**Optimized Design Criteria for MAGE Oligonucleotides**

Oligonucleotide-mediated allelic replacement was achieved in the modified *E. coli* strain EcNR2 (mutS\(^-\), \(\lambda\)-Red\(^+\)) by directing oligos to the lagging strand of the replication fork during DNA replication\(^1\). Targeting the lagging strand of replicating DNA with single-stranded oligonucleotides (ss-oligos) has been shown to be more efficient than targeting the leading strand in this mutS\(^-\) strain (EcNR2)\(^2\). The replacement efficiency was characterized by using oligos either to inactivate the *lacZ* gene and screen for white colonies on Xgal/IPTG-containing agar plates or to fix a defective *cat* gene and select for chloramphenicol resistant colonies.

In rare cases during MAGE experiments, we observed small genomic sequence changes in the oligo targeted region that are not by design, i.e., 7\(^+\) bp mutations when only 6 bp are targeted (Fig. 3). We hypothesize that these additional mutations are likely the result of allelic replacement by faulty oligos that arise from errors during oligo synthesis. Purification of oligos may reduce instances of such cases.

Supplementary Figure 2a shows that replacement efficiency was found to be dependent on oligo length, highest at 90 basepairs (bp). We hypothesize that the 90 bp oligo has the most optimal replacement efficiency for two main reasons. First, the \(\lambda\)-Red single-stranded DNA-binding protein \(\beta\), has been shown to require at least 30 bp to complex with oligos *in vitro*\(^3\). *In vivo*, shorter oligos have fewer basepairs of homology to hybridize to the targeted chromosomal site, thus decreasing the likelihood of replacement. Second, while oligos longer than 90 bp may have more regions of homology to the chromosome, they are also more likely to form secondary structures. Inhibitory secondary structures (e.g., hairpin loops) can lead to dramatically lower efficiencies of replacement since reducing the number of exposed bases on the oligo will decrease the frequency of hybridization to its chromosomal target. Along these lines, we observed that oligos with computationally predicted minimal folding energies\(^4\) of less than \(-12.5\) kcal/mol showed significantly reduced allelic replacement frequencies experimentally.
(Supplementary Fig. 2b). Phosphorothioate bonds located at the terminal bases may increase replacement efficiency by preventing in vivo degradation of synthetic oligonucleotide molecules by endogenous exonucleases in the cell. Phosphorothioated nucleotides increased replacement efficiency by more than 2-fold when placed at the 5' terminus, but showed no effect when placed at the 3' terminus (Supplementary Fig. 2c). Increasing the number of phosphorothioated bases at the 5' terminus increased the efficiency of replacement, which saturates to its highest level at four phosphorothioated bases (Supplementary Fig. 2d). Oligonucleotides, in which all bases contained phosphorothioated bonds, did not incorporate into the chromosome (data not shown). The replacement efficiency remains high across a wide range of oligo concentrations (0.05-50 µM), thus allowing for large and highly complex oligo pools (Supplementary Fig. 2e). Allelic replacement efficiency was low when low concentrations of oligos (<0.05 µM) were used, suggesting a dilution effect. In fact, at low oligo concentrations (i.e., 0.005 µM), there are on average three DNA molecules per volume of a cell (~10⁻¹⁸ m³), leading to drastically decreased likelihood of a replacement event. Therefore, increasing the amount of oligos available for allelic replacement by either increasing the oligo concentration during electroporation or by increasing the oligo half-life inside the cell (via terminal phosphorothioated nucleotides) will lead to higher efficiencies of replacement. Interestingly, we observed chromosomal deletions of up to 45 kbp with a single 90mer oligo using the EcNR2 (recA+) strain as well as a recA- EcNR2 derivative, suggesting a recA-independent β-mediated mechanisms.

**Design of MAGE Oligonucleotides for DXP Pathway**

The main text of this paper described the design criteria that were implemented to optimize the DXP pathway for lycopene production. Here, we provide additional details to clarify our oligo design criteria. The design of every oligo was based on optimization experiments such that oligo length, concentration, stability, secondary structure, strand bias and modification were
optimal (Supplementary Table 1 and Supplementary Fig. 2). Two main oligo design strategies were implemented: 1) oligos with specified sequences produced specific changes by making targeted modification that knocked out the expression of target genes ($ytjC$, $fdhF$, $aceE$, $gdhA$) and 2) oligos with degenerate sequences produced diverse changes tailored for exploring a vast sequence space of RBS strengths. Importantly, in both oligo designs, the location of the genetic modification is precise and well-defined based on homology arms of the oligos. As described in the main text, the degenerate oligos were designed to mutate RBS sequences to be more similar to the canonical Shine-Dalgarno sequence (TAAGAGGT)$^6$, giving rise to enhanced translation efficiency. More specifically, RBS optimization utilized 90mer oligo pools containing the DDRRRRDDDDDD degeneracy at the 41-51 bp position of the oligos (D = G, A, T and R = G, A). This mutation region targeted the -4 through -14 positions from the start codon of each gene with an optimal RBS spacing of 5 bp for replacement by one of the oligos from the degenerate pool. We also calculated the cost and maximum level of degeneracy that can be introduced into a single oligo. For 30 USD we obtain 50 nmol yield of a 90mer oligo, giving us $3 \times 10^{16}$ molecules, which can support full degeneracy of 27 bp.

**MAGE Automation**

Automation instrumentation was constructed using the following major components:

- **Electroporator**: ECM 630  BTX Technologies Inc (MA, USA)
- **Digital controllers**: RS-232 serial modules  Superlogics Inc. (MA, USA)
- **Syringe pumps**: Cavro XLP600 9-port  Tecan Group Ltd. (NC, USA)
- **Solenoid valves**: Miniature Rocker Isolation Valves  Central Distribution Sales (NH, USA)
- **Temperature controller**: CNI-3233-C24  Omega Engineering Inc. (CT, USA)
- **Orbital shaker**: Advanced 3500 Orbital Shaker  VWR International LLC (PA, USA)
- **Control system software**: LabView  National Instruments (TX, USA)
- **Growth chamber system**: custom manufactured  David Breslau Design Inc. (NH, USA)


Supplementary Figure 1

Supplementary Figure 1 | Diagram of the oligo-genome hybridization structure during mismatch, insertion, and deletion modifications.
Supplementary Figure 2

Supplementary Figure 2 | Characterization of the allelic replacement frequency in the MAGE strain (EcNR2) and its derivative (EcFI5) by screening for the introduction of a nonsense mutation in the lacZ gene or recovery of the cmR gene. 

a, Replacement efficiency as a function of oligonucleotide length. Oligos contain two phosphorothioated bonds at both the 3’ and 5’ termini. 

b, Predicted optimal folding energy ∆G of 90-mer oligos as a function of frequency of replacement. Oligos with ∆G < -12.5 kcal/mol are considered to have significant secondary structure that hinder allelic replacement. 

c, The effect of terminal phosphorothioated bonds on replacement efficiency. Oligos of 90 bp with 0, 1(*) or 2 (**) phosphorothioated bonds between bases at the 3’, 5’ or both 3’ and 5’ termini were tested. 

d, Replacement efficiency as a function of the number of consecutive terminal 5’ phosphorothioated bonds of a 90mer oligo as measured by the introduction of a nonsense mutation in the lacZ gene. 

e, Replacement efficiency as a function of the concentration of 90-mer oligos containing no phosphorothioated bonds (in black) or 4 phosphorothioated bonds at the 5’ terminus (in red). Solid lines represent data fitted using polynomial functions. Error bars, ±SD.
Supplementary Figure 3 | Predicted distribution of genetic variants in a population that has undergone simultaneous allelic manipulation at 10 different genomic locations at 30% overall replacement efficiency. In this case, 10 different genes are simultaneously targeted for inactivation. Each colored solid line represents the histogramic distribution of variants containing different numbers of knockouts (KO) across the population as a function of MAGE cycle number. As MAGE cycles increase, population evolves towards acquiring all 10 gene KO's. The population is binomially distributed according to the equation:

\[
P(K, N) = \sum_{j=0}^{K} \binom{K}{j} (1 - M)^N (1 - (1 - M)^N)^j,
\]

where \( K \) is the number of loci simultaneously targeted, \( N \) is the number of MAGE cycles, and \( M \) is the mutation rate at any individual locus.
Supplementary Table 1

<table>
<thead>
<tr>
<th>MAGE Parameters</th>
<th>Optimal Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oligo length</td>
<td>90 bp</td>
</tr>
<tr>
<td>Oligo concentration range</td>
<td>0.05 – 50 µM</td>
</tr>
<tr>
<td>Oligo stability</td>
<td>Four 5’ phosphorothioated bases</td>
</tr>
<tr>
<td>Oligo secondary structure</td>
<td>&gt; –12.5 kcal/mol</td>
</tr>
<tr>
<td>Strand bias</td>
<td>Target lagging strand</td>
</tr>
<tr>
<td>Size of genetic modification</td>
<td>Predict efficiency using hybridization energy</td>
</tr>
<tr>
<td>Cycle time</td>
<td>2 – 2.5 hours</td>
</tr>
</tbody>
</table>

Supplementary Table 1 | Optimized parameters for maximal allelic replacement efficiency.
### Supplementary Table 2

<table>
<thead>
<tr>
<th></th>
<th>Strain</th>
<th>Wild-type RBS</th>
<th>Optimized RBS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EcHW2a</strong></td>
<td>(KO’s: none)</td>
<td>idi</td>
<td>acatgtgagaattatg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>optimized</td>
<td>ggaagggatgattatg</td>
</tr>
<tr>
<td><strong>EcHW2b</strong></td>
<td>(KO’s: ΔgdhA, ΔytjC)</td>
<td>dxs</td>
<td>ttaataggcccctgatg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>optimized</td>
<td>taggaatgtctgatg</td>
</tr>
<tr>
<td><strong>EcHW2c</strong></td>
<td>(KO’s: none)</td>
<td>dxs</td>
<td>ttaataggcccctgatg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>optimized</td>
<td>aaaaaggaagagctgatg</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>ispA</strong></td>
<td>cccggaatggtatgatg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>optimized</td>
<td>ggagaaggaagtaatg</td>
</tr>
<tr>
<td><strong>EcHW2d</strong></td>
<td>(KO’s: ΔfdhF)</td>
<td>dxs</td>
<td>ttaataggcccctgatg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>optimized</td>
<td>gtaaggaagagctgatg</td>
</tr>
<tr>
<td><strong>EcHW2e</strong></td>
<td>(KO’s: none)</td>
<td>dxs</td>
<td>ttaataggcccctgatg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>optimized</td>
<td>ggaagggaggaactgatg</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>idi</strong></td>
<td>acatggtgagaattatg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>optimized</td>
<td>ttaggaatgaaatattatg</td>
</tr>
<tr>
<td><strong>EcHW2f</strong></td>
<td>(KO’s: ΔytjC)</td>
<td>dxs</td>
<td>ttaataggcccctgatg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>optimized</td>
<td>taggaagagagctgatg</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>rpoS</strong></td>
<td>gtagagccaccttattatg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>optimized</td>
<td>gagaggatggacttattatg</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>idi</strong></td>
<td>acatggtgagaattatg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>optimized</td>
<td>aaaaaggttgattatg</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>dxr</strong></td>
<td>actctgtgatgatgatg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>optimized</td>
<td>ttaaggtgtattctatg</td>
</tr>
</tbody>
</table>

**Supplementary Table 2** | Optimized RBS sequences of strains EcHW2a-f
Supplementary Information

Suppl Fig 2a introduce bp mismatches

car_f_dslv
GGCTUGAGAAGGCTTTGAGTATACGCGTTACCCAACTTAATC

car_f_re50*
GGCTUGAGAAGGCTTTGAGTATACGCGTTACCCAACTTAATC

car_f_re70*
GGCTUGAGAAGGCTTTGAGTATACGCGTTACCCAACTTAATC

car_f_re70**
GGCTUGAGAAGGCTTTGAGTATACGCGTTACCCAACTTAATC

Suppl Fig 2b introduce 1 bp mismatch

goO_oligo
GGCTUGAGAAGGCTTTGAGTATACGCGTTACCCAACTTAATC

tpeO_oligo
GGCTUGAGAAGGCTTTGAGTATACGCGTTACCCAACTTAATC

Suppl Fig 2c&d introduce bp mismatches

alZ_oligo_m1_v1(5')
GCAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATC

Suppl Fig 2e introduce bp mismatches

alZ_oligo_m1_v1(5')
GCAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATC

GCAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATC

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT

G*G*A*A*ACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTT