CHAPTER EIGHTEEN

Multiplexed Genome Engineering and Genotyping Methods: Applications for Synthetic Biology and Metabolic Engineering

Harris H. Wang[*,†] *and* George M. Church[*,†]

Contents

Abstract

Engineering at the scale of whole genomes requires fundamentally new molecular biology tools. Recent advances in recombineering using synthetic oligonucleotides enable the rapid generation of mutants at high efficiency and specificity and can be implemented at the genome scale. With these techniques, libraries of mutants can be generated, from which individuals with functionally useful phenotypes can be isolated. Furthermore, populations of cells can be evolved *in situ* by directed evolution using complex pools of

* Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA
† Wyss Institute for Biologically Inspired Engineering, Harvard University, Massachusetts, USA

Methods in Enzymology, Volume 498
ISSN 0076-6879, DOI: 10.1016/B978-0-12-385120-8.00018-8

oligonucleotides. Here, we discuss ways to utilize these multiplexed genome engineering methods, with special emphasis on experimental design and implementation.

1. Introduction

Construction of genomes with highly engineered genetic components is a hallmark challenge and opportunity for synthetic biologists in the postgenomics era. Decreased cost and rising demand for DNA sequencing and oligonucleotide synthesis have created an entire service industry dedicated to reading and writing DNA material (Lipshutz *et al.*, 1999; Shendure and Ji, 2008). DNA synthesized *in vitro* is now used efficiently to modify genomes (Yu *et al.*, 2000; Zhang *et al.*, 1998), plasmids (Swaminathan *et al.*, 2001; Wang *et al.*, 2009b; Warming *et al.*, 2005), and phages (Marinelli *et al.*, 2008; Thomason *et al.*, 2009) of an expanding list of organisms (Shanks *et al.*, 2009; Swingle *et al.*, 2010; van Kessel *et al.*, 2008) using homologous recombination-based genetic engineering, or recombineering, techniques (Sharan *et al.*, 2009). Large libraries of DNA constructs can be combinatorial incorporated into the genome to test $>10^9$ genetic designs in a highly multiplexed fashion (Wang *et al.*, 2009a). These techniques present opportunities to create organisms with optimally engineered metabolic pathways, regulatory, and protein modules, as well as new genetic codes.

The λ-Red (Datsenko and Wanner, 2000) and the similar *rac*-encoded RecET (Muyrers *et al.*, 2004) homologous recombineering systems have been widely used to introduce genomic modifications into *Escherichia coli*. The λ-Red system is based on three essential proteins, Exo, Beta, and Gam from the λ-bacteriophage (Court *et al.*, 2002). Exo is a 5′ to 3′ exonuclease that digests linear double-stranded DNA (dsDNA), leaving 3′ overhangs that then act as substrates for subsequent recombination events. Beta is a single-stranded DNA (ssDNA) binding protein that facilitates recombination via hybridization of the linear fragment to its genomic complement. Gam acts to inhibit RecBCD activity *in vivo* to protect the degradation of foreign linear dsDNA fragments. Heterologous expressions of other λ-Red protein homologs also lead to increased recombinagenicity in *E. coli*, suggesting the universality of this mode of genome integration (Datta *et al.*, 2008). Numerous other modified λ-Red constructs have been described and are reviewed elsewhere (Datta *et al.*, 2006; Sawitzke *et al.*, 2007).

Both ssDNA and dsDNA can be used with the λ-Red system to insert novel genetic sequences, introduce mismatches, or delete genes. In dsDNA-based recombineering, which requires Exo, Beta, and Gam, a linear dsDNA cassette with at least 50 bps of flanking homology to the target site is used. The efficiency of double-stranded homologous

recombination can be as high as 0.01% among cells that survive transformation. Isolation of cells harboring a cassette with a selectable phenotype (i.e., antibiotic resistance) is done easily on agar plates to obtain modified mutants at >95% efficiency using a strong selection.

In ssDNA-based recombineering where only Beta is required, the ssDNA integrates into the genome most efficiently by hybridizing to the exposed lagging strand at the replication fork (Wu *et al.*, 2005; Yu *et al.*, 2003). This manner of integration appears to mimic that of an Okazaki fragment of replicating DNA. Recent evidences suggest that linear dsDNA may be completely transformed into a ssDNA intermediate prior to integration into the genome (Maresca *et al.*, 2010; Mosberg *et al.*, 2010). The leading strand can also be targeted with ssDNA, but albeit at a 10- to 100-fold lower efficiency than for the lagging strand (Ellis *et al.*, 2001). The incorporation efficiency is highest for ssDNA in the 70–90 bps range, but can be as short as 30 bps, which is the minimum binding size for Beta (Erler *et al.*, 2009). The efficiency of ssDNA-based recombineering can be as high as 25% among cells that survive transformation when the native mismatch repair system is evaded (Costantino and Court, 2003). Based on these advances, a cyclical and shotgun approach called Multiplex Automated Genome Engineering (MAGE) was developed to simultaneously introduce many chromosomal changes in a combinatorial fashion across a population of cells to generate up to 4 billion genetic variants per day (Wang *et al.*, 2009a). This rapid chromosomal engineering method offers the opportunity to construct both highly modified genomes and explore large sequence landscapes by directed evolution in a semirational fashion. The general MAGE process (Fig. 18.1) will be detailed extensively in the sections below to provide a useful guide for designing and performing MAGE experiments. While the potential of MAGE is fully realized through automated instrumentations, they are not necessarily required to perform the MAGE protocols described here.

1.1. Iterative engineering of a single chromosomal site

The first aspect of MAGE is the iterative application of the ssDNA (or oligo) recombineering protocol on a cell population without the intermediate step of colony isolation for genotyping or phenotyping. While the efficiency of replacing the chromosomal alleles with synthetic oligonucleotides may be high in certain instances (e.g., 1-bp mismatches), the efficiency decreases markedly with increase in size of the replacement. To overcome low efficiency, the oligo-recombineering protocol is iterated on the same cell population over multiple cycles using the same oligo species. In this fashion, the population is enriched for mutants containing the desired sequence conversions. Typically, each full cycle takes ~2–3 h depending on the growth rate of the cells. The relative abundance of mutants in the

population M can be approximated by $M = 1 - (1 - RE)^N$, where N is the number of cycles and RE is the allelic replacement efficiency per cycle. RE is highly dependent on the type of target conversion (mismatch, insertion, deletion) and the size of the conversion. General exponential decay functions of empirically determined RE are shown in Table 18.1.

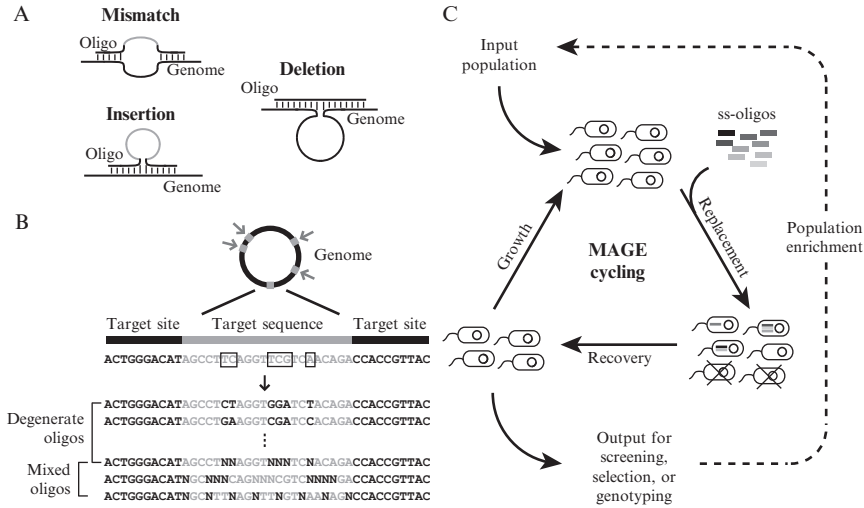

Figure 18.1 (A) Recombineering can be used to generate mismatches, insertions, and deletions up to 30 or more bps using a 90 bps oligonucleotide. Larger deletions (kbs) can be achieved at lower efficiency ($<10^{-3}$). (B) Many targets can be multiplexed in the same recombineering reaction using degenerate or mixed oligo pools. (C) General schematic of MAGE process with input population being continually cycled with MAGE. Subpopulations can be removed for assay by genotyping or phenotyping and used as enrichment inoculum for subsequent MAGE cycles.

Table 18.1 Allelic replacement efficiency prediction function based on fitting empirically determined efficiencies from Wang *et al.* (2009a) where *b* is the base-pair size of the modification

Replacement type	Replacement size in base-pairs (*b*)	Multiplier[a] (RE_0)	Predicted replacement efficiency (RE)
Mismatch	$b = 1$ to 30 bp	$RE_0 = 0.26$	$RE = RE_0 \times e\hat{}(-0.135(b-1))$
Insertion	$b = 1$ to 30 bp	$RE_0 = 0.15$	$RE = RE_0 \times e\hat{}(-0.075(b-1))$
Deletion	$b = 1$ to 30 bp	$RE_0 = 0.23$	$RE = RE_0 \times e\hat{}(-0.058(b-1))$

[a] The multiplier RE_0 may vary depending on the local contextual features of the target chromosomal site and the formation of secondary structures by the 90-bp oligonucleotide.

Thus, the relative abundance of desired mutants in the population can be easily estimated by defining the number of iterative cycles and the size and type of the desired mutation.

Often time, the required number of cycles is dictated by the throughput of the genetic screen. Genetic screens can be in the form of direct genotypic methods such as PCR or DNA sequencing, or phenotypic screening or selection methods such as colorimetry, growth rate, or antibiotic resistance. The number of cycles N needed to produce mutation size of b base-pairs at a frequency of at least F in the population can be estimated by

$$N = \log(1 - F)/\log(1 - \mathrm{RE}). \tag{18.1}$$

For example, the number of cycles needed to generate mutants with a 6 bp chromosomal mismatch to a frequency of 0.25 (i.e., 25%) in the population with an oligo folding energy of −5.4 kcal/mol (predicted through MFold; Markham and Zuker, 2005) is $N = \log(1 - 0.25)/\log(1 - 0.26 \times e^{-0.135 \times 5}) = 2.0$ cycles, and to a frequency of 0.50 (i.e., 50%) is $N = 4.9$ cycles. Thus, one would expect from a PCR screen that at least one in four cells would show conversion after two cycles and one in two would show conversion after five cycles of oligo-recombineering.

Another useful application is the generation of a large number of variants at one particular genomic site, such as to make promoter or ribosomal binding site (RBS) variants or to mutagenize the active site of an enzyme. Using oligos with the same flanking homology arms but different mutation sequences, the same chromosome site can be targeted across all cells in the population. At every MAGE cycle, the conversion frequency of the population to a new mutant genotype is determined by RE. For example, to introduce a 7-bp consecutive or nonconsecutive mismatch to a promoter region (RE = 0.1), we could potentially generate 10^8 promoter variants in a population of 10^9 cells (a typical MAGE population size) every cycle. In this example, the actual oligo pool complexity is $4^7 = 16,384$, so on average each variant is found in 6100 cells in the population after each cycle. After one cycle, however, 90% of the cells in the population still contain the wild-type promoter sequence. Iterative cycling of the same population with the degenerate oligo pool will reduce the abundance of the wild-type sequence, which is $(1 - \mathrm{RE})^N$. For high oligo pool complexities ($> 10^9$), the population should be cycled multiple times to generate all possible variants. It is important to note that because the population is constantly changing after each MAGE cycle, the total sequence space that can be explored is much greater than the carrying capacity (10^9) of the cycled population at any cycle. Therefore, the number of variants generated is dependent on the number of MAGE cycles. This feature of MAGE can be especially useful when simultaneously targeting different chromosomal sites, discussed in Section 1.2.

1.2. Multiplexed engineering of multiple chromosomal sites

In Section 1.1, we described how to assess MAGE cycling to target one chromosomal site. More frequently, one would want to simultaneously target multiple chromosomal sites. Several advantages arise by multiplexing. First, many different variants can be combinatorially generated and screened/selected all at once from a single population. Second, the mechanism of oligo-mediated allelic replacement allows multiple sites to be simultaneously converted during each MAGE cycle. For this shotgun approach, a mixed pool of oligo species that target different chromosomal sites is used. Multiplex engineering of up to 40 chromosomal sites can be easily done, while at higher pool diversity (100s–1000s of different species) oligo–oligo interactions may potentially begin to inhibit the reaction.

Simultaneous allelic manipulation of $k \geq 1$ different genomic locations, each with an average efficiency of replacement of $\mathrm{RE_{av}}$, can be modeled as a binomial process, assuming that replacement operates independently across all loci (no linkage association). Here, the probability of replacement at any one location is $p_N = 1 - (1 - \mathrm{RE_{av}})^N$, and the probability of finding exactly m variants is $P(m \text{ variants}) = \binom{k}{m} p_N^m (1 - p_N)^{k-m}$. Under typical conditions, this will be well approximated by the Gaussian distribution

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-(x-\mu)^2/2\sigma^2} \tag{18.2}$$

where the average number of mutations is $\mu = k(1 - (1 - \mathrm{RE_{av}})^N)$ and the variance of the distribution is $\sigma^2 = k(1 - \mathrm{RE_{av}})^N(1 - (1 - \mathrm{RE_{av}})^N)$. To estimate the frequency with which one can find cells with at least m mutations after N cycles, the Standard Normal Table or the Gaussian error function can be used to estimate the size of the tail to the right of m using the mean and variance above. To determine the number of cycles N needed to produce m mutants at a particular abundance in the population, we need to analyze $\mu + Z\sigma$, which is

$$m = k\left(1 - (1 - \mathrm{RE_{av}})^N\right) + Z\sqrt{k(1 - \mathrm{RE_{av}})^N\left(1 - (1 - \mathrm{RE_{av}})^N\right)} \tag{18.3}$$

where Z is the Z-score based on the Standard Normal Table. If one finds the m for which the tail size is $\geq 1/20$ (5%) of the entire distribution, one will on average find one cell among 20 in which there are at least m mutations. For a standard Gaussian distribution, the point at which the right tail of the distribution is 5% of the whole occurs at $Z = 1.645$. Therefore, using the Gaussian approximation, the value of m that meets this condition is estimated by $\mu + 1.645\sigma$.

We further illustrate these calculations in Table 18.2 for $N = 5$, 10, and 20 for a situation in which oligos are multiplexed to introduce Stop codon nonsense mutations to 10 target genes to knockout function ($k = 10$). Here, the overall RE is 0.26 and we assume RE per locus is $\mathrm{RE}_{\mathrm{av}} = \mathrm{RE}/k = 0.26/10 = 0.026$ because of the shared 10-plex oligo pool. This illustration shows how m increases with N. We find that five MAGE cycles ($N = 5$) would be sufficient to produce mutants with at least 2.9 knockouts ($m \geq 2.9$) at an abundance of 5% in the population (corresponding to $Z = 1.645$). Twenty cycles would be sufficient to enrich for mutants with at least 6.7 knockouts at the same abundance of 5% (also illustrated in Fig. 18.2). Note that a tail size of 1/20 or 5% means that one can have 95% confidence of finding a cell with at least m mutations among 59 cells as determined by P (not finding an m mutant among s cells) or $(1 - 0.05)^s < 0.05$, which implies $s > \log(0.05)/\log(0.95)$ or $s > 58.4$. Methods to screen for these mutants are discussed later.

Each locus in a multilocus-targeting reaction can also be multiplexed. For example, cells with multiple promoter variants for each gene of a multicomponent pathway can be combinatorially generated in the population. A mixture of knockouts, RBS changes, promoter modulation, and protein coding sequence modifications can be multiplexed through a single oligo pool. Economically, the cost of generating oligonucleotides with degenerate sequences by column-based DNA synthesis is same as the cost of generating oligo of a specific sequence. Coupled with automation

Table 18.2 A list of variables to consider for a 10-target MAGE reaction to introduce single base-pair mutations ($\mathrm{RE}_{\mathrm{av}} = \mathrm{RE}/k = 0.026$) as a function of the number of MAGE cycles, with p_N = abundance level of each of the 10 target locus, μ = average number of accumulated mutations in each cell, σ = variance of the mutations, and m = the number of mutations in the top 5%, 2%, and 1% of cells in the population

Number of MAGE cycles (N)		5	10	20
$p_N = 1 - (1 - \mathrm{RE}_{\mathrm{av}})^N = 1 - (1 - 0.026)^N$		0.12	0.23	0.41
$\mu = kp_N = 10p_N$		1.23	2.31	4.10
$\sigma = \sqrt{kp_N(1 - p_N)} = \sqrt{10p_N(1 - p_N)}$		1.04	1.33	1.56
Top 5% clones	$m = \mu + 1.645\sigma$	2.9	4.5	6.7
	95% screening confidence	59 cells to screen		
Top 2% clones	$m = \mu + 2.054\sigma$	3.4	5.1	7.3
	95% screening confidence	149 cells to screen		
Top 1% clones	$m = \mu + 2.326\sigma$	3.7	5.4	7.7
	95% screening confidence	298 cells to screen		

The number of cells need to isolate a clone with at least m mutations is provided at 95% confidence.

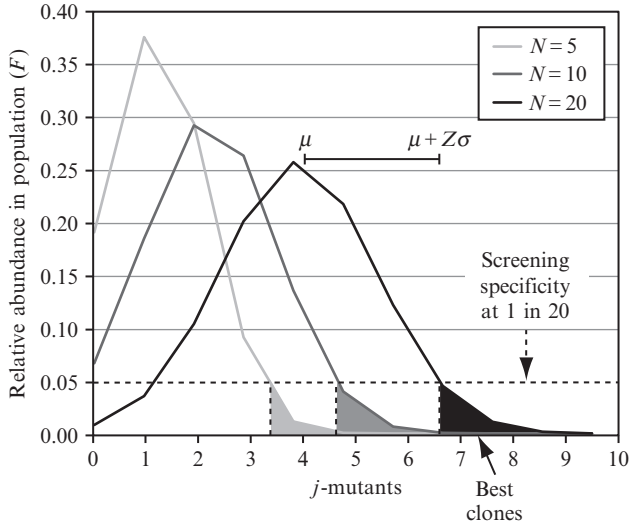

Figure 18.2 Relative abundance of cells containing j mutations in the population, where $0 \leq j \leq 10$, $k = 10$, and $RE_{av} = RE/k = 0.026$. For a cumulative distribution value of 0.95 (ability to screen and identify mutation in the top 5% of clones), the Z-score is 1.645, given a likely isolation of clones containing 6–7 mutations after 20 MAGE cycles.

systems to continuously cycle population of cells, MAGE holds the potential to turn genome engineering from a laboratory-based method to a scalable platform comparable in scale and throughput as large modern day DNA synthesis and sequencing services.

2. Design Protocol

2.1. Oligonucleotides: Design and procurement

Lagging strand targeting: Oligonucleotides should be designed to target the lagging strand of replicating DNA (Fig. 18.3a). Since replication in *E. coli* is bidirectional, care should be taken to ensure that the oligo sequence designed targets the lagging strand. The origin of replication (*oriC*) in *E. coli* is located at positions 3923767–3923998 (Blattner *et al.*, 1997) and the *dif* terminus is at 1588774–1588801. If the target chromosomal position is on replichore 1 (>3923998 or <1588774), then the oligo sequence should be the *complementary sequence* to the (+) strand sequence. If target chromosomal locus is on replichore 2 (>1588774 and <3923998), then the oligo sequence should be the *same sequence* as the (+) strand sequence (i.e., the *complementary sequence* to the (−) strand).

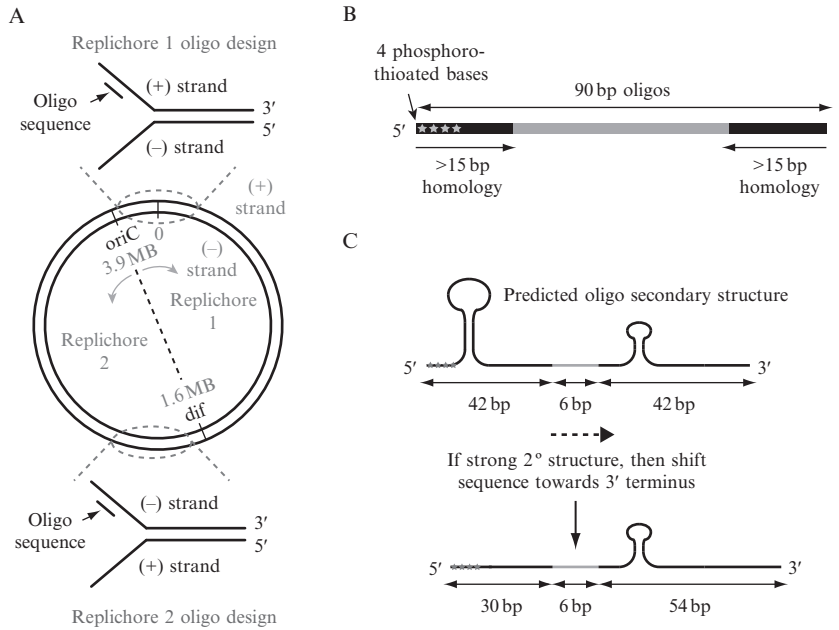

Figure 18.3 Optimal oligonucleotide design. (A) Design of oligos to target the lagging strand, based on the location of the target site on the chromosome. (B) Optimally efficient oligos should be 90 bps with at least 15 bps of homologous sequences to the target region on both the 5′ and 3′ ends. The target mutation sequence should be placed at the center of the oligo whenever possible. Four phosphorothioated bases should be used at the 5′ terminus of the oligo to reduce its degradation rate *in vivo*. (C) The secondary structure of the oligo should be assessed using MFold (Markham and Zuker, 2005) or other folding prediction algorithm. If folding energy $\Delta G_{ss} < -12.5$ kcal/mol, redesign oligo by shifting sequence toward the 3′ or 5′ terminus (3′ is preferred due to higher error rates of 5′ sequences during oligo synthesis). A minimum of 15 bps of homology should be left at the ends. Mutations nearing the termini are less frequently incorporated.

Secondary structure optimization: Oligonucleotides can often form hairpin structures that inhibit the allelic replacement because the homology arms are not available for hybridization. In general, we recommend ensuring that the oligo design has a folding energy that is no less than − 12.5 kcal/mol as predicted by MFold (Markham and Zuker, 2005) on default values. If the folding energy reaches this prohibitive value, the oligo can be redesigned by shifting the mutation site toward the 3′ terminus of the oligo thereby potentially disrupting the local hairpin structures (Fig. 18.3c). In general, a shift to the 3′ end is more desired than a shift to the 5′ end. Because oligo synthesis is 3′ to 5′, the likelihood of retaining the mutation sequence is higher toward the 3′ end, where truncations and errors are less prevalent. At least 15 bps of homology should be left on each end of the oligo as

mutations at the distant arms are less likely to be incorporated into the chromosome due to chew back of the oligo ends during integration (H.H. Wang, unpublished results).

Mismatch repair evasion: When active, the mismatch repair (MMR) machinery converts mutations generated by the oligos back to the wild-type sequence. To avoid reversion, the EcNR2 or EcHW24 strain can be used where the MMR system is inactivated through a *mutS* knockout. The drawback of this approach is the higher background mutation rate (10^{-8}) of a Δ*mutS* strain versus the wild type (10^{-10}). Alternatively, in the presence of MMR, incorporation of silent mutations near the mutation site that are poorly recognized by *mutS* (e.g., C–C pairs) can increase efficiency (Costantino and Court, 2003). Furthermore, utilization of modified bases not recognized by *mutS* can also increase efficiency in the presence of an intact MMR system (Wang *et al.*, 2011c). Since, the MMR can only repair short segments of mutations (< 6 bps), large mutations are also naturally avoided.

Synthetic oligonucleotides: In general, 90 bps oligos appear to produce the highest allelic replacement efficiency. Longer oligos tend to form more inhibitory hairpin structures and are more costly to synthesize. Shorter oligos are less efficient due to lower hybridization energy to the chromosomal target. Up to four phosphorothioated bases should be used at the 5′ terminus of the oligo to prevent exonuclease degradation of the oligos inside the cell (Fig. 18.3b). Absence of phosphorothioation protection can lead to a two- to threefold decrease in efficiency. For most applications, oligos with standard purification should suffice although in certain applications PAGE/HPLC purified oligos may be needed. Typically, oligos can be obtained through a commercial oligonucleotide synthesis vendor in 2–3 days (e.g., Integrated DNA Technologies, USA).

2.2. Designing appropriately scoped MAGE experiments

One needs to weigh several factors when determining the scope of a MAGE experiment. First, the size of the mutations determines the allelic replacement efficiency. Second the number of targets determines the complexity of the oligo pool. These two factors affect the overall number of MAGE cycles that will be required. Third, the throughput of the genotyping/phenotyping method affects the degree to which mutants in the population need to be enriched before they can be successfully isolated. In general, 100–200 colonies can be easily screened by multiplex allele specific PCR to query 10–20 target alleles simultaneously. Increased screening capacity decreases the number of MAGE cycles required for mutant enrichment.

Detailed example: Let us design an experiment in which we will attempt to explore 64 RBS variants xxxNNNxxxATG upstream of 5 genes of a biosynthesis pathway to tune gene expression. Our genotyping/phenotyping method is by plating on agar and observing a colorimetric indicator change. We can therefore distinguish one mutant from plate of 1000 cells. Using

Table 18.1, we first estimate the allelic replacement efficiency RE of this 3 bp mismatch to be at $RE = 0.26 \times e^{-0.135(3-1)} = 0.198$ or 19.8%. To produce mutants in >50% of our population, we use Eq. (18.1) to find that we require $N = \log(1 - 0.5)/\log(1 - 0.198) = 3$ cycles. The full complexity of the oligo pool is $64^5 = 1.07 \times 10^9$, therefore we will have to explore about half of these sequences after three MAGE cycles. So, we want to isolate mutants that contain at least m out of the five possible RBS locations on our indicator plate. Our cumulative distribution value is $(1 - 1/1000) = 0.999$, so our Z-score is 3.08. Using Eq. (18.3), we find that with 10 cycles, $m = 5\left(1 - (1 - 0.198/5)^{10}\right) + 3.08\sqrt{5(1 - 0.198/5)^{10}\left(1 - (1 - 0.198/5)^{10}\right)} = 3.4$, meaning that 1 in 1000 cells that are found on the plate will contain mutations in three or four RBS positions out of a possible five. Under these considerations, 10–15 MAGE cycles may be required in total to fully explore the sequence space to produce successful mutants.

2.3. Primer design for multiplex allele-specific colony (MASC) PCR

Allele-specific PCR can be used to directly query genotypes. Two forward primers are designed for each query target, a forward primer that is specific only to the wild-type sequence *primer_f(wt)* and another only to the mutant sequence *primer_f(mut)*. Both forward primers share the same reverse primer (*primer_r*). The specificity is designed into the 3′ terminus of the forward primer. For single nucleotide polymorphism (SNP) detection, the 3′ base is either the wild-type base or the mutant base. Thus, for a colony containing the wild-type SNP, the *f(wt)/r* and not the *f(mut)/r* primer pair should amplify a PCR product. Conversely, for a colony containing the mutant SNP, the *f(mut)/r* and not the *f(wt)/r* primer pair should amplify a PCR product. This reaction can be multiplexed across more than 10 sites in a single PCR by designing primer pairs with amplicons of different lengths to distinguish each SNP. In general, we advise designing primer pairs that amplify at sizes of 100, 150, 200, 250, 300, 400, 500, 600, 700, 850 bps, which can produce clearly distinguishable bands on a 1.5% agarose gel. The primer pairs should be designed to a T_m of 62 °C (Zhang), but the actual T_m of the MASC-PCR is determined through a gradient PCR.

3. Experimental Protocol

3.1. Strains and media

The protocol described here is optimized for *E. coli* MG1655 derivatives EcNR1, EcNR2, and EcHW24 (Wang *et al.*, 2009a). EcNR1 contains a chromosomally integrated λ-prophage construct (based on DY330

(Yu *et al.*, 2000)) fused to the *bla* gene for ampicillin resistance. The λ-Red construct (containing *exo*, *beta*, and *gam*) is integrated at the *bioA/bioB* gene locus and is temperature inducible by brief heat shock at 42 °C. EcNR2 is an EcNR1 derivative with Δ*mutS*::*cat* for chloramphenicol resistance. EcHW24 is an EcNR1 derivative with inactivated *mutS* by Stop codon mutations at amino acids 189 and 191. All strains must be grown at 30–32 °C. For rich media, LB-Lennox (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl) is used with the appropriate antibiotics, chloramphenicol (cat), kanamycin (kan), or carbenicillin (carb) at concentrations of 20, 30, or 50 μg/mL, respectively. Standard M9 minimal media supplemented with D-biotin (0.25 μg/mL) can also be used.

3.2. Supplies/reagents

- Rotator drum in 30–32 °C incubator
- Shaking water bath at 42 °C
- Ice bucket with ice/water mixture
- Distilled sterile water (chilled)
- Synthetic oligonucleotides (in 50 μL dH_2O at 0.05–50 μ*M*, chilled)
- Microcentrifuge tubes (chilled)
- 1-mL and 200-μL pipettes and pipette tips (chilled)
- Tabletop centrifuge (at 4 °C)
- Electroporation system and electroporation cuvettes or plates
- Glass culture tubes with prewarmed LB-Lennox.

3.3. MAGE cycling

In the receding day, streak out the appropriate strain (e.g., EcNR2) on agar plate and allow colonies to grow overnight at 32 °C.

If initiating a new MAGE experiment:

Step 1. Pick a colony into a glass tube with 3 mL of LB-Lennox media and place in a rotator drum spun at 300 rpm in a 32 °C incubator.

If continuing from a previously paused MAGE cycle:

Step 1. Take 100 μL of the overnight MAGE cell culture from 32 °C incubator or the 4 °C storage and dilute into a glass tube with 3 mL of LB-Lennox media and place in a rotator drum spun at 300 rpm in a 32 °C incubator. This step ensures that the cells, which are in stationary phase, can recover back into exponential phase growth. The remaining cell culture can be stored or discarded.

Step 2. Once cells have reached mid-exponential growth phase as determined by $OD_{600\ nm}$ of 0.6–0.7, place the culture tube in a 42 °C

shaking water bath for 15 min. This step ensures that the λ-Red system is properly induced. Lengthening the temperature induction is undesirable as the Gam protein is highly toxic to the cells when expressed for >20 min.

Step 3. After the 15-min 42 °C induction, immediately place cells in ice-water bucket and cool by gentle swirling for 30–60 s. Induced cells can stay on ice for up to 3 h prior to the next step.

Step 4. Making electrocompetent cells *(this step should be done at 4 °C if possible)*:

- **a.** Place 1 mL of culture in prechilled 1.5 mL microcentrifuge tubes. Spin tube in tabletop centrifuge at 13,000×*g* for 30 s. The remaining 2-mL culture can be discarded or frozen in 15% glycerol at −80 °C for future assays.
- **b.** Remove LB supernatant from tube and resuspend pellet in 1 mL of prechilled sterilized distilled H_2O by pipetting up and down, do not vortex the cells.
- **c.** Spin tube in tabletop centrifuge at 13,000×*g* for 30 s.
- **d.** Remove H_2O supernatant from tube and resuspend pellet in another 1 mL of prechilled dH_2O by pipetting up and down, do not vortex the cells.
- **e.** Spin tube in tabletop centrifuge at 13,000×*g* for 30 s.
- **f.** Remove H_2O supernatant from tube and add 50 μL of oligos to the pellet. A maximum amount of oligos that can be added to the cell pellet without arcing the electroporation reaction is 140 μg (50 μ*M*). Typically, 2 μ*M* in 50 μL is used. The lower range is 0.05 μ*M* while still producing detectable replacement efficiency. For highly complex oligo pools, a 20 μ*M* total oligo concentration is advised to ensure higher numbers of oligos are reaching each cell.
- **g.** Place cell–oligo mixture in a prechilled 1 mm gap electroporation cuvette.

Step 5. Remove cuvette from ice and dry sides with a paper towel prior to electroporation. Transform oligos into the cells by electroporation using a standard electroporation pulse generator (i.e., Bio-Rad MicroPulser, BTX ECM-830). For a 1 mm gap cuvette, use settings: 1.8 kV, 200 Ω, 25 μF. For a 2 mm gap cuvette use settings: 2.5 kV, 200 Ω, 25 μF. Time constant for the electroporation should be >4.0 ms.

Step 6. After electroporation, immediately add 1 mL of LB-Lennox to the cuvette and transfer to a glass tube containing 2 mL of LB-Lennox, resulting in the standard 3 mL growth volume.

Step 7. Allow cells to recover and grow back into midexponential growth phase. The bulk of the MAGE cycle time is dominated by this posttransformation recovery phase, which is where the oligos are

being incorporated into the chromosome. An adequate number of cell divisions (>4) are required for segregation of the mutant allele, which may take 2–3 h depending on the growth rate. Furthermore, only 1–5% of the cells in the population survive electroporation. Therefore, outgrowth after transformation is required to repopulate the culture to the appropriate density. The end of this recovery phase marks the end of one MAGE cycle.

a. If continuing MAGE cycles, go to Step 2 and wait for $OD_{600\ nm}$ to reach 0.6–0.7.

b. If pausing MAGE cycles, continue to grow the culture into stationary phase. Keep in 32 °C for <1 day storage and keep in 4 °C for ≥ 1 day storage. Paused cultures can be restarted by 1:30 dilution into fresh LB-Lennox (see Step 1)

c. If cultures are being plated for colony isolation, recover for at least 3 h to allow all recombinant genomes to segregate prior to plating.

In general, 3–4 MAGE cycles can be carried through per day. Multiple independent cultures can be run simultaneously. Up to 48 cultures can potentially be cycled at a time using 2.2 mL 96-well plates, 8-channel multichannel pipettes, and 96-well electroporation plates and pulse generators (BTX ECM 830, Model 45-0421 or Lonza Nucleofector 96-well Shuttle System).

3.4. Genotyping by multiplex allele specific colony PCR verification

A gradient MASC-PCR should first be run to determine the optimal T_m for the rest of the PCR. The multiplex primer mix should contain primers (up to 20) at individual primer concentration of 0.2 μ*M*. Two separate PCRs should be run, one containing *f*(*wt*)/*r* and the other containing *f*(*mut*)/*r*. An optimized multiplex PCR kit is recommended (Qiagen Cat #206143). The PCR is highly sensitive to template concentration. In general, using 1 μL of a 1:100 dH_2O dilution of a saturated culture or a single colony is recommended. The gradient PCR optimized melting temperature, T_m, is best used for the specific dilution and may need to be repeated for other dilutions or template preparations. Generally a gradient T_m ranging from 61 to 69 °C is used although finer ranges are also acceptable. An example of gel containing a gradient MASC-PCR is shown in Fig. 18.4. To choose the optimal T_m, we want to ensure that all bands can be adequately amplified, and that there is binary specificity of the *f*(*wt*)/*r* and *f*(*mut*)/*r*.

The MASC-PCR cycles are as follows (using a Taq polymerase):

Step 1: 95 °C for 15 min
Step 2: 94 °C for 30 s

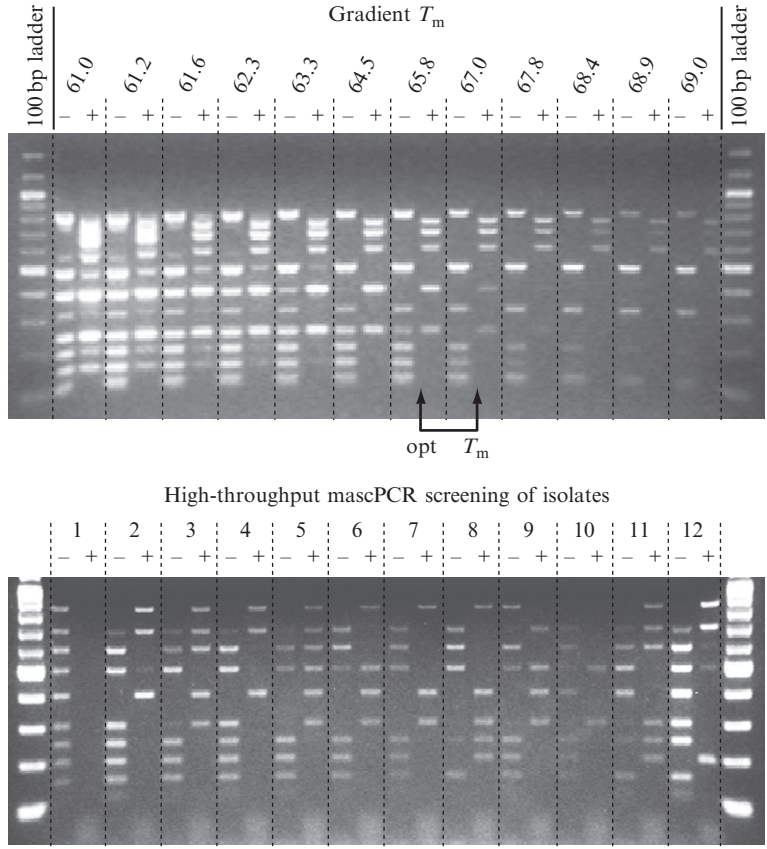

Figure 18.4 Example of a gradient MASC-PCR (top gel). Symbol (−) denotes amplification of the template with the *f(wt)/r* primer set and (+) with f(*mut*)/r primer set. The optimal melting temperature T_m is chosen based on high specificity (i.e., either *f(wt)/r* or *f(mut)/r* primer set amplify) and strong signal (i.e., visible bands). Here, we determined that 65.8 °C < optimal T_m < 67.0 °C (denoted by arrows). A large number of colonies can be screened directly using MASC-PCR to isolate variants generated combinatorially (bottom gel). Here, all clones except #2 and #4 have unique combinations of 10 targeted mutations.

Step 3: 61–69 °C (gradient) or T_m (optimal) for 30 s
Step 4: 72 °C for 80 s
Step 5: go to Step 2 for 26 times
Step 6: 72 °C for 5 min
Step 7: 4 °C for forever

Generally, 20 μL PCRs are suggested. Xylene cyanol loading dye is added to each reaction (high molecule weight dye to not interfere with <1000 bp bands) and 10 μL is loaded to each lane of an ethidium

bromide-strained 1.5% agarose gel. The gel is run by electrophoresis at 180 V for 60–70 min and analyzed subsequently on a Gel Documentation System. To increase the throughput of PCR screening, utilization of 96-well PCR blocks, 200-lane gel electrophoresis setups (e.g., BioRad Subcell Model 192), and multichannel pipetting and gel loading is highly recommended.

4. Concluding Remarks

Recombineering-based genome engineering provides a powerful approach for constructing and modifying chromosomes synthetically. As the cost of oligonucleotide synthesis continues to drop and automation capacities continue to expand, efficient "on-the-fly" manipulation of a living organism's genome will continue to improve. With the MAGE platform, existing genomic templates are used as scaffolds to produce newly engineered variants. An important aspect of template-based genome engineering is the benefit from the natural selection process as new genomes evolve by directed steps from existing functional genomes. Genome engineering approaches coupled with *de novo* synthesis methods (Chan *et al.*, 2005; Gibson *et al.*, 2010; Menzella *et al.*, 2005; Tian *et al.*, 2004) will continue to offer an expanding capability to engineer living organisms at the resolution of single nucleotides, but scaled across the entire genome and beyond.

ACKNOWLEDGMENTS

The authors wish to thank John Aach for helpful discussions and careful reading of the manuscript. This work was funded by the Wyss Institute for Biologically Inspired Engineering, the National Science Foundation, the U.S. Department of Energy, and the Defense Advanced Research Project Agency.

REFERENCES

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., *et al.* (1997). The complete genome sequence of Escherichia coli K-12. *Science* **277,** 1453–1462.

Chan, L. Y., Kosuri, S., and Endy, D. (2005). Refactoring bacteriophage T7. *Mol. Syst. Biol.* **1**(2005), 0018.

Costantino, N., and Court, D. L. (2003). Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc. Natl. Acad. Sci. USA* **100,** 15748–15753.

Court, D. L., Sawitzke, J. A., and Thomason, L. C. (2002). Genetic engineering using homologous recombination. *Annu. Rev. Genet.* **36,** 361–388.

Datsenko, K. A., and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **97,** 6640–6645.

Datta, S., Costantino, N., and Court, D. L. (2006). A set of recombineering plasmids for gram-negative bacteria. *Gene* **379,** 109–115.

Datta, S., Costantino, N., Zhou, X., and Court, D. L. (2008). Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. *Proc. Natl. Acad. Sci. USA* **105,** 1626–1631.

Ellis, H. M., Yu, D., DiTizio, T., and Court, D. L. (2001). High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl. Acad. Sci. USA* **98,** 6742–6746.

Erler, A., Wegmann, S., Elie-Caille, C., Bradshaw, C. R., Maresca, M., Seidel, R., Habermann, B., Muller, D. J., and Stewart, A. F. (2009). Conformational adaptability of Redbeta during DNA annealing and implications for its structural relationship with Rad52. *J. Mol. Biol.* **391,** 586–598.

Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie, M. M., *et al.* (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329,** 52–56.

Lipshutz, R. J., Fodor, S. P., Gingeras, T. R., and Lockhart, D. J. (1999). High density synthetic oligonucleotide arrays. *Nat. Genet.* **21,** 20–24.

Maresca, M., Erler, A., Fu, J., Friedrich, A., Zhang, Y., and Stewart, A. F. (2010). Single-stranded heteroduplex intermediates in lambda Red homologous recombination. *BMC Mol. Biol.* **11,** 54.

Marinelli, L. J., Piuri, M., Swigonova, Z., Balachandran, A., Oldfield, L. M., van Kessel, J. C., and Hatfull, G. F. (2008). BRED: A simple and powerful tool for constructing mutant and recombinant bacteriophage genomes. *PLoS ONE* **3,** e3957.

Markham, N. R., and Zuker, M. (2005). DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res.* **33,** W577–W581.

Menzella, H. G., Reid, R., Carney, J. R., Chandran, S. S., Reisinger, S. J., Patel, K. G., Hopwood, D. A., and Santi, D. V. (2005). Combinatorial polyketide biosynthesis by *de novo* design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* **23,** 1171–1176.

Mosberg, J. A., Lajoie, M. J., and Church, G. M. (2010). Lambda Red recombination in Escherichia coli occurs through a fully single-stranded intermediate. *Genetics*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20813883.

Muyrers, J. P., Zhang, Y., Benes, V., Testa, G., Rientjes, J. M., and Stewart, A. F. (2004). ET recombination: DNA engineering using homologous recombination in E. coli. *Methods Mol. Biol.* **256,** 107–121.

Sawitzke, J. A., Thomason, L. C., Costantino, N., Bubunenko, M., Datta, S., and Court, D. L. (2007). Recombineering: *In vivo* genetic engineering in E. coli, S. enterica, and beyond. *Meth. Enzymol.* **421,** 171–199.

Shanks, R. M., Kadouri, D. E., MacEachran, D. P., and O'Toole, G. A. (2009). New yeast recombineering tools for bacteria. *Plasmid* **62,** 88–97.

Sharan, S. K., Thomason, L. C., Kuznetsov, S. G., and Court, D. L. (2009). Recombineering: A homologous recombination-based method of genetic engineering. *Nat. Protoc.* **4,** 206–223.

Shendure, J., and Ji, H. (2008). Next-generation DNA sequencing. *Nat. Biotechnol.* **26,** 1135–1145.

Swaminathan, S., Ellis, H. M., Waters, L. S., Yu, D., Lee, E. C., Court, D. L., and Sharan, S. K. (2001). Rapid engineering of bacterial artificial chromosomes using oligonucleotides. *Genesis* **29,** 14–21.

Swingle, B., Markel, E., Costantino, N., Bubunenko, M. G., Cartinhour, S., and Court, D. L. (2010). Oligonucleotide recombination in Gram-negative bacteria. *Mol. Microbiol.* **75,** 138–148.

Thomason, L. C., Oppenheim, A. B., and Court, D. L. (2009). Modifying bacteriophage lambda with recombineering. *Methods Mol. Biol.* **501,** 239–251.

Tian, J., Gong, H., Sheng, N., Zhou, X., Gulari, E., Gao, X., and Church, G. (2004). Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* **432,** 1050–1054.

van Kessel, J. C., Marinelli, L. J., and Hatfull, G. F. (2008). Recombineering mycobacteria and their phages. *Nat. Rev. Microbiol.* **6,** 851–857.

Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., and Church, G. M. (2009a). Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460,** 894–898.

Wang, S., Zhao, Y., Leiby, M., and Zhu, J. (2009b). A new positive/negative selection scheme for precise BAC recombineering. *Mol. Biotechnol.* **42,** 110–116.

Wang, H. H., Xu, G., Vonner, A. J., and Church, G. M. (2011c). Modified bases enable high-efficiency oligonucleotide-mediated allelic replacement via mismatch repair evasion. *Nucleic Acids Res.* (in press).

Warming, S., Costantino, N., Court, D. L., Jenkins, N. A., and Copeland, N. G. (2005). Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res.* **33,** e36.

Wu, X. S., Xin, L., Yin, W. X., Shang, X. Y., Lu, L., Watt, R. M., Cheah, K. S., Huang, J. D., Liu, D. P., and Liang, C. C. (2005). Increased efficiency of oligonucleotide-mediated gene repair through slowing replication fork progression. *Proc. Natl. Acad. Sci. USA* **102,** 2508–2513.

Yu, D., Ellis, H. M., Lee, E. C., Jenkins, N. A., Copeland, N. G., and Court, D. L. (2000). An efficient recombination system for chromosome engineering in Escherichia coli. *Proc. Natl. Acad. Sci. USA* **97,** 5978–5983.

Yu, D., Sawitzke, J. A., Ellis, H., and Court, D. L. (2003). Recombineering with overlapping single-stranded DNA oligonucleotides: Testing a recombination intermediate. *Proc. Natl. Acad. Sci. USA* **100,** 7207–7212.

Zhang, K. "Oligonucleotide Tm Calculator." Web Resource. Sept 2010. http://arep.med.harvard.edu/kzhang/cgi-bin/myOligoTm.cgi.

Zhang, Y., Buchholz, F., Muyrers, J. P., and Stewart, A. F. (1998). A new logic for DNA engineering using recombination in Escherichia coli. *Nat. Genet.* **20,** 123–128.