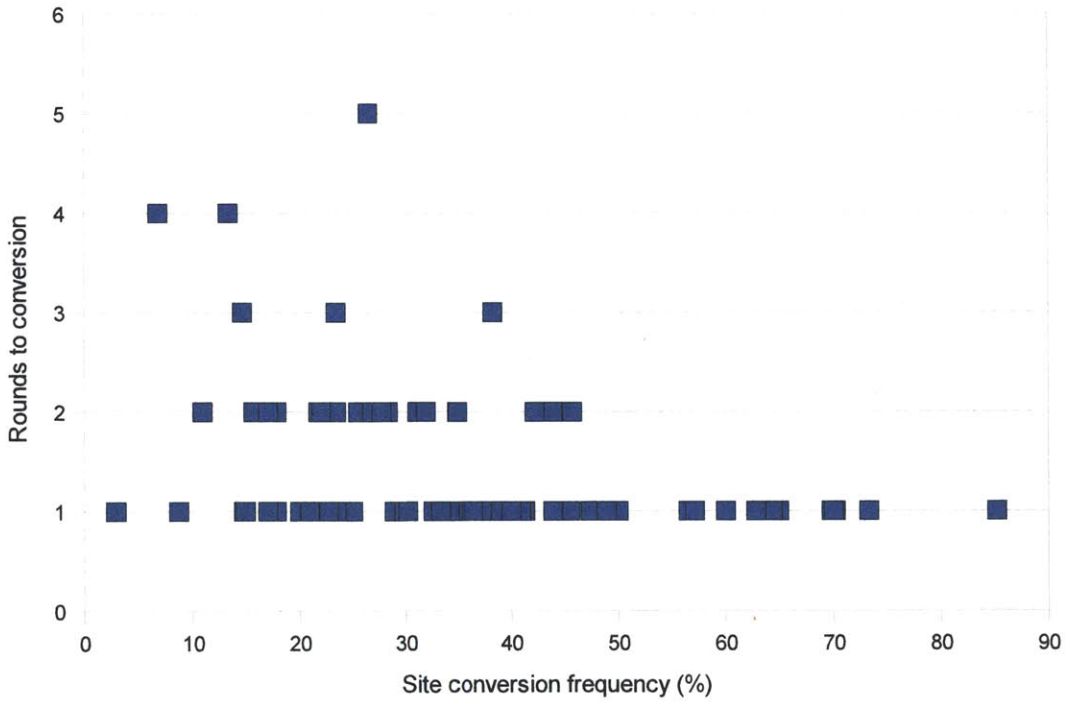



*Figure 5: Scatter plot of rounds to conversion versus reported frequency in 18 unselected MAGE cycles, as listed in Table 3. Of possible note, all sites with frequencies over 50% were converted on the first round that they were present.*