Harnessing *Saccharomyces cerevisiae* Genetics for Cell Engineering

Laura Michele Wingler

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2011

ABSTRACT

Harnessing *Saccharomyces cerevisiae* Genetics for Cell Engineering

Laura Michele Wingler

Cell engineering holds the promise of creating designer microorganisms that can address some of society's most pressing needs, ranging from the production of biofuels and drugs to the detection of disease states or environmental contaminants. Realizing these goals will require the extensive reengineering of cells, which will be a formidable task due both to our incomplete understanding of the cell at the systems level and to the technical difficulty of manipulating the genome on a large scale. In **Chapter 1**, we begin by discussing the potential of directed evolution approaches to overcome the challenges of cell engineering. We then cover the methodologies that are emerging to adapt the mutagenesis and selection steps of directed evolution for *in vivo*, multi-component systems.

Yeast hybrid assays provide versatile systems for coupling a function of interest to a high-throughput growth selection for directed evolution. In **Chapter 2**, we develop an experimental framework to characterize and optimize the performance of yeast two- and three-hybrid growth selections. Using the *LEU2* reporter gene as a model selectable marker, we show that quantitative characterization of these assay systems allows us to identify key junctures for optimization. In **Chapter 3**, we apply the same systematic characterization to the yeast three-hybrid counter selection, beginning with our previously reported *URA3* reporter. We further develop a screening approach to identify effective new yeast three-hybrid counter selection reporters.

Installing customized multi-gene pathways in the cell is arguably the first step of any cell engineering endeavor. **Chapter 4** describes the design, construction, and initial validation of Reiterative Recombination, a robust *in vivo* DNA assembly method relying on homing endonuclease-stimulated homologous recombination. Reiterative Recombination elongates constructs of interest in a stepwise manner by employing pairs of alternating, orthogonal endonucleases and selectable markers. We anticipate that Reiterative Recombination will be a valuable tool for a variety of cell engineering endeavors because it is both highly efficient and technically straightforward. As an initial application, we illustrate Reiterative Recombination's utility in the area of metabolic engineering in **Chapter 5**. Specifically, we demonstrate that we can build functional biosynthetic pathways and generate large libraries of pathways *in vivo*. The facility of pathway construction by Reiterative Recombination should expedite strain optimization for metabolic engineering.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

$(GSG)_2$	glycine-serine-glycine-glycine-serine-glycine linker
5-FOA	5-fluoroorotic acid
A	deoxyadenosine
AD	transcriptional activation domain
Amp^R	ampicillin resistance
A_x	absorbance at x nm
bp	base pair
°C	degrees Celsius
C	deoxycytoside
cAMP	adenosine 3′,5′-cyclic monophosphate
Cm^R	chloramphenicol resistance
DBD	DNA-binding domain
Dex	dexamethasone
DHFR	dihydrofolate reductase
DMF	dimethyl formamide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DSB	double-strand DNA break
E. coli	*Escherichia coli*
et al.	*et alia*
FACS	fluorescence-activated cell sorting
FPP	farnesyl pyrophosphate

g	gram
G	deoxyguanosine
G418	geneticin
gal	galactose
gDNA	genomic DNA
GFP	green fluorescent protein
GR	glucocorticoid receptor
h	hour
HRS	heritable recombination system
IPP	isopentenyl pyrophosphate
IPTG	isopropyl β-D-thiogalactopyranoside
Kan^R	kanamycin resistance
kb	kilobase
k_{cat}	catalytic turnover number
K_m	Michaelis constant
L	liter
LexAop	operator binding a dimer of the LexA DNA-binding domain
M	moles per liter
MAGE	multiplex automated genome engineering
Magenta-Gal	5-bromo-6-chloro-3-indolyl-β-D-galactopyranoside
MAPK	mitogen-activated protein kinase
Mb	megabase
MCS	multiple cloning site

mg	milligram
min	minute
mL	milliliter
Mtx	methotrexate
nm	nanometer
nM	nanomoles per liter
NRPS	nonribosomal peptide synthase
OD_x	optical density at x nm
OMP	orotidine-5'-phosphate
ONPG	ortho-nitrophenyl-β-galactoside
ori	origin of replication
PACE	phage-assisted continuous evolution
PCR	polymerase chain reaction
PKS	polyketide synthase
pM	picomoles per liter
raf	raffinose
RNA	ribonucleic acid
rpm	revolutions per minute
s	second
S. cerevisiae	*Saccharomyces cerevisiae*
SC	synthetic complete
SDS	sodium dodecyl sulfate
t	time

T	deoxythymidine
UV	ultraviolet
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
X-Gluc	5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid
Y2H	yeast two-hybrid
Y3H	yeast three-hybrid
YPD	yeast peptone dextrose media
μg	microgram
μL	microliter
μM	micromoles per liter

Acknowledgements

First, let me acknowledge funding from the NIH and the NSF, and the NSF Graduate Research Fellowship in particular.

I would like to thank my advisor, Professor Virginia Cornish, for being a mentor in the truest sense of the word. Thank you for teaching me both the hard and soft skills of science, for keeping the big picture up front, and for giving me challenging problems to work on and so much freedom in chasing them down. Thank you most of all for believing in my work and in me when I usually did not.

My thanks also go to Professor Samuel Danishefsky and Professor Ruben Gonzalez for serving on my graduate committee, for their encouragement throughout graduate school, and for kindly providing recommendation letters as I began preparing for the next step of my career. I am also grateful to Professor Lorraine Symington and Dr. George Ellestad for their willingness to serve on my thesis committee.

Thanks to all the members of the Cornish laboratory, past and present, for making this a dynamic and stimulating place to learn and work. Forgive me if I only call out a few of the wonderful people who passed through during my tenure here. Dr. Pamela Peralta-Yahya, who initiated projects that directly or indirectly led to both of my major lines of work, was a role model of everything a graduate student should be and taught me most of what I know in the lab. Vanessa Mondol taught me the rest while keeping the lab in line. Dr. Dante Romanini and Nili Ostrov provided many invaluable discussions and manuscript critiques. Nili also initiated the Stephanopoulos collaboration that led to much of Chapter 5, is taking ReRec in multiple new directions, and is my partner in crime for brainstorming. One way or another, we *will* make it through our (top-secret)

bucket list of experiments. Heather Horgan made these past five years happen and somehow shares an incredible number of my obscure interests. I'm saving 8% of my love for you. Undergraduate Chris Sosa and rotation students Cheng Wang and Timothy Kernan followed up on various aspects of Chapter 2, and Marie Harton is already doing a wonderful job of picking up the work of Chapter 3. Last but not least, let me thank the ad hoc bioethics and gender issues "journal club" for discussions both thought provoking and entertaining, for preserving my sanity, and of course, for many excellent meals.

I cannot thank the members of Calvary Baptist Church enough for adopting me into your family even before I was quite ready to be taken in. In particular, I am so grateful to each member of Crossroads for shattering my complacency week after week and for sharing the same struggles in so many different corners of this crazy city. My love goes to the Calvary Handbell Choir for getting me to Calvary in the first place and for teaching me much about music, worship, and fellowship. Finally, thank you to the vibrant and amazingly talented Youth Handbell Choir for teaching me to lead and putting up with me in the meantime, for energizing me with your enthusiasm, and for always demanding the next challenge. I've said it before and say it again—y'all are awesome.

Moving beyond the island of Manhattan, my love and gratitude go out to the members of Olive Chapel Baptist Church, the village that raised me and shaped my faith, that has continued to support me and my family through some very dark days these past five years, and that makes some mighty fine barbecue. *Real* barbecue.

Most of all, my thanks to all of my family for loving me in spite of myself. Most *most* of all, thank you to my parents, for everything and then some. I ought to elaborate, but if I did, this would be by far the longest chapter of my thesis. I love you so much.

Not that we are sufficient of ourselves to think of anything as being from ourselves,

but our sufficiency is from God . . .

2 Corinthians 3:5

And for Papa, who never lost a chance to inform me what a Ph.D. really means.

Chapter 1

Directed Evolution Strategies for Cell Engineering

---

*The content of this chapter will be published in

L.M. Wingler, V.W. Cornish. "Directed Evolution Strategies for Cell Engineering," *in preparation.*

1.0 Chapter outlook

Synthetic biology holds the promise of creating designer microorganisms that can address some of society's most pressing needs, ranging from the production of biofuels and drugs to the detection of disease states or environmental contaminants. Realizing these goals will require the extensive reengineering of cells at the systems level. However, the rational design of all but the simplest *in vivo* systems is difficult or impossible, even when extensively characterized systems and components are used. By relying on iterative cycles of mutagenesis and selection rather than complete understanding of systems' behavior, directed evolution has the potential to overcome the complexities of cell engineering. While early examples of directed evolution in the context of *in vivo* systems provide strong evidence for the promise of this approach, the scope of these directed evolution experiments has been quite limited due to the technical difficulties associated with effecting genome-wide, targeted mutagenesis and assaying for any given function of interest. Here we review existing and emerging methods that will allow both the mutagenesis and assay steps of directed evolution to be carried out at the systems level inside the cell. An exciting consequence of performing the entire directed evolution cycle *in vivo* is the fact that we can now creatively harness the cell's own processes and machinery for these tasks. The novel strategies enabled by these advances have the potential to surmount the challenges of cell engineering and bring its enormous promise to fruition.

1.1 Directed evolution for *in vivo* systems

Directed evolution has matured into a widely employed technology for engineering antibodies and antibody mimics, enzymes, and nucleic acid aptamers for use as therapeutics, diagnostics, and research reagents[1]. The power of directed evolution lies in its ability to bypass limitations in our understanding of the relationships among biomolecules' sequences, structures, and functions that make rational redesign difficult or impossible. Instead, large "libraries" of biomolecules are generated using various strategies to mutagenize the encoding DNA, and those with the desired function are identified through screens or selections (**Fig. 1-1A**). The mutagenesis and assay process can be iteratively repeated until the function of interest is obtained or optimized.

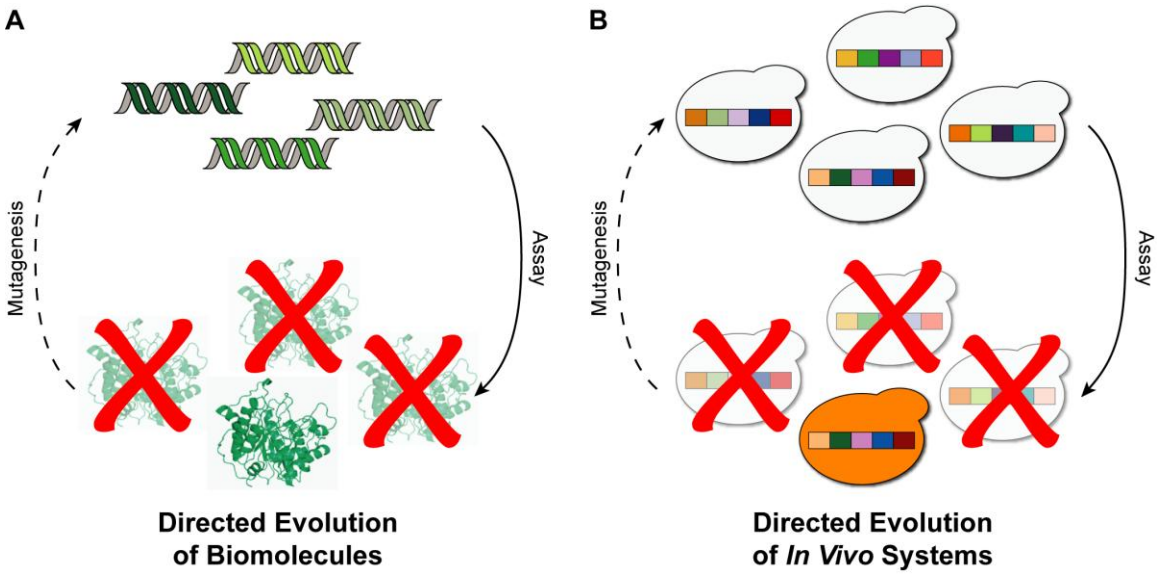


Figure 1-1. (A) The directed evolution of biomolecules versus **(B)** the directed evolution of *in vivo* systems. The directed evolution of *in vivo* systems will require advances in both the mutagenesis and assay steps.

With the advent of synthetic biology, the next challenge is to leap from engineering individual biomolecules *in vitro* to engineering multi-component systems

that function *in vivo*[2]. Under various appellations such as synthetic biologists, metabolic engineers, and cell engineers, researchers are already taking steps towards repurposing microorganisms as biosensors, "smart" therapeutics, and factories for biosynthesizing biofuels and drugs[3]. However, if modifying the function of individual molecules is difficult, modifying cells for these sophisticated applications will be orders of magnitude more challenging. Not only must the function of individual biomolecules within the system often be altered, but all components must interact productively with each other and with the intricate and incompletely understood network of existing cellular machinery. Just as directed evolution impacted the engineering of biomolecules by solving problems at the limits of rational design, it now has the potential to overcome the complexity of engineering *in vivo* systems (**Fig. 1-1B**).

Directed evolution approaches could help circumvent the intellectual bottlenecks impeding cell engineering, but implementing such strategies is hindered by the technical challenge of generating and effectively searching large libraries of variants *in vivo*. At the end of the 20th century, advances in recombinant DNA technology and high-throughput assays, which exponentially increased the number of variants that could be tested at a time, were critical to making directed evolution the robust tool for the engineering of individual biomolecules that it is today. However, most of these established directed evolution methods cannot be directly employed in contemporary efforts to carry out directed evolution on a larger scale and in the context of the cell. New technologies are needed to drive the transition from the engineering of biomolecules to the engineering of living cells. In spite of the associated complications, moving directed evolution *in vivo* presents us with an intriguing opportunity—we can now

envision harnessing the power and elegance of natural biological processes to reengineer cells for our own purposes.

In this chapter, we begin by presenting select examples from diverse fields that exemplify the rational design, directed evolution, and related library approaches being taken as we push the boundaries of our ability to reprogram cells for novel functions. We then discuss the technical advances that will be required to adapt mutagenesis and selection for *in vivo*, multi-component systems and the methodologies that are emerging to address these needs. We limit our discussion to the work that has been done on the relatively well-characterized microorganisms *Escherichia coli* and *Saccharomyces cerevisiae*, the hosts that are and will likely continue to be the workhorses of synthetic biology for the foreseeable future.

1.2 Approaches to the engineering of *in vivo* systems

1.2.1 Rational design

One major thrust of the cell engineering community has been to characterize and standardize the behavior of biological "parts" (e.g., promoters, transcription factors, ribosome-binding sites), with the goal of being able to computationally design *in vivo* systems with predictable behaviors[4]. An example where these efforts have borne fruit over the past decade is the refinement of artificial "genetic oscillators"[5-7]. These circuits use series of transcription factors to create positive or negative feedback loops to achieve regular temporal undulations in the expression of a reporter gene. The first-generation oscillator was far from ideal, as oscillations persisted for only three cycles and exhibited considerable noise in gene expression levels[5]. After years of refinement, Stricker *et al.* reported highly robust genetic oscillators that continued to function for hours and had

periods that could be adjusted by altering growth conditions[7]. The authors note that the key to the dramatic improvements in the oscillator design was modeling the circuit more comprehensively—incorporating not only genes' transcription rates, but also parameters such as the translation, folding, and oligomerization of proteins. They explicitly caution against "making simplifying assumptions in the design of engineered gene circuits"[7].

Nevertheless, the high profile of this and other[8-14] examples of designed *in vivo* systems that are functional, robust, and reliable is evidence that reprogramming cells at will has not yet become a routine reality. Notably, these successes have all been achieved for systems that were already exhaustively characterized, that involved only small numbers of genes (<10)[15,16], and that were generally intended to be insulated from other cellular processes as much as possible. As we push synthetic biology forward into more sophisticated applications such as metabolic engineering, exogenous components will need to be more deeply integrated with the host cells' endogenous machinery, and individual circuits and pathways will need to be productively interwoven into highly sophisticated networks. Given that we are only beginning to be able to design even simple systems, the routine design of engineered *in vivo* systems is not yet feasible.

1.2.2 Library approaches

In light of our inability to build fully descriptive, predictive models of the cell, library approaches such as directed evolution are needed to allow researchers to move cell engineering efforts forward rapidly. These strategies are already being implemented to optimize and diversify a variety of *in vivo* systems.

Optimizing regulatory circuits. Yokobayashi *et al.* used directed evolution to refine the performance of a genetic circuit[17]. They began with a very simple framework

that looked as though it should be functional on paper—two transcriptional repressors placed in series should lead to the transcription of a reporter gene, but when the small molecule IPTG is added to inactivate the first repressor (*lacI*), the reporter's expression should be turned off. However, due to the noise of biological systems, such as finite amounts of transcription from leaky promoters, the original circuit never expressed high levels of the GFP reporter under any conditions. The authors then mutated one of the repressors and its binding site to tune the sensitivity of the circuit and screened for cells that exhibited the desired change in GFP expression upon addition of IPTG. This strategy led to the identification of a number of functional circuits without requiring any biochemical information on the system's components to guide optimization attempts.

Peisajovich *et al*. created a library of pathways by focusing on the level of protein rather than transcriptional regulation[18]. They recombined the regulatory and catalytic domains from 11 proteins involved in the MAPK signaling pathway that leads to the mating response in *S. cerevisiae* following pheromone stimulation. Expressing this library of chimeric proteins with the endogenous pathway provided strains that exhibited a range of kinetics in the mating response and some strains that had improved mating efficiencies.

Generating new regulatory topologies. Guet and coworkers demonstrated that they could generate transcriptional regulatory circuits with a variety of topologies by combinatorially mixing and matching small sets of regulatable promoters (five) driving the expression of three transcriptional regulators[19]. The resulting networks exhibited diverse qualitative and quantitative patterns of reporter expression with or without two small-molecule inducers.

Mody *et al.* were able to manipulate the architecture of signaling pathways in yeast by creating chimeras of Fus3p and Hog1p, related MAP kinases from the yeast mating and osmolarity response pathways, respectively[20]. They found that when these chimeras were expressed in strains in which both endogenous MAP kinases were deleted, some chimeras were able to rewire the pathways (e.g., to activate the mating pathway in response to the osmolarity signal or to activate both the osmolarity and mating pathways in response to a single signal).

Optimizing metabolic pathways. The need to incorporate directed evolution approaches in to cell engineering efforts is being manifested even more clearly in the realm of metabolic engineering[21]. Recent, high-profile breakthroughs in metabolic engineering[22-28] have raised the tantalizing prospect of, ultimately, routinely biosynthesizing any desired biofuel, chemical feedstock, or natural product in tractable heterologous microorganisms. Metabolic engineering poses an additional layer of complexity to cell engineering because it inherently perturbs cellular metabolism. Even simply overexpressing a heterologous protein imposes a burden on cells[29,30]; pointedly disrupting metabolic flux can have even more profound, unanticipated effects, complicating rational optimization efforts[31]. Without exception, intensive optimization of strains has been required to attain high-level production, or often any detectable production, of the target compounds. To date, most of these optimization processes have proceeded in a laborious, stepwise fashion. However, several elegant examples of using combinatorial approaches to streamline metabolic engineering efforts have recently been presented[32,33].

For example, Ajikumar *et al*. recently achieved high-yield production of taxadiene, the first committed intermediate in the biosynthesis of the potent chemotherapeutic taxol (paclitaxel), in *E. coli* by dividing the biosynthetic pathway into two modules and combinatorially varying the expression levels of enzymes in the two modules[25]. This library approach pinpointed a promoter combination that gave a 15,000-fold increase in taxadiene yield. Analysis of other strains from the library also revealed that there was no simple relationship between expression levels and taxadiene production, and the global maximum for taxadiene yield fell within a very limited range of expression levels. When analyzing lower-performing strains, they discovered that a completely unexpected metabolic byproduct, indole, accumulated through unknown mechanisms in most members of the library, inhibiting taxadiene production. These results highlight our current inability to predict the cascading effects of disrupting delicately balanced cellular pathways.

While most efforts to optimize metabolic pathways' yields through library approaches have focused on tuning the expression levels of relevant genes[25,32,33], Leonard and coworkers also incorporated protein engineering into optimizing biosynthesis of the diterpenoid levopimaradiene in *E. coli*[24]. After maximizing flux towards universal terpenoid precursors by overexpressing genes in the non-mevalonate pathway, the authors discovered that geranylgeranyl diphosphate synthase and levopimaradiene synthase became the rate-limiting enzymes in the pathway. Separately, they mutagenized both enzymes and identified variants with increased activities. Replacing the wild-type proteins with these two improved variants, together with optimizing the production of

terpenoid precursors, increased levopimaradiene yields by a factor of approximately 2,600.

Diversifying metabolic pathways. In addition to increasing yields of natural products, library approaches have also been used to biosynthesize natural product analogs[34]. Perhaps the most impressive example has come from Kosan Biosciences, whose researchers placed the three modules of the small polyketide synthase (PKS) gene cluster for 6-deoxyerythronolide B on two plasmids for expression in *E. coli* and then replaced two of the modules with over ten variants each[35]. Co-transforming all possible combinations of plasmids led to the production of 154 different pathway variants, almost half of which produced detectable levels of the expected triketides.

While the modularity of assembly-line enzymes such as PKSs and nonribosomal peptide synthases (NRPSs) provides enormous potential for combinatorial biosynthesis, chimeric enzymes derived from mixing and matching modules often have poor or no activity. Fischbach *et al.* addressed this problem with directed evolution[36]. After replacing a valine-specific domain of the hybrid NRPS-PKS cluster that produces andrimid with a heterologous isoleucine-specific domain, the yield of the andrimid analog was only a seventh of andrimid yields from the wild-type pathway. Mutagenesis of the isoleucine domain by error-prone PCR led to the identification of a domain variant that increased the analog's yields 4.5-fold.

1.3 Mutagenesis technologies for systems-level directed evolution

Despite the widespread adoption of directed evolution approaches for biomolecule engineering, and despite the apparent advantages of analogous strategies for cell engineering, examples of carrying out controlled directed evolution experiments for

cell engineering applications remain relatively few. The reticence of the field to fully embrace this strategy undoubtedly stems from the technical challenges that accompany translating directed evolution approaches for biomolecules to *in vivo* systems. It is telling that in the small number of systems-level directed evolution experiments reported, such as those described above, researchers have typically only interrogated one or two components of a more complex system.

The first technical advance required for systems-level directed evolution will be the ability to mutagenize numerous, defined loci. Efficient methods for mutating individual genes have been developed for model organisms such as *E. coli* and *S. cerevisiae*[37,38]. However, these techniques are not useful for mutating multiple genes because 1) they are too time-consuming for or incompatible with iterative mutagenesis and/or 2) their efficiencies drop precipitously when used to mutagenize multiple loci in parallel. It is important to recognize that many cell engineering applications will require both modification of heterologous genes that are newly introduced into the host cell and of the cell's genetic background. The strategies that are emerging to effect mutagenesis at the systems level can be divided into two broad categories—*in vitro* mutagenesis of DNA, followed by subsequent introduction into the desired host cell, and *in vivo* mutagenesis of DNA directly in the cell (**Fig. 1-2**).

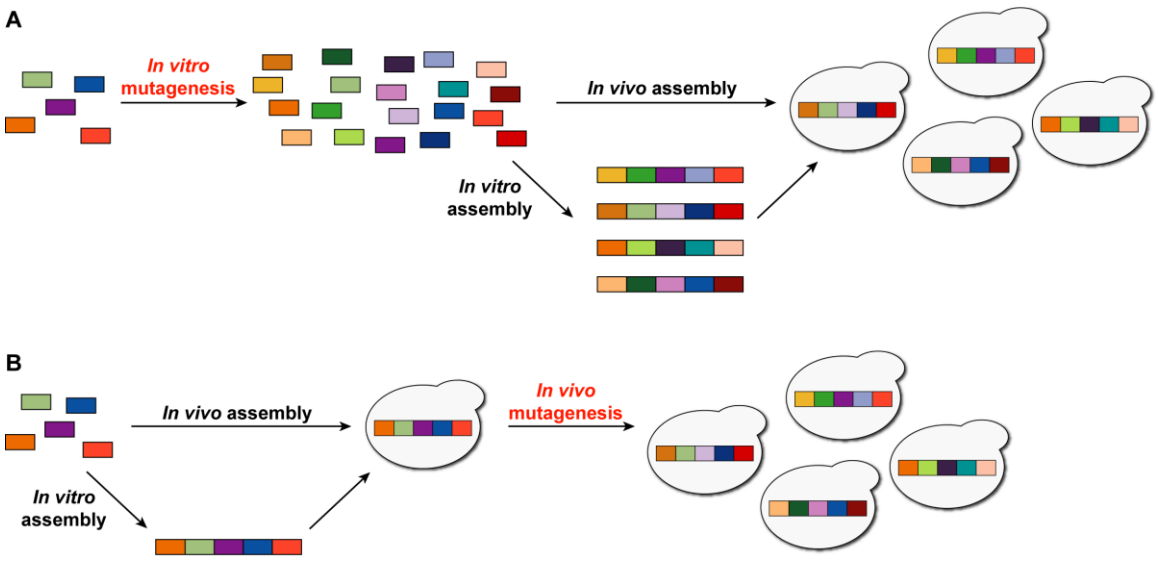


Figure 1-2. Mutagenesis strategies for systems-level directed evolution. DNA can either be **(A)** mutagenized *in vitro* and then introduced into the cell or **(B)** mutagenized *in vivo*.

1.3.1 *In vitro* mutagenesis of DNA

The first option, *in vitro* mutagenesis of DNA, is primarily applicable only to exogenous constructs introduced into host cells and not to mutation of the strain's genetic background. *In vitro* mutagenesis has the advantage of being able to utilize the wealth and wide variety of existing mutagenesis techniques that have been developed for the directed evolution of individual biomolecules[39,40]. A rich spectrum of diversification patterns can be achieved through these methods—ranging from entirely random mutations to site-specific modifications to the shuffling of fragments from related molecules. Importantly, many of these mutation techniques are very straightforward molecular biology protocols that could easily be multiplexed to mutate numerous genes in parallel. As the cost of DNA synthesis continues to drop, high-throughput gene synthesis is also becoming an increasingly feasible option for creating large libraries of defined mutants[41].

The main challenge associated with applying *in vitro* mutagenesis techniques to systems directed evolution is moving the mutated DNA into the cell in a manner that is efficient enough to cover large library sizes. As discussed above, the stepwise introduction of mutated genes into the cell one at a time using standard genetic techniques is arduous and impractical. To streamline this process, the mutated DNA of individual genetic components must somehow be assembled, either *in vitro* or *in vivo*.

A plethora of *in vitro* enzymatic methods specifically designed to accurately assemble multi-gene constructs have been introduced in recent years[35,42-51]. These techniques are typically derivatives of routine molecular biology procedures, but the robustness and stringency of assembly have been greatly improved to enforce accurate assembly for large constructs. A number of these protocols have been applied to the construction of impressively large DNA molecules from tens to hundreds of kilobases in length, and some of them have efficiencies high enough to theoretically be useful for making libraries, which we define as being able to generate at least 10^3, but preferably closer to $\geq 10^8$ variants at a time. However, the few methods that have actually been tested in the context of library construction have only been used to build rather small collections of pathways ($\leq 10^2$ variants)[52].

The main drawbacks of these *in vitro* assembly methods are the technical difficulty of manipulating large, fragile DNA pieces *in vitro*[16] and the fact that they do not inherently solve the problem of moving the assembled DNA into the cell. Even for constructs that are located on plasmids, as the size of DNA constructs increases, there is a corresponding decrease in the efficiency with which it can be transformed into cells, and thus the achievable *in vivo* library size[53,54]. For many applications, it is necessary to

integrate constructs into the chromosome to improve strains' stability and control the exogenous genes' copy number[55-57]. Integration protocols proceed with even lower efficiencies than plasmid transformations, especially for large DNA constructs[58].

The alternative to *in vitro* assembly is exploiting cells' endogenous homologous recombination machinery to stitch together numerous DNA fragments *in vivo*[57,59-63]. *In vivo* assembly is quite straightforward technically, as it only requires transformation of DNA pieces into the host cell, and has been used by expert laboratories to build even genome-sized (>100-kb) molecules[57,62,63]. Several methods further allow for the integration of large DNA constructs into the chromosome[57,59,60]. However, even in organisms with efficient homologous recombination machinery, such as *Bacillus subtilis* and *S. cerevisiae*, these assembly methods proceed with very low efficiencies, generating only tens to hundreds of recombinants at a time and making them infeasible for the generation of *in vivo* libraries. Robust and high-yielding methods for assembling DNA *in vivo* will be required before libraries of multiple heterologous genes mutated *in vitro* can installed in the cell efficiently enough to be useful for directed evolution library construction (see **Chapters 4 and 5**).

1.3.2 *In vivo* mutagenesis of DNA

A new wave of technologies is now emerging to mutagenize DNA directly in living cells in a targeted manner. These methods will provide a critical advance for the field by enabling mutagenesis not only of DNA newly introduced into the cell but also of multiple loci in the host cell's genome.

In some respects, *in vivo* mutagenesis has been utilized for decades. Chemical mutagens, UV radiation, and mutator strains are all classic methods for evolving strains'

phenotypes, but they suffer from obvious issues of toxicity and do not allow mutations to be targeted to specific loci[40,64]. Given that the number of possible combinations of 10 point mutations in the *E. coli* genome (~10^{72}) vastly exceeds the number of atoms on earth (~10^{50})[65], these fully random mutagenesis methods are an extremely inefficient way to introduce diversity for most applications. Furthermore, it is difficult to recover information on what mutations contributed to improvements in phenotype to guide future directed evolution experiments[66]. Loeb *et al.* have developed a next-generation mutator strain that employs a mutant DNA polymerase to focus enhanced mutagenesis to a plasmid of interest[67]. However, this strain still suffers from toxicity issues and does not allow at-will modification of the genome.

As with *in vivo* DNA assembly, homologous recombination seems to be presenting itself as the most general and broadly applicable solution for targeted *in vivo* mutagenesis. Wang *et al*. recently described multiplex automated genome engineering (MAGE), a powerful platform for introducing modifications throughout the chromosome of *E. coli*[32]. MAGE takes advantage of the body of work from the Court laboratory showing that expressing the recombination machinery from the λ-Red bacteriophage in a mismatch repair-deficient *E. coli* strain promotes high-frequency mutagenesis of the chromosome by single-stranded oligonucleotides transformed into the cell[68,69]. Wang *et al*. optimized the conditions for this "recombineering" system so that they could reliably introduce deletions, replace sequences, or make short insertions (<~30 bp) with efficiencies of up to ~30%. By then automating the procedure to allow repeated cycles of transformation and mutagenesis, they mutagenized the ribosome binding sites of twenty genes that were known to affect yields of the terpenoid pigment lycopene in *E. coli*,

calculating that they generated $>10^{10}$ variants in the course of 35 rounds of mutagenesis. Screening a small fraction of this diversity (10^5 colonies) yielded variants with up to a ~5-fold increase in lycopene production over the parental strain, and sequencing of a number of mutants provided insight into key parameters affecting yields.

As part of a Heritable Recombination System (HRS) (see **Section 1.5**), our laboratory has developed an efficient *in vivo* mutagenesis method for *S. cerevisiae*[70]. We place "cassettes," consisting of a mutagenic region flanked on either side by short (~30 bp) homology regions targeting the gene of interest, between two homing endonuclease recognition sites in a plasmid. Following induction of endonuclease expression to cleave the cassette plasmids, we observe high-efficiency (~5%) homologous recombination with the target DNA sequence. Since not all cassette plasmids are cleaved and yeast can undergo sexual reproduction, mutations and cassette plasmids can subsequently be exchanged among populations of cells to accumulate mutations in multiple genes. Though we initially demonstrated this technology using plasmid-based target genes, this same iterative mutagenesis strategy should be extendable to chromosomal modifications in subsequent versions of the HRS.

The breadth of types of genetic modifications that can be accessed via *in vivo* mutagenesis is also beginning to rival that of *in vitro* mutagenesis. For example, Bikard *et al.* have harnessed another natural mechanism for diversifying genomes, the integron of *Vibrio cholera*, to shuffle genes within a defined pathway in *E. coli*[71]. Integrons employ an integrase, a site-specific recombinase enzyme, to excise members of an array of DNA cassettes flanked by recombination sites and integrate the cassettes into a primary recombination site[72]. Adapting this system to create a "synthetic integron,"

Bikard *et al.* placed the genes for tryptophan biosynthesis between recombination sites, initially preventing their expression by placing transcriptional terminators within the pathway. Expression of the integrase to induce recombination, followed by selection for growth in the absence of tryptophan, led to the identification of pathways in which the terminators had been excised, some of which also had rearrangements and duplications in the biosynthetic genes. While the frequency of cassette shuffling events ($\sim 10^{-4}$ for a single reordering event) is currently too low to be useful as the sole mutagenesis strategy in most directed evolution experiments, optimization of recombination or coupling of this method to a robust selection strategy capable of searching very large library sizes (see **Section 1.4**) could reach into new areas of genetic diversity.

1.4 Assay methods for systems-level directed evolution

The second broad challenge in systems-level directed evolution will be developing suitable high-throughput assays, either screens or selections, to evaluate the performance of variants. In screens, each library member must be assayed individually, thus limiting the library size to 10^3-10^6 even with automation techniques. Selections, by contrast, are designed such that only the desired variants survive; much larger libraries ($\geq 10^8$) can be tested because the vast majority of variants do not have to be analyzed explicitly. It is worth noting that a screen of 10^5 would have to be carried out every day for over two and a half years to test a library of 10^8. Considering the tremendous number of variables that need to be interrogated in the context of *in vivo* directed evolution experiments, the use of selections whenever possible will be essential.

At their most basic level, the diverse objectives being pursued by cell engineers can roughly be divided into two categories: obtaining variants with defined response

patterns and obtaining variants with a defined output. These manifold applications need to be linked to assays that have both high *throughput* (the ability to identify one cell with the desired output from 10^n lacking it) and *sensitivity* (the ability to distinguish n-fold differences in the output) (**Fig. 1-3**).

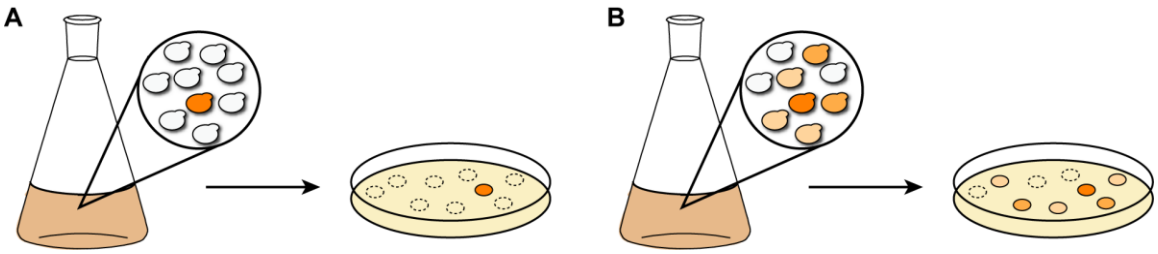


Figure 1-3. (A) Throughput and **(B)** sensitivity in *in vivo* selection systems.

1.4.1 Assays for defined output patterns

The first common scenario is that researchers are trying to create cells that exhibit a certain response pattern; examples of this include the creation of regulatory circuits (**Section 1.2**) and the longer-term goal of fashioning cell-based biosensors. In these cases, the actual identity of the output is secondary and a matter of choice, meaning that reporters providing the best throughput and sensitivity can be utilized (see **Section 1.4.3**). The challenge for these systems will be customizing and calibrating these reporters to ensure that they perform optimally in a given system.

1.4.2 Assays for defined outputs

For other applications, such as metabolic engineering, the objective is to engineer cells to provide a specific output, such as the production of a particular chemical compound (**Section 1.2**). The challenge for these endeavors is finding means to couple the output of interest to a high-throughput assay. Most small molecules, for example, do not have chromophores or other distinctive, measurable properties that allow them to be

detected directly even in medium-throughput microtiter plate assays. Instead, their production must be analyzed using liquid or gas chromatography-mass spectrometry, placing an absolute cap of $<10^3$ on the number of variants that can be tested[21]. The frustration with this bottleneck in screening throughput has actually altered the course of research in the field of metabolic engineering. Instead of pursuing the highest-impact targets, researchers continually re-optimize the production of a few readily assayable compounds[31,32,73], such as the terpenoid pigment lycopene that confers an orange phenotype on colonies, hoping that they will at least serve as imperfect surrogates for more valuable molecules[24,74].

Linking sundry outputs to robust assays is not a new problem; researchers have long struggled with how to assay individual biomolecules both *in vitro* and *in vivo*. Some of the many creative solutions developed for biomolecules should be able to translate to or inspire solutions for providing high-throughput readouts for *in vivo* systems—for example, chemical[75] or enzymatic[76] conversion of the biosystem's initial output to create a more readily detectable product, or interaction of the output with another molecule that induces a readout[77-79].

Ideally, these assay strategies should be as general and modular as possible both to maximize the scope of their utility and to enable coupling to the readout of choice. Yeast hybrid systems (and their derivatives in other organisms) provide elegant examples of such versatility in assays for biomolecules (**Fig. 1-4**). Based on the artificial coupling of a DNA-binding and transcription activation domain to create a transcriptional activator, yeast hybrid assays allow arbitrary protein-DNA (one-hybrid)[80], protein-protein (two-hybrid)[81], protein-RNA (three-hybrid)[82], or protein-small molecule (three-hybrid)[83]

interactions to be coupled to the expression of a reporter gene of choice. Our laboratory further elaborated the yeast three-hybrid assay to permit the detection of enzymatic activity via "chemical complementation"[84], allowing enzymatic reactions that previously could only be detected with low-throughput screens, such as glycosidic bond formation[85] and cleavage[86], to be coupled to an *in vivo* growth selection. Devising equally flexible assay methods should be an invigorating pursuit for the cell engineering community.

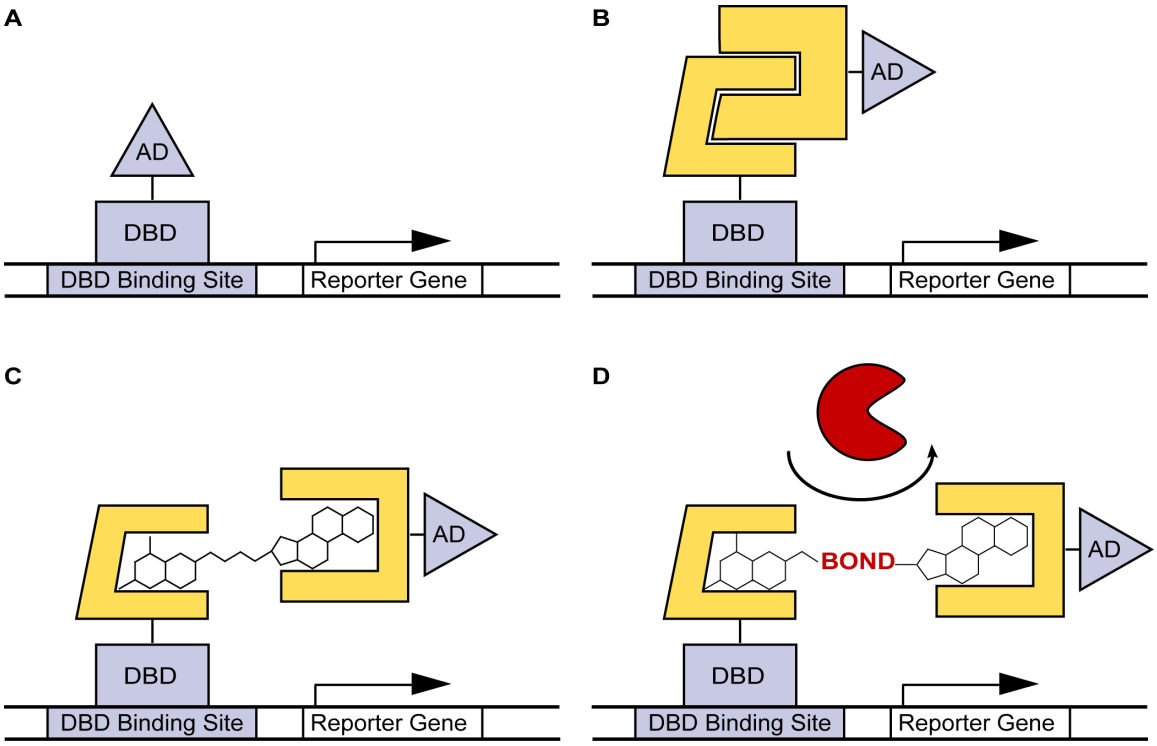


Figure 1-4. Yeast n-hybrid systems. The yeast one-hybrid **(A)**, two-hybrid **(B)**, three-hybrid **(C)**, and chemical complementation **(D)** systems provide a model for coupling diverse functions to a readout of choice.

1.4.3 Achieving throughput and sensitivity in *in vivo* assays

What readouts can provide the throughput and sensitivity needed for systems-level directed evolution experiments? The most difficult criterion to satisfy will be throughput. Two assay methods that are compatible with *in vivo* systems have the

potential to search libraries $\geq 10^8$ in size—fluorescence activated cell sorting (FACS) and growth selections.

Though technically a screen rather than a selection, FACS can analyze and sort cells at rates of up to 10^7 per hour, meaning that library sizes of 10^8-10^9 are achievable in directed evolution experiments[87-89]. The fluorescence readout that can most readily be coupled to diverse functions of interest is the production[87,88], reconstitution[90], or activation[79] of a fluorescent protein such as GFP. GFP is already commonly used as a reporter gene for many cell engineering applications such as the refinement of genetic circuits (**Section 1.2**). In addition, emulsion techniques have expanded the range of small molecule-based fluorescent readouts that are compatible with FACS[89]. The quantitative fluorescence signal measured by FACS means that sensitivity is inherently built into the assay. However, for some applications, the analysis of single cells can be problematic since stochastic cell-to-cell variations are not averaged out over a population[21]. The primary drawback of FACS that will limit its utility the most is the high cost of the required instrumentation, which is not yet ubiquitous, and of performing each experiment[91].

Growth selections are a second means to achieve very high assay throughputs *in vivo*; the majority of directed evolution experiments that have achieved the most impressive functional changes (e.g., large shifts in enzyme substrate specificity) in individual molecules have employed such selections[92-94]. As with GFP, the most versatile means to link an arbitrary output of interest to a growth selection is by the expression or activation of a protein required for growth under certain conditions, typically auxotrophic markers or antibiotic resistance markers.

The sheer simplicity and accessibility of growth selections advocates for their broader use in directed evolution for cell engineering applications, but there will be several challenges associated with appropriating these classic genetic assays. These reporter genes have historically been employed by geneticists as digital "on/off" markers for purposes such as plasmid maintenance or gene knockouts. When placed in the context of more sophisticated systems, where issues such as basal reporter gene transcription will be present, they may not maintain the throughputs of $\geq 10^8$ that can be achieved in genetic screens (see **Chapters 2 and 3**). In addition, the sensitivity of these growth selections to intermediate levels of activation is underdetermined.

The first step towards overcoming these issues will be rigorous characterization and optimization of the throughput and sensitivity of these growth selections. The collective experience from the biomolecule directed evolution field has drilled home the lesson that careful analysis of *in vitro* assays' performance parameters is critical, but this step has often been neglected in early examples of the application of *in vivo* growth selections. Second, if growth selections are not sufficiently effective, simple changes to the strain background, growth conditions, or the reporter gene itself (e.g., destabilizing the protein product) can often drastically improve or modulate growth selections' throughput and sensitivity[95-97]. Thinking more creatively, our increasing knowledge of how to construct simple genetic circuits could be put to use to construct feedback loops and other regulatory mechanisms to decrease basal transcription, amplify small differentials in transcription, and adjust the threshold of the output required to obtain robust growth[11,98,99]. We have further suggested identifying new selectable markers, beyond those traditionally used by geneticists, that might provide throughputs and

sensitivities better matched to contemporary needs (**Chapter 3**). Here again, directed evolution can be an effectual strategy to refine and optimize the performance of these selection systems.

1.5 Continuous *in vivo* directed evolution

The developing technologies to mutagenize and assay living cells at the systems level are moving us closer towards the ultimate goal of achieving continuous *in vivo* directed evolution. Fully integrating *in vivo* mutagenesis and assay techniques will obviate the need to move DNA in and out of the cell at every round of directed evolution, eliminating the transformation barrier that is currently the absolute cap on library size, while permitting us to retain molecular control over the evolutionary process. The Liu laboratory and our own have recently made strides towards this vision, describing frameworks that are compatible with performing multiple consecutive rounds of *in vivo* directed evolution.

Liu and coworkers' phage-assisted continuous evolution (PACE) system[100] uses an automated bioreactor system to evolve a bacteriophage population. After the phage library infects *E. coli*, expression of a phage-encoded gene with the activity of interest is coupled to the expression of a phage coat protein. This coat protein required for the phage to be infectious, but it is encoded only by a gene in the host. As the phage replicate in *E. coli*, high rates of mutagenesis are achieved using a non-targeted, mutator strain-type approach. Cells continually flow through the reactor; therefore the phage's progeny must receive sufficient coat protein to rapidly infect a new host cell before they are "washed out." Though the authors evolved only an individual protein, T7 RNA polymerase, in their initial report, PACE could conceivably be applied to the directed

evolution of pathways. However, since fresh *E. coli* cells are infected in each round of phage evolution, PACE will be incompatible with mutation of the host's chromosome.

Our laboratory has developed a Heritable Recombination System (HRS) in yeast permitting continuous *in vivo* directed evolution that also incorporates genetic exchange among evolving populations[70] (**Fig. 1-5A**). As described above (**Section 1.3.2**), mutagenesis is effected by endonuclease cleavage of cassette plasmids. Since some cassette plasmids survive cleavage, cells can exchange cassette plasmids between each round of mutagenesis by utilizing yeast's sexual reproduction cycle, invoking natural evolutionary mechanisms. In our first-generation HRS, we demonstrated that we could sequentially mutagenize two selectable markers carried on a plasmid, repairing stop codons engineered into the sequence that inactivated the gene products, without ever removing the DNA from the cell. Furthermore, even if we transformed cells with libraries of cassette plasmids containing a 10^6-fold excess of stop codons relative to wild-type codons, we were able to recover the doubly-repaired plasmid, demonstrating that we were generating very large library sizes *in vivo*.

Combining *in vivo* mutagenesis with genetic exchange opens up intriguing possibilities for the directed evolution field. Since libraries of variants can be crossed via sexual reproduction after they are initially installed by transformation, library sizes exceeding the transformation limit can be accessed. Perhaps even more significantly, winnowing of individual libraries through selection before sexually combining them could allow the "virtual search" of exceptionally large libraries that surpass even the number of cells that can be grown in high-density culture (**Fig. 1-5B**). This winnowing

approach would be analogous to the use of pruning algorithms in computational biology to enable breakthroughs in sequence space sampling[101,102].

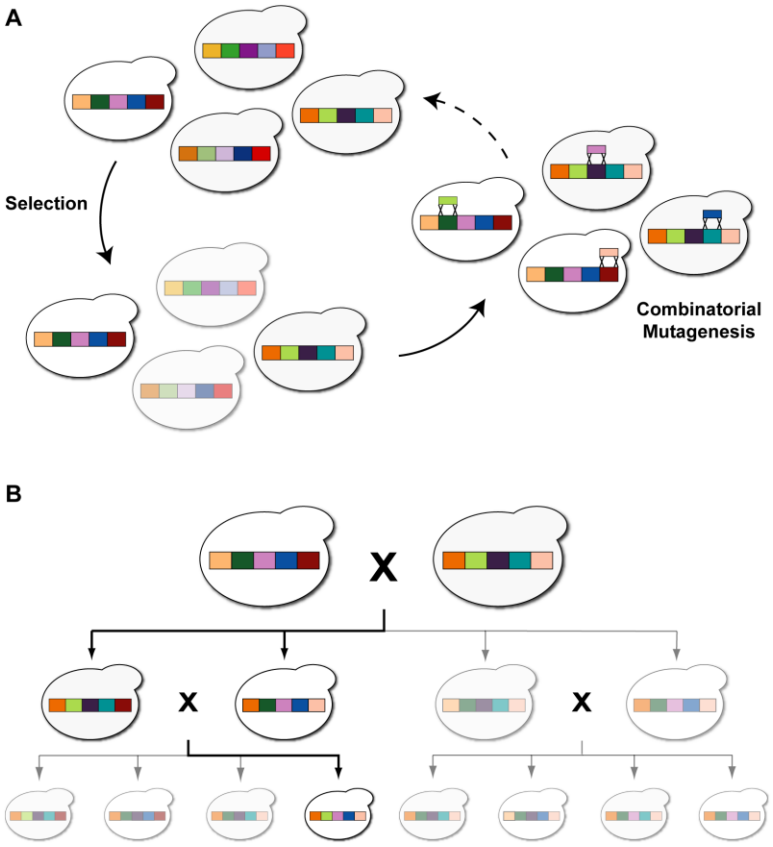


Figure 1-5. Continuous *in vivo* directed evolution incorporating genetic exchange. **(A)** Our heritable recombination system (HRS) allows the crossing of beneficial mutations through mating and sporulation. **(B)** Winnowing and then crossing individual libraries could allow the virtual search of extremely large library sizes. Grayed cells represent strains never explicitly generated and tested.

1.6 Conclusions

By analogy to the directed evolution of biomolecules, the directed evolution of *in vivo* systems can energize the field of cell engineering. While computational approaches have been making steady advances in recent years, the rational design of *in vivo* systems is far from a solved problem even for simple, exhaustively studied systems. A direct parallel can be seen here to the field of computational enzyme design, where after decades of work experts are just now finally beginning to be able to design enzymes *de novo* for very well-characterized reactions[103-105]. Even here, small libraries (~10^2) of designed enzymes must be tested to identify functional catalysts, and the resulting

enzymes have low activities[106]. As we now attempt to engineer not only individual but also entire systems of biomolecules to function in living cells, it is clear that directed evolution approaches that circumvent intellectual bottlenecks will be essential for rapid progress.

Elegant new approaches are already emerging to address many of the technical challenges associated with creating and searching large libraries of variants *in vivo.* In spite of the complications that accompany moving into the cell, this transition offers us the opportunity to appropriate and redirect biological mechanisms that directed evolution practitioners have long attempted to mimic *in vitro*, such as mutagenesis based on homologous recombination and sexual reproduction. As technologies for the systems-level *in vivo* directed evolution mature, we will be able to weave mutagenesis and assay techniques into continuous, streamlined processes. This combination of faithfully recapitulating natural evolution while still retaining molecular control over the process will allow us to explore biological diversity in creative new ways and help us realize the potential of cell engineering.

1.7 References

1. Arnold, F.H. Combinatorial and computational challenges for biocatalyst design. *Nature* **409**, 253-257 (2001).

2. Baker, D. et al. Engineering life: Building a fab for biology. *Sci. Am.* **294**, 44-51 (2006).

3. Khalil, A.S. & Collins, J.J. Synthetic biology: Applications come of age. *Nat. Rev. Genet.* **11**, 367-379 (2010).

4. Canton, B., Labno, A. & Endy, D. Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol.* **26**, 787-793 (2008).

5. Elowitz, M.B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335-338 (2000).

6. Atkinson, M.R., Savageau, M.A., Myers, J.T. & Ninfa, A.J. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli. *Cell* **113**, 597-607 (2003).

7. Stricker, J. et al. A fast, robust and tunable synthetic gene oscillator. *Nature* **456**, 516-519 (2008).

8. Fung, E. et al. A synthetic gene-metabolic oscillator. *Nature* **435**, 118-122 (2005).

9. Friedland, A.E. et al. Synthetic gene networks that count. *Science* **324**, 1199-1202 (2009).

10. Gardner, T.S., Cantor, C.R. & Collins, J.J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**, 339-342 (2000).

11. Basu, S., Gerchman, Y., Collins, C.H., Arnold, F.H. & Weiss, R. A synthetic multicellular system for programmed pattern formation. *Nature* **434**, 1130-1134 (2005).

12. You, L., Cox, R.S., 3rd, Weiss, R. & Arnold, F.H. Programmed population control by cell-cell communication and regulated killing. *Nature* **428**, 868-871 (2004).

13. Tamsir, A., Tabor, J.J. & Voigt, C.A. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature* **469**, 212-215 (2011).

14. Regot, S. et al. Distributed biological computation with multicellular engineered networks. *Nature* **469**, 207-211 (2011).

15. Purnick, P.E.M. & Weiss, R. The second wave of synthetic biology: From modules to systems. *Nat. Rev. Mol. Cell Biol.* **10**, 410-422 (2009).

16. Baker, M. Synthetic genomes: The next step for the synthetic genome. *Nature* **473**, 403-408 (2011).

17. Yokobayashi, Y., Weiss, R. & Arnold, F.H. Directed evolution of a genetic circuit. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16587-16591 (2002).

18. Peisajovich, S.G., Garbarino, J.E., Wei, P. & Lim, W.A. Rapid diversification of cell signaling phenotypes by modular domain recombination. *Science* **328**, 368-372 (2010).

19. Guet, C.C., Elowitz, M.B., Hsing, W. & Leibler, S. Combinatorial synthesis of genetic networks. *Science* **296**, 1466-1470 (2002).

20. Mody, A., Weiner, J. & Ramanathan, S. Modularity of MAP kinases allows deformation of their signalling pathways. *Nat. Cell Biol.* **11**, 484-491 (2009).

21. Dietrich, J.A., McKee, A.E. & Keasling, J.D. High-throughput metabolic engineering: Advances in small-molecule screening and selection. *Annu. Rev. Biochem.* **79**, 563-590 (2010).

22. Ro, D.K. et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940-943 (2006).

23. Hawkins, K.M. & Smolke, C.D. Production of benzylisoquinoline alkaloids in Saccharomyces cerevisiae. *Nat. Chem. Biol.* **4**, 564-573 (2008).

24. Leonard, E. et al. Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 13654-13659 (2010).

25. Ajikumar, P.K. et al. Isoprenoid pathway optimization for taxol precursor overproduction in Escherichia coli. *Science* **330**, 70-74 (2010).

26. Atsumi, S., Hanai, T. & Liao, J.C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451**, 86-89 (2008).

27. Steen, E.J. et al. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* **463**, 559-562 (2010).

28. Schirmer, A., Rude, M.A., Li, X., Popova, E. & del Cardayre, S.B. Microbial biosynthesis of alkanes. *Science* **329**, 559-562 (2010).

29. Klumpp, S., Zhang, Z. & Hwa, T. Growth rate-dependent global effects on gene expression in bacteria. *Cell* **139**, 1366-1375 (2009).

30. Tan, C., Marguet, P. & You, L. Emergent bistability by a growth-modulating positive feedback circuit. *Nat. Chem. Biol.* **5**, 842-848 (2009).

31. Alper, H., Miyaoku, K. & Stephanopoulos, G. Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. *Nat. Biotechnol.* **23**, 612-616 (2005).

32. Wang, H.H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-898 (2009).

33. Pfleger, B.F., Pitera, D.J., Smolke, C.D. & Keasling, J.D. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat. Biotechnol.* **24**, 1027-1032 (2006).

34. Zhou, H., Xie, X. & Tang, Y. Engineering natural products using combinatorial biosynthesis and biocatalysis. *Curr. Opin. Biotechnol.* **19**, 590-596 (2008).

35. Menzella, H.G. et al. Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* **23**, 1171-1176 (2005).

36. Fischbach, M.A., Lai, J.R., Roche, E.D., Walsh, C.T. & Liu, D.R. Directed evolution can rapidly improve the activity of chimeric assembly-line enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 11951-11956 (2007).

37. Smolke, C.D., Martin, V.J.J. & Keasling, J.D. Tools for metabolic engineering in Escherichia coli. in *Protein Expression Technologies: Current Status and Future Trends* (ed. Baneyx, F.) 149-197 (Horizon Bioscience, Norfolk, 2004).

38. Adams, A., Gottschling, D., Kaiser, C. & Stearns, T. (eds.). *Methods in Yeast Genetics* (Cold Spring Harbor Laboratory Press Plainview, NY 1998).

39. Arnold, F.H. & Georgiou, G. (eds.). *Methods in Molecular Biology. Directed Evolution Library Creation: Methods and Protocols* (Humana, Totowa, NJ, 2003).

40. Neylon, C. Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: Library construction methods for directed evolution. *Nucleic Acids Res.* **32**, 1448-1459 (2004).

41. Quan, J. et al. Parallel on-chip gene synthesis and application to optimization of protein expression. *Nat. Biotechnol.* **29**, 449-452 (2011).

42. Bieniossek, C. et al. Automated unrestricted multigene recombineering for multiprotein complex production. *Nat. Methods* **6**, 447-450 (2009).

43. Gibson, D.G., Smith, H.O., Hutchison, C.A., 3rd, Venter, J.C. & Merryman, C. Chemical synthesis of the mouse mitochondrial genome. *Nat. Methods* **7**, 901-903 (2010).

44. Gibson, D.G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343-345 (2009).

45. Li, M.Z. & Elledge, S.J. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods* **4**, 251-256 (2007).

46. Naesby, M. et al. Yeast artificial chromosomes employed for random assembly of biosynthetic pathways and production of diverse compounds in Saccharomyces cerevisiae. *Microb. Cell Fact.* **8**, 45 (2009).

47. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A modular cloning system for standardized assembly of multigene constructs. *PLoS One* **6**, e16765 (2011).

48. Kodumal, S.J. et al. Total synthesis of long DNA sequences: Synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 15573-15578 (2004).

49. Kodumal, S.J. & Santi, D.V. DNA ligation by selection. *Biotechniques* **37**, 34-40 (2004).

50. Reisinger, S.J., Patel, K.G. & Santi, D.V. Total synthesis of multi-kilobase DNA sequences from oligonucleotides. *Nat. Protoc.* **1**, 2596-2603 (2006).

51. Blake, W.J. et al. Pairwise selection assembly for sequence-independent construction of long-length DNA. *Nucleic Acids Res.* **38**, 2594-2602 (2010).

52. Ramon, A. & Smith, H.O. Single-step linker-based combinatorial assembly of promoter and gene cassettes for pathway engineering. *Biotechnol. Lett.* **33**, 549-555 (2011).

53. Farrar, K. & Donnison, I.S. Construction and screening of BAC libraries made from Brachypodium genomic DNA. *Nat. Protoc.* **2**, 1661-1674 (2007).

54. Zhu, H. & Dean, R.A. A novel method for increasing the transformation efficiency of Escherichia coli-application for bacterial artificial chromosome library construction. *Nucleic Acids Res.* **27**, 910-911 (1999).

55. Karlsson, E.N., Johansson, L., Holst, O. & Liden, G. Escherichia coli as a well-developed host for metabolic engineering. in *The Metabolic Pathway Engineering Handbook: Fundamentals* (ed. Smolke, C.D.) (CRC Press, Boca Raton, FL, 2010).

56. Flagfeldt, D.B., Siewers, V., Huang, L. & Nielsen, J. Characterization of chromosomal integration sites for heterologous gene expression in Saccharomyces cerevisiae. *Yeast* **26**, 545-551 (2009).

57. Itaya, M., Tsuge, K., Koizumi, M. & Fujita, K. Combining two genomes in one cell: Stable cloning of the Synechocystis PCC6803 genome in the Bacillus subtilis 168 genome. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15971-15976 (2005).

58. Kuhlman, T.E. & Cox, E.C. Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Res.* **38**, e92 (2010).

59. Itaya, M., Fujita, K., Kuroki, A. & Tsuge, K. Bottom-up genome assembly using the Bacillus subtilis genome vector. *Nat. Methods* **5**, 41-43 (2008).

60. Shao, Z.Y., Zhao, H. & Zhao, H.M. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.* **37**, e16 (2009).

61. Shao, Z., Luo, Y. & Zhao, H. Rapid characterization and engineering of natural product biosynthetic pathways via DNA assembler. *Mol. Biosyst.* **7**, 1056-1059 (2011).

62. Gibson, D.G. et al. Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science* **319**, 1215-1220 (2008).

63. Gibson, D.G. et al. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20404-20409 (2008).

64. Sauer, U. Evolutionary engineering of industrially important microbial phenotypes. *Adv. Biochem. Eng. Biotechnol.* **73**, 129-169 (2001).

65. Baaquie, B.E. & Willeboordse, F.H. *Exploring Integrated Science*, 334 (CRC Press, Boca Raton, FL, 2010).

66. Otero, J.M. et al. Whole genome sequencing of Saccharomyces cerevisiae: From genotype to phenotype for improved metabolic engineering applications. *BMC Genomics* **11**, 723 (2010).

67. Camps, M., Naukkarinen, J., Johnson, B.P. & Loeb, L.A. Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9727-9732 (2003).

68. Costantino, N. & Court, D.L. Enhanced levels of lambda red-mediated recombinants in mismatch repair mutants. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 15748-15753 (2003).

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

70. Romanini, D.W., Peralta-Yahya, P., Mondol, V. & Cornish, V.W. A heritable recombination system for synthetic darwinian evolution in living cells. *Submitted*.

71. Bikard, D., Julie-Galau, S., Cambray, G. & Mazel, D. The synthetic integron: An in vivo genetic shuffling device. *Nucleic Acids Res.* **38**, e153 (2010).

72. Mazel, D. Integrons: Agents of bacterial evolution. *Nat. Rev. Microbiol.* **4**, 608-620 (2006).

73. Farmer, W.R. & Liao, J.C. Improving lycopene production in Escherichia coli by engineering metabolic control. *Nat. Biotechnol.* **18**, 533-537 (2000).

74. Ajikumar, P.K. et al. Terpenoids: Opportunities for biosynthesis of natural product drugs using engineered microorganisms. *Mol. Pharmaceutics* **5**, 167-190 (2008).

75. Otey, C.R. & Joern, J.M. High-throughput screen for aromatic hydroxylation. in *Methods in Molecular Biology. Directed Enzyme Evolution: Screening and Selection Methods*, Vol. 230 (eds. Arnold, F.H. & Georgiou, G.) 141-148 (Humana, Totowa, NJ, 2003).

76. Joo, H., Lin, Z. & Arnold, F.H. Laboratory evolution of peroxide-mediated cytochrome P450 hydroxylation. *Nature* **399**, 670-673 (1999).

77. Doi, N. & Yanagawa, H. Design of generic biosensors based on green fluorescent proteins with allosteric sites by directed evolution. *FEBS Lett.* **453**, 305-307 (1999).

78. Tucker, C.L. & Fields, S. A yeast sensor of ligand binding. *Nat. Biotechnol.* **19**, 1042-1046 (2001).

79. Pavoor, T.V., Cho, Y.K. & Shusta, E.V. Development of GFP-based biosensors possessing the binding properties of antibodies. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 11895-11900 (2009).

80. Wang, M.M. & Reed, R.R. Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast. *Nature* **364**, 121-126 (1993).

81. Fields, S. & Song, O. A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245-246 (1989).

82. SenGupta, D.J. et al. A three-hybrid system to detect RNA-protein interactions in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 8496-8501 (1996).

83. Licitra, E.J. & Liu, J.O. A three-hybrid system for detecting small ligand-protein receptor interactions. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 12817-12821 (1996).

84. Baker, K. et al. Chemical complementation: A reaction-independent genetic assay for enzyme catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16537-16542 (2002).

85. Lin, H., Tao, H. & Cornish, V.W. Directed evolution of a glycosynthase via chemical complementation. *J. Am. Chem. Soc.* **126**, 15051-15059 (2004).

86. Peralta-Yahya, P., Carter, B.T., Lin, H., Tao, H. & Cornish, V.W. High-throughput selection for cellulase catalysts using chemical complementation. *J. Am. Chem. Soc.* **130**, 17446-17452 (2008).

87. Santoro, S.W. & Schultz, P.G. Directed evolution of the site specificity of Cre recombinase. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 4185-4190 (2002).

88. Santoro, S.W., Wang, L., Herberich, B., King, D.S. & Schultz, P.G. An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat. Biotechnol.* **20**, 1044-1048 (2002).

89. Agresti, J.J. et al. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 4004-4009 (2010).

90. Ozawa, T., Sako, Y., Sato, M., Kitamura, T. & Umezawa, Y. A genetic approach to identifying mitochondrial proteins. *Nat. Biotechnol.* **21**, 287-293 (2003).

91. Link, A.J., Jeong, K.J. & Georgiou, G. Beyond toothpicks: New methods for isolating mutant bacteria. *Nat. Rev. Microbiol.* **5**, 680-688 (2007).

92. Yano, T., Oue, S. & Kagamiyama, H. Directed evolution of an aspartate aminotransferase with new substrate specificities. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 5511-5515 (1998).

93. Wang, L., Brock, A., Herberich, B. & Schultz, P.G. Expanding the genetic code of Escherichia coli. *Science* **292**, 498-500 (2001).

94. Castle, L.A. et al. Discovery and directed evolution of a glyphosate tolerance gene. *Science* **304**, 1151-1154 (2004).

95. Bartel, P., Chien, C.T., Sternglanz, R. & Fields, S. Elimination of false positives that arise in using the two-hybrid system. *Biotechniques* **14**, 920-924 (1993).

96. Kleeb, A.C. et al. Metabolic engineering of a genetic selection system with tunable stringency. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 13907-13912 (2007).

97. Neuenschwander, M., Butz, M., Heintz, C., Kast, P. & Hilvert, D. A simple selection strategy for evolving highly efficient enzymes. *Nat. Biotechnol.* **25**, 1145-1147 (2007).

98. Ellis, T., Wang, X. & Collins, J.J. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat. Biotechnol.* **27**, 465-471 (2009).

99. Yu, R.C. et al. Negative feedback that improves information transmission in yeast signalling. *Nature* **456**, 755-761 (2008).

100. Esvelt, K.M., Carlson, J.C. & Liu, D.R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499-503 (2011).

101. Dahiyat, B.I. & Mayo, S.L. De novo protein design: Fully automated sequence selection. *Science* **278**, 82-87 (1997).

102. Looger, L.L. & Hellinga, H.W. Generalized dead-end elimination algorithms make large-scale protein side-chain structure prediction tractable: Implications for protein design and structural genomics. *J. Mol. Biol.* **307**, 429-445 (2001).

103. Jiang, L. et al. De novo computational design of retro-aldol enzymes. *Science* **319**, 1387-1391 (2008).

104. Rothlisberger, D. et al. Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190-195 (2008).

105. Siegel, J.B. et al. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* **329**, 309-313 (2010).

106. Baker, D. An exciting but challenging road ahead for computational enzyme design. *Protein Sci.* **19**, 1817-1819 (2010).

Chapter 2

An Experimental Framework for the Characterization and Optimization of Yeast Hybrid Selections

---

*The content of this chapter will be published in

L.M. Wingler, V.W. Cornish. "An Experimental Framework for the Characterization and Optimization of Yeast Hybrid Selections," *in preparation*.

2.0 Chapter outlook

Growth selections can provide powerful, very high-throughput assays for *in vivo* directed evolution experiments. Most growth selections used for directed evolution are adapted from selectable markers used for conventional genetic applications, but it has been underexplored whether these markers retain their high throughputs—their ability to search libraries of $\geq 10^6$—in more complex, engineered systems. Rigorous characterization of the performance of these standard markers in the context of the specific application of interest will be imperative for their effective use in directed evolution. In this chapter, we develop an experimental framework to quantitatively assess the performance of yeast two- and three-hybrid growth selections and apply it to optimization of the throughput of the yeast two-hybrid *LEU2* growth selection. This framework is based on 1) comparison of the efficacy of the selectable marker in systems of increasing complexity, 2) detailed analysis of strains' growth, and 3) mock selections for enrichment, which are facilitated by a colorimetric enrichment assay that can be performed in medium throughput. For yeast two- and three-hybrid systems utilizing the *LEU2* reporter, throughputs were found to be comparable to each other but lower than when *LEU2* was used as a digital "on/off" marker, as it is in conventional genetic selections. Based on information obtained by analyzing the growth of these strains, we were able to make very simple modifications to the yeast two-hybrid selection protocol that improved its enrichment factor by two orders of magnitude and enabled large libraries of up to 10^7 to be searched. These results underscore the importance and value of quantitatively characterizing and optimizing growth selections as they are harnessed for directed evolution applications.

2.1 Introduction

Directed evolution holds promise for creating designer proteins for use as reagents and biosensors and for industrial applications such as bioenergy production. Progress in the directed evolution field over the past two decades shows that the key to achieving ever more dramatic changes in function is library size[1]. Screens of ~10^5 variants can be carried out with miniaturization of traditional assays for protein functions in a microtiter plate format. To test larger numbers of variants, however, selections are needed, where only the "winners" survive.

Yeast two-hybrid (Y2H) and three-hybrid (Y3H) systems represent potential methods for linking protein-protein and protein-small molecule interactions, respectively, to *in vivo* growth selections for directed evolution applications (**Chapter 1**, **Fig. 1-4)**[2,3]. Functional interactions involving the protein of interest reconstitute a transcriptional activator that leads to the expression of a reporter gene, which can be a selectable marker. Our laboratory's version of the yeast three-hybrid system uses the chemical dimerizer Dexamethasone-Methotrexate (Dex-Mtx) to reconstitute a transcriptional activator from LexA-dihydrofolate reductase (DHFR) and B42-glucocorticoid receptor (GR) fusion proteins[4]. By elaborating upon this three-hybrid scaffold, we further developed chemical complementation, which provides a general platform for linking enzymatic activity to reporter transcription via covalent coupling of Dexamethasone- and Methotrexate-derivatized substrates[5]. Using chemical complementation, we can now couple arbitrary enzyme activities to an *in vivo* growth assay for directed evolution[6].

For yeast n-hybrid-based selections to be useful for demanding applications in the field of directed evolution, they must be able to enrich active variants from a pool of ≥10^6

less active variants. Selectable markers that complement strains' auxotrophies or confer drug resistance exhibit these very high throughputs in the context of traditional genetic applications, such as plasmid maintenance or gene knockouts. However, it is underdetermined whether these markers retain their throughputs when appropriated for more complex synthetic biology applications such as yeast hybrid systems. Typically only qualitative analysis of growth selections is performed, and the focus of these studies has been on determining the sensitivity (i.e., the correlation between binding affinity and growth for yeast two- and three-hybrid systems) rather than the throughput of these assays[7,8].

Here, we develop an experimental framework to quantitatively monitor enrichment in yeast hybrid-based assays and identify key junctures for optimization. Our characterization strategy is based on three principles. First, using a reductionist approach, we evaluate the performance of the growth selection as layers of complexity are added in a stepwise manner—testing the performance of the reporter gene when used as a digital ("on or off") selectable marker as in standard genetic applications, then as a yeast two-hybrid and finally a yeast three-hybrid reporter. Second, we collect detailed data on the growth of these strains to identify potential ways to optimize enrichment. Third, we use mock selections to explicitly determine the throughput of these selections, developing a medium-throughput colorimetric assay to easily monitor enrichment. We apply this methodology to characterize and optimize the performance of the *LEU2* reporter gene in the yeast two- and three-hybrid assays, resulting in a >300-fold increase in the enrichment factor. This basic characterization framework should be useful not

only for yeast hybrid assays but also more generally for the application of reporter genes to directed evolution and synthetic biology.

2.2 Results

2.2.1 An experimental framework for characterizing yeast hybrid *LEU2* growth selections

We chose the *LEU2* reporter gene as a model system to develop our experimental framework for characterization of yeast hybrid selections. The *LEU2* gene encodes the β-isopropylmalate dehydrogenase enzyme required for leucine biosynthesis in *Saccharomyces cerevisiae*[9] and is a commonly used auxotrophic marker for routine yeast manipulations[10] and for the yeast two-hybrid assay[8]. We previously used this selectable marker to convert our Dex-Mtx yeast three-hybrid system and chemical complementation into growth selections, and we demonstrated that we could search small libraries (10^2) of glycosynthase enzyme variants using plate-based growth selections[6].

We wanted to adapt yeast hybrid *LEU2* selections for liquid culture, which should have a number of advantages for more demanding directed evolution experiments. First, larger libraries can be searched in high-density liquid culture selections than in selections on plates[11]. Minimizing selection volumes will be particularly important for the expanded use of the yeast three-hybrid and chemical complementation selections, which require compounds that can be challenging to synthesize in large quantities[12]. Second, members of the library are forced to directly compete against each other for resources. Third, small differences in growth between strains can be amplified through multiple rounds of selection[13].

We began by constructing a series of strains to evaluate the performance of *LEU2* as a reporter for the yeast two- and three-hybrid systems. We also constructed analogous strains allowing us to determine the maximum attainable efficacy of the *LEU2* selection under the conditions of the hybrid assays—that is, when *LEU2* is simply used as a marker gene that is expressed from its endogenous promoter. For each system, two strains were constructed—a "positive" strain containing all constructs required to activate *LEU2* transcription under selective conditions and a "null" strain lacking one of the requisite constructs (i.e., the B42 activation domain (AD) fusion protein for the hybrid systems or *LEU2* under the control of its endogenous promoter for the *LEU2* marker system) (**Fig. 2-1**). Each strain also contained a constitutively expressed colorimetric marker gene, either *gusA* (β-glucuronidase), which causes colonies to turn blue in the presence of the compound X-Gluc, or *lacZ* (β-galactosidase), which causes colonies to turn red in the presence of the compound Magenta-Gal. These colorimetric markers gave us a medium-throughput method to monitor enrichment in each system, as described below.

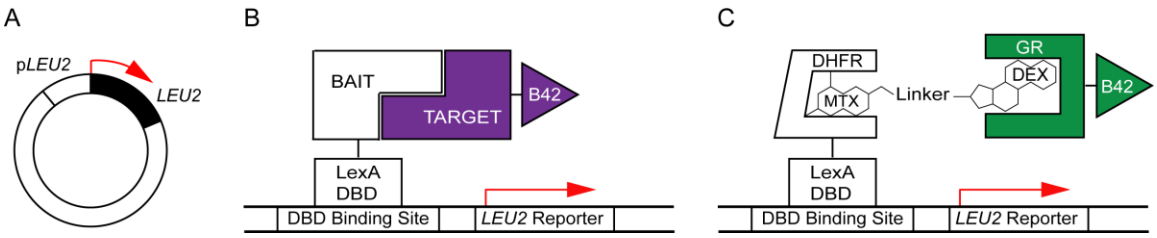


Figure 2-1. Strains for the characterization of yeast hybrid *LEU2* selections. In "positive" strains for the **(A)** plasmid-based *LEU2* marker **(B)** the yeast two-hybrid **(C)**, and the yeast three-hybrid systems, all constructs required to activate *LEU2* transcription were present. In otherwise isogenic "null strains," the shaded construct for each system was absent so that *LEU2* transcription should not be activated. In the *LEU2* marker system, strains also contained LexA-DHFR and the DNA-binding domain (DBD) binding site-*LEU2* reporter gene to facilitate a more direct comparison to the hybrid systems. For the yeast two-hybrid system, "BAIT" and "TARGET" are strongly interacting proteins provided with the commercial DupLEX-A yeast two-hybrid kit as positive controls.

2.2.2 Characterization of enrichment in *LEU2* selection systems

Next, we characterized the baseline throughput of the *LEU2* marker, two-hybrid, and three-hybrid systems. We performed mock selections in which the positive strain for each system was diluted with increasing excesses of the null strain and subjected the cell mixtures to selective conditions. We performed these mock selections at relatively high cell densities (~$3x10^7$ cells/mL) to push the limits of the *LEU2* selection and to enable us to search large libraries of up to 10^7 in even in small (<5 mL) culture volumes. As shown in **Figure 2-2**, the addition of a constitutively expressed colorimetric marker gave us a convenient assay to easily track enrichment in these mock selections. At various time points, aliquots of selections were plated on non-selective media, and the proportion of positive and null cells in the population could be determined simply by counting the number of blue and red colonies. Importantly, since several hundred colonies from the

non-selective plates could be counted, this colorimetric assay allowed us to detect when the positive strain comprises >~1% of the population. In a real directed evolution experiment, this would be the minimum enrichment required to detect true hits in a secondary screen of ~100 colonies, the maximum number it would be feasible to analyze with most low-throughput assays.

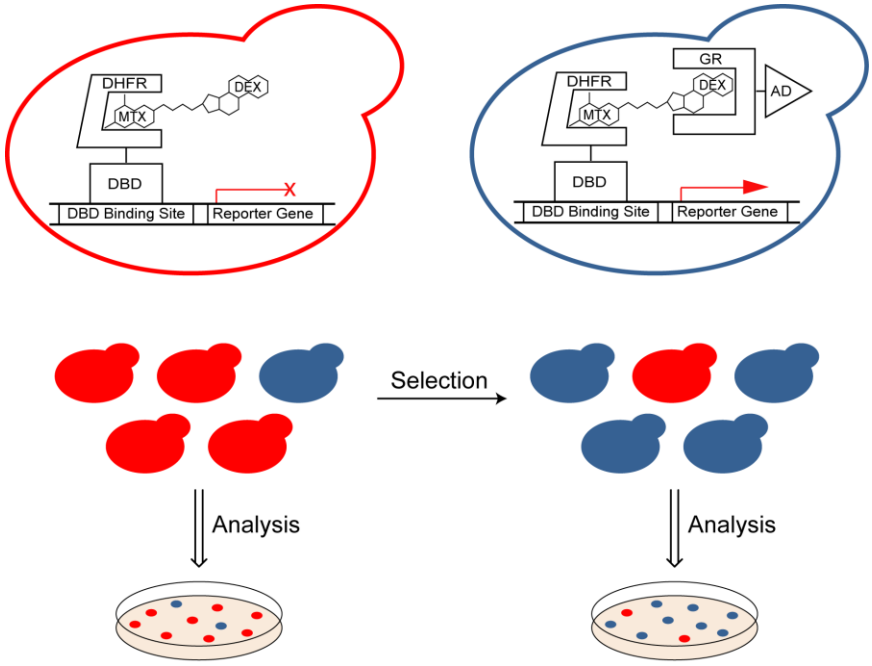


Figure 2-2. A colorimetric assay to monitor enrichment. As shown for the yeast three-hybrid system as an example, null and positive strains constitutively express β-galactosidase (*lacZ*, red) and β-glucuronidase (*gusA*, blue), respectively. To analyze the progress of mock selections, aliquots of cells from selections were plated on non-selective media at various time points. After colonies grew, the compounds X-Gluc and Magenta-Gal were used to assay for β-glucuronidase and β-galactosidase, respectively, and red and blue colonies were counted to determine the percentage of null and positive cells in the selection.

As shown in **Figure 2-3**, the ability of positive strains to overtake selections decreases in the more complex hybrid systems. When *LEU2* was used as a marker gene expressed from its constitutive promoter, the positive strain readily overtook the

population, representing approximately ~40% of the population after six days of selection under these conditions, even when initially diluted by a factor of up to 10^7. The yeast two-hybrid and three-hybrid positive strains, by contrast, always comprised a lower proportion of the population than the *LEU2* marker positive strain at the same time points and initial library sizes, and they could not recover the positive strain from libraries of 10^7.

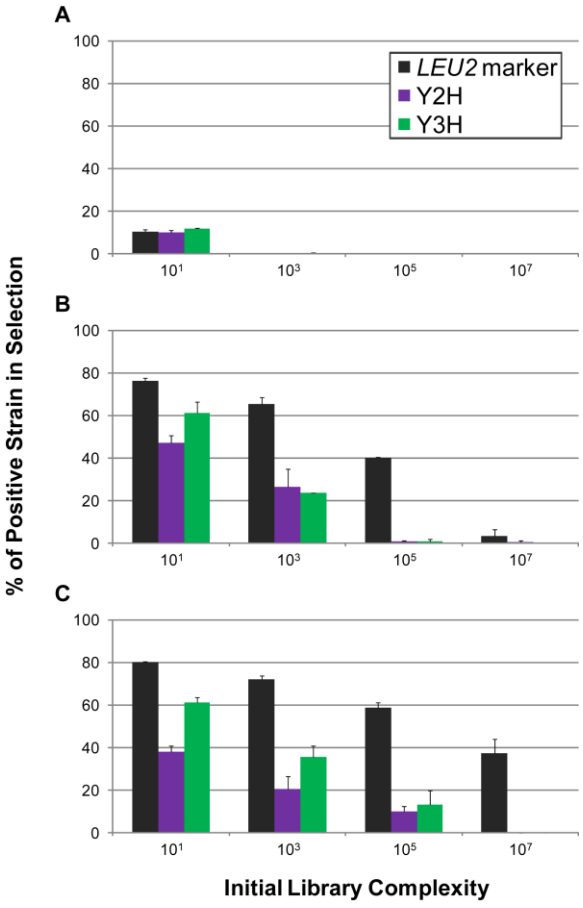


Figure 2-3. Assessment of the throughput of *LEU2* selection systems. Mock selections from increasing library sizes (10^1-10^7) for the *LEU2* marker, yeast two-hybrid, and yeast three-hybrid systems were assayed to determine the percentage of null and positive cells in the population after **(A)** 0, **(B)** 3, **(C)** and 6 days of selection using the colony color assay described above. Mock selections were performed in synthetic media containing 2% galactose, 2% raffinose, lacking tryptophan, histidine, and leucine. Yeast three-hybrid selective media also contained 1 μM Dexamethasone-Methotrexate. Error bars are the standard error of three replicate selections.

2.2.3 Characterization of strains' growth in *LEU2* selection systems

We then collected detailed growth curves for the null and positive strains under selective conditions to identify potential sources of the observed differences among these three systems. The difference in growth for the three positive strains lies in the length of

the lag phase before growth begins (**Fig. 2-4**). All positive strains have essentially identical growth rates, with doubling times of 3.5-4 hours during the exponential growth phase. This is consistent with previous reports that *LEU2* behaves as a "threshold reporter"[8]; a minimum level of transcription is required for growth to begin, but growth rates are independent of transcription levels above this threshold. As expected, none of the null strains grew significantly.

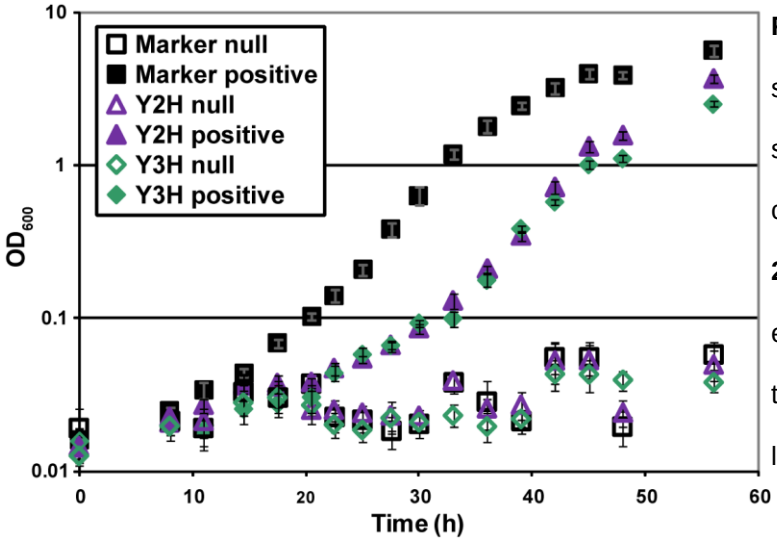


Figure 2-4. Growth of *LEU2* selection system strains under selective conditions. Media composition was as in **Figure 2-3**. Error bars are the standard error of six replicates. Note that the y-axis is plotted on a logarithmic scale.

The LexA and B42 fusion proteins are driven by the galactose-inducible and glucose-repressed *GAL1* promoter. The observed lag phase for the positive hybrid strains relative to the *LEU2* marker strain could be due to the time required to synthesize the fusion proteins after moving cells into galactose-containing media at the beginning of the selection, before which *LEU2* transcription cannot even begin. Accordingly, we looked at the effect of inducing fusion protein expression before beginning selections. After testing a wide variety of pre-induction conditions, we found that growing the yeast two-hybrid positive strain in non-selective galactose media for 24 hours before beginning the selection essentially eliminated the lag phase (**Fig. 2-5**).

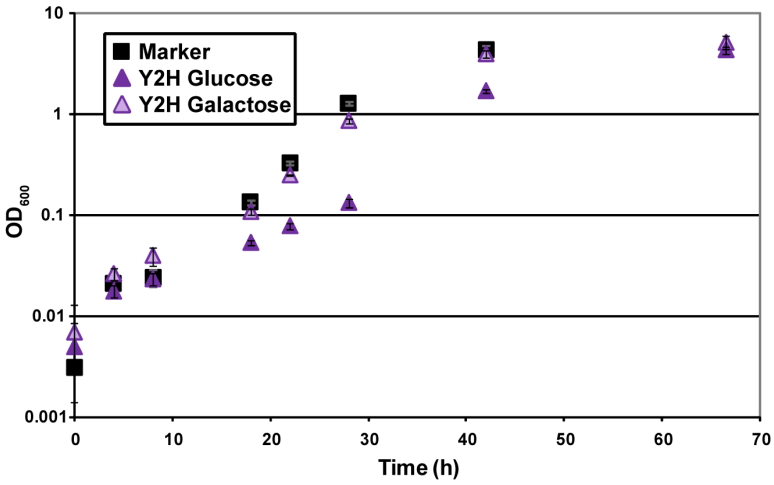


Figure 2-5. Growth curve optimization of the yeast two-hybrid selection. The yeast two-hybrid positive strain was grown in synthetic media lacking histidine and tryptophan and containing either 2% glucose or 2% galactose, 2% raffinose for 24 hours prior to beginning the selection at time = 0 hours. The *LEU2* marker strain was grown in the same way as the yeast two-hybrid glucose sample. Selective media was as in **Figure 2-3**. Error bars are the standard error of four replicates. Note that the y-axis is plotted on a logarithmic scale.

2.2.4 Optimization of enrichment in the yeast two-hybrid *LEU2* selection

Finally, we used the information from the growth curves to optimize the throughput of the yeast two-hybrid *LEU2* selection. Applying the galactose pre-induction of fusion protein expression to mock selections increased the percentage of positive cells in the selection population under most conditions (**Fig. 2-6**). In particular, the positive strain from a 10^5 library was enriched to a significant proportion of the population (>10%) at earlier time points (3 days), and it could be reproducibly enriched to detectable levels even from libraries of 10^7. This equates to an over 70-fold increase in the enrichment factor for the galactose pre-induction as compared to the glucose pre-induction at the same time point, where the enrichment factor is defined as[14]:

$$\text{Enrichment factor} = \frac{(\#\text{ positive cells}/\#\text{ negative cells})_{\text{final}}}{(\#\text{ positive cells}/\#\text{ negative cells})_{\text{initial}}}$$

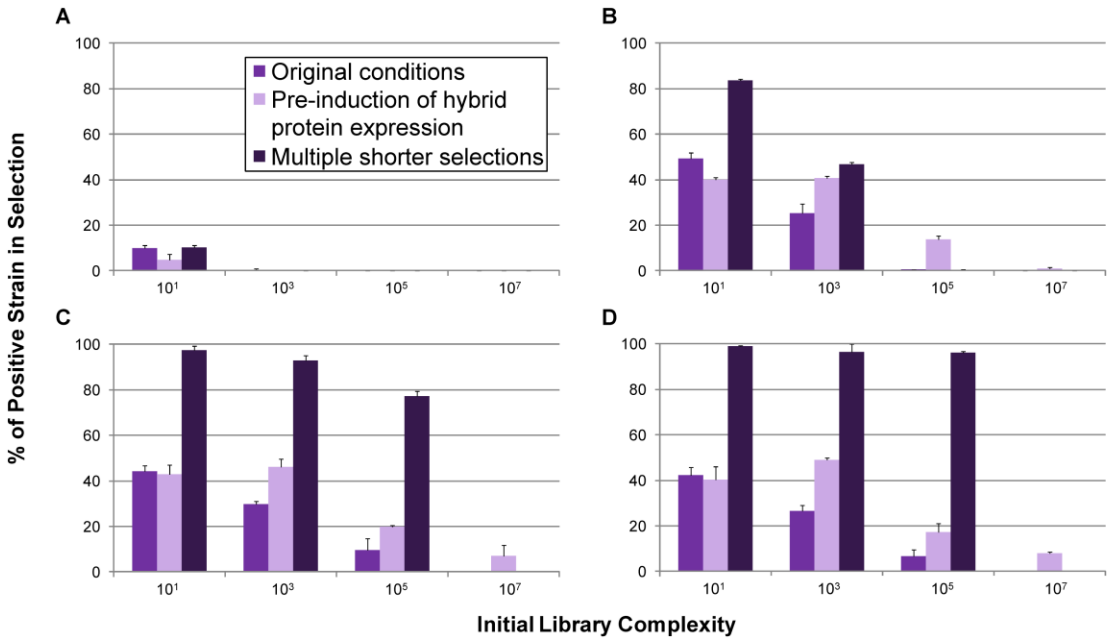


Figure 2-6. Optimization of yeast two-hybrid mock selection conditions. Mock selections from increasing library sizes (10^1-10^7) for the *LEU2* yeast two-hybrid system were assayed to determine the percentage of null and positive cells in the population after **(A)** 0, **(B)** 3, **(C)** 6 , and **(D)** 9 days of selection. The "original conditions" were as described in **Figure 2-3**. "Pre-induction of hybrid protein expression" involved a 24-hour induction in non-selective galactose media, as described in **Figure 2-5**. For "multiple shorter selections," the OD_{600} of the selection was adjusted to 1.0 once a day, removing cells and adding fresh media to maintain a constant culture volume. Error bars represent the standard error of two replicate selections.

Given that the growth curves indicated that the positive strain began growing rapidly, within 24 hours of beginning the selection (**Fig. 2-5**), we also tested the effect of performing multiple, shorter rounds of selection rather than a single extended selection (**Fig. 2-6**). Maintaining selection cultures at a constant cell density and volume allowed the positive strain to dominate the population, comprising >95% of cells, at the end of nine days of selection for mock libraries of up to 10^5. Compared to the original selection conditions, this improvement amounts to an over 300-fold increase in the enrichment factor for the 10^5 library. While this procedural modification did not recover the desired

strain from libraries of 10^7, combining multiple rounds of selection with the galactose pre-induction should allow the positive strain to overtake the population even from a library of 10^7.

2.3 Discussion

In this work, we have established an experimental framework that enables us to rigorously characterize and optimize enrichment in yeast n-hybrid growth selections. The development of quantitative, medium-throughput assays (i.e., detailed growth curves and a colorimetric enrichment assay) to evaluate these selections permitted us to analyze a number of conditions to improve their performance. Comparing systems of increasing complexity facilitated the identification of factors that diminished selections' efficacy. This analysis allowed us to make very simple alterations to the yeast two-hybrid selection that improved the enrichment factors and the searchable library size by two orders of magnitude.

Here we only looked at the throughput of yeast hybrid selections, or their ability to recover one cell with the desired function from a 10^n-fold excess of cells lacking it. Another important aspect of selections is their sensitivity to differences in the function of interest; this will be key for directed evolution experiments where researchers are endeavoring to maximize the affinity of an interaction. The methodology described in this chapter will be readily applicable to assessing this parameter. Strains expressing mutant fusion proteins with known, varying affinities for a protein (yeast two-hybrid) or small molecule (yeast three-hybrid) can be constructed[8,15], and their growth and performance in mock selections can be evaluated, as we did for the null and positive strains described above. Previous qualitative analyses of the correlation between

interaction strength and growth in the yeast two-hybrid assay suggest that the *LEU2* reporter may only be capable of determining the presence or absence of an interaction above a certain affinity threshold rather than gauging its strength more quantitatively[8]. Fortunately, this threshold seems to be adjustable if features of the *LEU2* reporter gene, such as the number of DNA-binding sites in its promoter, are altered[8]. This would argue for continuing to focus on optimizing the throughput of the *LEU2* selection, as described in this chapter, for a variety of reporter constructs so that increasingly stringent strains could be employed as a function of interest gradually improves over multiple rounds of directed evolution.

Beyond yeast hybrid systems, this same type of comprehensive characterization of other growth selections should facilitate the expansion of their use for directed evolution applications. Growth selections have already proven to be a powerful method for searching large libraries of mutants for directed evolution applications. Indeed, the directed evolution experiments that have achieved the most substantial changes in the function of biomolecules have almost invariably employed such selections in the evolution process[16-18]. However, there seems to be a general hesitancy in the field to embrace growth selections as the assay of choice, likely because their performance parameters are still ill-defined. Advances in directed evolution utilizing *in vitro* display selections and microtiter plate screens have repeatedly illustrated that systematic calibration and tuning of assays is critical for the success of directed evolution applications[19-23], yet the equivalent characterization experiments[11,24] have only rarely been performed for *in vivo* growth selections. Most researchers only provide a qualitative analysis of growth selections[8,25] or report endpoint data (e.g., the cell density at one time

point, or the number of days for colonies to appear in a plate assay)[26]. Our results indicated that far more detailed, quantitative analyses of growth selections are crucial for developing strategies to optimize their performance.

As directed evolution is increasingly harnessed to refine *in vivo* systems for synthetic biology rather than only individual biomolecules *in vitro*, having robust, well-characterized growth selections to connect a variety of functions to high-throughput assays will no longer be optional. Our experimental approach should provide a useful model for the rigorous assessment of growth selections for diverse applications.

2.4 Experimental methods

General materials and methods. Standard methods for molecular biology in *Saccharomyces cerevisiae* and *Escherichia coli* were used[10,27]. *S. cerevisiae* strains were grown at 30°C in media containing 2% glucose unless otherwise noted. Restriction enzymes and Vent DNA polymerase were purchased from New England Biolabs. Vent polymerase was used for all PCR reactions except yeast or *E. coli* colony PCR unless otherwise noted. For yeast colony PCR, cells were prepared according to a reported protocol (http://labs.fhcrc.org/hahn/Methods/mol_bio_meth/pcr_yeast_colony.html), and amplifications were performed with GoTaq DNA polymerase (Promega). The dNTPs used for PCR were purchased from GE Healthcare Life Sciences. Oligonucleotides were purchased from Invitrogen or Integrated DNA Technologies. DNA sequencing was performed by Genewiz. Plasmid DNA was purified using QIAprep miniprep kits (Qiagen); for yeast minipreps, cells were vortexed with acid-washed glass beads (Sigma) for five minutes before cell lysis. PCR products were purified with agarose gel electrophoresis and QIAquick spin columns purchased from Qiagen. Yeast genomic

DNA was purified using a YeaStar Genomic DNA Kit (Zymo Research). For the overlay assays, 5-bromo-4-chloro-3-indolyl-beta-D-glucuronic acid, cyclohexylammonium salt (X-Gluc) was obtained from Gold Biotechnology, and 5-bromo-6-chloro-3-indolyl-β-D-galactopyranoside (Magenta-Gal) from Biosynth International. DNA concentrations were determined by absorption at 260 nm, and all absorbance measurements were taken on a Molecular Devices SpectraMax Plus 384 instrument. All aqueous solutions were made with distilled water prepared from a Milli-Q Water System. For PCR, a MJ Research PTC-200 Pellier Thermal Cycler was employed. Transformation of *E. coli* was carried out by electroporation using a Bio-Rad *E. coli* Pulser. Yeast electroporation was carried out using a Bio-Rad Gene Pulser Xcell and a previously reported protocol[28].

Construction of the *LEU2* marker, yeast two-hybrid, and yeast three-hybrid "null strains." The *ADH* promoter and *lacZ* gene were amplified from vectors pMW103 (primers LMW203 and LMW204) and pMW112 (primers LMW201 and LMW202), respectively. Following gel purification, the fragments were combined in a fusion PCR reaction to create the p*ADH-lacZ* construct (primers LMW202 and LMW203, Accutaq polymerase). The purified construct was amplified with primers LMW205 and LMW206 to install an additional 30 bp of homology to the vector, in addition to the 30 bp of homology already provided by the fusion PCR primers. The plasmid pJG4-5 was digested with KpnI and EcoRI to remove the B42 fusion cassette and its promoter. A 1:10 molar mixture of cut plasmid (1.2 μg) and the p*ADH-lacZ* fusion product (8.0 μg) was transformed into V760Y via electroporation, and transformants were selected on SC(HT^-) (lacking histidine and tryptophan) plates. Transformants were analyzed via an X-Gal overlay assay, and one of the transformants that turned blue in the presence of X-

Gal was used to miniprep plasmid pLW2570. The plasmid pLW2570 was retransformed into V760Y, and pLW2570 and pBAIT were co-transformed into EGY48. One of the V760Y transformants that tested positive in the X-Gal overlay assay was used as the null strain for the *LEU2* marker and yeast three-hybrid assays (LW2630Y); one of the EGY48 transformants that tested positive was used as the null strain for the yeast two-hybrid assays (LW2631Y).

Construction of the *LEU2* marker "positive strain." The *ADH* promoter and *gusA* gene were amplified from vectors pMW103 (primers LMW209 and LMW210) and pDR8 (primers LMW207 and LMW208), respectively. Following gel purification, the fragments were combined in a fusion PCR reaction to create the p*ADH-gusA* construct (primers LMW208 and LMW209, Accutaq polymerase). The purified construct was digested with KpnI and EcoRI, ligated to pJG4-5 that had also been digested with these enzymes, and transformed into *E. coli*. The resulting plasmid pLW2569 was digested with ScaI and PvuII. The *LEU2* marker and its upstream regions were amplified from pRS425 with primers LMW211 and LMW213, which incorporated 30 bp of homology to the vector, and the purified PCR product was amplified with primers LMW212 and LMW214 to incorporate an additional 30 bp of homology. A 1:10 mixture of cut plasmid (1.1 μg) and the *LEU2* PCR product (3.4 μg) was transformed into V760Y, and transformants were selected on SC(HTL$^-$) (lacking histidine, tryptophan, and leucine) plates. Transformants were analyzed via an X-Gluc overlay assay, and one of the transformants that turned blue in the presence of X-Gluc was used to miniprep plasmid pLW2627. The plasmid pLW2627 was retransformed into V760Y, and one of the

transformants that tested positive in the X-Gluc overlay assay was used as the positive strain for the *LEU2* marker (LW2632Y).

Construction of the yeast two-hybrid "positive strain." The *ADH* promoter and *gusA* gene were amplified from vectors pMW103 (primers LMW216 and LMW210) and pDR8 (primers LMW207 and LMW215), respectively. Following gel purification, the fragments were combined in a fusion PCR reaction to create the p*ADH-gusA* construct (primers LMW215 and LMW216, Accutaq polymerase). The purified construct was amplified with primers LMW217 and LMW218 to install an additional 30 bp of homology to the vector, in addition to the 30 bp of homology already provided by the fusion PCR primers. The plasmid pTARGET was digested with ScaI. A ~1:10 mixture of cut plasmid (2.4 μg) and p*ADH-gusA* fusion (7.5 μg) was transformed into V760Y via electroporation, and transformants were selected on SC(HT$^-$) plates. Transformants were analyzed via an X-Gluc overlay assay, and one of the transformants that turned blue in the presence of X-Gluc was used to miniprep plasmid pLW2628. The plasmid pLW2628 and pBAIT were co-transformed into EGY48. One of the EGY48 transformants that tested positive in the X-Gluc overlay assay was used as the positive strain for the yeast two-hybrid assays (LW2633Y).

Construction of the yeast three-hybrid "positive strain." The *ADH* promoter and *gusA* gene were amplified from vectors pMW103 (primers LMW220 and LMW210) and pDR8 (primers LMW207 and LMW225), respectively. Following gel purification, the fragments were combined in a fusion PCR reaction to create the p*ADH-gusA* construct (primers LMW220 and LMW225, Accutaq polymerase). The purified construct was amplified with primers LMW221 and LMW226 to install an additional 30

bp of homology to the vector, in addition to the 30 bp of homology already provided by the fusion PCR primers. The plasmid pV398E was digested with PvuII. A ~1:10 mixture of cut plasmid (2.4 μg) and p*ADH-gusA* fusion (7.5 μg) was transformed into V760Y via electroporation, and transformants were selected on SC(HT^-) plates. Transformants were analyzed via an X-Gluc overlay assay, and one of the transformants that turned blue in the presence of X-Gluc was used to miniprep plasmid pLW2629. The plasmid pLW2629 was retransformed into V760Y, and one of the transformants that tested positive in the X-Gluc overlay assay was used as the positive strain for the yeast three-hybrid assay (LW2634Y).

Growth curves for *LEU2* selections. A representative experiment testing the effect of an induction before selection is described. Glycerol stocks of the strains LW2630Y, LW2631Y, LW2632Y, LW2633Y, and LW2634Y were used to inoculate 1-mL overnight cultures (SC(HT^-), 2% glucose). Cells were harvested (5000 rpm, 5 min, room temperature) and washed twice with sterile water. For each strain, the cells from 333 μL of the original culture were resuspended in SC(HT^-), 2% glucose, and cells from an equal volume of culture were resuspended in SC(HT^-), 2% galactose, 2% raffinose. After 24 hours of induction, 2 μL of cells were used to inoculate 198 μL of selection media (SC(HTL^-), 2% galactose, 2% raffinose). For yeast three-hybrid selections, 1 μM Dex-Mtx was also included in the selection media. At least 3 replicate wells were inoculated for each condition. Growth was monitored by absorption at 600 nm.

***LEU2* mock selections.** A representative experiment is described. One mL of SC(HT^-, 2% glucose) was inoculated from glycerol stocks of strains LW2630Y, LW2631Y, LW2632Y, LW2633Y, and LW2634Y. These starter cultures were used

inoculate overnight cultures (10-50 mL) of the same media. Cells were harvested, washed 3x with sterile water, and resuspended in selection media (SC(HTL$^-$), 2% galactose, 2% raffinose). The OD_{600} of the each strain was determined and cells were mixed to give 10:1, 10^3:1, 10^5:1, and 10^7:1 ratios of the null:positive strain for the *LEU2* marker selection (LW2630Y:LW2632Y), the yeast two-hybrid selection (LW2631Y:LW2633Y), and the yeast three-hybrid selection (LW2630Y:LW2634Y). For yeast three-hybrid selections, Dex-Mtx was added to a final concentration of 1 μM. Each selection had a final volume of 3 mL and a calculated initial OD_{600} of 1. All selections were set up in triplicate. Selections were shaken at 30°C. On days 0, 3, 6, and 9, cells were plated on SC(HT$^-$, 2% glucose), and after 2 days of growth, plates were assayed using the overlay assay described below. After the plates developed, the number of red, blue, and white/ambiguous colonies on each plate was counted.

Magenta-Gal/X-Gluc overlay assay. The overlay procedure was adapted from a reported protocol (http://biochemistry.ucsf.edu/labs/herskowitz/xgalagar.html). Potassium phosphate buffer (300 mL; 0.5 M, pH 7.0), 20 mL of DMF, 3.3 mL of 10% SDS, and 3.3 g low-melting agarose were combined in an Erlenmeyer flask. The solution was microwaved until the SDS and agarose went into solution. The flask was cooled in a 65°C water bath. β-mercaptoethanol (165 μL), X-Gluc (165 mg dissolved in 1 mL DMF), and Magenta-Gal (50 mg dissolved in 1 mL DMF) were added. After gentle mixing, a pipette was used to carefully cover each plate with approximately 10 mL of the agarose solution. The color typically developed sufficiently for plates to be counted within a few hours.

2.5 Strains, plasmids, and oligonucleotides

Table 2-1 Strains used in this study

Name	Genotype	Source/Reference
EGY48	*MATa trp1 his3 ura3 6LexAop-LEU2 GAL+*	R. Brent/8
V760Y	*MATα trp1 ura3 6LexAop-LEU2 ade4::pGAL1-LexA-eDHFR(HIS3) GAL+*	K. Baker/29
LW2630E	V760Y with pLW2570	This study
LW2631E	EGY48 with pLW2570 and pBAIT	This study
LW2632E	V760Y with pLW2627	This study
LW2633E	EGY48 with pLW2628 and pBAIT	This study
LW2634E	V760Y with pLW2629	This study

Table 2-2. Plasmids used in this study

Name	Details	Source/Reference
pBAIT	*pADH-LexA-BAIT 2μ HIS3* pBR ori amp^R	Origene
pDR8	*8LexAop-lacZ 3cIop-gusA 2μ URA3* colEI ori kan^R	I. Serebriiskii/30
pJG4-5	*pGAL1-B42 2μ TRP1* pUC ori amp^R	R. Brent/31
pMW103	*pGAL1-B42 2μ TRP1* pUC ori kan^R	R. Brent/32
pMW112	*8LexAop-lacZ 2μ URA3* pBR ori kan^R	R. Brent/32
pRS425	*2μ LEU2* pBIISK ori amp^R	ATCC #77106
pTARGET	*pGAL1-B42-TARGET 2μ TRP1* pUC ori amp^R	Origene
pBC398	*pGAL1-B42-(GSG)$_2$-rGR2 2μ TRP1* pUC ori kan^R	B. Carter/33
pLW2569	p*ADH1-gusA 2μ TRP1* pUC ori amp^R	This study
pLW2570	*pADH1-lacZ 2μ TRP1* pUC ori kan^R	This study
pLW2627	*pADH1-gusA 2μ LEU2 TRP1*	This study
pLW2628	pTARGET with p*ADH-gusA* replacing amp^R	This study
pLW2629	pBC398E with p*ADH-gusA* replacing (pUC ori kan^R)	This study

Table 2-3 Oligonucleotides used in this study

Name	Sequence (5'-3')
LMW201	GGGCGGAATGACTAAATCTCATTCAGAA
LMW202	GAAGAAGTCCAAAGCTTCTCGAGTCGGCCGTTATTTTTGACACCAGACCAA
LMW203	ATGACATGATTACGAATTAATTCGAGCTCGCAACTTCTTTTCTTTTTTTTTCT
LMW204	AGATTTAGTCATTCCGCCCGGAATTAAAGC
LMW205	ACAATTTCACACAGGAAACAGCTATGACATATGACATGATTACGAATTAATT
LMW206	CTTGATTGGAGACTTGACCAAACCTCTGGCGAAGAAGTCCAAAGCTTCT
LMW207	GGGCGGAATGTTACGTCCTGTAGAAAC
LMW208	GACGTGAATTCTTATCATTGTTTGCCTCCCTG
LMW209	ACGTCGGTACCCAACTTCTTTTCTTTTTTTTTCT
LMW210	GACGTAACATTCCGCCCGGAATTAAAGC

LMW211	TACACTATTCTCAGAATGACTTGGTTGAGTACAGGGGCGCTATCGCA
LMW212	ACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACT
LMW213	GCTTTCCAGTCGGGAAACCTGTCGTGCCAGTCGACTACGTCGTAAGGC
LMW214	CTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAAC
LMW215	CTCAGCGATCTGTCTATTTCGTTCATCCATTCATTGTTTGCCTCCCTG
LMW216	ATTCCCTTTTTTGCGGCATTTTGCCTTCAACTTCTTTTCTTTTTTTTTCT
LMW217	ATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATT
LMW218	TTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTT
LMW220	GCTTTCCAGTCGGGAAACCTGTCGTGCCAGCAACTTCTTTTCTTTTTTTTTCT
LMW221	CTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAAC
LMW225	AAGGAAGAGTCCTGAGGCGGAAAGAACCAGTCATTGTTTGCCTCCCTG
LMW226	ACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTCCTGAGGC

2.6 References

1. Aharoni, A., Griffiths, A.D. & Tawfik, D.S. High-throughput screens and selections of enzyme-encoding genes. *Curr. Opin. Chem. Biol.* **9**, 210-216 (2005).

2. Koide, A., Abbatiello, S., Rothgery, L. & Koide, S. Probing protein conformational changes in living cells by using designer binding proteins: Application to the estrogen receptor. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 1253-1258 (2002).

3. Bickle, M.B., Dusserre, E., Moncorge, O., Bottin, H. & Colas, P. Selection and characterization of large collections of peptide aptamers through optimized yeast two-hybrid procedures. *Nat. Protoc.* **1**, 1066-1091 (2006).

4. Lin, H.N., Abida, W.M., Sauer, R.T. & Cornish, V.W. Dexamethasone-methotrexate: An efficient chemical inducer of protein dimerization in vivo. *J. Am. Chem. Soc.* **122**, 4247-4248 (2000).

5. Baker, K. et al. Chemical complementation: A reaction-independent genetic assay for enzyme catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16537-16542 (2002).

6. Lin, H., Tao, H. & Cornish, V.W. Directed evolution of a glycosynthase via chemical complementation. *J. Am. Chem. Soc.* **126**, 15051-15059 (2004).

7. Hook, B., Bernstein, D., Zhang, B. & Wickens, M. RNA-protein interactions in the yeast three-hybrid system: Affinity, sensitivity, and enhanced library screening. *RNA* **11**, 227-233 (2005).

8. Estojak, J., Brent, R. & Golemis, E.A. Correlation of two-hybrid affinity data with in vitro measurements. *Mol. Cell. Biol.* **15**, 5820-5829 (1995).

9. Kohlhaw, G.B. Beta-isopropylmalate dehydrogenase from yeast. *Methods Enzymol.* **166**, 429-435 (1988).

10. Ausubel, F. et al. *Current Protocols in Molecular Biology*, (Wiley, New York, 1995).

11. Muranaka, N., Sharma, V., Nomura, Y. & Yokobayashi, Y. An efficient platform for genetic selection and screening of gene switches in Escherichia coli. *Nucleic Acids Res.* **37**, e39 (2009).

12. Tao, H., Peralta-Yahya, P., Lin, H. & Cornish, V.W. Optimized design and synthesis of chemical dimerizer substrates for detection of glycosynthase activity via chemical complementation. *Bioorg. Med. Chem.* **14**, 6940-6953 (2006).

13. Pelletier, J.N., Arndt, K.M., Pluckthun, A. & Michnick, S.W. An in vivo library-versus-library selection of optimized protein-protein interactions. *Nat. Biotechnol.* **17**, 683-690 (1999).

14. Beutel, B.A. & Gold, L. In vitro evolution of intrinsically bent DNA. *J. Mol. Biol.* **228**, 803-812 (1992).

15. de Felipe, K.S., Carter, B.T., Althoff, E.A. & Cornish, V.W. Correlation between ligand-receptor affinity and the transcription readout in a yeast three-hybrid system. *Biochemistry* **43**, 10353-10363 (2004).

16. Yano, T., Oue, S. & Kagamiyama, H. Directed evolution of an aspartate aminotransferase with new substrate specificities. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 5511-5515 (1998).

17. Wang, L., Brock, A., Herberich, B. & Schultz, P.G. Expanding the genetic code of Escherichia coli. *Science* **292**, 498-500 (2001).

18. Castle, L.A. et al. Discovery and directed evolution of a glyphosate tolerance gene. *Science* **304**, 1151-1154 (2004).

19. Salazar, O. & Sun, L. Evaluating a screen and analysis of mutant libraries. *Methods Mol. Biol.* **230**, 85-97 (2003).

20. Hegmans, J.P. et al. A model system for optimising the selection of membrane antigen-specific human antibodies on intact cells using phage antibody display technology. *J. Immunol. Methods* **262**, 191-204 (2002).

21. Levitan, B. Stochastic modeling and optimization of phage display. *J. Mol. Biol.* **277**, 893-916 (1998).

22. Mandecki, W., Chen, Y.C.J. & Grihalde, N. A mathematical model for biopanning (affinity selection) using peptide libraries on filamentous phage. *J. Theor. Biol.* **176**, 523-530 (1995).

23. Irvine, D., Tuerk, C. & Gold, L. Selexion - systematic evolution of ligands by exponential enrichment with integrated optimization by nonlinear-analysis. *J. Mol. Biol.* **222**, 739-761 (1991).

24. Yokobayashi, Y. & Arnold, F.H. A dual selection module for directed evolution of genetic circuits. *Nat. Comput.* **4**, 245-254 (2005).

25. Neuenschwander, M., Butz, M., Heintz, C., Kast, P. & Hilvert, D. A simple selection strategy for evolving highly efficient enzymes. *Nat. Biotechnol.* **25**, 1145-1147 (2007).

26. Kleeb, A.C. et al. Metabolic engineering of a genetic selection system with tunable stringency. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 13907-13912 (2007).

27. Adams, A., Gottschling, D., Kaiser, C. & Stearns, T. (eds.). *Methods in Yeast Genetics* (Cold Spring Harbor Laboratory Press, Plainview, NY 1998).

28. Pirakitikulr, N., Ostrov, N., Peralta-Yahya, P. & Cornish, V.W. PCRless library mutagenesis via oligonucleotide recombination in yeast. *Protein Sci.* **19**, 2336-2346 (2010).

29. Baker, K., Sengupta, D., Salazar-Jimenez, G. & Cornish, V.W. An optimized dexamethasone-methotrexate yeast 3-hybrid system for high-throughput screening of small molecule-protein interactions. *Anal. Biochem.* **315**, 134-137 (2003).

30. Serebriiskii, I.G. et al. Detection of peptides, proteins, and drugs that selectively interact with protein targets. *Genome Res.* **12**, 1785-1791 (2002).

31. Gyuris, J., Golemis, E., Chertkov, H. & Brent, R. Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. *Cell* **75**, 791-803 (1993).

32. Watson, M.A., Buckholz, R. & Weiner, M.P. Vectors encoding alternative antibiotic resistance for use in the yeast two-hybrid system. *Biotechniques* **21**, 255-259 (1996).

33. Abida, W.M., Carter, B.T., Althoff, E.A., Lin, H. & Cornish, V.W. Receptor-dependence of the transcription read-out in a small-molecule three-hybrid system. *ChemBioChem* **3**, 887-895 (2002).

Chapter 3

Characterization and Discovery of

Yeast Three-Hybrid Counter Selections

---

*The content of this chapter has been partially published in

L.M. Wingler, V.W. Cornish. "A Library Approach for the Discovery of Customized Yeast Three-Hybrid Counter Selections," *ChemBioChem*, **12**, 715-717 (2011).

3.0 Chapter outlook

Adapting counter selections, or selections *against* activation of a reporter gene, for directed evolution applications is particularly challenging due to the need to match basal and activated levels of reporter gene expression to the gene product's toxicity level. Our laboratory previously developed a yeast three-hybrid counter selection based on the standard *URA3* counter selection from yeast genetics; elaboration of this yeast three-hybrid framework enabled the directed evolution of cellulases via our chemical complementation technology. However, the throughput of the *URA3* counter selection in our original system was too low for it to be utilized in demanding directed evolution experiments requiring large libraries to be searched. In this chapter, we describe several approaches we have taken to improve our yeast three-hybrid counter selection. We began by rigorously characterizing enrichment and growth of strains in the *URA3* counter selection for both the yeast three-hybrid system and analogous simpler systems, in which *URA3* is used simply as a digital "on/off" marker. We found that the poor performance of the *URA3* reporter in the yeast three-hybrid assay can be attributed to insufficient activation of the reporter under selection conditions rather than high basal transcription or inherent limitations in the efficacy of the selection. Given the difficulty of appropriately modulating the reporter's transcription levels, we developed a screening approach to empirically identify novel counter selectable reporter genes whose thresholds for toxicity match the basal and activated transcription levels already achieved in our system. A screen of only eleven candidate genes yielded a counter selection reporter that, without any optimization, performed more effectively than the extensively optimized *URA3* reporter in our yeast three-hybrid system.

3.1 Introduction

The rise of synthetic biology is creating a heightened demand for methods to connect complex *in vivo* circuitry to a readily assayable cellular phenotype, such as cell growth. To date, cell engineers have almost exclusively borrowed the standard selections for antibiotic resistance and complementation of auxotrophies long used by geneticists. However, these traditional selections, historically employed as digital "on/off" markers for purposes such as plasmid maintenance or gene knockouts, can be less effective when used for applications that require, for example, the detection of small changes in transcription levels or the observation of graded, analog responses.

The yeast three-hybrid (Y3H) assay is an example of a synthetic biology system that would benefit from customized selectable markers. Our laboratory recently developed a yeast three-hybrid counter selection based on the classic yeast *URA3* counter selection[1], in which the gene product OMP decarboxylase converts 5-fluoroorotic acid (5-FOA) to the toxic compound 5-fluorouracil. Without further modification, this yeast three-hybrid *URA3* selection strain was able to detect cellulase activity upon addition of a gene encoding a cellulase and a Methotrexate-Cellotetraose-Dexamethasone (Mtx-Cel-Dex) dimerizer substrate (**Fig. 3-1**)[2]. We then used this cellulase chemical complementation system to select improved cellulase variants from a library generated by family DNA shuffling[2]. However, adapting the *URA3* counter selection to provide the desired growth phenotype in the yeast three-hybrid system was nontrivial, requiring a multi-step strain construction, extensive optimization of growth conditions, and the screening of numerous strains[2]. Even after this optimization, the *URA3* reporter provided relatively small differences in growth in the activated versus the unactivated states in the

context of the yeast three-hybrid and cellulase chemical complementation counter selections[2]. While the performance of this counter selection was sufficient for initial proof-of-principle studies, the current *URA3* counter selection is unlikely to enable the effective search of large libraries for demanding directed evolution applications.

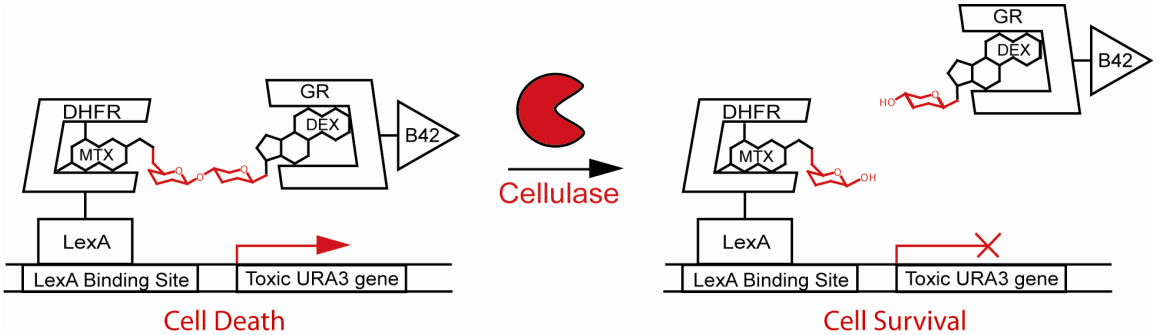


Figure 3-1. The chemical complementation counter selection provides a growth selection for cellulase catalysts. The Mtx-Cel-Dex heterodimer links a LexA DNA-binding domain fused to dihydrofolate reductase (DHFR) and a B42 activation domain fused to the glucocorticoid receptor (GR), activating transcription of a downstream toxic *URA3* reporter gene, thus leading to cell death. Active cellulases cleave the β-1,4-glucosidic bond in the cellotetraose linker between Mtx and Dex, halting transcription of the toxic reporter gene and leading to cell survival.

Historically, developing counter selections has been more difficult than developing positive selections (e.g., resistance to antibiotics or complementation of strain auxotrophies). Only a limited number of yeast counter selections have been reported[3-9], and even fewer have proven to be dependable enough to be widely employed even for routine genetic manipulations[10]. The *URA3* marker is considered to be the "gold standard" of these established selections because it is the most robust and reliable[7], and as such, it has been the reporter of choice for converting yeast hybrid systems into counter selections[11-13]. The two yeast two-hybrid systems that utilize this reporter did not quantitatively characterize its performance[11,12]. Using a yeast three-hybrid system with a

LexA DNA-binding domain and a *GAL4* activation domain, Chidley *et al*. recently reported a *URA3* counter selection designed to reduce the number of false positives identified in a positive three-hybrid selection[13]. However, mock selections indicated that their counter selection only provided 10- to 100-fold enrichment factors for large library sizes (i.e., enriching cells that do not activate reporter transcription from 1 in 10^4 to less than 1 in 10^2)[13]. While these very low enrichments were satisfactory for their purpose of eliminating low-frequency false positives, they would be entirely inadequate for systems such as chemical complementation, in which the counter selection is the primary selection method.

One explanation of why optimizing the most robust known counter selection for yeast hybrid systems has proven to be so difficult is that there are fundamental differences between the yeast three-hybrid assay and typical counter selection applications. When used for purposes such as curing cells of plasmids or knocking out genes, the counter selectable marker is either "on" and being expressed from its endogenous promoter, or "off" and completely deleted from the cell. In the yeast three-hybrid assay, however, the functional reporter gene is present in all cells, and the "on" and "off" states reflect activated and basal transcription of the gene, respectively. If the expression level at which the reporter gene begins to inhibit cell growth happens to fall outside this window, the counter selection will be unable to discern between activated and basal transcription, and low levels of basal transcription of the gene could inhibit cell growth even in the absence of the reconstituted transcription factor.

In this chapter, we take multiple approaches to improving our laboratory's Dexamethasone-Methotrexate-based yeast three-hybrid counter selection. First, we apply

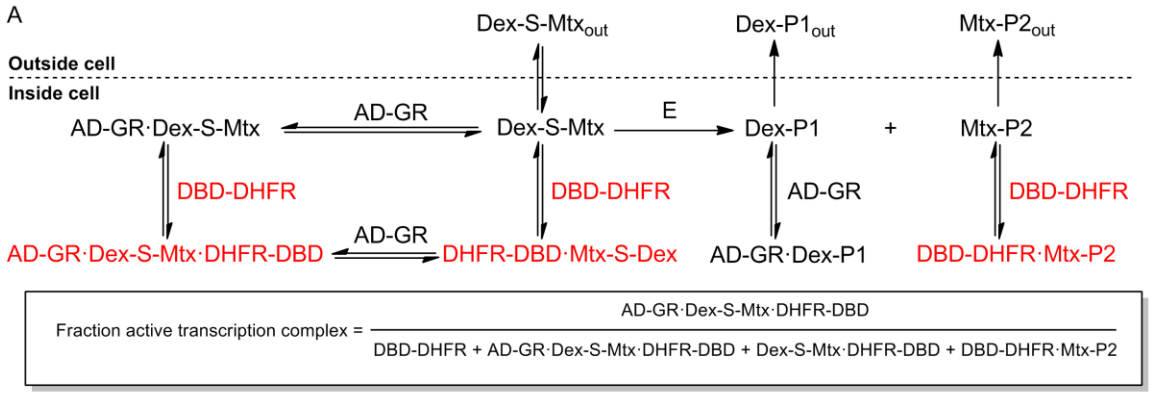
the framework for characterizing genetic selection systems outlined in **Chapter 2** to the *URA3* counter selection reporter. We identify the key issues impairing its performance and describe our efforts to optimize its regulation accordingly. In addition, we develop a library screening approach to discover new counter selection reporters for the yeast three-hybrid assay. A small screen yielded a reporter construct that, without any optimization, was more effective than the *URA3* reporter in our system.

3.2 Results

3.2.1 Model of the cellulase chemical complementation system

We first sought to confirm that our existing chemical complementation framework should be able to produce discernable differences in reporter gene expression in response to cellulase activity. We built a simple model to predict how changes in the system components should affect assembly of the active transcription complex (**Fig. 3-2**). Current transcriptional activation mechanism models[14] suggest that reporter gene transcription levels should correlate with the ratio of active transcription complex to total DNA-binding domain. Briefly, our chemical complementation model predicts that the dynamic range of the assay is determined by rate of conversion of substrate to product by the enzyme relative to the rate at which the heterodimer substrate and products can cross the plasma membrane and exchange with the extracellular media. While this model is oversimplified and leaves out major events such as transport of the transcription complex to the nucleus, it is compelling that the model correctly predicts that the assay should detect cellulases with activities on the order of 10^5–10^6 $M^{-1}s^{-1}$—the activity levels of the cellulases in our first directed evolution experiment[2]. These results suggest that with the

appropriate reporter gene, we should be able to distinguish between 1) active and inactive cellulase enzymes and 2) cellulases with high and intermediate levels of activity.



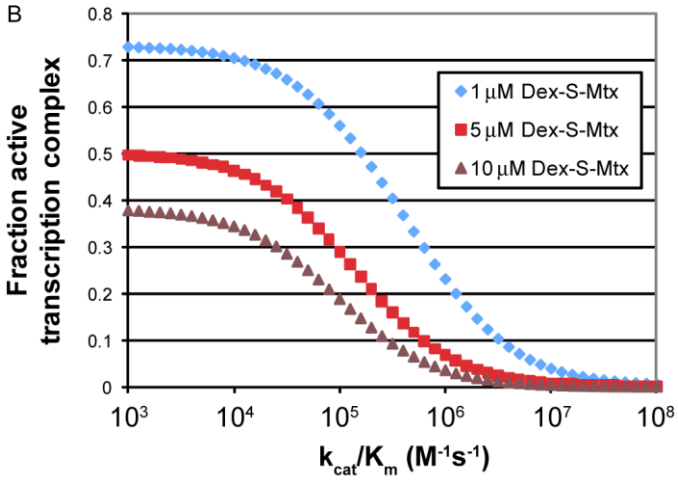


Figure 3-2. Model of the cellulase chemical complementation system predicting how the components of the system control the levels of reporter gene transcription and hence the dynamic range of the assay. **(A)** The fraction of DBD-DHFR that is part of an active transcription complex is expressed as a fraction of total DBD-DHFR; components of the model containing DBD-DHFR are highlighted in red. S represents the tetrasaccharide substrate; P1 and P2 represent the disaccharide products. Based on previous quantification of protein levels in our system by Western blotting[15], the concentrations of the DBD and AD fusion proteins are assumed to be 10 μM. Rate constants for small molecule transport in and out of the cell, believed to occur by diffusion, are estimated at 0.004 s^{-1} based on values for the cellular uptake of Dex[16]. Dissociation constants for the binding of Dex to GR and Mtx to DHFR are 1 nM and 1 pM, respectively. This model was built using COPASI[17]. **(B)** The fraction of active transcription complex varies with the catalytic efficiency of the enzyme and the concentration of the heterodimer substrate in our model. The enzyme concentration is assumed to be 10 nM, again based on previously reported quantification of enzyme levels in our system[18].

3.2.2 Characterization of the 8LexAop-p*SPO13-URA3* reporter gene

Having verified that chemical complementation could theoretically provide a usable selection for cellulase enzymes, we next tested our hypothesis that the suboptimal growth differences observed in our original cellulase chemical complementation selection were due to the performance of the *URA3* counter selection reporter. The reporter gene in our original chemical complementation selection strain contains the *URA3* gene under the control of the tightly regulated *SPO13* promoter downstream of eight LexA operators[2] (four complete *colEI* operators, each of which binds two LexA dimers[19]). Taking a reductionist approach analogous to that described in **Chapter 2**, we constructed a series of isogenic strains that differed only in the *URA3* allele they contained, allowing the performance of the *URA3* counter selection in the yeast three-hybrid system to be directly compared to its performance as when used as a digital selectable marker, as it is in conventional genetic applications.

To determine the maximum achievable dynamic range of the *URA3* counter selection under the yeast hybrid selection conditions, we first compared the growth of strains containing either a wild-type *URA3* gene expressed from its endogenous promoter or the inactive *ura3-52* allele, which produces no functional Ura3 protein. As shown in **Figure 3-3**, growth curves under counter selection conditions showed that strains containing the wild-type allele never grow to saturation, while cultures of strains with the inactive allele reach saturation (OD_{600}>1). Mock selections confirmed that cells with the inactive allele can be enriched to a significant proportion of the population (>20%) from an excess of at least 10^5 cells with the wild-type gene (**Fig. 3-4**). These observations indicate that *URA3* should be an effective reporter for the chemical complementation

counter selection if sufficient differences in basal versus activated *URA3* transcription can be obtained.

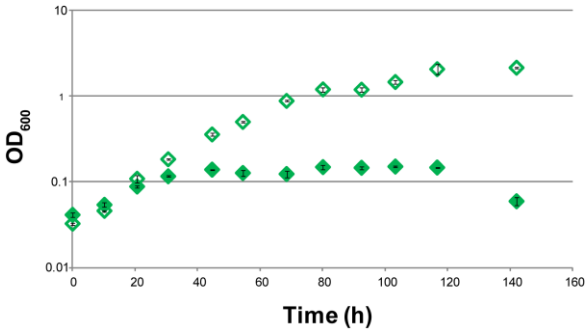


Figure 3-3. Growth of *ura3* and *URA3* strains under yeast three-hybrid counter selection conditions. Growth curves were performed in synthetic media containing 2% galactose, 2% raffinose, and 0.2% 5-FOA, lacking tryptophan and histidine. Open and closed symbols represent strains with the inactive *ura3-52* allele and the active *URA3* allele, respectively. A total of four colonies from each transformation were analyzed with similar results; representative data from a single colony from each transformation are shown. Error bars represent the standard error in the OD_{600} readings for triplicate cultures. Note that the y-axis is plotted on a logarithmic scale.

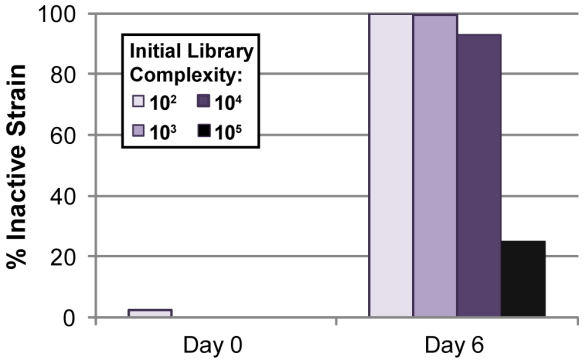


Figure 3-4. Enrichment of *ura3* strains under yeast three-hybrid counter selection conditions. Mock selections were performed in which the *ura3-52* strain, with the inactive selectable marker, was mixed with increasing excesses (10^2-10^5) of the *URA3* strain. After 0 and 6 days of growth in selective media (same composition as **Figure 3-3**), the percentage of *ura3* cells in the population was determined using the colorimetric assay described in **Chapter 2**.

Next, we compared the growth of yeast three-hybrid strains containing the integrated 8LexAop-p*SPO13-URA3* reporter with the LexA-DHFR fusion protein required to activate *URA3* transcription (active yeast three-hybrid strain), or with only LexA (inactive yeast three-hybrid strain). Growth curves with yeast three-hybrid strains that were otherwise isogenic to the control strains described above confirmed that they did not grow slower than *ura3-52* strains in the presence of 5-FOA (compare **Figs. 3-3**

and **3-5**, open symbols), demonstrating that basal reporter transcription was not problematic. As expected, for inactive yeast three-hybrid strains with only LexA, growth curves with and without the chemical dimerizer Dex-Mtx were superimposable (**Fig. 3-5B**). For the active yeast three-hybrid strain, addition of Dex-Mtx resulted in a reproducible but small degree of growth inhibition (**Fig. 3-5A**). Mock selections indicated that these slight differences in growth were insufficient to enrich an inactive yeast three-hybrid strain from only a small excess (10-fold) of active yeast three-hybrid cells even after nine days of selection (**Fig. 3-6**). These data indicate that 1) the poor performance of the cellulase chemical complementation selection can be traced back to its yeast three-hybrid framework and 2) the low throughput is due to insufficient activation of the *URA3* reporter under activated conditions rather than high basal activation levels under non-activated conditions.

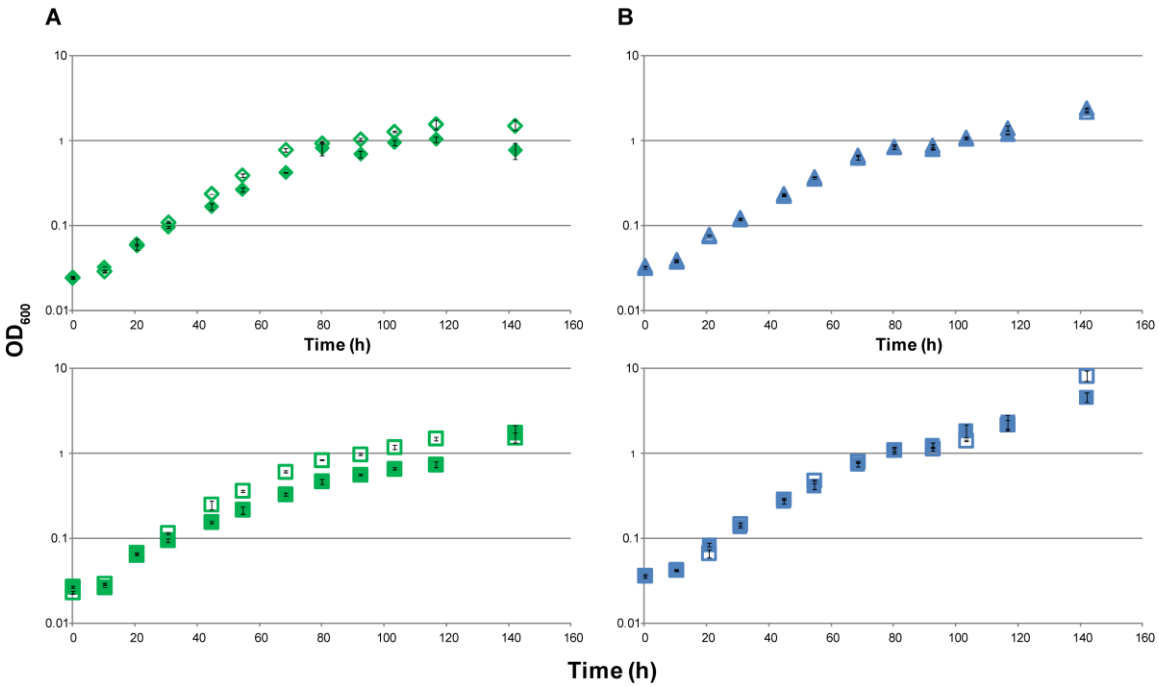


Figure 3-5. Growth of **(A)** active and **(B)** inactive yeast three-hybrid strains under *URA3* counter selection conditions. Media composition was the same as in **Figure 3-3**; open and closed symbols represent cultures grown in the absence and presence of 5 μM Dex-Mtx, respectively. Each graph represents the results for a unique, randomly selected colony. A total of four colonies from each transformation were analyzed with similar results. Error bars represent the standard error in the OD_{600} readings for duplicate or triplicate cultures. Note that the y-axis is plotted on a logarithmic scale.

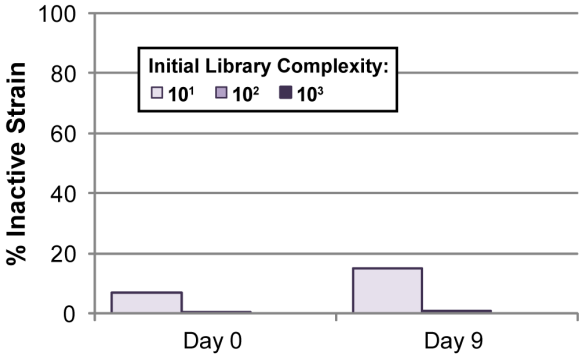


Figure 3-6. Enrichment of inactive yeast three-hybrid strains in the *URA3* counter selection. Mock selections were performed in which the inactive yeast three-hybrid strain was mixed with increasing excesses (10^1-10^3) of the active yeast three-hybrid strain. After various amounts of time in selective media (same composition as **Figure 3-3**, with the addition of 5 μM Dex-Mtx), the percentage of inactive yeast three-hybrid cells in the population was determined using the colorimetric assay described in **Chapter 2**.

3.2.3 Efforts to optimize the *URA3* yeast-three hybrid reporter gene

Based on these results, we tried several alternative *URA3* reporter constructs that could potentially increase the expression of the reporter under activated conditions. First, since the original reporter gene was integrated in the chromosome and only present in a single copy in cells, we increased the reporter's copy number by placing the LexAop-p*SPO13-URA3* construct on a high-copy plasmid. However, these strains grew just as robustly as the *ura3-52* strain even in the presence of Dex-Mtx, indicating that transcription of *URA3* was still too low under activated conditions (data not shown).

Next, we tested alternative promoter constructs known to provide large differences in reporter transcription levels. Specifically, we placed *URA3* under the control of the 2- and 8LexAop-p*GAL*-based promoters on low- or high-copy plasmids. In these promoter constructs, the upstream activation sequences of the *GAL* promoter are replaced with LexA operators, and in our Dex-Mtx yeast three-hybrid system, they exhibit up to a 100-fold difference in basal versus activated transcription (see **Fig. 3-8** in **Section 3.2.4**). However, basal transcription from all of these promoters was evidently too high. Most colonies did not grow with or without Dex-Mtx, and colonies that did grow either did not display small molecule-dependent growth inhibition or exhibited variation among replicates, suggesting that growth resulted from strain reversion (data not shown).

3.2.4 A library approach for the discovery of yeast-three hybrid counter selection reporter genes

Given the difficulty of rationally redesigning the *URA3* yeast three-hybrid system to appropriately modulate the reporter gene's expression, we hypothesized that the most

straightforward route to improve the counter selection would be to screen a library of novel candidate reporter genes in the context of the desired application. We could thereby empirically identify one whose threshold for toxicity corresponds to the expression levels achieved in our existing system. Since reporter expression is not activated in the absence of the chemical dimerizer, a conditionally lethal reporter (e.g., only toxic in the presence of a compound such as 5-FOA) is unnecessary. This broadens the pool of candidate counter selection reporters greatly, as a wealth of genetics studies have identified numerous yeast genes that inhibit growth or affect the cell cycle when simply overexpressed[20-25]. Interestingly, in spite of the paucity of effective yeast counter selections, this information has seldom been exploited to develop new counter selections[9,20,26]. We thought that an endogenous yeast gene could be a particularly suitable yeast three-hybrid counter selection reporter, as the gene's mere presence in the genome implies that the cell is able to tolerate some basal level of expression. Accordingly, we drew a list of eleven potential candidate counter selection reporter genes (**Table 3-1**) from overexpression studies in *S. cerevisiae*. Since each previous study used different expression conditions, we focused on genes that had been identified in multiple screens to increase the likelihood that the gene product would be toxic under the specific conditions of our assay.

Table 3-1. Candidate counter selection reporter genes.

Gene	Reported Studies
ACT1 (YFL039C)	Refs. 21-23,25
AMN1 (YBR158W)	Ref. 24
CDH1 (YGL003C)	Refs. 21,25
GIS1 (YDR096W)	Refs. 21,24,25
HSF1 (YGL073W)	Refs. 21,22,24
MSC1 (YML128C)	Refs. 21,24,26
NSR1 (YGR159C)	Refs. 21-23,25
SPC42 (YKL042W)	Refs. 24,26
TPK3 (YKL166C)	Refs. 24,26
TUB2 (YFL037W)	Refs. 22,23,25
WWM1 (YFL010C)	Refs. 25

We then constructed a system to quickly and efficiently evaluate candidate reporters by constructing libraries of reporters directly in a yeast three-hybrid strain and screening for growth inhibition effects under yeast three-hybrid counter selection conditions (**Fig. 3-7**). We elected to put the reporter genes on low-copy centromeric plasmids, allowing us to use plasmid gap repair techniques to generate large libraries of reporter constructs *in vivo*[27,28] while minimizing cell-to-cell variation in expression levels[29]. To coarsely adjust the levels of basal and activated reporter expression in the yeast three-hybrid assay, we built a family of six parental plasmids containing LexA operator-promoter constructs that should provide varying expression. Candidate reporter genes, PCR amplified with appropriate homology regions, can be readily inserted into these plasmids downstream of the promoter by homologous recombination.

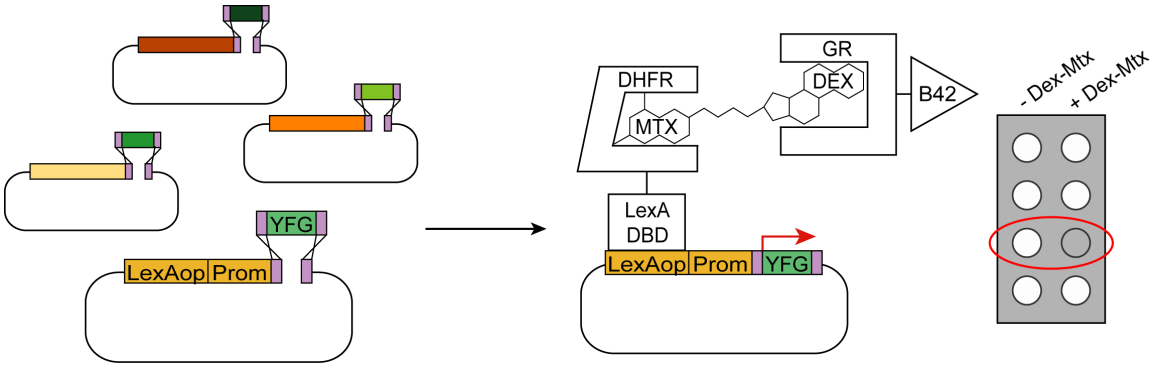


Figure 3-7. Screen for alternative reporters for the yeast three-hybrid counter selection. Yeast three-hybrid strains harboring a library of candidate reporters are constructed *in vivo* in one step by co-transforming pools of candidate reporter genes (green; YFG or "your favorite gene") and plasmids containing various LexA operator-promoter constructs (orange; LexAop-Prom). Hundreds of transformants can then be screened in parallel for growth in the presence and absence of Dex-Mtx to identify strains whose growth is inhibited by the chemical dimerizer.

Two or eight LexA operators were placed upstream of three different parental promoters (p*KEX2*, p*CYC1*, and p*TEF1*), which have been shown to provide a range of gene expression of over three orders of magnitude[30]. In contrast to the promoter constructs typically used for hybrid system reporters[31], the LexA operator sites of our promoter library do not replace the parental promoters' upstream activating sequences. Therefore, our promoter library should provide higher basal levels of reporter expression and smaller fold increases in transcription when activated. We confirmed this prediction by placing *lacZ* under the control of each promoter and quantifying *lacZ* expression in our yeast three-hybrid system in the absence and presence of Dex-Mtx (**Fig. 3-8**). All of the promoters showed Dex-Mtx-dependent increases in reporter expression ranging in magnitude from 1.5- to 9-fold. By comparison, the standard *lacZ* reporter plasmids, containing p*GAL* promoters with the upstream activating sequences entirely replaced by LexA operators, gave 100- to 1000-fold increases in reporter expression in the yeast

three-hybrid assay. This should be advantageous for our purposes since we would expect to see smaller changes in transcription during chemical complementation than observed in the yeast three-hybrid assay. Therefore, the performance of a counter selection reporter driven by a member of the promoter library in the yeast three-hybrid assay might be comparable to that of the same reporter driven by an optimized promoter (e.g., LexAop-p*GAL*) in chemical complementation. Surprisingly, in our system the "weak" *KEX2* and "intermediate" strength *CYC1* promoters[30] provide comparable levels of expression, but each promoter construct gave a somewhat different temporal pattern of expression. In addition, *lacZ* is an imperfect proxy for a counter selection reporter, which would likely have manifold effects on the cell depending on the mechanism of growth inhibition.

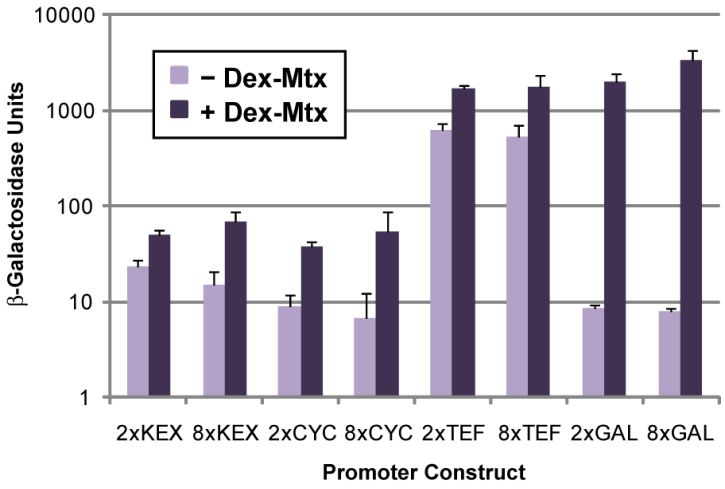


Figure 3-8. Characterization of *lacZ* expression from the promoter library. ONPG assays were performed after 24, 48, and 96 hours of growth in synthetic media containing 2% galactose, 2% raffinose, lacking tryptophan and histidine, with or without 1 μM Dexamethasone-Methotrexate. The 48-hour time point is shown here. 2x and 8x represent the number of LexA operators upstream of the parental promoter. Error bars represent the standard error of measurements from three or four unique colonies transformed with the promoter-*lacZ* constructs. Note that the y-axis is plotted on a logarithmic scale.

We constructed our counter selection reporter library by co-transforming the promoter library plasmids and the candidate reporter genes as a pool into our yeast three-hybrid strain, conveniently generating the 66 potential reporter plasmids *in vivo* by plasmid gap repair. With the exception of the reporter gene, we used a strain containing our previously reported yeast three-hybrid framework, which has been optimized to ensure a consistent reporter readout[32]. To ensure full coverage of the reporter library, we picked two hundred transformants and monitored their growth in the presence and absence of Dex-Mtx. As shown in **Figure 3-9**, almost 20% of the colonies exhibited chemical dimerizer-dependent growth inhibition. Five unique constructs were identified from the ten best colonies, and four of these continued to provide some degree of growth inhibition reproducibly upon retransformation into the yeast three-hybrid strain (**Figs. 3-10 and -11**).

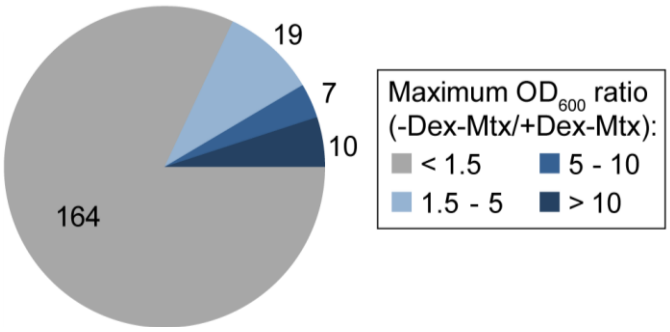


Figure 3-9. Representation of 200 colonies' performance in the counter selection reporter screen. Colonies were grown in synthetic media containing 2% galactose, 2% raffinose, lacking tryptophan and histidine, with or without 1 μM Dexamethasone-Methotrexate. Since colonies exhibited a variety of growth patterns, Dex-Mtx-dependent growth inhibition was scored by the maximum observed ratios of the cell densities (OD_{600}) of the –Dex-Mtx culture to the +Dex-Mtx culture for each colony, and the number of colonies that fell within each range is shown. Reporters from the ten best colonies (dark blue) were further characterized.

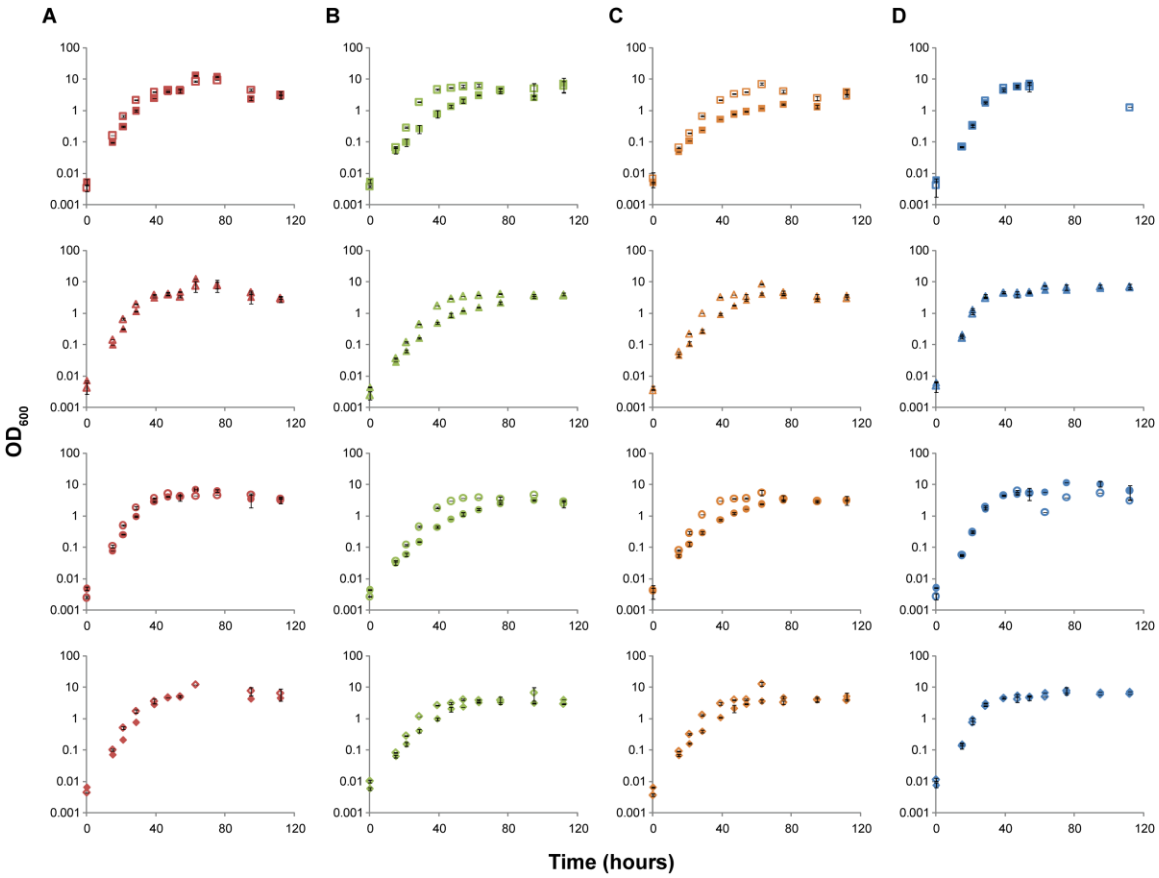


Figure 3-10. Growth curves for retransformed yeast three-hybrid counter selection reporter candidates **(A)** 8LexAop-p*CYC1-TUB2*, **(B)** 2LexAop-p*TEF1-GIS1*, **(C)** 8LexAop-p*TEF1-ACT1*, and **(D)** 8LexAop-p*KEX2-GIS1*. Each graph represents the results for a unique, randomly selected colony. Media composition was as in **Figure 3-9**. Open and closed symbols represent cultures grown in the absence and presence of 1 μM Dex-Mtx, respectively. Error bars represent the standard error in the OD_{600} readings for duplicate cultures. All reporters except 8LexAop-p*KEX2-GIS1* again provided some degree of growth inhibition in the yeast three-hybrid assay.

3.2.5 Characterization of the *GIS1* counter selection reporter

The most promising reporter construct, 8LexAop-p*TEF1-GIS1*, was selected for further characterization. After retransformation of the reporter plasmid into the yeast three-hybrid strain, 26 colonies were individually assayed for growth in the presence and absence of Dex-Mtx. Significantly, using our unoptimized selection conditions, several

GIS1 reporter clones demonstrated superior growth inhibition relative to our extensively optimized *URA3* counter selection strain[2] when it was tested under analogous conditions (**Fig. 3-11**).

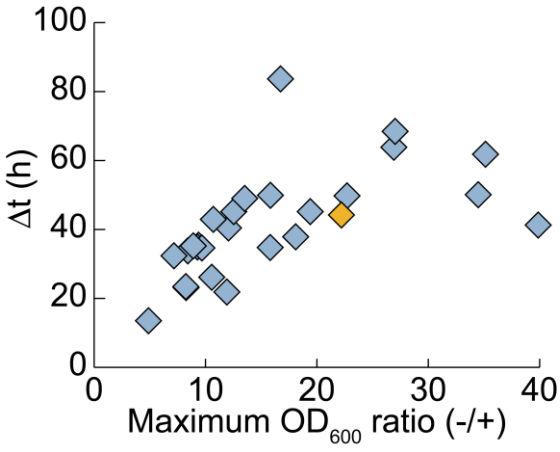


Figure 3-11. Characterization of 8LexAop-p*TEF1-GIS1* as a yeast three-hybrid counter selection reporter. Twenty-six randomly selected Y3H colonies retransformed with the reporter plasmid were assayed for growth with or without 1 μM Dex-Mtx. Selective media was as in **Figure 3-9**. Two metrics were used to evaluate each colony's performance: the maximum observed ratio of the OD_{600} reading for the –Dex-Mtx culture and the +Dex-Mtx culture (Maximum OD_{600} ratio (–/+)), and the difference in time required for the –Dex-Mtx and +Dex-Mtx cultures to reach an OD_{600} of 1 (Δt). Each blue data point represents one colony, and the performance of our published *URA3* Y3H counter selection strain[2] in this assay is shown in orange.

We explicitly tested the ability of the *GIS1* reporter to provide enrichment in the yeast three-hybrid counter selection by attempting to enrich an inactive three-hybrid strain, containing only the B42 activation domain, from an excess of active three-hybrid strains, containing the requisite B42-GR fusion protein. Cells were mixed to provide an initial ratio of 100:1 or 1000:1 active:inactive strains and subjected to yeast three-hybrid selection conditions. After four days of growth, the inactive yeast three-hybrid cells comprised the majority of the population for both selections (**Fig. 3-12**). Diluting the 1000:1 selection into fresh media on the second day, or "seeding" the selection, was even more effective, allowing the inactive yeast three-hybrid strain essentially to overtake the culture.

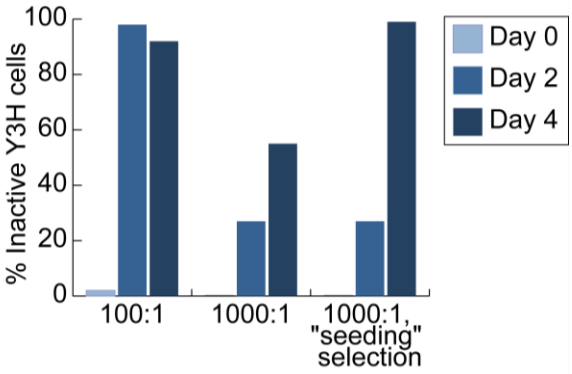


Figure 3-12. *GIS1* mock selection results. Active and inactive yeast three-hybrid cells were mixed in 100:1 or 1000:1 ratios and grown in selective media (as in **Fig. 3-9**) in the presence of 1 μM Dex-Mtx. The percentage of inactive cells was determined after 0, 2, and 4 days of selection using the colorimetric enrichment assay. For the "seeding" selection, 2 rounds of 2-day selections, rather than one longer 4-day selection, were performed.

We then tested the efficacy of the counter selection reporter in the cellulase chemical complementation assay. Since there is significant clonal variation in the *GIS1* yeast three-hybrid strains (**Fig. 3-11)**, we screened a number of colonies and selected two with the greatest growth difference with and without Dex-Mtx to test in chemical complementation. These strains were transformed with plasmids containing cellulases and their corresponding active-site nucleophile mutants. When growth was tested with under chemical complementation selection conditions (**Fig. 3-1**), no Dex-Cel-Mtx or cellulase-dependent growth effects were observed (data not shown).

3.2.6 Efforts to optimize the *GIS1* counter selection reporter

Efforts to optimize the *GIS1* yeast three-hybrid counter selection, with the goal of translating these improvements into our chemical complementation system, were not immediately successful. We first tested to see if the *GIS1* counter selection could be improved simply by optimizing the promoter. Since the 8LexAop-p*GAL* promoter gave comparable activated expression to the 8LexAop-p*TEF* promoter, but exhibited lower basal transcription (**Fig. 3-8**), we tested the 8LexAop-p*GAL*-*GIS1* reporter construct. The new *GIS1* reporter construct also served as a counter selection reporter, but its performance was no better than the original reporter plasmid (data not shown). This

could be at least partially due to issues of timing; having a higher level of basal transcription may slow cells' growth until the toxic gene product has time to accumulate.

GIS1 encodes a cAMP-dependent kinase-regulated transcription factor involved in the response to nutrient limitation; it is believed to activate the expression of a gene or genes that inhibit proliferation[33]. Attenuation of cAMP-dependent kinase activity has been reported to potentiate *GIS1*-mediated growth inhibition[33]. Accordingly, we reconstructed the *GIS1* yeast three-hybrid system in a strain with a temperature-sensitive allele of *CDC25*[34], which encodes a guanine nucleotide-exchange factor that indirectly regulates cAMP levels, to see if this would enhance the counter selection's efficacy. Since the altered strain background might require lower levels of basal *GIS1* expression, the *GIS1* reporter was again screened with the entire promoter library (**Section 3.2.4**). However, no constructs exhibited an improved growth difference with and without Dex-Mtx as compared to the wild-type strain (data not shown).

3.3 Discussion

The yeast three-hybrid counter selection exemplifies the problem of adapting classic genetic selections for advanced synthetic biology applications. As for the *LEU2* positive selection described in **Chapter 2**, we found that rigorous, quantitative characterization of the yeast three-hybrid *URA3* counter selection lent new insights into the limitations of this selection system and their root causes. Our data indicate that the *URA3* counter selection has the potential to serve as a robust reporter. The challenge will now be to appropriately regulate the reporter's expression to achieve essentially zero basal transcription while still attaining high levels of activated transcription. One intriguing possibility would be to exploit the synthetic biology community's recent

successes in the design of genetic circuits to build feedback loops into our system, thereby maximizing the differences between background and induced reporter gene transcription. Alternatively, the stability of the protein gene product could be modulated to adjust its toxicity.

The difficulty of rationally redesigning genetic selections to function robustly in engineered *in vivo* systems argues for the more widespread use of directed evolution and other library approaches in their optimization. We found that a straightforward screen of only a small library of candidate genes yielded multiple alternate counter selection reporter genes for the yeast three-hybrid assay. Furthermore, one of these proved to be as effective as the ubiquitous *URA3* counter selection in our system. Employing a library approach allowed us to define the conditions we wanted to use in our assay while still circumventing the laborious optimization process required to adapt the *URA3* reporter for the yeast three-hybrid assay.

As synthetic biologists endeavor to develop an arsenal of effective parts that will function in increasingly complex and diverse systems, our results underscore the importance of looking beyond the standard components historically used by geneticists, which were selected for their functionality in a different context. Rather, we should think broadly and creatively as we design the next generation of bioengineering tools. Furthermore, we should acknowledge that there will not always be "one-size-fits-all" solutions when designing sophisticated *in vivo* applications, and we should embrace the use of directed evolution and screening strategies to optimize systems' components and performance.

3.4 Experimental methods

General materials and methods. General materials and methods were as in **Chapter 2**.

Plasmid construction. The LexAop-promoter library was constructed by *in vivo* plasmid gap repair in *S. cerevisiae* as follows. The 2x- and 8xLexA operators were amplified from plasmids pMW109 and pMW112, respectively, using primers LMW328 and LMW329, which add 30 bp of homology to pRS416GAL before the GAL promoter. The *KEX2*, *CYC1*, and *TEF1* promoters were amplified from FY251 genomic DNA with primers LMW330/LMW331, LMW332/LMW333, and LMW334/LMW335, respectively, which add 30 bp of homology to the LexAop fragments at the 5' end and 30 bp of homology to the pRS416GAL multiple cloning site (MCS) at the 3' end. Plasmid pRS416GAL was digested with SacI and XbaI to cut out the *GAL* promoter. Each of the six combinations of the LexAop and promoter fragments was separately co-transformed with digested pRS416GAL in a 100:100:1 ratio into yeast. Transformants were miniprepped to recover plasmid DNA, which was retransformed into *E. coli*. The plasmids used for the final promoter library were pLW2571, pLW2572, pLW2573, pLW2574, pLW2575, and pLW2576.

Plasmids containing B42 constructs and a constitutively expressed colorimetric marker were constructed as follows. Plasmids pBC398 and pJG4-5 were digested with SalI to cleave at the 3' end of the *ADH1* terminator following the p*GAL1*-B42 constructs. Constructs containing p*ADH1-gusA* and p*ADH1-lacZ* were amplified from plasmids pLW2569 and pLW2570 with primers LMW323/LMW324 and LMW325/LMW324, respectively. These primers incorporated 30 bp of homology to the digested plasmids

and oriented the colorimetric markers so that they would be transcribed towards the bidirectional *ADH1* terminator. Digested pBC398 and the *gusA* PCR product, and digested pJG4-5 and the *lacZ* PCR product were co-transformed into yeast in 1:100 ratios and moved to *E. coli* for plasmid maintenance, as described above. The final plasmids used in the active and inactive yeast three-hybrid mock selection strains were pLW2578 and pLW2577, respectively.

To make an integration plasmid to integrate p*GAL*-LexA-eDHFR into the *HO* locus, pKB521 was digested with SacI/EaeI/PvuII to obtain a p*GAL*-LexA-eDHFR-t*ADH* fragment that could be isolated by gel purification, and this insert was cloned into SacI/SmaI-digested pRS423 to place this construct next to the *HIS3* marker, giving pLW2665. The insert was expected to have SacI/EaeI ends, but evidently PvuII exhibited star activity, resulting in a slightly truncated *ADH* terminator. Since 335 bp of the terminator still remained, pLW2665 was carried forward. A p*GAL*-LexA-eDHFR-t*ADH-HIS3* fragment was obtained by digesting pLW2665 with SacI/SmaI and cloning it into SacI/AfeI-digested pV2265 (integration vector for the *HO* locus), giving pLW2666.

Evaluation of the *URA3* reporter with isogenic strains. Strains were transformed with the plasmids required for the yeast three-hybrid assay as shown in **Table 3-2**. All strains were derivatives of FY251. LW2635Y was obtained by growing PPY2240Y under non-selective conditions until colonies that had been cured of all plasmids were identified.

Table 3-2. Construction of *URA3* counter selection strains.

	Parental Strain	Plasmids Transformed
Strains for growth curves		
URA3 with endogenous promoter (p*URA3*)	V2169Y	pKB521, pBC398
ura3-52 (inactive allele)	FY251	pKB521, pBC398
Active Y3H (LexAop-p*SPO13-URA3*)	LW2635Y	pKB521, pBC398
Inactive Y3H (LexAop-p*SPO13-URA3*)	LW2635Y	pMW103, pBC398
Strains for Mock Selections		
p*URA3-URA3* blue strain (*URA3* marker)	V2169Y	pKB521, pLW2578
ura3-52 red strain (inactive *URA3* marker)	FY251	pKB521, pLW2577
LW2636Y (*URA3* active Y3H blue strain)	LW2635Y	pKB521, pLW2629
LW2637Y (*URA3* inactive Y3H red strain)	LW2635Y	pKB521, pLW2570

For growth curves, 4 unique doubly transformed colonies from each transformation were inoculated into SC(HT^-) (lacking histidine and tryptophan) media and grown to saturation. One μL of cells were inoculated into 199 μL of (SC(HT^-), 2% galactose, 2% raffinose, 0.2% 5-FOA) (with or without 5 μM Dex-Mtx* for the active and inactive yeast three-hybrid strains) to begin the growth curves.

For the mock selections, see **Chapter 2** for a general description of selection set-up. For evaluation of *URA3* as a marker gene, the selective media was (SC(HT^-), 2% galactose, 2% raffinose, 0.2% 5-FOA); for evaluation of *URA3* as a yeast three-hybrid reporter, the selective media was (SC(HT^-), 2% galactose, 2% raffinose, 0.2% 5-FOA media, 5 μM Dex-Mtx*).

*A stock of Dex-Mtx that works significantly better at 5 μM than at 1 μM was used.

Construction of strains with different promoters for the *URA3* reporter. Strains with multiple copies of the LexAop-p*SPO13-URA3* reporter were made by transforming pPPY2176 into a colony resulting from FY251/pKB521/pBC398 transformation described above. Low- or high-copy plasmids containing LexAop-p*GAL-URA3* reporters were made by *in vivo* gap repair as follows. The 2- and 8LexAop-p*GAL* promoters were amplified with appropriate homology regions from pMW109 and

pMW112, respectively, using primers LMW328 and LMW541. The *URA3* coding sequence was amplified from plasmid pRS416 with appropriate homology regions using primers LMW542 and LMW543. The *MET* promoters of pRS415MET and pRS425MET were removed by digestion with SacI and BamHI, and the digested plasmids were co-transformed with a promoter PCR and the *URA3* PCR into an FY251/pKB521/pBC398 colony. The resulting transformants were tested directly in growth assays.

Characterization of the promoter library in ONPG assays. The *lacZ* gene was amplified from plasmid pLW2570 with primers LMW505 and LMW506, and the unpurified PCR product was amplified with primers LMW339 and LMW340. The PCR product was co-transformed with the individual members of the promoter library (plasmids pLW2571, pLW2572, pLW2573, pLW2574, pLW2575, and pLW2576 digested with ClaI) and pBC398 into strain V704Y. Plasmids pMW109 and pMW112 were also co-transformed with pBC398 into the same strain for comparison. Four colonies from each transformation were inoculated into 100 μL SC(HTU^- (lacking histidine, tryptophan, and uracil), 2% galactose, 2% raffinose) (inducing) media. After 24 hours, 1 μL of the cultures were inoculated into 199 μL of the same media with or without 1 μM Dex-Mtx. ONPG assays were conducted on the cultures at various time points. After 24, 48, and 96 hours of growth, the cell densities of the cultures were measured, and 50 μL were harvested. The cells were washed once with Z buffer[35], resuspended in 100 μL YPER, and lysed for 30 minutes at room temperature. A solution of ONPG in Z buffer (8.5 μL of a 10 mg/mL solution) was added to the lysates, and the reactions were quenched with 110 μL of 1M Na_2CO_3 after most wells turned visibly yellow. The reactions were centrifuged at 2000 rpm for 5 min to pellet cellular debris,

and the A_{420} of 100 μL of the supernatant was measured. β-galactosidase units were calculated as (1000* A_{420})/(time*volume*A_{600}) (time in minutes, volume in mL).

Construction of a yeast three-hybrid counter selection reporter library. Equal amounts of the six promoter library plasmids listed above were combined and digested in the MCS with ClaI. Eleven candidate reporter genes were amplified from FY251 genomic DNA with the primers indicated in **Table 3-3**. The unpurified PCR products were then amplified with LMW339 and LMW340; the two PCR reactions added a total of 30 bp of homology to the promoter library MCS at the 5' end and 30 bp of homology to the *CYC1* terminator at the 3' end. The PCR products were mixed in an equimolar ratio and co-transformed with the digested promoter library in a 100:1 ratio and with pV398E into V704Y. Double transformants were selected on SC(HTU$^-$) plates.

Table 3-3. PCR amplification of candidate counter selection reporter genes.

Gene	Primers
ACT1 (YFL039C)	LMW391/LMW392
AMN1 (YBR158W)	LMW393/LMW394
CDH1 (YGL003C)	LMW395/LMW396
GIS1 (YDR096W)	LMW397/LMW398
HSF1 (YGL073W)	LMW399/LMW400
MSC1(609-1359) (YML128C)	LMW401/LMW402
NSR1 (YGR159C)	LMW403/LMW404
SPC42 (YKL042W)	LMW405/LMW406
TPK3 (YKL166C)	LMW407/LMW408
TUB2 (YFL037W)	LMW409/LMW410
WWM1 (YFL010C)	LMW411/LMW412

Counter selection reporter library screening. Two hundred double transformants from the counter selection reporter library were inoculated into SC(HTU$^-$) media containing 2% galactose and 2% raffinose. After 24 hours of growth, cultures were diluted into nonselective (SC(HTU$^-$), 2% gal, 2% raf) and selective (SC(HTU$^-$), 2% gal, 2% raf, 1 μM Dex-Mtx) media (200 μL). Cell density was periodically monitored by

measuring the absorbance of cultures at 600 nm. Forty-four colonies whose growth appeared to be inhibited by Dex-Mtx in the initial screen were retested, and 36 again demonstrated significant small molecule-dependent growth inhibition. Reporter plasmids from the ten best-performing colonies were miniprepped, transformed into *E. coli*, and sequenced to determine the constructs' identities.

Growth curves. Growth curves for reporter plasmids retransformed into the yeast three-hybrid strain (pLW2579, pLW2580, pLW2581, pLW2582, and pLW2583) were carried out as described for the counter selection library screening, except that all cultures were set up at least in duplicate. For *URA3* counter selection strain growth curve (PPY2240Y), the media used was SC(HT^-), 0.2% 5-FOA, 2% galactose, 2% raffinose, with or without 1 μM Dex-Mtx. During exponential growth, OD_{600} readings were taken approximately every 12 hours. The ratio of the averaged OD_{600} value readings of the –Dex-Mtx and +Dex-Mtx cultures was calculated for each time point. The difference in time for the –Dex-Mtx and +Dex-Mtx cultures to reach an OD_{600} of 1 was calculated from the averaged data points immediately before and after this OD_{600} was reached, assuming exponential growth during the entire interval:

$$OD_{600(final)} = OD_{600(initial)} \times 2^{(time(final)-time(initial))/doubling\ time}$$

For analysis of chemical complementation using the *GIS1* reporter, two yeast three-hybrid strains retransformed with pLW2579 were selected after their performance in the yeast three-hybrid counter selection was characterized (**Fig. 3-11**). These strains, LW2672Y and LW2673Y, were transformed with plasmids pV2230 and pV2557. Growth curves for transformants were performed as for the *GIS1* yeast three-hybrid

counter selection, but with media that also lacked leucine and with 1 μM Dex-Cel-Mtx rather than Dex-Mtx.

Mock selections. Plasmids pLW2578 and pLW2577 were co-transformed with the reporter plasmid pLW2579 into V704Y to generate active and inactive yeast three-hybrid strains. Sixteen colonies from each transformation were tested in growth curves to verify their performance in the yeast three-hybrid assay and to ensure that they had the same growth rates under non-selective conditions. One active (LW2585Y) and one inactive (LW2584Y) yeast three-hybrid strain that best met both conditions were glycerol stocked for use in mock selections. To begin the mock selections, cultures were inoculated directly from the glycerol stocks into SC(HTU$^-$), 2% glucose media, and once these cultures reached saturation, they were inoculated into SC(HTU$^-$), 2% galactose, 2% raffinose media (1 mL). After 24 hours, the cultures were harvested (2000 rpm, 5 minutes) and resuspended in fresh SC(HTU$^-$), 2% galactose, 2% raffinose media (1 mL). The OD_{600} of the cells was measured, and the cultures were mixed and diluted with media to give 3 mL of the desired ratio of strains at an OD_{600}=0.1. Dex-Mtx was added to 1 mL of the culture to a concentration of 1 μM, and another 1 mL was used to start a parallel non-selective (–Dex-Mtx) culture. For the "seeding" selection, the OD_{600} of the 1000:1 selections were measured, and a sample of the culture was used to inoculate fresh media (+ or – 1 μM Dex-Mtx) to an OD_{600}=0.1. On days 0, 2, and 4, a sample of each culture was plated on non-selective SC(HTU$^-$), 2% glucose media. After 3 days of growth, colonies (>100 for each sample) were assayed for *lacZ* and *gusA* expression using an agarose overlay assay[36], replacing X-Gal with X-Gluc and Magenta-Gal. Colonies' colors developed within hours, and the numbers of red and blue colonies were counted to

score the percentage of active and inactive yeast three-hybrid cells in the culture. For cultures without Dex-Mtx, the percentage of inactive strain always remained below 10%.

Construction of an alternate reporter plasmid for the *GIS1* counter selection. *GIS1* was amplified from pLW2579 with homology to pMW112 5' and 3' to the open reading frame of *lacZ* using primers LMW527 and LMW528. The PCR product was co-transformed with MluI-digested pMW112 into V704Y to construct a 2μ plasmid containing the reporter construct 8LexAop-p*GAL-GIS1* by gap repair. Transformants were miniprepped, retransformed into *E. coli*, and analyzed by colony PCR. Three of the resulting plasmids that had the reporter construct and pLW2579 were co-transformed with pBC398 into V704Y. Growth assays were performed as described above for the *GIS1* reporter.

Construction of Y3H strain with a *cdc25-2* background. Strain V2668Y, containing a temperature-sensitive *cdc25-2* allele, was obtained from the Elledge laboratory[37], and the genotype was verified by PCR and restriction analysis of the locus. This strain and all of its derivatives were grown at room temperature rather than at 30°C. The plasmid pLW2666 was digested with ApaLI and transformed into V2668Y. A *HIS3* colony was analyzed for correct integration of the LexA-DHFR construct by PCR and sequencing of its genomic DNA, and it was glycerol stocked as LW2667Y. Yeast three-hybrid strain was constructed by co-transformation of pBC398 and the appropriate reporter plasmids. Growth assays were performed as described above for the *GIS1* reporter.

3.5 Strains, plasmids, and oligonucleotides

Table 3-4. Strains used in this study

Name	Genotype	Source/Reference
FY251	*MATa trp1Δ63 his3Δ200 ura3-52 leu2Δ 1 Gal+*	M. Carlton
PPY2240Y	*MATa trp1Δ63 his3Δ200 8LexAop-pSPO13-URA3 leu2Δ1 Gal+* pKB521 pBC398	P. Peralta-Yahya/2
V2169Y	*MATa trp1Δ63 his3Δ200 URA3 leu2Δ1 Gal+*	P. Peralta-Yahya/2
V2668Y	*MATα ura3 lys2 leu2 trp1 cdc25-2 his3Δ200 ade101 GAL1*	S. Elledge/37
V704Y	*MATa trp1Δ63 his3Δ200 ura3-52 leu2Δ1 ade4::pGAL1-LexA-eDHFR(HIS3) GAL+*	K. Baker/15
LW2584Y	VC704Y with pLW2577 and pLW2579	This study
LW2585Y	VC704Y with pLW2578 and pLW2579	This study
LW2635Y	*MATa trp1Δ63 his3Δ200 8lexAop-Spo13-URA3 leu2Δ1 Gal+* (PPY2240Y cured of plasmids)	This study
LW2636Y	LW2635Y with pLW2629 and pKB521	This study
LW2637Y	LW2635Y with pLW2570 and pKB521	This study
LW2667Y	V2668Y with pLW2666 integrated	This study
LW2672Y	V704Y with pBC398 and pLW2579	This study
LW2673Y	V704Y with pBC398 and pLW2579	This study

Table 3-5. Plasmids used in this study

Name	Details	Source/Reference
pBC398	*pGAL1-B42-(GSG)$_2$-rGR2 2μ TRP1* pUC ori kan^R	B. Carter/38
pJG4-5	*pGAL1-B42 2μ TRP1* pUC ori amp^R	R. Brent/31
pKB521	*pGAL1-LexA-eDHFR 2μ HIS3* pBR ori kan^R	K. Baker/32
pMW103	*pGAL1-B42 2μ TRP1* pUC ori kan^R	R. Brent/39
pMW109	*2LexAop-lacZ 2μ URA3* pBR ori kan^R	R. Brent/39
pMW112	*8LexAop-lacZ 2μ URA3* pBR ori kan^R	R. Brent/39
pPPY2176	pRS425MET carrying *8LexAop-pSPO13-URA3* fusion	P. Peralta-Yahya/2
pRS415MET	*pMET25 CEN6/ARSH4 LEU2* pBIISK ori amp^R	ATCC #87322
pRS416	*CEN6/ARSH4 URA3* pBIISK ori amp^R	ATCC #87521
pRS416GAL	*pGAL1 CEN6/ARSH4 URA3* pBIISK ori amp^R	ATCC #87332
pRS423	*2μ HIS3* pBIISK ori amp^R	ATCC #77104
pRS425MET	*pMET25 2μ LEU2* pBIISK ori amp^R	ATCC #87323
pVC2230	pRS425MET-*Erwinia carotovora* CelN	P. Peralta-Yahya/2
pVC2265	HO-polylinker-KanMX4-HO	D. Stillman/40
pVC2557	pRS425MET-CelN:E226G	V. Mondol
pLW2569	*pADH1-gusA 2μ TRP1* pUC ori amp^R	This study
pLW2570	*pADH1-lacZ 2μ TRP1* pUC ori kan^R	This study
pLW2571	pRS416 *2xLexAop-pKEX2-MCS-tCYC*	This study
pLW2572	pRS416 *8xLexAop-pKEX2-MCS-tCYC*	This study
pLW2573	pRS416 *2xLexAop-pCYC1-MCS-tCYC*	This study

pLW2574	pRS416 *8xLexAop-pCYC1-MCS-tCYC*	This study
pLW2575	pRS416 *2xLexAop-pTEF-MCS-tCYC*	This study
pLW2576	pRS416 *8xLexAop-pTEF-MCS-tCYC*	This study
pLW2577	pJG4-5 with *pADH-lacZ* inserted	This study
pLW2578	pBC398E with *pADH-gusA* inserted	This study
pLW2579	pLW2576 with *GIS1* in the MCS	This study
pLW2580	pLW2574 with *TUB2* in the MCS	This study
pLW2581	pLW2575 with *GIS1* in the MCS	This study
pLW2582	pLW2576 with *ACT1* in the MCS	This study
pLW2583	pLW2572 with *GIS1* in the MCS	This study
pLW2629	pBC398E with p*ADH-gusA* replacing (pUC ori kan^{R})	This study
pLW2665	pRS423 with *pGAL-LexA-eDHFR-tADH*	This study
pLW2666	pVC2265 with *pGAL-LexA-eDHFR-tADH-HIS3*	This study

Table 3-6. Oligonucleotides used in this study

Name	**Sequence (5'-3')**
LMW323	GTGTGTATTTTATGTCCTCAGAGGACAACACCTGTTGTAATCATTGTTTGCCTCCCTG
LMW324	GTTTTAGGACTGGTTCAGAATTGCTGCAGGTCGAACAACTTCTTTTCTTTTTTTTTCT
LMW325	TATGTCCTCAGAGGACAACACCTGTTGTAATTATTTTTGACACCAGACCAA
LMW328	ATTAACCCTCACTAAAGGGAACAAAAGCTGCATATCTAATCTTACCTCGA
LMW329	CTAATCGCATTATCATCCCT
LMW330	CAGTACGTCGAGGGATGATAATGCGATTAGTCAGCAGCTCTGATGTAGA
LMW331	GAATTCCTGCAGCCCGGGGGATCCACTAGTCTGATAATGGGTTAGTAGTTT
LMW332	CAGTACGTCGAGGGATGATAATGCGATTAGTTTGGAAAACCAAGAAATGAAT
LMW333	GAATTCCTGCAGCCCGGGGGATCCACTAGTTATTAATTTAGTGTGTGTATTTG
LMW334	CAGTACGTCGAGGGATGATAATGCGATTAGATAGCTTCAAAATGTTTCTACT
LMW335	GAATTCCTGCAGCCCGGGGGATCCACTAGTTTTGTAATTAAAACTTAGATTAGA
LMW339	CCCGGGCTGCAGGAATTCGATATCAAGCTT
LMW340	TAACTAATTACATGACTCGAGGTCGACGGT
LMW391	AGGAATTCGATATCAAGCTTATGGATTCTGGTATGTTCTA
LMW392	GACTCGAGGTCGACGGTTTAGAAACACTTGTGGTGAA
LMW393	AGGAATTCGATATCAAGCTTATGAAACTAGAACGCGTAAG
LMW394	GACTCGAGGTCGACGGTCTAGTCCACATTATTCTCTA
LMW395	AGGAATTCGATATCAAGCTTATGTCCACAAACCTGAACC
LMW396	GACTCGAGGTCGACGGTCTAACGTATTTGATTAAATGC
LMW397	AGGAATTCGATATCAAGCTTATGGAAATCAAGCCAGTTG
LMW398	GACTCGAGGTCGACGGTCTATGATTCAGCTAATTTAGTA
LMW399	AGGAATTCGATATCAAGCTTATGAATAATGCTGCAAATACA
LMW400	GACTCGAGGTCGACGGTCTATTTCTTAGCTCGTTTGG
LMW401	AGGAATTCGATATCAAGCTTATGTCGGGACAAATAAAAGACAC

LMW402	GACTCGAGGTCGACGGTTTATTCTTGAGACCAGCTCTG
LMW403	AGGAATTCGATATCAAGCTTATGGCTAAGACTACTAAAGT
LMW404	GACTCGAGGTCGACGGTTTAATCAAATGTTTTCTTTGAAC
LMW405	AGGAATTCGATATCAAGCTTATGAACGGATCTCCCACT
LMW406	GACTCGAGGTCGACGGTTCATCGATTATTGGGAGTG
LMW407	AGGAATTCGATATCAAGCTTATGTATGTTGATCCGATGAA
LMW408	GACTCGAGGTCGACGGTTTAAAATTCTTTCATTAAATCCAT
LMW409	AGGAATTCGATATCAAGCTTATGAGAGAAATCATTCATATC
LMW410	GACTCGAGGTCGACGGTTTATTCAAAATTCTCAGTGATT
LMW411	AGGAATTCGATATCAAGCTTATGGCTCAAAGTAAAAGTAAT
LMW412	GACTCGAGGTCGACGGTTTAAAAGTCACTACCGTCAAA
LMW505	AGGAATTCGATATCAAGCTTATGACTAAATCTCATTCAGAA
LMW506	GACTCGAGGTCGACGGTTTATTTTTGACACCAGACCAA
LMW527	ATACTTTAACGTCAAGGAGAAAAAACTATAATGGAAATCAAGCCAGTTG
LMW528	TACGGGCAGACATGGCCTGCCCGGTTATTACTATGATTCAGCTAATTTAGTA
LMW541	CATTATAGTTTTTTCTCCTTGA
LMW542	ATACTTTAACGTCAAGGAGAAAAAACTATAATGTCGAAAGCTACATATAAG
LMW543	TAACTAATTACATGACTCGAGGTCGACGGTTTAGTTTTGCTGGCCGCAT

3.6 References

1. Boeke, J.D., LaCroute, F. & Fink, G.R. A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol. Gen. Genet.* **197**, 345-346 (1984).

2. Peralta-Yahya, P., Carter, B.T., Lin, H., Tao, H. & Cornish, V.W. High-throughput selection for cellulase catalysts using chemical complementation. *J. Am. Chem. Soc.* **130**, 17446-17452 (2008).

3. Chattoo, B.B. et al. Selection of *lys2* mutants of the yeast *Saccharomyces cerevisiae* by the utilization of α-aminoadipate. *Genetics* **93**, 51-65 (1979).

4. Singh, A. & Sherman, F. Association of methionine requirement with methyl mercury resistant mutants of yeast. *Nature* **247**, 227-229 (1974).

5. Hoffmann, W. Molecular characterization of the CAN1 locus in Saccharomyces cerevisiae. A transmembrane protein without N-terminal hydrophobic signal sequence. *J. Biol. Chem.* **260**, 11831-11837 (1985).

6. Fried, H.M. & Warner, J.R. Molecular cloning and analysis of yeast gene for cycloheximide resistance and ribosomal protein L29. *Nucleic Acids Res.* **10**, 3133-3148 (1982).

7. Toyn, J.H., Gunyuzlu, P.L., White, W.H., Thompson, L.A. & Hollis, G.F. A counterselection for the tryptophan pathway in yeast: 5-fluoroanthranilic acid resistance. *Yeast* **16**, 553-560 (2000).

8. Regenberg, B. & Hansen, J. GAP1, a novel selection and counter-selection marker for multiple gene disruptions in Saccharomyces cerevisiae. *Yeast* **16**, 1111-1119 (2000).

9. Kawahata, M., Amari, S., Nishizawa, Y. & Akada, R. A positive selection for plasmid loss in Saccharomyces cerevisiae using galactose-inducible growth inhibitory sequences. *Yeast* **15**, 1-10 (1999).

10. Ausubel, F. et al. *Current Protocols in Molecular Biology*, (Wiley, New York, 1995).

11. Vidal, M., Brachmann, R.K., Fattaey, A., Harlow, E. & Boeke, J.D. Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 10315-10320 (1996).

12. Huang, Q., Roessner, C.A., Croteau, R. & Scott, A.I. Engineering Escherichia coli for the synthesis of taxadiene, a key intermediate in the biosynthesis of taxol. *Bioorg. Med. Chem.* **9**, 2237-2242 (2001).

13. Chidley, C., Haruki, H., Pedersen, M.G., Muller, E. & Johnsson, K. A yeast-based screen reveals that sulfasalazine inhibits tetrahydrobiopterin biosynthesis. *Nat. Chem. Biol.* **7**, 375-383 (2011).

14. Boeger, H. et al. Structural basis of eukaryotic gene transcription. *FEBS Lett.* **579**, 899-903 (2005).

15. de Felipe, K.S., Carter, B.T., Althoff, E.A. & Cornish, V.W. Correlation between ligand-receptor affinity and the transcription readout in a yeast three-hybrid system. *Biochemistry* **43**, 10353-10363 (2004).

16. Kok, R.J., Everts, M., Asgeirsdottir, S.A., Meijer, D.K. & Molema, G. Cellular handling of a dexamethasone-anti-E-selectin immunoconjugate by activated endothelial cells: Comparison with free dexamethasone. *Pharm. Res.* **19**, 1730-1735 (2002).

17. Hoops, S. et al. COPASI--a COmplex PAthway SImulator. *Bioinformatics* **22**, 3067-3074 (2006).

18. Sengupta, D., Lin, H., Goldberg, S.D., Mahal, J.J. & Cornish, V.W. Correlation between catalytic efficiency and the transcription read-out in chemical complementation: A general assay for enzyme catalysis. *Biochemistry* **43**, 3570-3581 (2004).

19. Estojak, J., Brent, R. & Golemis, E.A. Correlation of two-hybrid affinity data with in vitro measurements. *Mol. Cell. Biol.* **15**, 5820-5829 (1995).

20. Akada, R., Hirosawa, I., Kawahata, M., Hoshida, H. & Nishizawa, Y. Sets of integrating plasmids and gene disruption cassettes containing improved counter-

selection markers designed for repeated use in budding yeast. *Yeast* **19**, 393-402 (2002).

21. Boyer, J. et al. Large-scale exploration of growth inhibition caused by overexpression of genomic fragments in Saccharomyces cerevisiae. *Genome Biol.* **5**, R72 (2004).

22. Espinet, C., Delatorre, M.A., Aldea, M. & Herrero, E. An efficient method to isolate yeast genes causing overexpression-mediated growth arrest. *Yeast* **11**, 25-32 (1995).

23. Liu, H., Krizek, J. & Bretscher, A. Construction of a GAL1-regulated yeast cDNA expression library and its application to the identification of genes whose overexpression causes lethality in yeast. *Genetics* **132**, 665-673 (1992).

24. Sopko, R. et al. Mapping pathways and phenotypes by systematic gene overexpression. *Mol. Cell* **21**, 319-330 (2006).

25. Stevenson, L.F., Kennedy, B.K. & Harlow, E. A large-scale overexpression screen in Saccharomyces cerevisiae identifies previously uncharacterized cell cycle genes. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 3946-3951 (2001).

26. Akada, R., Yamamoto, J. & Yamashita, I. Screening and identification of yeast sequences that cause growth inhibition when overexpressed. *Mol. Gen. Genet.* **254**, 267-274 (1997).

27. Cherry, J.R. et al. Directed evolution of a fungal peroxidase. *Nat. Biotechnol.* **17**, 379-384 (1999).

28. Colby, D.W. et al. Engineering antibody affinity by yeast surface display. *Methods Enzymol.* **388**, 348-358 (2004).

29. Ishii, J. et al. A simple and immediate method for simultaneously evaluating expression level and plasmid maintenance in yeast. *J. Biochem.* **145**, 701-708 (2009).

30. Nacken, V., Achstetter, T. & Degryse, E. Probing the limits of expression levels by varying promoter strength and plasmid copy number in Saccharomyces cerevisiae. *Gene* **175**, 253-260 (1996).

31. Gyuris, J., Golemis, E., Chertkov, H. & Brent, R. Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. *Cell* **75**, 791-803 (1993).

32. Baker, K., Sengupta, D., Salazar-Jimenez, G. & Cornish, V.W. An optimized dexamethasone-methotrexate yeast 3-hybrid system for high-throughput screening of small molecule-protein interactions. *Anal. Biochem.* **315**, 134-137 (2003).

33. Pedruzzi, I., Burckert, N., Egger, P. & De Virgilio, C. Saccharomyces cerevisiae Ras/cAMP pathway controls post-diauxic shift element-dependent transcription through the zinc finger protein Gis1. *EMBO J.* **19**, 2569-2579 (2000).

34. Petitjean, A., Hilger, F. & Tatchell, K. Comparison of thermosensitive alleles of the CDC25 gene involved in the cAMP metabolism of Saccharomyces cerevisiae. *Genetics* **124**, 797-806 (1990).

35. Adams, A., Gottschling, D., Kaiser, C. & Stearns, T. (eds.). *Methods in Yeast Genetics* (Cold Spring Harbor Laboratory Press Plainview, NY 1998).

36. Petsko, G. *X-gal agarose overlay assay*. *http://biochemistry.ucsf.edu/labs/herskowitz/xgalagar.html*, (1996).

37. Aronheim, A., Zandi, E., Hennemann, H., Elledge, S.J. & Karin, M. Isolation of an AP-1 repressor by a novel method for detecting protein-protein interactions. *Mol. Cell. Biol.* **17**, 3094-3102 (1997).

38. Abida, W.M., Carter, B.T., Althoff, E.A., Lin, H. & Cornish, V.W. Receptor-dependence of the transcription read-out in a small-molecule three-hybrid system. *ChemBioChem* **3**, 887-895 (2002).

39. Watson, M.A., Buckholz, R. & Weiner, M.P. Vectors encoding alternative antibiotic resistance for use in the yeast two-hybrid system. *Biotechniques* **21**, 255-259 (1996).

40. Voth, W.P., Richards, J.D., Shaw, J.M. & Stillman, D.J. Yeast vectors for integration at the HO locus. *Nucleic Acids Res.* **29**, e59 (2001).

Chapter 4

Reiterative Recombination for the *In Vivo* Assembly of Multi-Gene Pathways

---

*The content of this chapter has or will be partially published in

V.W. Cornish, L.M. Wingler. "In Vivo Assembly Of DNA Via Homologous Recombination." US Patent Application # 61/174,272.

L.M. Wingler, V.W. Cornish. "Reiterative Recombination for the *In Vivo* Assembly of Libraries of Multi-Gene Pathways," *submitted.*

4.0 Chapter outlook

Installing customized multi-gene pathways in the cell is arguably the first step of any synthetic biology experiment. Currently, building such constructs requires choosing between two unsatisfactory alternatives: high-yielding but resource-intensive *in vitro* DNA assembly methods, or straightforward but low-yielding *in vivo* methods. We envisioned that we could exploit the known efficiency of double-strand break repair by homologous recombination in yeast to develop a robust platform for *in vivo* DNA assembly. Our system, Reiterative Recombination, elongates a construct of interest in a stepwise manner by employing pairs of alternating, orthogonal endonucleases and selectable markers. In this chapter, we present the design, development, and first demonstration of Reiterative Recombination. First, we designed and constructed a system for Reiterative Recombination in *Saccharomyces cerevisiae*. Next, we verified that endonuclease cleavage of the chromosome led to high-efficiency recombination that was amenable to cyclical repetition. Finally, we challenged Reiterative Recombination to build multi-gene constructs *in vivo* by integrating readily assayable reporter genes into the yeast chromosome, creating a three-gene, 8.5-kilobase "pathway," and rigorously characterizing the resulting strains both phenotypically and genotypically. To our knowledge, Reiterative Recombination is the first high-yielding technology for the assembly of multi-gene constructs *in vivo*. This attribute, together with its technical straightforwardness, should make it a robust and accessible methodology for a broad spectrum of researchers to build large, custom DNA pathways and should allow it to generate large libraries *in vivo* for pathway optimization and directed evolution.

4.1 Introduction

A key bottleneck to reengineering cells for diverse synthetic biology applications is the technical difficulty of constructing optimized, multi-gene pathways *in vivo*. The advent of synthetic biology has raised the tantalizing prospect of reprogramming cells at will for purposes ranging from the biosynthesis of high-value feedstocks and natural product analogs to the development of cell-based sensors and therapeutics[1]. Engineering cells for such tasks requires the introduction of numerous exogenous genes into the genome to create novel "pathways." However, standard molecular biology and genetic techniques, developed for the manipulation of single genes, become unwieldy or ineffective when applied to much larger multi-gene constructs[2]. A new generation of robust, accessible tools for building pathways inside the cell is needed.

The difficulty of rationally designing complex systems that operate as desired in the cellular milieu[3] further argues that the ability to construct not only individual pathways but also libraries of pathways *in vivo* will be essential. Precedent has indicated that multi-component systems introduced into the cell typically require refinement to function optimally[4,5]. By analogy to the directed evolution approaches that empowered the routine discovery of proteins and nucleic acids with prescribed functions, generating large numbers of variant pathways in parallel and screening for those that exhibit the required behavior could streamline optimization efforts[6]. Library-based approaches could circumvent the gaps in our knowledge, immediately yielding functional systems, but they will also require DNA assembly methods that can reliably generate sizable collections of pathways ($>10^3$) inside of the cell, an especially high standard of efficiency.

4.1.1 Methods for assembling multi-gene pathways *in vitro*

Though several technologies for building large DNA constructs have been reported, none has yet emerged as the transformative solution that will be required for the routine assembly of large, customized DNA for the era of synthetic biology. Initial approaches to custom DNA assembly have sought to stretch the limits of standard molecular biology tools that were designed to manipulate single genes[7-12]. While these *in vitro* technologies have allowed entire biosynthetic pathways to be assembled, often in high yield, they are inherently resource intensive, relying on expensive enzymatic reagents and repeated cycles of multi-step DNA manipulation.

Early efforts in this area were pioneered by researchers at Kosan Biosciences, who introduced "ligation by selection" for the convergent construction of polyketide synthase genes in 2004[13-16]. Fully synthetic gene "synthons" are cloned into donor and acceptor plasmids designed to present compatible restriction sites and unique selective marker combinations; correct ligation of the digested donor and acceptor generates a plasmid with a different marker pair combination.

Related strategies have proliferated in more recent years[17,18]. For example, Codon Devices introduced "pairwise selection assembly" for convergent, large-scale DNA assembly[19]. Fragments to be assembled are cloned into vectors so that they are flanked on both sides with "activation tags" that turn on expression of two antibiotic resistance genes. Two fragments are then excised with only one activation tag each, ligated to create a fragment with tags on both ends, and cloned into a vector with a different pair of resistance genes that require the activation tags for expression. This creates a stringent selection for correct assembly. However, all of these types of strategies necessitate

repeated cycles of moving the growing constructs in and out of *E. coli* and handling large, unwieldy DNA fragments *in vitro*.

Other researchers have focused on developing "*in vitro* recombination protocols," which have the advantage of not relying on restriction sites that may also appear in the DNA being assembled. First developed by the Elledge group[10], *in vitro* recombination imitates *in vivo* homologous recombination by using a 5' to 3' exonuclease to reveal complementary single-stranded DNA at the ends of two fragments. After fragments anneal, polymerase and ligase enzymes repair the DNA to generate the intact product. Gibson *et al.* have extensively optimized *in vitro* recombination to develop a one-step "isothermal assembly" protocol that uses only commercially available reagents[8,9]. Impressively, the authors are able to use it to assemble even genome-sized DNA molecules (>500 kb). Despite its reported high efficiency and the field's intense interest in combinatorial pathway construction, the Venter Institute has only published a single example of using this technology to construct a "library"[20] (see **Chapter 5**). Moreover, this approach does not inherently address the issue of efficiently moving the resulting large DNA constructs into the desired host, particularly for applications that require the stable integration of pathways into the chromosome.

4.1.2 Methods for assembling multi-gene pathways *in vivo*

Several other laboratories have begun to exploit homologous recombination for DNA assembly directly in cells[21-24]. While attractive for their simplicity, these assembly protocols have basically consisted of co-transforming multiple DNA fragments and determining the limit of the cell's ability to join them together correctly. Not surprisingly, these methods are inherently low yielding. For example, Zhao and co-

workers reported the assembly of up to a 19-kb, two-pathway cluster by co-transformation of nine overlapping DNA fragments into yeast[23]. Zhao recovered only ~100 recombinants under conditions that would typically yield ~10^7 transformants (0.001% efficiency), making this approach completely impracticable for library applications. This method also does not allow subsequent modification of constructs.

The next breakthrough will be to combine the control and high efficiency of *in vitro* methods with the technical ease of performing recombination directly in the cell. Itaya and co-workers have performed pioneering work in this arena, using homologous recombination in conjunction with elegant marker recycling strategies to integrate constructs ranging in size from 16 kb to 3.5 Mb into the *Bacillus subtilis* genome[25,26]. However, their systems require over 700 bp of overlapping homology and still only yield dozens to hundreds of colonies per round.

4.2 Design of Reiterative Recombination

We envisioned that we could overcome the critical shortcoming of existing *in vivo* DNA assembly methods—their very low efficiencies—by coupling assembly to the repair of double-strand DNA breaks (DSBs). Building on an established technique for targeted gene disruption[27,28], Reiterative Recombination introduces these defined DSBs, which stimulate homologous recombination, using homing endonucleases with large recognition sequences (~20 bp) that will only cleave at engineered sites. As shown in **Figure 4-1**, these cleavage sites are placed between fragments of the construct of interest and selectable markers in "donor" and "acceptor" modules. Upon endonuclease cleavage of the acceptor module, the donor module provides homology to repair the DSB through a short region of overlap between the fragments to be assembled on one side of the break,

and a homology region upstream of the selectable markers on the other side. Repair by homologous recombination adds the donor module's fragment to the acceptor module's growing construct and replaces the acceptor module's endonuclease cleavage site and selectable marker. Since only the acceptor module's marker is actively transcribed, recombinants can be readily identified. During the next round of elongation, the endonuclease cleavage site and selectable marker return to the original configuration, allowing assembly to proceed in a cyclical format.

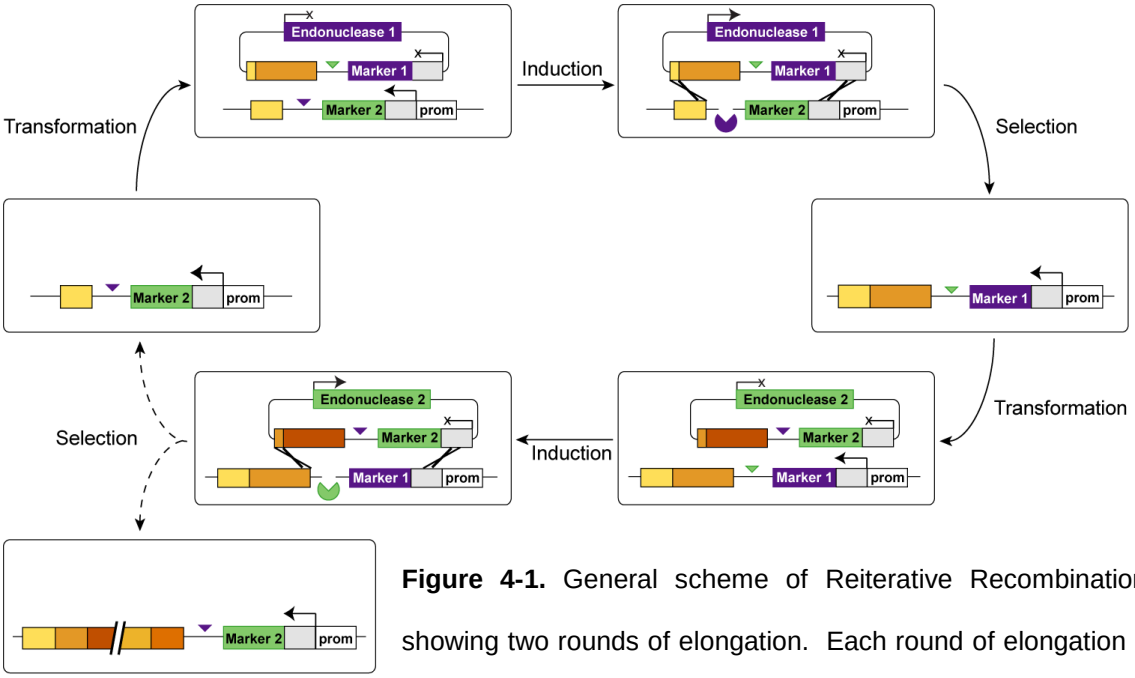


Figure 4-1. General scheme of Reiterative Recombination, showing two rounds of elongation. Each round of elongation is achieved by recombination between an "acceptor module" (in the linear chromosome) and a "donor module" (in the circular plasmid). The two modules contain orthogonal homing endonuclease cleavage sites (triangles) adjacent to different selectable markers (purple and green). Both markers are downstream of a homology region (gray), but only the acceptor module contains a promoter (white) driving marker expression. Endonuclease cleavage of the acceptor module stimulates recombination, joining the fragments being assembled (orange) and replacing the acceptor module's endonuclease site and expressed selectable marker with those of the donor module. Repeating this procedure with a donor module of the opposite polarity returns the acceptor module to its original state, allowing the assembly to be elongated indefinitely.

Several lines of evidence from studies of homologous recombination in *S. cerevisiae* supported our hypothesis that pathways could be constructed with high accuracy and efficiency using this Reiterative Recombination strategy. Paques *et al.* demonstrated that repair of an HO endonuclease-induced chromosomal DSB by homologous recombination from a plasmid template proceeded with high efficiency (>5%) even when fragments up to 9.1 kb in length were inserted in the repair process, demonstrating that we should be able to integrate gene-sized fragments in our analogous system[29]. Inbar and Kupiec found that when two templates were available for the repair of an HO endonuclease-induced break, one containing homology immediately adjacent to the DSB and one containing homology distant from the break, donors with homology 1 to 6 kb from the DSB were utilized with high frequency (≥40% of cells)[30]. This suggests that the presence of selectable markers between the endonuclease cleavage site and the homology regions upstream of the markers should not impede efficient repair. Finally, homology regions as short as ~30 bp are sufficient to effect gene conversion or accurate repair of DSBs in systems expected to have varying degrees of mechanistic similarity to Reiterative Recombination[31-34]. Using such short homology regions between assembled subfragments would be advantageous for Reiterative Recombination, as it would allow homology to be incorporated by PCR, eliminating the need to use more sophisticated *in vitro* techniques (e.g., overlap extension PCR) to add homology to adjacent subfragments from different sources.

4.3 Results

4.3.1 Construction of a system for Reiterative Recombination

We constructed our initial Reiterative Recombination system in *S. cerevisiae*. For the orthogonal endonucleases, we turned to the two well-studied *S. cerevisiae* enzymes employed throughout the homologous recombination literature, HO and SceI. HO specifically cleaves the *MAT* locus to stimulate mating-type switching[35,36]; SceI is encoded by an intron in yeast mitochondrial DNA and has no recognition sites at all in yeast nuclear DNA[37,38]. These enzymes were placed under the *GAL1* promoter, the most widely used inducible promoter in yeast genetics (**Fig. 4-2**).

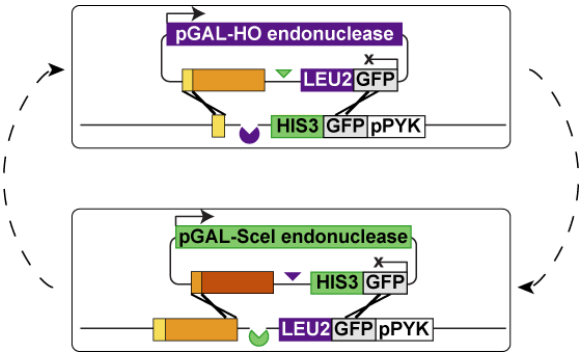


Figure 4-2. Details of Reiterative Recombination. Donor plasmids contain the HO or SceI endonucleases under the control of the galactose-inducible *GAL* promoter and GFP-*HIS3* or GFP-*LEU2* genes that lack a promoter. Acceptor modules have a GFP-*HIS3* or GFP-*LEU2* gene downstream of the constitutive p*PYK1* promoter.

We then created a pair of orthogonal selectable markers with appropriate homology regions (**Fig. 4-2**). We chose *HIS3* and *LEU2*, which provide robust, widely used growth selections by complementing the histidine and leucine auxotrophies of many common yeast strains. To provide an upstream homology region, we constructed *N*-terminal GFP fusions of both markers, and we inserted an HO or SceI recognition site downstream of their terminators. We placed the GFP-*HIS3* construct under a constitutive *PYK1* promoter to create an actively expressed acceptor module marker, and we placed both GFP-marker fusions into centromeric (low-copy) shuttle vectors without promoters

to create donor modules. The donor plasmids also contain a positive and negative selectable *URA3* marker, allowing cells to be cured of donor plasmids after each elongation round by growth on 5-fluoroorotic acid (5-FOA)[39].

To build an initial strain for Reiterative Recombination, we began with BY4733, a background strain with full deletions of all markers used in our system[40], eliminating the potential for unwanted homologous recombination events. We used standard "pop-in/pop-out" gene replacement[41] to put a silent mutation in the *MAT* allele to eliminate its HO cleavage site[42] (**Fig. 4-3**). Then we placed the acceptor module into an integration vector targeting the *HO* locus[43], simultaneously integrating the construct and eliminating homology to the endonuclease gene in the donor plasmid. This basic parental acceptor strain can be used for the assembly of any desired DNA construct. For some applications, it may be desirable to use a different background strain. Now that the appropriate integration plasmids have been constructed, it will only require two integration steps, known to proceed efficiently, to convert any strain with the appropriate auxotrophies into an acceptor cell.

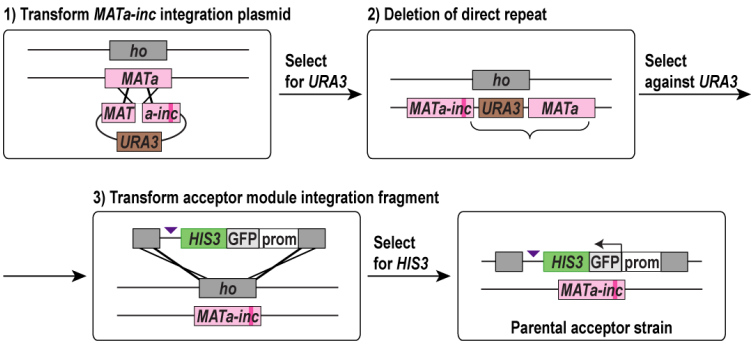


Figure 4-3. Construction of the parental acceptor Reiterative Recombination strain.

4.3.2 Validation of the Reiterative Recombination system

First, we sought to verify that endonuclease-stimulated recombination occurred as expected, leading to both 1) conversion between the alternating *HIS3* and *LEU2* markers

and 2) the insertion of exogenous DNA adjacent to the endonuclease cleavage site. We therefore transformed our Reiterative Recombination parental acceptor strain with a donor plasmid that contained a 950-bp region of homology to the *HO* locus adjacent to the KanMX gene, which confers resistance to G418 in yeast, as shown in **Figure 4-4A**.

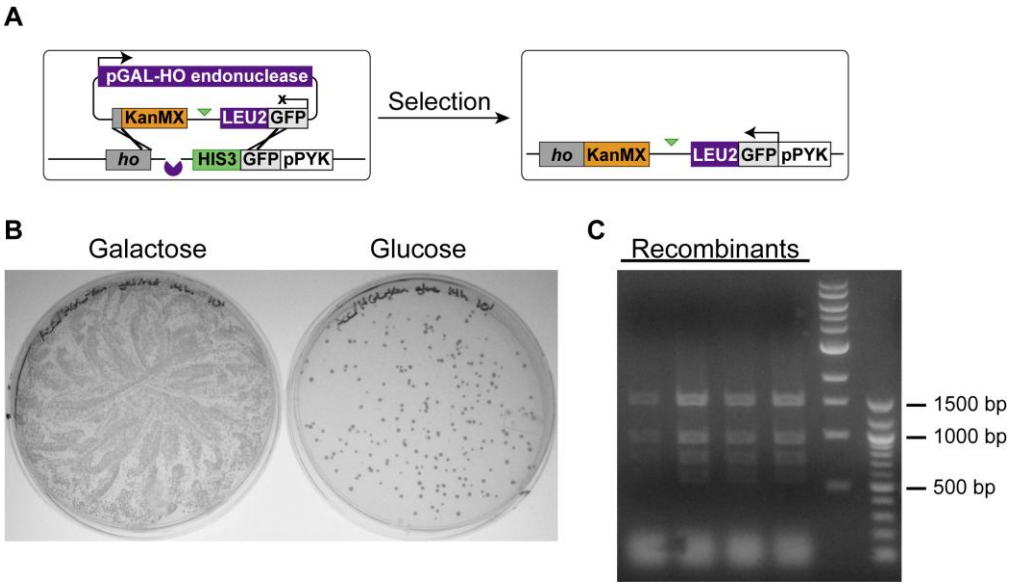


Figure 4-4. Validation of endonuclease-stimulated integration of DNA via Reiterative Recombination. **(A)** The Reiterative Recombination parental acceptor strain was transformed with a donor plasmid containing the features shown. Homologous recombination stimulated by HO endonuclease cleavage of the chromosome is expected to lead to integration of the KanMX marker and conversion of the expressed alternating marker from *HIS3* to *LEU2*. **(B)** Transformants were grown in synthetic media lacking uracil (to select for the donor plasmid) and containing 2% galactose/2% raffinose or in 2% glucose for 14 hours, and 100 μL was plated on synthetic media lacking leucine to assay for acquisition of actively expressed *LEU2*. **(C)** Genomic DNA was purified from 4 *LEU2* colonies cured of the donor plasmid and analyzed by PCR and restriction analysis. Primers amplified from the *ho* allele to GFP. PCR products were digested with ClaI and BsrGI; restriction fragments 1560, 971, 749, 571, and 172 bp in length were expected.

As shown in **Figure 4-4B**, induction of transformants with galactose, which activates HO endonuclease expression, instead of glucose, which represses endonuclease expression, resulted in a large increase in the number of recombinants as assayed by selection for acquisition of the *LEU2* marker. Correct integration of the KanMX gene was validated by curing the recombinants of the donor plasmid and showing that the cells were still resistant to G418. The genomic DNA from four colonies was purified, and the acceptor module was PCR amplified and analyzed by restriction mapping (**Fig. 4-4C**); all recombinants had the expected digestion pattern.

Next, we tested whether efficient homologous recombination could still occur in our system when increasingly shorter homology regions were utilized. As shown in **Figure 4-5**, galactose induction of transformants with donor plasmids containing regions of homology to the *ho* locus as short as 41 bp still resulted in high-efficiency acquisition of the *LEU2* phenotype. As homology regions decreased in length, background levels of recombination in glucose media due to leaky endonuclease expression or unstimulated homologous recombination also decreased (**Table 4-1**); induction of the endonuclease became increasingly important for obtaining maximal recombination efficiency. Importantly, 40-bp homology regions are short enough to be readily incorporated with PCR primers, meaning that Reiterative Recombination can easily be used to assemble fragments obtained from different sources.

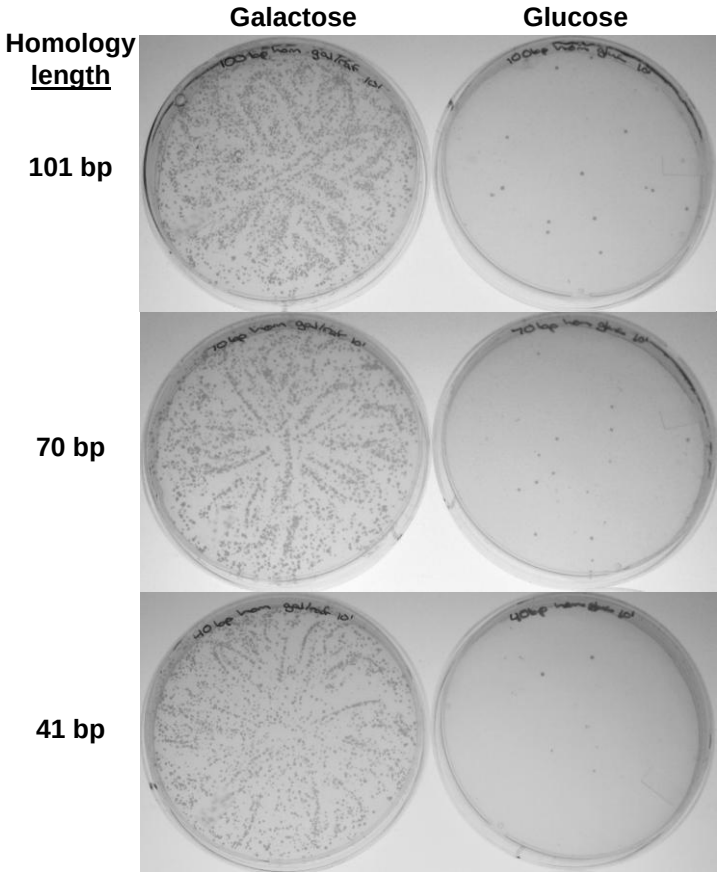


Figure 4-5. Marker conversion efficiencies in Reiterative Recombination with decreasing lengths of homology. Experiments were performed as in **Figure 4-4B** but using donor plasmids containing various lengths of homology to the *ho* locus. Twelve-hour inductions in galactose or glucose were performed before plating on leucine-deficient media.

Table 4-1. Marker conversion efficiencies in Reiterative Recombination with decreasing lengths of homology.

Homology length (bp)	Induction media	Recombinants/ mL/OD_{600}	Fold-induction by galactose
101	Glucose Galactose	65 33000	500
70	Glucose Galactose	62 40000	650
41	Glucose Galactose	12 31000	2700

Finally, we confirmed that the recombination process could be continued for multiple consecutive rounds. We transformed the cured recombinants obtained in the above experiments with a donor plasmid of the opposite polarity that contained a 134-bp homology region and then induced expression of the orthogonal SceI endonuclease with

galactose (**Fig. 4-6A**). Galactose induction resulted in a high rate of acquisition of the *HIS3* marker (**Fig. 4-6B**), returning recombinants to the parental acceptor's strain phenotype. Using these cured round 2 recombinants, we then verified that we could continue the cyclical Reiterative Recombination process for a third round, as shown in **Figure 4-6C,D**.

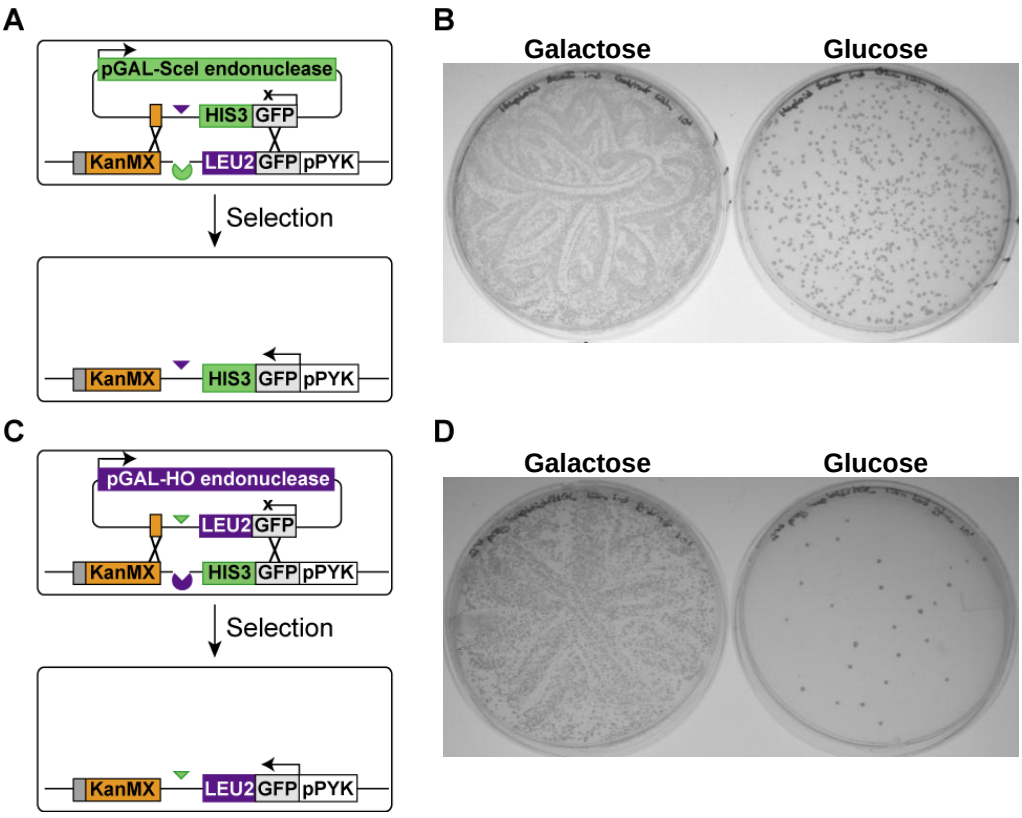


Figure 4-6. Consecutive rounds of Reiterative Recombination. Cured recombinants from the experiments in **(A,B) Figure 4-4** or **(C,D) Figure 4-6A,B** were transformed with donor plasmids containing the features shown. Homologous recombination stimulated by endonuclease cleavage of the chromosome is expected to lead to conversion of the expressed alternating marker from **(A)** *LEU2* to *HIS3* or **(C)** *HIS3* to *LEU2*. **(B,D)** Transformants were grown for 12 hours in galactose or glucose media as in **Figure 4-4**, and 100 μL was plated on synthetic media lacking **(B)** histidine or **(D)** leucine to assay for acquisition of actively expressed *HIS3* or *LEU2*, respectively.

4.3.3 Construction of a "pathway" of genes via Reiterative Recombination

Having validated the basic machinery of Reiterative Recombination, we then employed our system to build a "pathway" of genes, forcing our system to cycle through multiple rounds of elongation. We sequentially integrated the reporter genes *lacZ* (β-galactosidase; red when assayed with Magenta-Gal), *gusA* (β-glucuronidase; blue when assayed with X-Gluc), and *MET15* (O-acetylserine and O-acetylhomoserine sulfhydrylase; complements methionine auxotrophy) using three rounds of assembly, creating an 8.5-kb construct (**Fig. 4-7**).

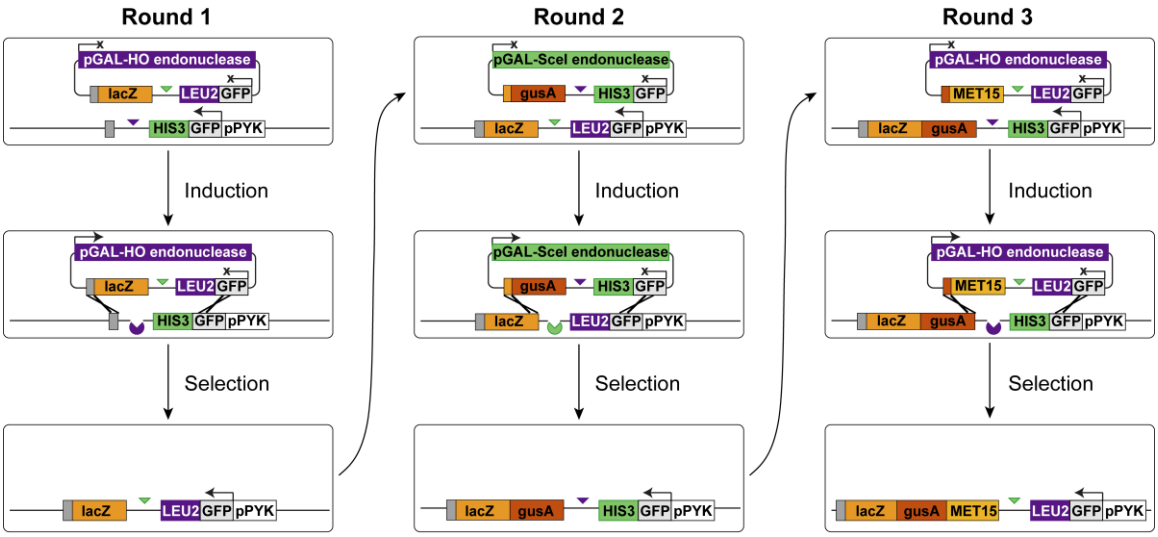


Figure 4-7. Construction of a reporter gene "pathway" by Reiterative Recombination. Details of the assembly process in which the three reporter genes *lacZ*, *gusA*, and *MET15* were sequentially integrated into the chromosome.

Subfragments for integration were PCR amplified as one or two overlapping pieces using primers that incorporated short regions of homology (30-40 bp) 1) to the preceding piece of the growing assembly and 2) to the donor plasmid. PCR products were co-transformed with a digested, generic donor plasmid into the acceptor strain to

generate intact donor plasmids by plasmid gap repair (**Fig. 4-8**)[44]. Our procedure thus eliminates any *in vitro* manipulation (e.g., subcloning) other than basic PCR.

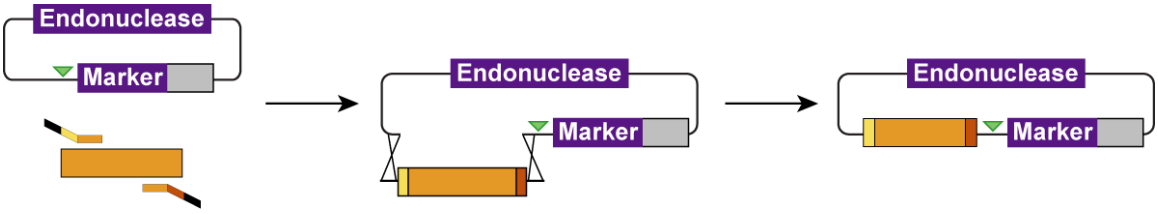


Figure 4-8. Construction of donor plasmids by plasmid gap repair. A digested universal donor plasmid and PCR fragments with appropriate homology regions are co-transformed into the Reiterative Recombination strain and assembled via homologous recombination.

As negative controls, we transformed donor plasmids lacking the endonuclease gene and/or homology to the previously integrated subfragment in the pathway at every round. These transformants were induced with galactose and glucose in media in parallel to the intact donor plasmid. Galactose induction of endonuclease expression in the transformants led to a high rate of marker conversion only when both the endonuclease gene and the homology on both sides of the endonuclease cut site were present (**Fig. 4-9**).

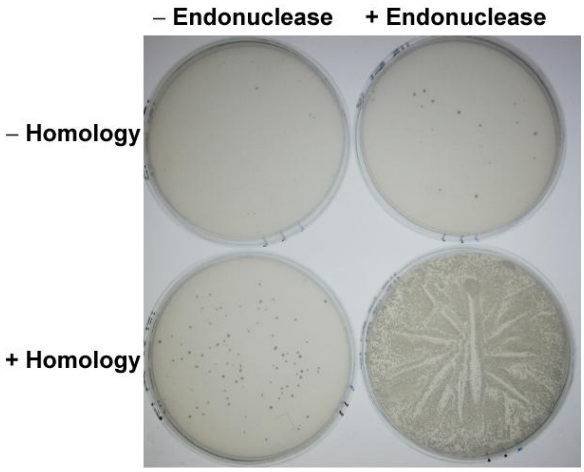


Figure 4-9. Donor plasmid controls in the reporter gene "pathway" assembly. Results of the round 2 induction step are shown as a representative example. Cells containing identical donor plasmids lacking the SceI endonuclease gene and/or the *gusA* fragment with *lacZ* homology were induced in parallel. A calculated $6x10^6$ cells were plated on synthetic media lacking histidine to assay for selective marker conversion after a 12-hour galactose induction. Homologous recombination stimulated by SceI endonuclease cleavage of the chromosome is expected to lead to conversion of the expressed alternating marker from *LEU2* to *HIS3*.

Phenotypic analysis of recombinants following donor plasmid curing indicated that auxotrophies for histidine and leucine alternated with each round of elongation, as expected (**Fig. 4-10A**, columns). Each newly integrated reporter (*lacZ*, *gusA*, or *MET15*) was functional in 75-100% of recombinants when >40 individual colonies from each round were assayed (**Fig. 4-10B,C,D**), and previously integrated reporters were maintained (**Fig. 4-10A**, rows). We also confirmed that integration occurred in the expected manner by analyzing the purified genomic DNA of cured recombinants by PCR and restriction digestion (**Fig. 4-11**).

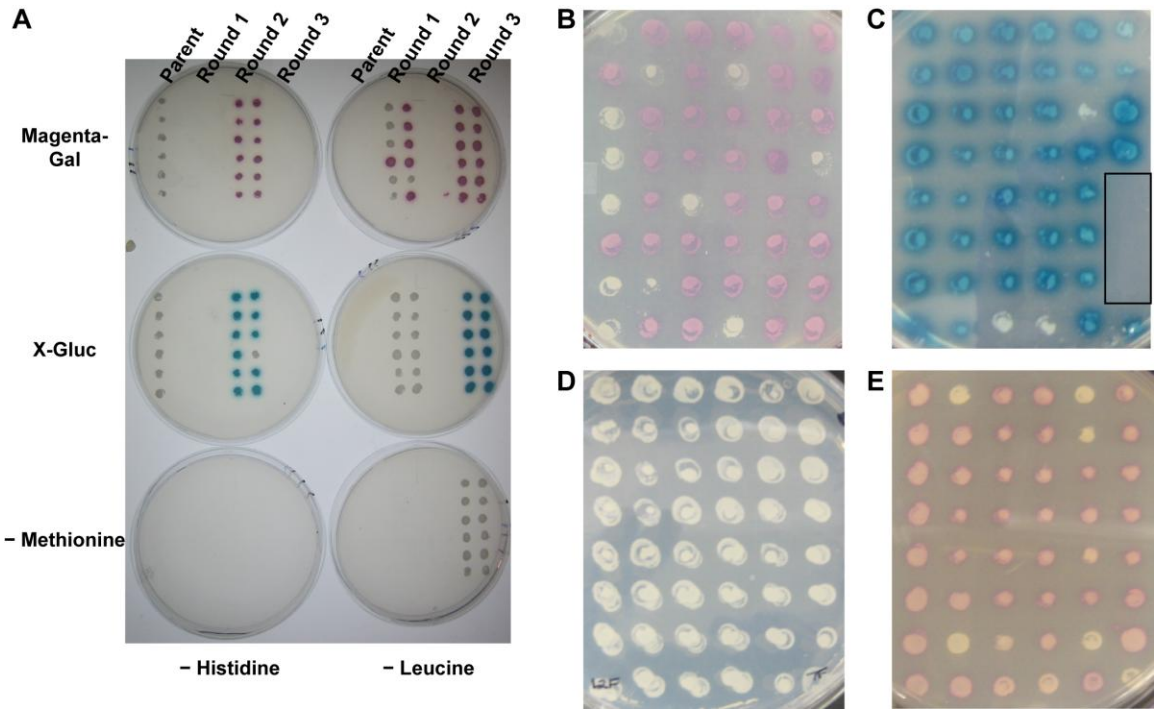
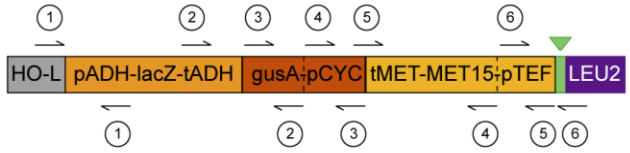


Figure 4-10. Phenotypic analysis of cured recombinants from the reporter gene "pathway" assembly. **(A)** Phenotypes of 12 unique cured colonies from each round of assembly. In columns, recombinants are assayed for the *HIS3* (synthetic media lacking histidine) and *LEU2* (synthetic media lacking leucine) markers. In rows, recombinants are assayed for *lacZ* (Magenta-Gal), *gusA* (X-Gluc), and *MET15* (synthetic media lacking methionine). **(B-E)** Phenotypic analysis of a larger number of phrogged cured recombinants from **(B)** round 1, **(C)** round 2, and **(D)** round 3 of the reporter gene proof-of-principle system assayed with Magenta-Gal, X-Gluc, and methionine-deficient media, respectively. No colonies were phrogged in the boxed area of **(C)**. **(E)** Phenotypic analysis with Magenta-Gal of cured round 1 recombinants resulting when *lacZ* was amplified as two overlapping subfragments.



	Amplicon	Total length (bp)	Enzyme	Expected fragments (bp)
1	HO-L - pADH	455	BfuAI	320,135
2	tADH - gusA	1932	BsaWI	823,682,427
			BsrGI	1132,426,374
			BsrGI/HindIII	1132,426,266,108
3	gusA - pCYC	2336	BsaWI	1163,569,427,177
4	pCYC - MET	2179	BsaWI	1289,377,277,236
5	MET - pTEF	2114	BsmAI	1505,290,195,124
6	pTEF - LEU	472	BsmAI	277,195

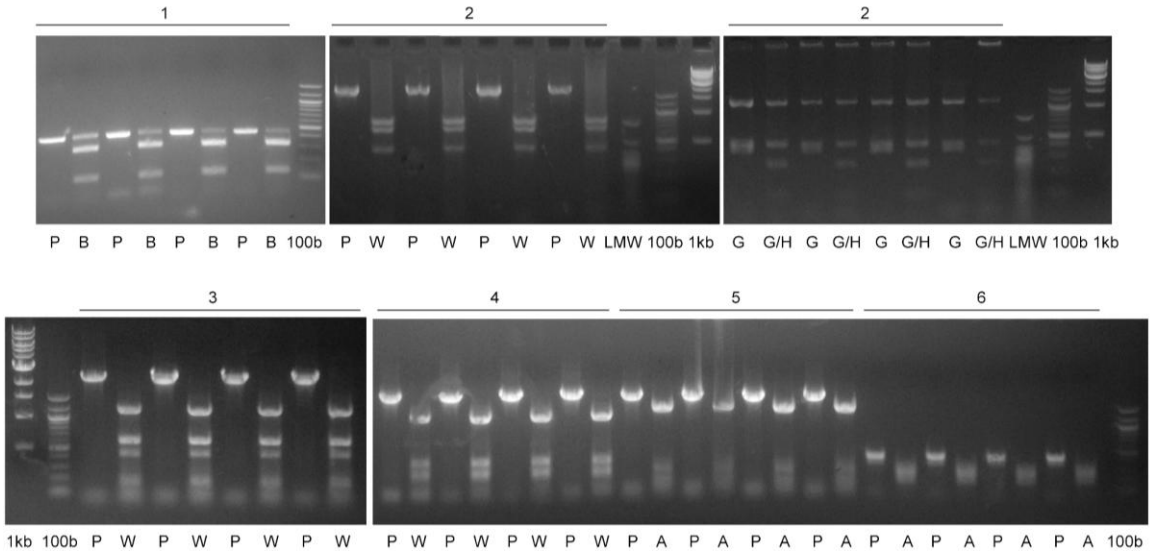


Figure 4-11. Genotypic analysis of cured recombinants from the reporter gene "pathway" assembly. Genomic DNA was purified from four to six colonies from each round and analyzed by PCR and restriction digestion. Representative data from round 3 colonies are shown. In the diagram of the integrated construct, solid lines between regions of different color represent new junctions that were created during endonuclease-stimulated integration; dashed lines indicate new junctions that were created between PCR fragments by plasmid gap repair. (P=undigested PCR product, B=BfuAI digest, W=BsaWI digest, G=BsrGI digest, G/H=BsrGI/HindIII digest, A=BsmAI digest, LMW=low molecular weight ladder, 100b=100 bp ladder, 1kb= 1 kb ladder)

To determine the source of the recombinants with inactive *lacZ* and *gusA* reporters in the reporter proof-of-principle system, we purified the genomic DNA of three of the white round 1 colonies and the single white round 2 colony shown in **Figure 4-10A**. PCR and restriction analysis of the white round 2 colony indicated correct construction of the pathway. Sequencing of the integrated *gusA* gene showed that it had two mutations in the amino acid sequence, D436G and F551V. For the white round 1

colonies, we could amplify the 5' end of the expected construct (*HO(L)*→p*ADH*) but not the 3' end (t*ADH*→*LEU2*), indicating that the complete fragment had not integrated as expected. We have not observed this result for any of the other constructs tested. Analysis of the original *lacZ* PCR product transformed during round 1 revealed that there was a truncated DNA fragment that could not be removed by gel purification. We subsequently repeated round 1, amplifying *lacZ* as two shorter, overlapping subfragments rather than as a single subfragment. As shown in **Figure 4-10E**, a higher percentage of colonies (42 out of 48, or 87.5%) tested positive for *lacZ* using this modified protocol, and PCR analysis of the white colonies indicated that all contained the complete fragment.

4.4 Discussion

In this work, we have designed, implemented, and characterized a high-yielding, robust system for installing multi-gene pathways in the yeast chromosome. Coupling the integration of DNA fragments to endonuclease cleavage of the chromosome and the conversion of a selectable marker enforces accurate and efficient DNA assembly. The Reiterative Recombination framework is not construct specific, meaning that essentially any desired pathway of genes can be built with our existing strain. In addition, if the use of a strain with a specific genetic background is required, most standard *S. cerevisiae* strains can easily be prepared for Reiterative Recombination in less than two weeks' time. Finally, DNA assembly systems analogous to our initial Reiterative Recombination method could be built in any organism with efficient homologous recombination machinery (e.g., *Bacillus subtilis*).

By providing a high-yielding method for the assembly of multi-gene constructs *in vivo,* Reiterative Recombination opens the door to the routine construction of gene circuits, pathways, and libraries thereof in the cell. Reiterative Recombination's high efficiency, together with its technical straightforwardness, makes it a reliable method for building pathways that is accessible to non-experts without specialized equipment. While a handful of laboratories that are experts in the field have described landmark achievements in the realm of large-scale DNA assembly, these techniques have not yet been widely adopted by the scientific community. Reiterative Recombination distills the construction of individual pathways into a user-friendly process that can be carried out by any laboratory equipped for basic molecular biology.

The introduction of what we now consider "basic" molecular biology tools revolutionized our ability to study the function of individual genes and proteins; in the same way, we will need new, equally empowering technologies as the scale of our ambitions increases and our applications move into living cells. Technologies such as Reiterative Recombination will contribute to the advance of synthetic biology by allowing cell engineers to easily build and refine new pathways *in vivo* so that reprogramming the cell can become a routine reality rather than a rare success.

4.5 Experimental methods

General materials and methods. General materials and methods were as in **Chapter 2**.

Construction of odd donor plasmid (pLW2592). A 200-bp region of *Humicola insolens* Cel7B flanked by SfiI sites between SpeI and XbaI sites was amplified from pHL1262 with primers LMW244 and LMW245. The PCR product and pRS416 were

digested with SpeI and XbaI and ligated to create vector pLW2639. *LEU2* was amplified from pRS425 with primers LMW250 and LMW255, and the product was amplified with primers LMW287 and LMW288; the 2 rounds of PCR added 20 bp of homology to the p*PYK* promoter at the 5' end and an SceI cleavage site plus 30 bp of homology to pLW2639 at the 3' end. The p*PYK* promoter was amplified from FY251 genomic DNA with primers LMW284 and LMW286, adding 30 bp of homology to pLW2639 at the 5' end and 20 bp of homology to *LEU2* at the 3' end. The vector pLW2639 was digested with BstXI and co-transformed with the p*PYK* and *LEU2* fragments into yeast to create vector pLW2641, containing *LEU2* under control of the *PYK* promoter. The yEGFP gene was amplified from pJEB2289E with primers LMW304 and LMW308, and the product was amplified with primers LMW305 and LMW308, adding an XbaI restriction site and 32 bp of homology to pLW2641 at the 5' end, and a $(GSG)_2$ linker followed by 40 bp of homology to *LEU2* at the 3' end. Vector pLW2641 was digested in the p*PYK* region with SnaBI and co-transformed with the GFP fragment into yeast to create vector pLW2646, containing a promoterless GFP-$(GSG)_2$-*LEU2* fusion protein. The HO endonuclease gene was amplified from the genomic DNA of V2237Y with primers LMW310 and LMW311, containing 30 bp of homology to the *GAL* promoter and *CYC1* terminator of pRS426GAL. The vector pLW2638, containing *H. insolens* Cel7B in pRS426GAL, was digested in the Cel7B gene and multiple cloning site with SalI and BstXI and co-transformed with the HO PCR product into yeast to create vector pLW2649, containing the HO endonuclease gene under the control of the galactose promoter. The entire p*GAL-HO*-t*CYC* construct was amplified from pLW2649 with primers LMW300 and LMW301, containing 30 bp of homology to pLW2646 3' to the

GFP-*LEU2* fusion and SfiI stuffer region. The vector pLW2646 was digested with SalI and co-transformed with the HO PCR product into yeast to create vector pLW2592. The parental odd donor plasmid pLW2592 contains (promoterless) GFP-*LEU2*, followed by an SceI cleavage site, followed by an SfiI stuffer, followed by t*CYC*-*HO*-p*GAL* (transcribed towards the SfiI stuffer) (**Fig. 4-12**). The plasmid pLW2592 can be digested with HindIII, BsaBI, NotI, EagI, AleI, Eco53kI, or SacI to prepare it for plasmid gap repair.

Construction of round 1 donor plasmid (pLW2594). The *HO(L)*-KanMX region was amplified from pVC2265 with primers LMW336 and LMW337 and co-transformed with BsaBI-digested pLW2592 into yeast to generate pLW2594 (**Fig. 4-12**). This plasmid can be used in place of the universal odd donor plasmid pLW2592 in round 1. The *HO(L)* region provides homology to facilitate integration of the first subfragment. The KanMX region can be cut out by restriction digestion (SmaI, XmaI, TspMI, BsoBI, or AvaI) before gap repair to make the desired linearized round 1 donor plasmid. Uncut pLW2594 can be used as a positive control during the first round of Reiterative Recombination, as cured recombinants acquire resistance to G418 due to integration of the KanMX marker.

Construction of even donor plasmid (pLW2593). *HIS3* was amplified from pRS423 with primers LMW253 and LMW256, and the product was amplified with primers LMW289 and LMW290; the 2 rounds of PCR added 20 bp of homology to the p*PYK* promoter at the 5' end and an HO cleavage site plus 30 bp of homology to pLW2639 at the 3' end. The p*PYK* promoter was amplified from FY251 genomic DNA with primers LMW284 and LMW285, adding 30 bp of homology to pLW2639 at the 5'

end and 20 bp of homology to *HIS3* at the 3' end. The vector pLW2639 was partially digested with TfiI and co-transformed with the p*PYK* and *HIS3* fragments into yeast to create plasmid pLW2640, containing *HIS3* under the control of the *PYK* promoter. An SfiI site was removed from the *HIS3* gene by site-directed mutagenesis with primers LMW302 and LMW303 (Stratagene QuikChange Lightning) to generate vectors pLW2642 and pLW2643. Vector pLW2642 was later found to have an insert in *HIS3* that appeared to be the result of concatamerization of the mutagenesis primers. The yEGFP gene was amplified from pJEB2289 with primers LMW306 and LMW308, and the product was amplified with primers LMW307 and LMW308, adding an XbaI restriction site and 32 bp of homology to pLW2642 at the 5' end, and a $(GSG)_2$ linker followed by 40 bp of homology to *HIS3* at the 3' end. Vectors pLW2642 and pLW2643 were digested in the p*PYK* region with SnaBI and co-transformed with the GFP fragment into yeast to create vectors pLW2644 and pLW2645, respectively, containing a promoterless GFP-$(GSG)_2$-*HIS3* fusion protein. The p*GAL-SceI*-t*CYC* construct was amplified with primers LMW300 and LMW301, containing 30 bp of homology to pLW2644 3' to the GFP-*HIS3* fusion and SfiI stuffer region. The vector pLW2644 was digested with SalI and co-transformed with the SceI PCR product into yeast to create the vector pLW2648. To eliminate the insertion in *HIS3* in pLW2648, the GFP-*HIS3* construct from pLW2650 was amplified with primers LMW290 and LMW308 and co-transformed with pLW2648, which had been digested with MscI and NheI to cut out part of the GFP-*HIS3* gene, into yeast to create the vector pLW2652. The MCS from pUC18 was amplified with primers LMW360 and LMW361 and co-transformed with NotI-digested pLW2652 into yeast. The resulting generic even donor plasmid pLW2593

contains (promoterless) GFP-*HIS3*, followed by an HO cleavage site, followed by a multiple cloning site, followed by t*CYC*-*SceI*-p*GAL* (transcribed towards the multiple cloning site) (**Fig. 4-12**). The plasmid pLW2593 can be digested with SphI, SalI, TspMI, XmaI, SmaI, Eco53kI, SacI, or EcoRI to prepare it for plasmid gap repair.

Construction of donor plasmid negative controls (pLW2595 and pLW2596). The p*GAL*-MCS-t*CYC* construct from pRS416GAL was amplified using primers LMW300 and LMW301. The plasmids pLW2646 and pLW2593 were digested with SalI and BseRI, respectively, and co-transformed with the PCR product into yeast to create the desired plasmids via homologous recombination. The resulting plasmid pLW2595 (from pLW2646) is equivalent to the odd donor plasmid pLW2592 except that there is no HO endonuclease gene under the control of the *GAL* promoter; the resulting plasmid pLW2596 (from pLW2593) is identical to the even donor plasmid pLW2593 except that there is no SceI endonuclease gene under the control of the *GAL* promoter.

Construction of acceptor module integration plasmid (pLW2590). The p*PYK* promoter was amplified from pLW2641 with primers LMW284 and LMW309, adding 30 bp of homology to pLW2644 at the 5' end and 36 bp of homology to pLW2644 (5' end of GFP) at the 3' end. The vector pLW2644 was digested with XbaI and co-transformed with the p*PYK* fragment into yeast to create vector pLW2650, containing the p*PYK*-GFP-*HIS3*-HO site construct. (As transformants were selected on media lacking both histidine and uracil, the insert in *HIS3* evidently looped out via homologous recombination to restore a functional *HIS3* gene.) Plasmids pRS416 and pVC2265 were digested with SpeI and ligated in order to place the *HO(L)*-KanMX-*HO(R)* integration module into a centromeric vector. The desired plasmid pLW2653 was recovered by selecting for the

KanMX marker via G418 resistance in yeast and kanamycin resistance in *E. coli*. The p*PYK*-GFP-*HIS3*-HO site acceptor module was amplified from pLW2650 with primers LMW326 and LMW327, containing 30 bp of homology to *HO(L)* (HO cleavage site end) and *HO(R)* (p*PYK* end). The vector pLW2653 was digested in KanMX with EcoNI and BseRI and co-transformed with the acceptor module PCR into yeast to create vector pLW2590. A fragment to integrate the odd acceptor module can be created by digesting pLW2590 with SpeI (**Fig. 4-12**).

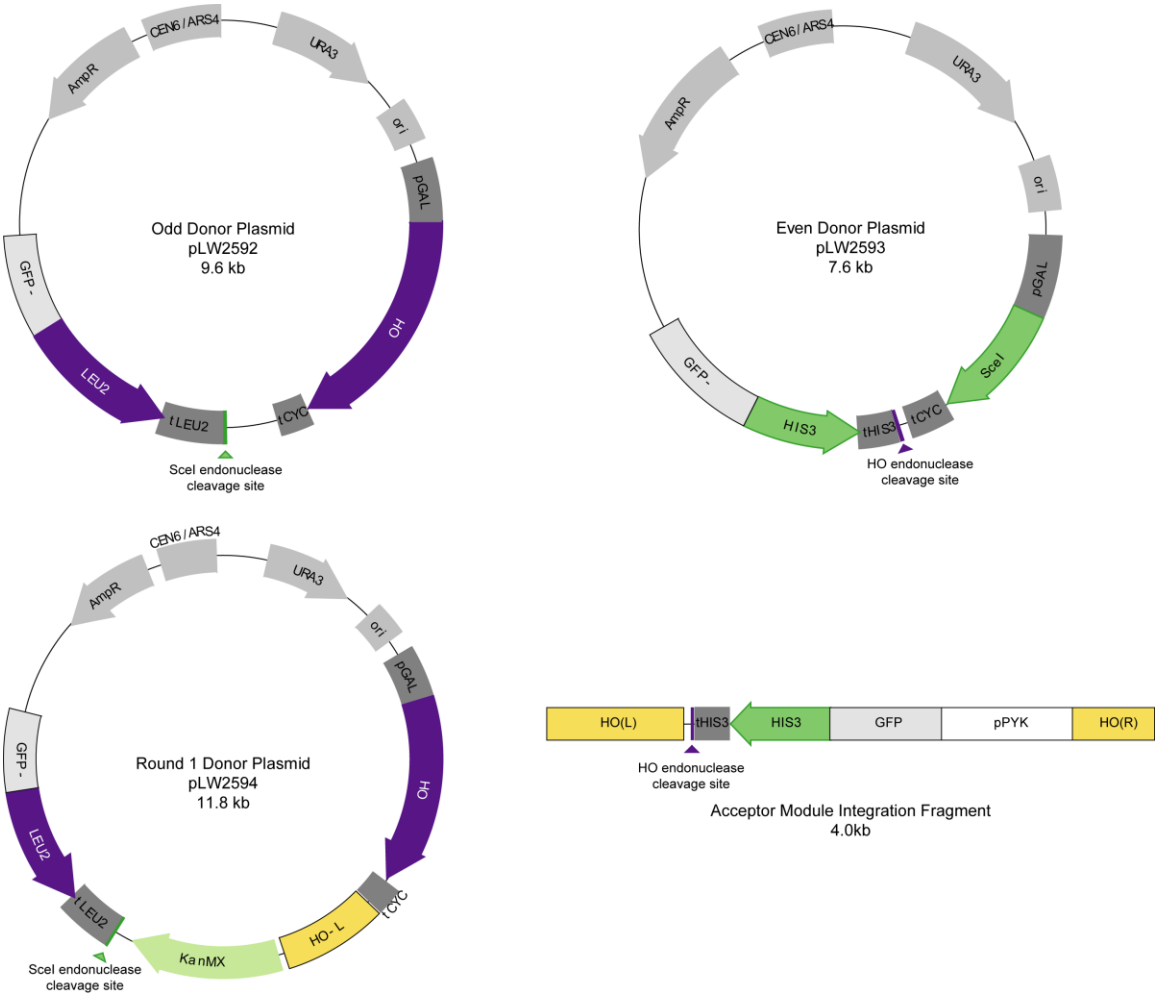


Figure 4-12. Maps of donor plasmids and the acceptor module integration fragment for Reiterative Recombination.

Construction of acceptor strain. The *MATa*-inc allele was amplified from DY3025 genomic DNA with primers LMW312 and LMW313, incorporating homology to pRS426GAL, and co-transformed with SalI- and BstXI-digested pLW2638 into yeast to create plasmid pLW2651. The 2μ origin of replication was cut out of plasmid pLW2651 with AfeI, and the vector was religated and transformed into *E. coli* to create the integration vector pLW2588. The vector pLW2588 was digested with BglII in the *MATa*-inc allele, transformed into BY4733, and selected on SC(−Uracil) media. All transformants analyzed by colony PCR had the *MATa*-inc allele 5' (in the W to Z direction[36]) to the plasmid sequences and the wild-type *MATa* allele 3' to the plasmid sequences. Transformants were grown non-selectively in YPD to allow loop-out of the duplicated gene and then plated on synthetic media containing 0.1% 5-FOA. Five out of 32 5-FOA-resistant colonies analyzed by colony PCR and restriction mapping with AciI contained the *MATa*-inc allele, including LW2589Y. The genotype of this colony was further verified by sequencing of the *MAT* locus. The vector pLW2590 was digested with SpeI and transformed into LW2589Y to integrate the odd acceptor module at the *HO* locus. Correct integration was confirmed by colony PCR of HIS^+ transformants, including the acceptor strain LW2591Y.

Basic Reiterative Recombination protocol. The protocol for an even round (e.g., round 2) of Reiterative Recombination is described.

1) *Preparation of subfragments:* Subfragments were amplified with primers that added appropriate homology to adjacent fragments and to the donor plasmid (see below). All PCR products were gel purified.

2) *Transformation:* The PCR products were co-transformed with the digested even donor plasmid (pLW2593) in a 100:1 ratio into the cured round 1 strain. Transformants were selected on synthetic complete media lacking leucine and uracil (SC(−Leucine, −Uracil)).
3) *Pre-Induction:* (Optional) After two days of growth, transformants were lifted from the transformation plates, washed once with sterile water, resuspended in pre-induction media (SC(Lactate, −Leucine, −Uracil)) to an OD_{600} of 1, and shaken at 30°C for three hours. (All experiments in this chapter incorporated a pre-induction step, but we have subsequently found that the pre-induction does not significantly improve the efficiency of marker conversion in Reiterative Recombination.)
4) *Induction:* Cells were then harvested, washed once with sterile water, and resuspended in induction media (SC(−Uracil, 2% galactose, 2% raffinose)) to an OD_{600} of 0.1. Cells were shaken at 30°C for 12 hours.
5) *Selection:* For control experiments, aliquots of the induction cultures were immediately plated on selective media SC(−Histidine) to determine the efficiency of marker switching. Colonies were counted after two days of growth. The remaining cells were inoculated into SC(−Histidine) liquid media, shaken at 30°C for one day, and plated on SC(−Histidine, 0.1% 5-FOA) to cure recombinants of the donor plasmid.
6) *Reiteration:* To begin the next round of Reiterative Recombination, after two days of growth, a single cured colony from the SC(−Histidine, 0.1% 5-FOA) plates

was inoculated into SC(−Histidine) liquid media to begin an overnight culture for the next transformation.

For odd rounds of Reiterative Recombination, pLW2592 was used as the donor plasmid, and the use of histidine and leucine in dropout media was reversed. All other aspects of the protocol remained the same.

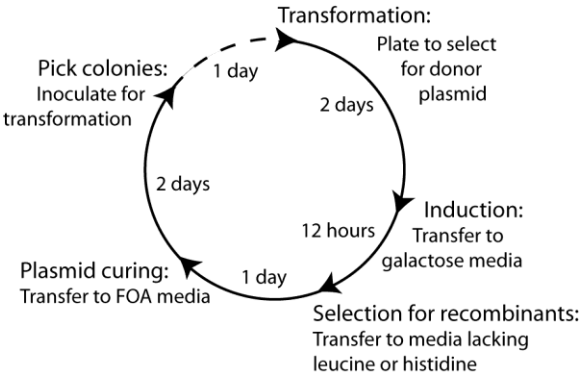


Figure 4-13. Reiterative Recombination timeline. Arrows represent procedural steps; times indicate periods of growth.

Fragment design. For clarity, a "fragment" refers to the total pathway-specific region of each donor plasmid, shown in orange in the figures. When convenient, fragments were divided into "subfragments" that were PCR amplified from different templates and assembled into the full fragment by plasmid gap repair upon transformation into yeast. Fragments contain 30 bp of homology to the donor plasmid and 20 bp of homology to the adjacent fragments (to provide a total of 40 bp of homology for each integration event) (**Fig. 4-14**). Overlapping ends of subfragments (within fragments) contained a total of 40 bp of homology. All regions of homology were incorporated with PCR primers.

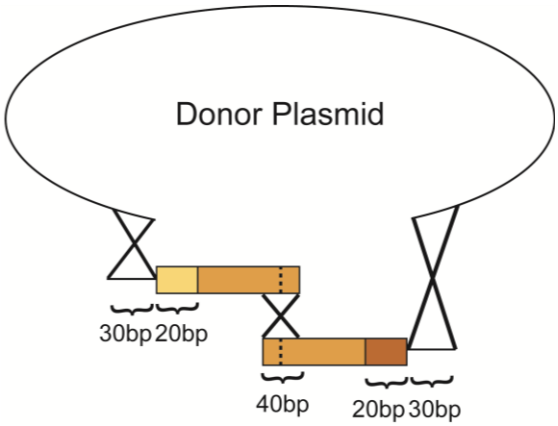


Figure 4-14. General design of subfragment homology regions for plasmid gap repair and Reiterative Recombination. The recombination events that create a construct-specific donor plasmid from a universal donor plasmid and subfragment PCR products are shown.

The sequence of the assembled reporter gene pathway is provided in the **Appendix**. The primers used to amplify constructs for donor plasmids made by plasmid gap repair in this chapter are shown in **Table 4-2**. The numbers (1) and (2) indicate the primers and templates used for the first and second rounds of amplification of the subfragments if more than one round was necessary. We used two rounds of PCR so that we could create universal outer primers that added homology to the donor plasmids. This decreased the cost of primers and kept them short enough to require only standard desalting purification. Primers LMW374 and LMW367 would be used as the outer primers for all subsequent odd rounds of Reiterative Recombination, and primers LMW374 and LMW375 serve as the outer primers for all even rounds of Reiterative Recombination. From 5' to 3', inner primers contain an annealing region for the outer primers (if necessary), 20 bp of homology to the adjacent fragment or subfragment, and a priming region for the subfragment being amplified. Typically we could use the unpurified PCR from first reaction as a template for the second round. For the round 1 reporter gene pathway subfragment, we added homology to *HO(L)* and used the round 1 donor plasmid pLW2594 rather than the universal odd donor plasmid pLW2592.

Table 4-2. PCR amplification of subfragments for Reiterative Recombination.

Elongation Round	Subfragment	Construct	Primers	Template
Homology length test				
1 (pLW2592 donor)	41 bp homology	*HO(L)*-KanMX	LMW353/337	pVC2265
	70 bp homology	*HO(L)*-KanMX	LMW354/337	pVC2265
	101 bp homology	*HO(L)*-KanMX	LMW355/337	pVC2265
	950 bp homology	*HO(L)*-KanMX	LMW336/337	pVC2265
Reporter gene pathway				
1 (pLW2594 donor)	1	p*ADH-lacZ*-t*ADH*	1) LMW419/420 2) LMW419/367	1) pLW2577 2) PCR (1)
2 (pLW2593 donor)	2a	*gusA* (reverse complement)	1) LMW421/422 2) LMW374/422	1) pLW2655 2) PCR (1)
	2b	p*CYC* (reverse complement)	1) LMW423/424 2) LMW423/375	1) FY251 gDNA 2) PCR (1)
3 (pLW2592 donor)	3a	*MET15*-t*MET15* (reverse complement)	1) LMW425/426 2) LMW374/426	1) FY251 gDNA 2) PCR (1)
	3b	p*TEF* (reverse complement)	1)LMW427/428 2)LMW427/367	1) FY251 gDNA 2) PCR (1)

Validation of Reiterative Recombination system. The *HO(L)*-KanMX region of pVC2265 was amplified with 5' primers that annealed to different positions on *HO(L)* to obtain homology regions of different lengths as shown in **Table 4-2**. The primers also added 30 bp of homology to pLW2592. The vector pLW2592 was digested in the SfiI stuffer region with BsaBI and co-transformed with the *HO(L)*-KanMX insert into LW2591Y. Colonies were lifted from the transformation plates (SC(−Histidine, −Uracil)), washed with sterile water, resuspended in 5 mL of pre-induction media (SC(Lac, −Histidine, −Uracil)), and shaken at 30ºC for 3 hours. Two 2-mL aliquots were harvested separately, washed with sterile water, resuspended in 5 mL of induction media (SC(−Uracil, 2% galactose, 2% raffinose)) or non-inducing media (SC(−Uracil, 2% glucose)), and shaken at 30ºC for 14 hours. Cells were plated on SC(−Leucine) to

identify successful round 1 recombinants, which were counted after 2 days of growth at 30ºC. Colonies scraped from the SC(−Leucine) plates (from the galactose induction) were plated on SC(−Leucine, 0.1% 5-FOA) to cure the cells of remaining donor plasmid. Pooled 5-FOA-resistant colonies were plated on YPD, YPD/G418, and SC(−Uracil) to verify successful integration of the KanMX marker and loss of the donor plasmid. Genomic DNA from four round 1 recombinants was purified and analyzed by PCR and restriction analysis to verify that recombination occurred as expected, and the acceptor module from one of these colonies was sequenced as further verification.

To begin round 2, these 5-FOA-resistant round 1 recombinants were transformed with pLW2652. Colonies were lifted from the transformation plates (SC(−Leucine, −Uracil)), washed with sterile water, resuspended in 3 mL pre-induction media (SC(Lactate, (−Leucine, −Uracil)), and shaken at 30ºC for 3 hours. Two 1-mL aliquots were harvested separately, washed with sterile water, resuspended in 1.5 mL of induction media (SC(−Uracil, 2% galactose, 2% raffinose)) or non-inducing media (SC(−Uracil, 2% glucose)), and shaken at 30ºC for 12 hours. Cells were plated on SC(−Histidine) to identify successful round 2 recombinants, which were counted after 2 days of growth at 30ºC. Colonies scraped from the SC(−Histidine) plates (from the galactose induction) were plated on SC(−Histidine, 0.1% 5-FOA) to cure the cells of the remaining donor plasmids.

To begin round 3, these 5-FOA-resistant round 2 recombinants were transformed with pLW2592. The procedure for round 1 was repeated.

4.6 Strains, plasmids, and oligonucleotides

Table 4-3. Strains used in this study

Name	Genotype	Source/Reference
BY4733	*MATa his3Δ200 leu2Δ0 met15Δ0 trp1Δ63 ura3Δ0*	ATCC #200895/40
DY3025	*MATa-inc ade2-101 his3-200 leu2Δ1 lys2-801::pUCGALHO::LYS2 trp1Δ1*	ATCC #MYA-2358/45
FY251	*MATa trp1Δ63 his3Δ200 ura3-52 leu2Δ 1 GAL+*	M. Carlson
V2237Y	*MATa pGAL-HO hmla trp1 leu2 his3 ura3 ade2-1 can1-100 SWI+*	K. Nasmyth/46
LW2589Y	BY4733 *MATa-inc*	This study
LW2591Y	Reiterative Recombination parental acceptor strain BY4733 *MATa-inc* pLW2590 integrated	This study
LW2659Y	Round 1 recombinant from reporter gene pathway	This study
LW2660Y	Round 2 recombinant from reporter gene pathway	This study
LW2661Y	Round 3 recombinant from reporter gene pathway	This study

Table 4-4. Plasmids used in this study

Name	Details	Source/Reference
pHL1262	pRS426MET- *H. insolens* Cel7B	H. Lin/47
pJEB2289	Plasmid containing 8LexAop-pGAL-yEGFP	B. Petersen/48
pRS416	*CEN6/ARSH4 URA3* pBIISK ori amp^R	ATCC #87521
pRS416GAL	*pGAL1 CEN6/ARSH4 URA3* pBIISK ori amp^R	ATCC #87332
pRS423	*2μ HIS3* pBIISK ori amp^R	ATCC #77104
pRS425	*2μ LEU2* pBIISK ori amp^R	ATCC #77106
pRS426GAL	*pGAL1 2μ URA3* pBIISK ori amp^R	ATCC #87333
pUC18	pMB1 ori amp^R	
pVC2265	HO-polylinker-KanMX4-HO	D. Stillman/43
pLW2577	pJG4-5 with p*ADH-lacZ* inserted	This study
pLW2588	Integrating version of pLW2651	This study
pLW2590	Plasmid containing acceptor module integration fragment for Reiterative Recombination pRS416 with *HO(R)*-p*PYK*-GFP-*HIS3*-HO site-*HO(L)*	This study
pLW2592	Universal odd donor plasmid for Reiterative Recombination *pGAL1-HO-tCYC GFP-LEU2-SceI cleavage site CEN6/ARSH4 URA3* pBIISK ori amp^R	This study
pLW2593	Universal even donor plasmid for Reiterative Recombination *pGAL1-SceI-tCYC GFP-HIS3-HO cleavage site CEN6/ARSH4 URA3* pBIISK ori amp^R	This study
pLW2594	Round 1 donor plasmid for Reiterative Recombination *pGAL1-HO-tCYC GFP-LEU2-SceI cleavage site-KanMX-HO(L) CEN6/ARSH4 URA3* pBIISK ori amp^R	This study
pLW2595	*pGAL1-tCYC GFP-LEU2-SceI cleavage site CEN6/ARSH4 URA3* pBIISK ori amp^R	This study

pLW2596	*pGAL1-tCYC GFP-HIS3-HO cleavage site CEN6/ARSH4 URA3* pBIISK ori amp^R	This study
pLW2638	pRS426GAL- *H. insolens* Cel7B	This study
pLW2639	pRS416 with Cel7B stuffer between SfiI sites	This study
pLW2640	*pPYK-HIS3-HO site* in pLW2639	This study
pLW2641	*pPYK-LEU2-SceI site* in pLW2639	This study
pLW2642	*pPYK-HIS3*-HO site* (SfiI stuffer removed) in pLW2639	This study
pLW2643	*pPYK-HIS3-HO site* (SfiI stuffer removed) in pLW2639	This study
pLW2644	*GFP-HIS3*-HO site* in pLW2639	This study
pLW2645	*GFP-HIS3-HO site* in pLW2639	This study
pLW2646	*GFP-LEU2-SceI site* in pLW2639	This study
pLW2648	*pGAL1-SceI-tCYC GFP-HIS3*-HO site* in pLW2639	This study
pLW2649	pRS426GAL-HO	This study
pLW2650	*pPYK-GFP-HIS3-HO site* in pLW2639	This study
pLW2651	pRS426GAL-*MATa-inc*	This study
pLW2652	*pGAL1-SceI-tCYC GFP-HIS3-HO site* in pLW2639	This study
pLW2653	pRS416 with *HO(L)-KanMX-HO(R)*	This study
pLW2655	pBC398 with *pADH-LacZ* inserted	This study

Table 4-5. Oligonucleotides used in this study

Name	Sequence (5'-3')
LMW244	TGGTACACGCTGTGCGACCGCTACGGCCTGTATGTGGTGGATGAAGCCAATATTGA
LMW245	GGTCGACTAGTGGCCCCCGGGGCCCCTGTCCGAGATGCACC
LMW250	CATCACAATGTCTGCCCCTAAGA
LMW253	ATCATCACAATGACAGAGCAGAAAGCC
LMW255	TGGTAGCGGCCGCATTACCCTGTTATCCCTAAGATTGTACTGAGAGTGCA
LMW256	TGGTAGCGGCCGCTTTCAGCTTTCCGCAACAGTATAACTGTGCGGTATTTCACAC
LMW284	GCCGAACGACCGAGCGCAGCGAGTCAGTGATAGCCGCCATGACCCC
LMW285	AGGGCTTTCTGCTCTGTCATTGTGATGATGTTTTAT
LMW286	ATCTTCTTAGGGGCAGACATTGTGATGATGTTTTAT
LMW287	ACAAATAAAACATCATCACA
LMW288	AATCATGGTCATAGCTGTTTCCTGTGTGAAATTACCCTGTTATCCCTAAG
LMW289	ACAAATAAAACATCATCACAATGACAGAGCAG
LMW290	AATCATGGTCATAGCTGTTTCCTGTGTGAATTTCAGCTTTCCGCAACAG
LMW300	CGATTAAGTTGGGTAACGCCAGGGTTTTCCACGGATTAGAAGCCGCC
LMW301	CTGCAGGAATTCGATATCAAGCTTATCGATACCAAAGCCTTCGAGCGTCC
LMW302	TTTTAAAGAGGCCCTAGGAGCAGTGCGTGGAGTAAAAAGG
LMW303	CCTTTTTACTCCACGCACTGCTCCTAGGGCCTCTTTAAAA
LMW304	TTAGGGGCAGAACCAGAACCACCAGAACCTTTGTACAATTCATCCATACC
LMW305	CGTGGTCACCTGGCAAAACGACGATCTTCTTAGGGGCAGAACCAGAA
LMW306	TTCTGCTCTGTACCAGAACCACCAGAACCTTTGTACAATTCATCCATACC
LMW307	TGGTTTCATTTGTAATACGCTTTACTAGGGCTTTCTGCTCTGTACCAGAAC

LMW308	CAGCCGAACGACCGAGCGCAGCGAGTCAGTGATCTAGAATGTCTAAAGGTGAAGAATTAT
LMW309	GACAACACCAGTGAATAATTCTTCACCTTTAGACATTGTGATGATGTTTTA
LMW310	CTTTAACGTCAAGGAGAAAAAACCCCGGATATGCTTTCTGAAAACACGAC
LMW311	AATGTAAGCGTGACATAACTAATTACATGATTAGCAGATGCGCGCAC
LMW312	ATACCTCTATACTTTAACGTCAAGGAGAAAAAACCCAGAGGTCCGCTAATTCTG
LMW313	AATGTAAGCGTGACATAACTAATTACATGAGATTGTTTGCTTGAGTCTG
LMW326	TCGACGGATCCCCGGGTTAATTAAGGCGCGTTTCAGCTTTCCGCAACAG
LMW327	ACAAAACATTCTGTGAAGTTGTTCCCCCAGTAGCCGCCATGACCCC
LMW336	ACGCTCGAAGGCTTTGGTATCGATAAGCTTAATTATCCTGGGCACGAGT
LMW337	GGCCCCCGGGGCCCCTGTCCGAGATGCACCGTTTTCGACACTGGATGG
LMW353	ACGCTCGAAGGCTTTGGTATCGATAAGCTTTTGATCTTTACCGTTTAGTTC
LMW354	ACGCTCGAAGGCTTTGGTATCGATAAGCTTATTGTGCCTTTGGACTTAAAA
LMW355	ACGCTCGAAGGCTTTGGTATCGATAAGCTTACGCTGCAGGTCGACG
LMW360	TGAGAAGGTTTTGGGACGCTCGAAGGCTTTGCATGCCTGCAGGTCGA
LMW361	GCACAGTTATACTGTTGCGGAAAGCTGAAAGAATTCGAGCTCGGTACC
LMW367	TCAGTACAATCTTAGGGATAACAGGGTAAT
LMW374	TGAGAAGGTTTTGGGACGCTCGAAGGCTTT
LMW375	GCACAGTTATACTGTTGCGGAAAGCTGAAA
LMW419	AAAATTGTGCCTTTGGACTTAAAATGGCGTCAACTTCTTTTCTTTTTTTTTCT
LMW420	CTTAGGGATAACAGGGTAATAGCAGGGAGGCAAACAATGAAAGCTTTGGACTTCTTCGC
LMW421	GGACGCTCGAAGGCTTTGGCGAAGAAGTCCAAAGCTTTCATTGTTTGCCTCCCTGCTG
LMW422	ATACACACACTAAATTAATAATGTTACGTCCTGTAGAAAC
LMW423	GTTTCTACAGGACGTAACATTATTAATTTAGTGTGTGTATTTG
LMW424	CTGTTGCGGAAAGCTGAAAAAGAGGATGATGGAGGTTTCTTTGGAAAACCAAGAAATGAA
LMW425	GGACGCTCGAAGGCTTTTCATTTCTTGGTTTTCCAAAGAAACCTCCATCATCCTC
LMW426	ATCTAAGTTTTAATTACAAAATGCCATCTCATTTCGATAC
LMW427	GTATCGAAATGAGATGGCATTTTGTAATTAAAACTTAGATTAGA
LMW428	CTTAGGGATAACAGGGTAATATAGCTTCAAAATGTTTCTACT

4.7 References

1. Baker, D. et al. Engineering life: Building a fab for biology. *Sci. Am.* **294**, 44-51 (2006).

2. Baker, M. Synthetic genomes: The next step for the synthetic genome. *Nature* **473**, 403-408 (2011).

3. Purnick, P.E.M. & Weiss, R. The second wave of synthetic biology: From modules to systems. *Nat. Rev. Mol. Cell Biol.* **10**, 410-422 (2009).

4. Leonard, E. et al. Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 13654-13659 (2010).

5. Yokobayashi, Y., Weiss, R. & Arnold, F.H. Directed evolution of a genetic circuit. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16587-16591 (2002).

6. Haseltine, E.L. & Arnold, F.H. Synthetic gene circuits: Design with directed evolution. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 1-19 (2007).

7. Bieniossek, C. et al. Automated unrestricted multigene recombineering for multiprotein complex production. *Nat. Methods* **6**, 447-450 (2009).

8. Gibson, D.G., Smith, H.O., Hutchison, C.A., 3rd, Venter, J.C. & Merryman, C. Chemical synthesis of the mouse mitochondrial genome. *Nat. Methods* **7**, 901-903 (2010).

9. Gibson, D.G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343-345 (2009).

10. Li, M.Z. & Elledge, S.J. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods* **4**, 251-256 (2007).

11. Naesby, M. et al. Yeast artificial chromosomes employed for random assembly of biosynthetic pathways and production of diverse compounds in Saccharomyces cerevisiae. *Microb. Cell Fact.* **8**, 45 (2009).

12. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A modular cloning system for standardized assembly of multigene constructs. *PLoS One* **6**, e16765 (2011).

13. Kodumal, S.J. et al. Total synthesis of long DNA sequences: Synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 15573-15578 (2004).

14. Kodumal, S.J. & Santi, D.V. DNA ligation by selection. *Biotechniques* **37**, 34-40 (2004).

15. Menzella, H.G. et al. Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* **23**, 1171-1176 (2005).

16. Reisinger, S.J., Patel, K.G. & Santi, D.V. Total synthesis of multi-kilobase DNA sequences from oligonucleotides. *Nat. Protoc.* **1**, 2596-2603 (2006).

17. Wang, X., Sa, N., Tian, P.F. & Tan, T.W. Classifying DNA assembly protocols for devising cellular architectures. *Biotech. Adv.* **29**, 156-163 (2011).

18. Ellis, T., Adie, T. & Baldwin, G.S. DNA assembly for synthetic biology: From parts to pathways and beyond. *Integr. Biol.* **3**, 109-118 (2011).

19. Blake, W.J. et al. Pairwise selection assembly for sequence-independent construction of long-length DNA. *Nucleic Acids Res.* **38**, 2594-2602 (2010).

20. Ramon, A. & Smith, H.O. Single-step linker-based combinatorial assembly of promoter and gene cassettes for pathway engineering. *Biotechnol. Lett.* **33**, 549-555 (2011).

21. Gibson, D.G. et al. Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science* **319**, 1215-1220 (2008).

22. Gibson, D.G. et al. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20404-20409 (2008).

23. Shao, Z.Y., Zhao, H. & Zhao, H.M. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.* **37**, e16 (2009).

24. Shao, Z., Luo, Y. & Zhao, H. Rapid characterization and engineering of natural product biosynthetic pathways via DNA assembler. *Mol. Biosyst.* **7**, 1056-1059 (2011).

25. Itaya, M., Tsuge, K., Koizumi, M. & Fujita, K. Combining two genomes in one cell: Stable cloning of the Synechocystis PCC6803 genome in the Bacillus subtilis 168 genome. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15971-15976 (2005).

26. Itaya, M., Fujita, K., Kuroki, A. & Tsuge, K. Bottom-up genome assembly using the Bacillus subtilis genome vector. *Nat. Methods* **5**, 41-43 (2008).

27. Storici, F., Durham, C.L., Gordenin, D.A. & Resnick, M.A. Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 14994-14999 (2003).

28. Lartigue, C. et al. Creating bacterial strains from genomes that have been cloned and engineered in yeast. *Science* **325**, 1693-1696 (2009).

29. Paques, F., Leung, W.Y. & Haber, J.E. Expansions and contractions in a tandem repeat induced by double-strand break repair. *Mol. Cell. Biol.* **18**, 2045-2054 (1998).

30. Inbar, O. & Kupiec, M. Homology search and choice of homologous partner during mitotic recombination. *Mol. Cell. Biol.* **19**, 4134-4142 (1999).

31. Ahn, B.Y., Dornfeld, K.J., Fagrelius, T.J. & Livingston, D.M. Effect of limited homology on gene conversion in a Saccharomyces cerevisiae plasmid recombination system. *Mol. Cell. Biol.* **8**, 2442-2448 (1988).

32. Hua, S.B., Qiu, M., Chan, E., Zhu, L. & Luo, Y. Minimum length of sequence homology required for in vivo cloning by homologous recombination in yeast. *Plasmid* **38**, 91-96 (1997).

33. Ira, G. & Haber, J.E. Characterization of RAD51-independent break-induced replication that acts preferentially with short homologous sequences. *Mol. Cell. Biol.* **22**, 6384-6392 (2002).

34. Sugawara, N., Ira, G. & Haber, J.E. DNA length dependence of the single-strand annealing pathway and the role of Saccharomyces cerevisiae RAD59 in double-strand break repair. *Mol. Cell. Biol.* **20**, 5300-5309 (2000).

35. Strathern, J.N. et al. Homothallic switching of yeast mating type cassettes is initiated by a double-stranded cut in the MAT locus. *Cell* **31**, 183-192 (1982).

36. Haber, J.E. Mating-type gene switching in Saccharomyces cerevisiae. *Annu. Rev. Genet.* **32**, 561-599 (1998).

37. Colleaux, L., Dauriol, L., Galibert, F. & Dujon, B. Recognition and cleavage site of the intron-encoded omega transposase. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 6022-6026 (1988).

38. Plessis, A., Perrin, A., Haber, J.E. & Dujon, B. Site-specific recombination determined by I-SceI, a mitochondrial group-I intron-encoded endonuclease expressed in the yeast nucleus. *Genetics* **130**, 451-460 (1992).

39. Boeke, J.D., LaCroute, F. & Fink, G.R. A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol. Gen. Genet.* **197**, 345-346 (1984).

40. Brachmann, C.B. et al. Designer deletion strains derived from Saccharomyces cerevisiae S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14**, 115-132 (1998).

41. Scherer, S. & Davis, R.W. Replacement of chromosome segments with altered DNA sequences constructed in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 4951-4955 (1979).

42. Weiffenbach, B. et al. Deletions and single base pair changes in the yeast mating type locus that prevent homothallic mating type conversions. *Proc. Natl. Acad. Sci. U. S. A.* **80**, 3401-3405 (1983).

43. Voth, W.P., Richards, J.D., Shaw, J.M. & Stillman, D.J. Yeast vectors for integration at the HO locus. *Nucleic Acids Res.* **29**, e59 (2001).

44. Ma, H., Kunes, S., Schatz, P.J. & Botstein, D. Plasmid construction by homologous recombination in yeast. *Gene* **58**, 201-216 (1987).

45. Sweetser, D.B., Hough, H., Whelden, J.F., Arbuckle, M. & Nickoloff, J.A. Fine-resolution mapping of spontaneous and double-strand break-induced gene conversion tracts in Saccharomyces cerevisiae reveals reversible mitotic conversion polarity. *Mol. Cell. Biol.* **14**, 3863-3875 (1994).

46. Nasmyth, K. The determination of mother cell-specific mating type switching in yeast by a specific regulator of HO transcription. *EMBO J.* **6**, 243-248 (1987).

47. Lin, H., Tao, H. & Cornish, V.W. Directed evolution of a glycosynthase via chemical complementation. *J. Am. Chem. Soc.* **126**, 15051-15059 (2004).

48. Clark, D.D. & Peterson, B.R. Fluorescence-based cloning of a protein tyrosine kinase with a yeast tribrid system. *ChemBioChem* **6**, 1442-1448 (2005).

Chapter 5

Reiterative Recombination for Metabolic Engineering

---

*The content of this chapter will be partially published in

L.M. Wingler, V.W. Cornish. “Reiterative Recombination for the *In Vivo* Assembly of Libraries of Multi-Gene Pathways,” *submitted.*

5.0 Chapter outlook

Metabolic engineering holds the promise of ultimately offering a general and elegant solution for the production of complex natural product therapeutics and their analogs. However, simply transplanting the genes required to biosynthesize a given natural product into a heterologous organism "as is" is rarely sufficient. Obtaining natural product production at all, much less in competitively high yields, typically requires multiple changes to the biosynthetic pathway and the host strain's genetic background. Modifying and rebuilding biosynthetic pathways to optimize compound production has become a rate-limiting step in the metabolic engineering field. Here we show that our Reiterative Recombination system for *in vivo* DNA assembly can address several technical challenges associated with customizing biosynthetic pathways. First, Reiterative Recombination provides a robust and straightforward method for building individual pathways. We explicitly demonstrate the construction first of a functional minimal pathway for biosynthesis of the terpenoid pigment lycopene in yeast and then of an extended pathway over 20 kilobases in length that was designed to optimize lycopene yields. Second, Reiterative Recombination's high efficiency makes it uniquely suitable for generating large collections of pathway variants inside of the cell for combinatorial optimization. We show we can construct libraries of $\geq 10^4$ different pathways, using a mock screen of active to inactive lycopene pathways that explicitly tests library size. Third, pathways built in yeast via Reiterative Recombination can be shuttled to other organisms if a different host is preferred, as we show by moving an operon directing carotenoid biosynthesis into *E. coli*. The development of next-generation versions of Reiterative Recombination should further expand its utility for metabolic engineering.

5.1 Introduction

5.1.1 Multi-gene pathways for metabolic engineering

Metabolic engineering has the potential to transform the production of the natural products and derivatives that dominate the pharmaceutical market[1,2]. Though expert organic chemists have made enormous strides in synthesizing these structurally complex compounds[3,4], each total synthesis is truly its own *tour de force*, limiting the number and quantity that can be chemically synthesized. An attractive alternative is to use cells' natural biosynthetic machinery to produce natural products or advanced intermediates. While most natural products cannot be obtained in substantial amounts from their native hosts[5], a handful of recent, high-profile breakthroughs in metabolic engineering[6-9] have raised the tantalizing prospect of, ultimately, routinely biosynthesizing any desired natural product or analog in tractable heterologous microorganisms[10-12].

Our ability to engineer cells to produce natural products, however, currently is obstructed by the technical difficulty of constructing optimized, multi-gene biosynthetic pathways in the host cell. Simply transplanting a given set of biosynthetic genes into a new organism "as is" is rarely productive. For optimal performance, the original DNA sequence typically must be modified by, for example, replacing promoters, optimizing genes' codon usage, and adding or overexpressing genes for precursor production (**Fig. 5-1A**)[12,13]. In addition, researchers have endeavored to produce novel analogs by deleting, mutating, exchanging, or adding pathway genes (**Fig. 5-1B**)[14-19]. Making these multiple, defined changes within a pathway or rebuilding pathways from smaller fragments increases in difficulty with pathway size[5,18]. As sequences become longer, fewer unique restriction sites exist, and simply handling the DNA *in vitro* becomes cumbersome. The crux of the problem is that standard molecular biology tools, intended

for the manipulation of single genes, are simply not intended to address these larger-scale applications.

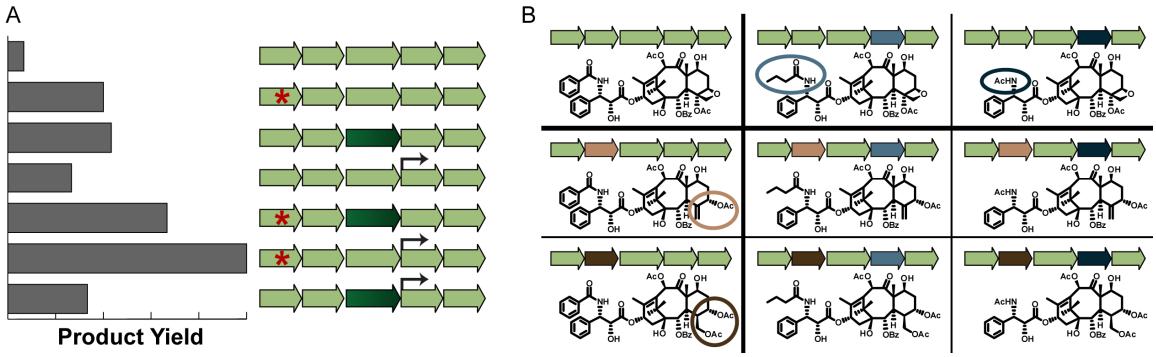


Figure 5-1. Modifying biosynthetic pathways for metabolic engineering. **(A)** Metabolite yields can be improved more rapidly by testing multiple variables (e.g. mutating pathway enzymes, testing isozymes, or adjusting expression levels) in parallel. **(B)** Generating libraries of modified pathways can enable the biosynthesis of natural product analogs. Both applications can benefit from combinatorial approaches.

Many of the multi-gene DNA assembly techniques reported in recent years have been developed with an eye towards introducing defined modifications within large biosynthetic pathways. Researchers at Kosan Biosciences, focusing on polyketide gene clusters, solve this problem by constructing entirely synthetic DNA sequences *in vitro*[20]. They begin with chemically synthesized, 40-bp oligonucleotides, allowing them to fully specify the desired sequence of the final pathway. These oligonucleotides are hierarchically assembled into pathways tens of kilobases in length using a series of PCR-based assembly methods and "ligation by selection," a cyclical, stringent subcloning method (see **Section 4.1.1**). However, the expense of synthesizing such a large number of oligonucleotides and the time-consuming nature of the assembly process make this *in vitro* construction process impractical for most researchers.

Shao *et al.* have used DNA assembler to assemble biosynthetic pathway variants *in vivo*[21]. In this method, multiple overlapping DNA fragments are co-transformed into yeast to assemble pathways by homologous recombination[22] (see **Section 4.1.2**). Point mutations can be incorporated into the pathway through the appropriate design of PCR primers used to amplify the overlapping fragments, and the authors demonstrated this by mutating a conserved motif in a gene in the aureothin biosynthetic pathway, thereby inactivating the enzyme and generating a different polyketide product[21]. They also used DNA assembler to create a hybrid pathway derived from the closely related aureothin and spectinabilin gene clusters[21]. However, the authors report needing to use long (400-bp) regions of overlap between fragments whenever possible to promote assembly. This constraint will limit the ability of this method to piece together fragments arising from different sources; overlap regions would need to be added on using *in vitro* methods (e.g., overlap extension PCR).

5.1.2 Multi-gene libraries for metabolic engineering

Despite enormous advances in our understanding of systems-level biology in the past decade, our ability to rationally predict the effects of perturbations to the cell's metabolism remains limited[23,24]. Thus, in addition to the technical difficulty of modifying biosynthetic pathways, there is an intellectual bottleneck hindering the optimization of heterologous natural product biosynthesis. Library approaches can bypass these gaps in our knowledge, and metabolic engineers have repeatedly improved natural product yields and synthesized analogs by searching libraries of isozymes[18,25], mutant biosynthetic enzymes[9,26], or promoters and regulatory regions that modulate the expression levels of genes that alter pathway flux[27,28]. However, the scope of these

experiments has been very limited due to the technical difficulty of building libraries of pathways *in vivo*. Typically only one or two components of a pathway can be manipulated in a single experiment, potentially missing synergistic effects, and only very small libraries can be generated, often using very labor-intensive routes.

In spite of the proliferation of DNA assembly techniques in recent years, and in spite the often-professed desire of researchers in fields such as metabolic engineering to generate libraries of pathways, there are very few reports of using these methods to generate libraries of optimized or modified biosynthetic pathways. For example, Kosan Biosciences, the creators of "ligation by selection," surprisingly did not employ this method when they later constructed a library of pathways to make polyketide natural products[18]. Rather, using a very labor-intensive strategy, the researchers replaced two modules in a polyketide synthase pathway with 11 and 14 variants in separate plasmids *in vitro* and individually co-transformed each combination of plasmids to generate 154 different triketide synthase pathways.

The Venter Institute has only published a single example of constructing a library using their "*in vitro* isothermal assembly" method[29] (see **Section 4.1.1**), even though they report that it proceeds with very high efficiency[30,31]. The authors made a small, 144-member library of two-gene pathways designed to enable mutant *E. coli* strains to utilize acetate as their sole carbon source[29]. Though the researchers argued that their transformation efficiency (10,000 clones) is more than sufficient to cover their library, they did not explicitly demonstrate that all possible library constructs were made. After selection for acetate utilization, which would eliminate any non-functional or weakly functional pathways, they characterized just 37 clones and showed that the 30 out of 37

clones that actually had the correctly assembled pathway represented ~10% of the theoretical library.

There are no reports of making libraries of pathways using *in vivo* assembly methods. It is important to note that these techniques yield only tens to hundreds of recombinants at a time[22,32-35], and these numbers represent the maximum library complexity that these methods can achieve, even if the method allows iterative pathway elongation[34,35]. Given that library sizes rapidly swell when exploring multiple variables in the context of pathway optimization (e.g., testing 100 mutants of enzyme 1 against 100 mutants of enzyme 2 is already 10^4 combinations), much higher assembly efficiencies are needed to even begin sampling the potential diversity available.

Having established that Reiterative Recombination provides a facile and efficient method for the assembly of multi-gene constructs (**Chapter 4**), here we show that Reiterative Recombination can begin to address several challenges faced by metabolic engineers. First, we demonstrate that Reiterative Recombination can build functional biosynthetic pathways, including constructs that are over 20 kilobases in length. Compared to conventional strain construction techniques and alternative DNA assembly methods, this process is robust, modular, and user-friendly. Next, we explicitly show that Reiterative Recombination can generate large libraries of pathways, containing at least 10^4 variants, inside of the cell using a mock screen. Finally, we establish that biosynthetic pathways built by Reiterative Recombination in the yeast chromosome can be recovered into a plasmid, shuttled to a different host organism, and functionally expressed.

5.2 Results

5.2.1 Application of Reiterative Recombination to the construction of the lycopene biosynthetic pathway

Our first objective was to demonstrate that Reiterative Recombination could be used to construct functional biosynthetic pathways *in vivo*, thereby establishing its generality and its utility for metabolic engineering. We strategically selected the terpenoid pigment lycopene for this purpose. Most valuable terpenoids (e.g., taxol, artemisinin) lack high-throughput assays for measuring production, and lycopene is widely used as a convenient colorimetric screen to optimize yields of common terpenoid precursors (**Fig. 5-2A**; see **Section 5.2.2**). Coupling lycopene biosynthesis to primary yeast metabolism requires the addition of three exogenous genes[36]: *crtE* (geranylgeranyl diphosphate synthase), *crtB* (phytoene synthase), and *crtI* (phytoene desaturase).

Using Reiterative Recombination, we integrated codon-optimized versions of *Erwinia herbicola crtE* (round 1), *crtB* (round 2), and *crtI* (round 3) to generate a yeast strain capable of producing lycopene (**Fig. 5-2B**). We also integrated the selectable marker *TRP1* during round 3 to provide further verification of correct pathway assembly. After the third round of assembly, 99% of the resulting recombinants exhibited the expected orange phenotype that is indicative of lycopene production (**Fig. 5-2F**). In parallel, as negative controls, we built pathways containing nonsense mutations in *crtB* and/or *crtI*, and the resulting strains did not produce lycopene (**Fig. 5-2C,D,E**).

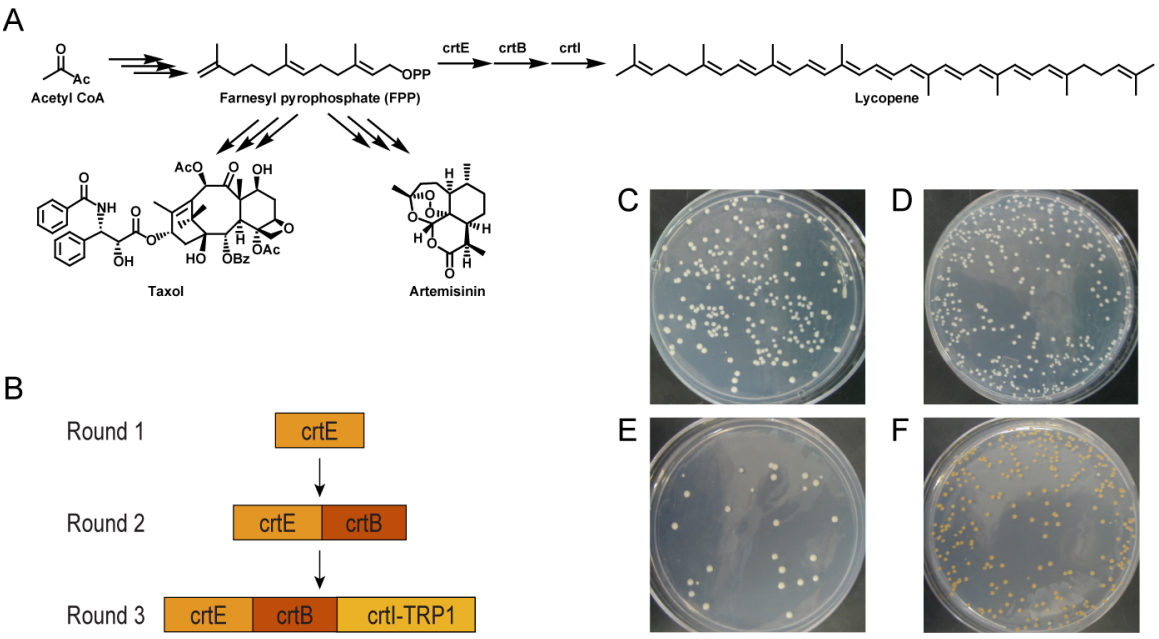


Figure 5-2. Assembly of the lycopene biosynthetic pathway using Reiterative Recombination. **(A)** Terpenoids such as taxol, artemisinin, and lycopene share the same common precursor FPP. Therefore, lycopene production can be used as a colorimetric screen to optimize flux towards the terpenoid pathway. **(B)** Order of *crt* gene insertion. Phenotypes of cured round 3 colonies containing wild-type *crtE* and **(C)** *crtB*-stop + *crtI*-stop, **(D)** *crtB*-stop + *crtI*-silent, **(E)** *crtB*-silent + *crtI*-stop, and **(F)** *crtB*-silent + *crtI*-silent. For **(F)**, 315 out of 317 colonies had an orange phenotype; none of the other plates contained any orange colonies.

5.2.2 Construction of an optimized biosynthetic pathway via Reiterative Recombination

Our next objective was to challenge our technology, not just to build pathways, but 1) to build significantly longer pathways (>10 kb in length) requiring numerous rounds of elongation and 2) to build defined pathways designed to optimize yields of natural products, specifically terpenoids. Terpenoids are a particularly prominent class of metabolic engineering targets. Though they are all derived from the universal precursor isopentenyl pyrophosphate (IPP), the >55,000 terpenoids isolated to date exhibit vast

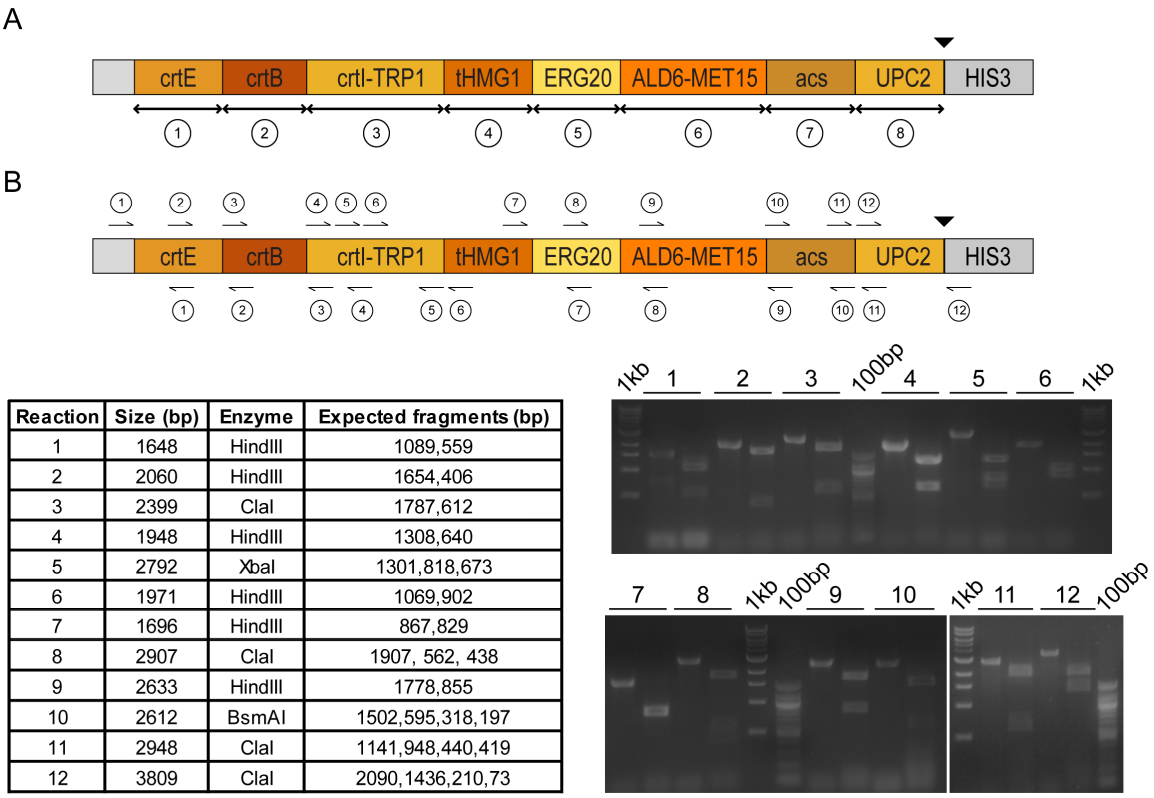
structural and functional diversity[37]. Many terpenoids, such as the chemotherapeutics taxol (paclitaxel) and vinblastine and the antimalarial artemisinin, are frontline therapeutics but cannot be sustainably supplied to the commercial market from their natural sources[38,39]. Routine expression of these structurally complex small molecules in robust, heterologous hosts such as yeast and bacteria would be (and in the case of artemisinin, is[6,40,41]) a breakthrough for the production of known pharmaceuticals and the discovery of potent new terpenoids.

Yeast metabolism, however, is not inherently geared toward high-titer terpenoid production, and terpenoid precursors (e.g., IPP and farnesyl pyrophosphate (FPP)) are largely channeled into sterol production[39]. Metabolic engineers have repeatedly improved terpenoid yields in yeast by overexpressing endogenous, heterologous, or mutant genes encoding regulatory proteins or biosynthetic enzymes that divert metabolic flux towards FPP (*ALD6*[42], *S. cerevisiae* or *Salmonella enterica ACS1*[42], truncated *HMG1*[6,43-46], *ERG20*[6], *BTS1*[44], *upc2-1*[6,46,47]). However, there are few examples of searching for additive effects by overexpressing multiple genes[6,23,42,48], undoubtedly because of the difficulty of constructing such strains.

Our collaborators in the Stephanopoulos group at the Massachusetts Institute of Technology previously made a yeast strain containing the three *E. herbicola crt* genes required for lycopene production along with five additional genes known to improve terpenoid yields when overexpressed individually (P. Ajikumar and G. Stephanopoulos, personal communication). Introducing this unusually large number of overexpressed genes into a single strain took two students almost a year using conventional genetic

techniques. The five "extra" genes indeed increased lycopene production 4-fold (to 14 mg/L).

Clearly, faster strain construction routes are needed to achieve truly high-titer terpenoid production in yeast. Thus, we explicitly demonstrated the utility of Reiterative Recombination for such tasks by constructing a strain containing the same eight overexpressed genes. Beginning with the minimal lycopene-producing strain constructed in **Figure 5-2**, we continued the Reiterative Recombination process for an additional five rounds, giving a final construct 21 kb in length resulting from a total of eight rounds of elongation (**Fig. 5-3A**). Significantly, reconstructing the entire pathway took less than two months using a protocol that required minimal effort in its design and execution. We verified the pathway's integrity by PCR, restriction analysis, and sequencing of strains' genomic DNA (**Fig. 5-3B**). To recreate the Stephanopoulos laboratory's strain as faithfully as possible, we used the same three strong, constitutive promoters to drive the expression of all eight genes. We eliminated the possibility of having these repeated elements "loop out" by homologous recombination and delete portions of the pathway by separating repeats with the selectable markers *TRP1* and *MET15*. Since metabolic engineers frequently reuse a small number of preferred regulatory elements, this now-validated design strategy should expand the utility of Reiterative Recombination.



Reaction	Size (bp)	Enzyme	Expected fragments (bp)
1	1648	HindIII	1089,559
2	2060	HindIII	1654,406
3	2399	ClaI	1787,612
4	1948	HindIII	1308,640
5	2792	XbaI	1301,818,673
6	1971	HindIII	1069,902
7	1696	HindIII	867,829
8	2907	ClaI	1907, 562, 438
9	2633	HindIII	1778,855
10	2612	BsmAI	1502,595,318,197
11	2948	ClaI	1141,948,440,419
12	3809	ClaI	2090,1436,210,73

Figure 5-3. Assembly of an optimized lycopene biosynthetic pathway using Reiterative Recombination. **(A)** Order of gene insertion using eight rounds of Reiterative Recombination. **(B)** The purified genomic DNA of a colony from the eighth round of Reiterative Recombination was analyzed by PCR and restriction mapping. The subfragments lifted by each PCR reaction and the expected sizes of the PCR products and restriction fragments are shown in the table. For each pair of lanes in the gel, the left lane is the undigested PCR reaction; the right lane is the digested PCR reaction. (100bp=100 bp DNA ladder, 1kb= 1 kb DNA ladder)

5.2.3 Construction of libraries of biosynthetic pathways via Reiterative Recombination

A distinctive advantage of Reiterative Recombination is that every step of the cyclical elongation process proceeds with very high efficiency. Using a basic yeast electroporation protocol, we can readily obtain as many as 10^6-10^8 transformants per transformation[49]; the induction, which is readily scalable, typically gives $\gg 10^4$

recombinants per milliliter of culture. We therefore expected that we could generate larger libraries of pathways than attainable with other *in vivo* DNA assembly techniques, which generate only tens to hundreds of constructs at a time. To test this hypothesis, we used the lycopene biosynthesis pathway to explicitly challenge Reiterative Recombination's ability to construct large libraries.

We repeated rounds 2 and 3 of the minimal lycopene pathway assembly, this time transforming various ratios of *crtB* and *crtI* alleles that contained either nonsense or silent mutations with diagnostic restriction sites (**Fig. 5-4**). Initially, we did not recover lycopene-producing colonies from our libraries at the expected frequencies. Further analysis of the pool of cured recombinants obtained from various Reiterative Recombination rounds, both from the reporter proof-of-principle system (**Section 4.3.3**) and the lycopene pathway assembly, revealed that a small percentage of cured recombinants (≤0.2%; **Table 5-1**) acquired both the *HIS3* and *LEU2* markers. This subpopulation of cells was sufficient to skew the observed ratios of orange colonies after carrying the library forward for multiple rounds.

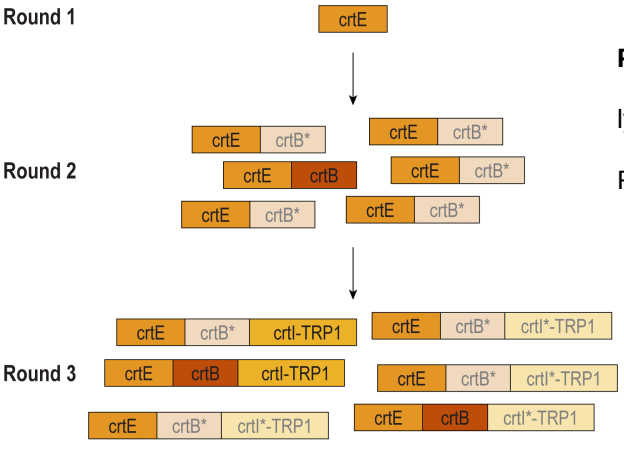


Figure 5-4. Construction of mock libraries of lycopene biosynthetic pathways via Reiterative Recombination.

Table 5-1. Percentage of cells with the *HIS LEU* phenotype in cured recombinant pools from various rounds of Reiterative Recombination.

Round	Percentage of recombinants with *HIS LEU* phenotype
Reporter proof-of-principle	
Round 1	0.2%
Round 2	0.00007%
Lycopene library round 2	
10:1 *crtB* stop:silent	0.006%
100:1 *crtB* stop:silent	0.01%

While we are developing a next-generation Reiterative Recombination system that eliminates this problem entirely (**Section 5.2.5**), we were immediately able to construct large libraries in this first-generation system by simply selecting for the *TRP1* marker at the end of the pathway (**Fig. 5-2A**) after the last round of assembly. This additional selection served as a stringent final purification step for our libraries and, importantly, is a general solution that could be used for any desired library application. As shown in **Table 5-2**, we were readily able to recover lycopene-producing colonies at the expected frequencies from mock libraries of up to 10^4. These colonies contained the expected silent mutations in *crtB* and *crtI*, demonstrating that they arose from the silent alleles rather than from mutation of the genes with nonsense mutations (**Fig. 5-5**).

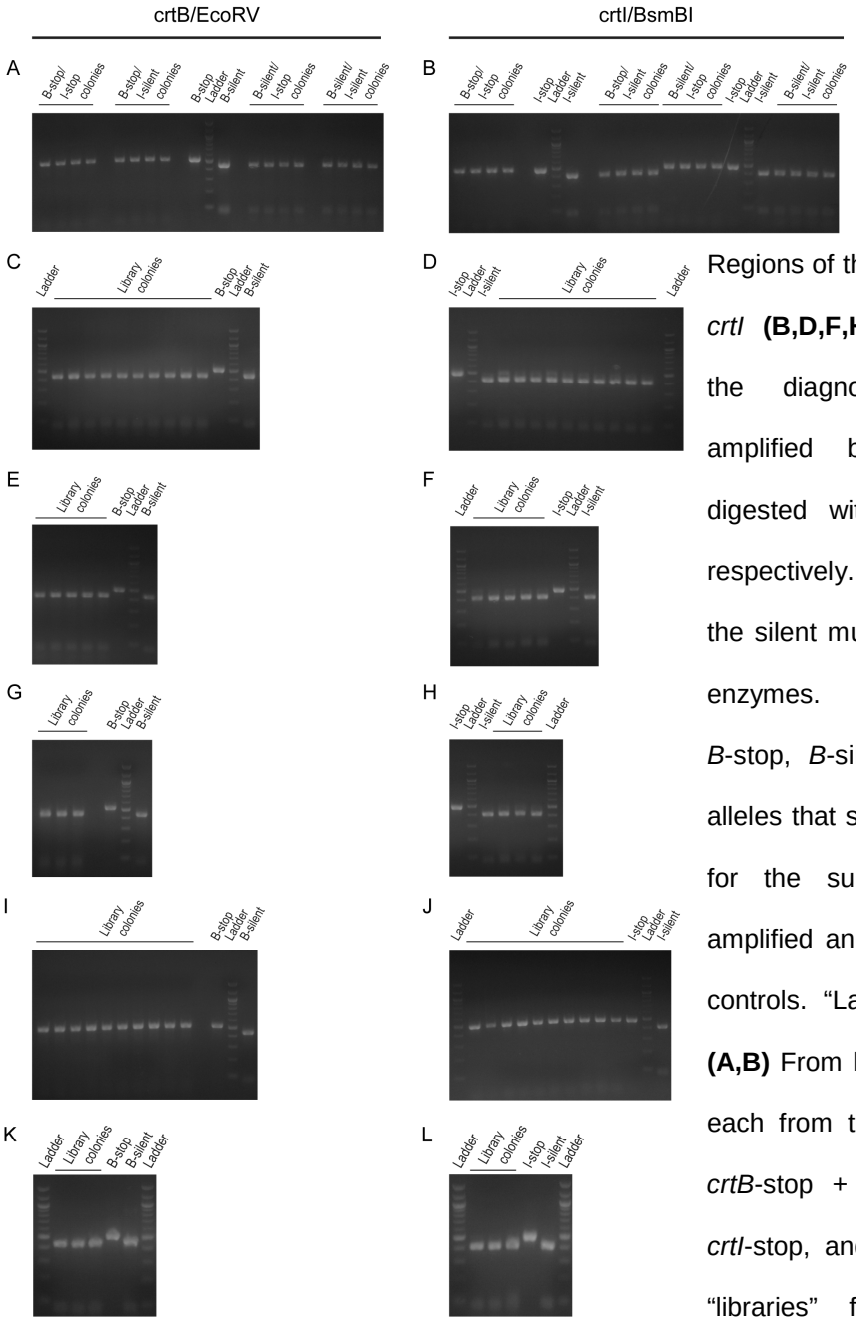


Figure 5-5. Restriction analysis of cured recombinants from the lycopene library screen. Regions of the *crtB* **(A,C,E,G,I,K)** and *crtI* **(B,D,F,H,J,L)** genes containing the diagnostic mutations were amplified by colony PCR and digested with EcoRV and BsmBI, respectively. Only alleles containing the silent mutations are cut by these enzymes. The plasmids with the *B*-stop, *B*-silent, *I*-stop, and *I*-silent alleles that served as PCR templates for the subfragments were PCR amplified and digested in parallel as controls. "Ladder" is a 100 bp ladder. **(A,B)** From left to right, four colonies each from the *crtB*-stop + *crtI*-stop, *crtB*-stop + *crtI*-silent, *crtB*-silent + *crtI*-stop, and *crtB*-silent + *crtI*-silent "libraries" from **Figure 5-2(C-F)**. **(C,D)** Ten orange colonies from the 10:1 *crtB* stop:silent + 0:1 *crtI* stop:silent library. **(E,F)** The five orange colonies from the 100:1 *crtB* stop:silent + 0:1 *crtI* stop:silent library. **(G,H)** The three orange colonies from the 100:1 *crtB* stop:silent + 10:1 *crtI* stop:silent library. **(I,J)** Ten white colonies from the 100:1 *crtB* stop:silent + 100:1 *crtI* stop:silent library. **(K,L)** The three orange colonies from the 100:1 *crtB* stop:silent + 100:1 *crtI* stop:silent library.

Table 5-2. Mock screen for lycopene-producing strains via Reiterative Recombination.

Transformed DNA ratios		Library complexity	Colonies assayed	Orange colonies	Observed percentage of orange colonies	P[a]
crtB stop:silent	*crtI* stop:silent					
10:1	0:1	10^1	2360	225	10%	0.5
100:1	0:1	10^2	587	5	0.9%	0.7
100:1	10:1	10^3	2079	3	0.1%	0.4
100:1	100:1	10^4	18450	3	0.02%	0.4

[a]Since the plated cells represented a randomly selected aliquot (<0.1%) of the population, a 1-proportion z-test was used to test if the observed percentages of orange colonies were significantly different than the expected percentages. All P-values were greater than α=0.1, indicating that none were significantly different.

5.2.4 Transfer of a Reiterative Recombination pathway to a heterologous organism

While *S. cerevisiae* is gaining popularity as a host organism for metabolic engineering applications, other organisms such as *E. coli*, *Streptomyces coelicolor*, and *Streptomyces lividans* are the preferred heterologous hosts for many natural products[10,11]. Only very simple polyketides[50,51] and non-ribosomal peptides[52], for example, have been successfully produced in yeast. To demonstrate that Reiterative Recombination is not limited to the assembly of DNA for yeast, we reconstructed a previously described, three-gene pathway for tetradehydrolycopene synthesis in *E. coli* (**Fig. 5-6A**)[53]. Since our first Reiterative Recombination system can only be used to assemble genes in the yeast chromosome, we attempted to move the construct onto an *E. coli* shuttle vector using plasmid gap repair from the chromosome[54]. Though gap repair is an established technique for applications such as the retrieval of mutant alleles[55], recovery of the much larger carotenoid pathway was inefficient, so we used a fourth round of elongation to add

a gene for kanamycin resistance, giving a 5-kb construct, and created a self-cleaving recovery vector to stimulate homologous recombination (**Fig. 5-6B**). *E. coli* retransformed with the recovery vector were selected for kanamycin resistance, leading to identification of a plasmid with the intact construct. Colonies with this plasmid had the same colorimetric phenotype due to tetradehydrolycopene production as those with the previously reported plasmid (**Fig. 5-6C**)[53]. The shuttle vector containing the recovered plasmid was 15 kb, near the size limit for pMB origins of replication, and poor stability in *E. coli* likely contributed to our difficulty moving the construct.

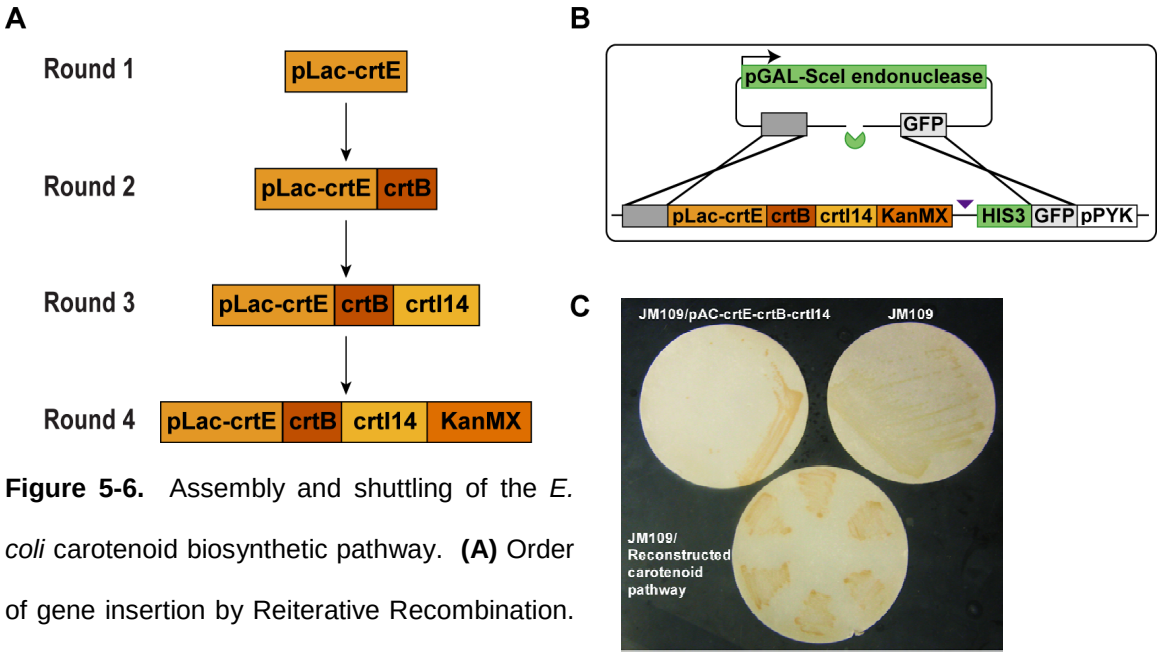


Figure 5-6. Assembly and shuttling of the *E. coli* carotenoid biosynthetic pathway. **(A)** Order of gene insertion by Reiterative Recombination. **(B)** The assembled pathway was recovered from the yeast chromosome into an *E. coli* shuttle vector using a self-cleaving plasmid as shown. **(C)** Following retransformation of the plasmid into *E. coli*, colonies that exhibited the expected orange phenotype indicative of tetradehydrolycopene production were identified. Colonies were transferred to filter paper to show the colorimetric phenotype more clearly. Positive and negative control strains are shown in the top left and right corners, respectively. The bottom filter contains strains with the pathway built by Reiterative Recombination.

5.2.5 Design of next-generation Reiterative Recombination systems

To expand the scope and utility of Reiterative Recombination, we have designed and are constructing several next-generation systems. First, we are developing methodology to assemble pathways on plasmids rather than on the yeast chromosome to facilitate their transfer to other organisms. Our preliminary results showed that recovery and transfer of even a short, 5-kb pathway from the yeast chromosome was difficult and inefficient (**Section 5.2.4**). We anticipate that the most effective strategy for DNA transfer will be to build constructs directly in shuttle vectors. The ideal acceptor plasmid for transfer to heterologous hosts will 1) stably maintain large inserts in both yeast and the heterologous organism and 2) lack homology to the yeast chromosome and to donor plasmids that could lead to unwanted recombination. As an example, to transfer Reiterative Recombination constructs to *E. coli*, we plan to construct an acceptor plasmid with the features shown in **Figure 5-7A**.

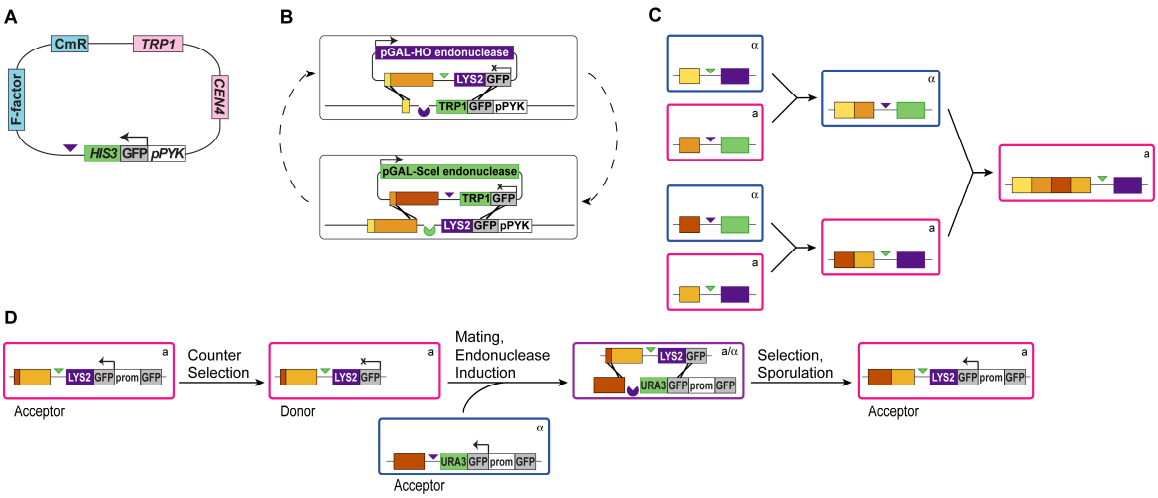


Figure 5-7. Next-generation Reiterative Recombination systems. **(A)** The *S. cerevisiae/E. coli* shuttle vector will contain 1) the Reiterative Recombination acceptor module, 2) sequences to propagate the plasmid as a bacterial artificial chromosome (F-factor) and centromeric vector (CEN4) in *E. coli* and yeast, respectively, 3) and selectable markers for *E. coli* (Cm^R) and yeast (*TRP1*). **(B)** Reiterative Recombination using positive- and negative-selectable alternating markers. **(C)** Conversion of acceptor modules into donor modules enables convergent DNA assembly, allowing combination of existing pathways and decreasing the number of cycles needed to build long clusters. **(D)** Conversion of acceptor modules into donor modules is accomplished by adding a second copy of GFP upstream of the acceptor module's promoter, creating a direct repeat. We will use *URA3* and *LYS2*, which have both positive and negative selections, as the GFP-marker fusions. Counter selection against *URA3* and *LYS2* can be used to identify cells in which recombination between the GFP repeats has led to deletion of the promoter (frequency ~10^{-4}, [56]), effectively converting the acceptor module into a chromosomal donor module. At all other times, selection for expression of the GFP-marker will eliminate cells that excise the promoter. Haploid cells of opposite mating type (a and α) will be mated to generate diploids with both acceptor and donor modules.

Second, we plan to replace the alternating selectable *HIS3* and *LEU2* selectable markers in Reiterative Recombination with markers that have both positive and negative selections, namely *TRP1*[57] and *LYS2*[58] (**Fig. 5-7B**). As shown in **Table 5-1**, a small

minority of colonies acquire both alternating selectable markers during each round of Reiterative Recombination. This phenomenon does not interfere with the construction of individual pathways, where single colonies can be carried forward, but it does become problematic when libraries are carried forward through multiple rounds. Having counter selectable alternating markers will allow libraries to be purified of cells that did not undergo correct marker conversion at every round. Even though the *TRP1* and *LYS2* counter selections are less robust than the favored *URA3* counter selection, which is already being employed for donor plasmid curing, they should be adequate to eliminate the already low fraction of cells with both selectable markers.

Finally, an intriguing strategy that we ultimately want to incorporate into strain optimization is combinatorially combining pathways. Towards this end, we plan to exploit sexual reproduction for convergent DNA assembly (**Fig. 5-7C**). The key features of this variation on Reiterative Recombination are 1) mating and sporulation cycles and 2) conversion of acceptor modules into donor modules (**Fig. 5-7D**). Conversion is achieved by excising acceptor modules' promoters with direct repeats and by employing counter selectable markers. We will use mating and sporulation, rather than transformation, to bring donor and acceptor modules together. Other elements of Reiterative Recombination will be the same as in the original system.

Convergent assembly will add exciting new dimensions to Reiterative Recombination. First, we envision both rationally and combinatorially building "super strains" combining the best pathways identified in separate experiments. Second, convergent pathway assembly will dramatically decrease the time and effort required to build long pathways (# linear assembly rounds = $2^{(\#\text{ convergent assembly rounds} - 1)}$). Concretely,

the time to construct even the relatively modestly-sized 21-kb optimized lycopene pathway (**Fig. 5-3A**) would be cut in half (one versus two months). In the longer term, this advance will be a *sine qua non* as we anticipate assembling pathways such as polyketide synthase gene clusters that can be >100 kb (e.g., assembling a 100-kb pathway in 5-kb increments would take 4-5 months with Reiterative Recombination via linear assembly, but only one month with convergent assembly).

5.3 Discussion

We have demonstrated that Reiterative Recombination addresses multiple technical bottlenecks being encountered by the metabolic engineering community, namely, the construction of individual defined biosynthetic pathways, the generation of large libraries of pathway variants, and the transfer of constructed pathways into preferred host organisms. The construction of a functional biosynthetic pathway in yeast, specifically the lycopene biosynthesis pathway, establishes the generality of Reiterative Recombination for DNA assembly and its application to metabolic engineering. We also demonstrated that biosynthetic pathways built in yeast via Reiterative Recombination can be transferred to and functionally expressed in other organisms, though alternative strategies will need to be developed to make the transfer step reliably efficient.

The use of recyclable markers and endonucleases in Reiterative Recombination renders it useful for the assembly and integration of very large DNA constructs. We have demonstrated that elongation can be continued indefinitely (at least eight rounds) to build pathways tens of kilobases in length. Recycling markers in Reiterative Recombination eliminates the perennial problem of running out of selectable markers during complex

strain constructions, and the modular protocol minimizes the effort needed to design each integration step.

Reiterative Recombination's robustness makes it capable of building sizable libraries of pathways ($\geq 10^4$) containing diversity at multiple loci. Though methods for constructing multi-gene pathways have proliferated in recent years, there are surprisingly few examples of using them to build libraries. Our mock library experiment is key because it explicitly tests the library sizes Reiterative Recombination can generate and shows that members of the library are present in the expected proportions. To our knowledge, this is the first DNA assembly method whose ability to create such libraries *in vivo* has been rigorously characterized in this way. In addition, though we only attempted to build libraries of up to 10^4—due to the limits of our ability to visually screen large numbers of colonies for lycopene production—the high efficiency and straightforward scalability of the recombination step suggests that it is only the transformation efficiency of yeast ($\sim 10^6$-10^8) that will limit library size in Reiterative Recombination. The development of highly efficient DNA assembly methods is an essential first step towards the combinatorial optimization of pathways *in vivo*, and efficient technologies for pathway construction such as Reiterative Recombination should ultimately allow the power of directed evolution and other library optimization approaches to be brought to bear on metabolic engineering and synthetic biology problems to an unprecedented extent.

Reiterative Recombination is part of a growing toolbox of techniques for making large-scale modifications to the genome. One strategy is *de novo* genome synthesis, which allows complete customization of the genome. However, in spite of recent heroic

feats in this field[32,59] and the falling price of chemical DNA synthesis[60], such ambitious undertakings are neither technically nor economically feasible for most researchers. The alternative is to reprogram well-characterized host organisms, such as *E. coli* and *S. cerevisiae*, for novel functions through genetic engineering, which will require 1) the modification of strains' genetic background and 2) the introduction of multiple exogenous genes into the chromosome. To meet the first of these needs, classic mutagenesis techniques such as mutator strains can be useful for phenotypic optimization, but they do not provide control over the extent and location of mutations[61]. The recently reported multiplex genome engineering (MAGE) platform can efficiently introduce specified deletions, point mutations, and short insertions (<~30bp) throughout the chromosome in *E. coli*, generating up to billions of variant strains[62]. Reiterative Recombination is one of several techniques that tackle the second issue, integrating exogenous pathways of genes into the chromosome[22,34,35]. However, Reiterative Recombination is uniquely able to integrate pathways in a highly efficient manner to access large numbers of variant strains.

In conclusion, we foresee Reiterative Recombination becoming a powerful addition to the 21st-century molecular biology toolkit. Its simplicity and robustness will make it a user-friendly option for any lab equipped for basic molecular biology to assemble multi-gene constructs. Its cyclical format means that it can be used to build pathways of indefinite length. Since it is highly efficient, in contrast to other *in vivo* DNA assembly technologies, it can be used to assemble libraries of $\geq 10^4$ pathways directly in the chromosome. Reiterative Recombination, as part of the expanding arsenal of cutting-edge cell engineering tools, will ensure the continued rapid development of

synthetic biology as the scale of our ambitions increases and our applications move into the cell.

5.4 Experimental methods

General materials and methods. General materials and methods were as in **Chapter 2**.

Plasmid construction. Donor plasmids containing *crtB* and *crtI* alleles with stop or silent mutations and diagnostic restriction sites were constructed as follows. The 3' region of *crtB* was amplified from LW2670Y genomic DNA with primers LMW374 and LMW529 (*crtB-stop*) or LMW374 and LMW530 (*crtB-silent*). The 5' region of *crtB* and the *PGK* promoter were amplified from LW2670Y genomic DNA with primers LMW491 and LMW375. These fragments (either *crtB-stop* or *crtB-silent*) were co-transformed with SmaI-digested pLW2593 into yeast. The p*TEF-crtI-tACT* construct was amplified from LW2671Y genomic DNA with primers LMW454 and LMW459. The unpurified reaction was used as a template to amplify the *TEF* promoter and the 5' region of *crtI* with primers LMW374 and LMW531 (*crtI-stop*) or primers LMW374 and LMW532 (*crtI-silent*). Using the same template, the 3' region of *crtI* and the *ACT* terminator were amplified with primers LMW452 and LMW459. *TRP1* was amplified from pRS414GAL with primers LMW460 and LMW461, followed by a second round of amplification with LMW460 and LMW367. These fragments (either *crtI-stop* or *crtI-silent*) were co-transformed with HindIII-digested pLW2592 into yeast. Following donor plasmid construction by plasmid gap repair, plasmids were retransformed into *E. coli*. Plasmids with the desired mutations in *crtB* and *crtI* were identified by PCR and

restriction analysis and confirmed by sequencing (pLW2597, *crtB-stop*; pLW2598, *crtB-silent*; pLW2599, *crtI-stop*; pLW2600, *crtI-silent*).

A plasmid to recover Reiterative Recombination pathways from the chromosome (pLW2662) was constructed as follows. The p*GAL-SceI*-t*CYC* construct was amplified using primers LMW300 and LMW301. Plasmid pLW2654 was digested with PspXI. The SceI PCR and the pLW2654 digest were co-transformed into yeast. Transformants were miniprepped from yeast without analysis, retransformed into *E. coli*, and then verified by colony PCR and sequencing.

Pathway construction via Reiterative Recombination. Reiterative Recombination was performed using the protocol described in **Chapter 4**, beginning with LW2591Y as the parental acceptor strain. The primers used to amplify constructs for donor plasmids made by plasmid gap repair in this chapter are shown in **Table 5-3**. Sequences of the assembled pathways are provided in the **Appendix**. The minimal lycopene pathway (**Section 5.2.1**) consists of rounds 1-3 of the lycopene pathway. For the complete, 8-gene lycopene pathway, the *TRP1* and *MET15* markers were selected for at all times after they were added in rounds 3 and 6, respectively.

Table 5-3. PCR amplification of subfragments for Reiterative Recombination.

Elongation Round	Subfragment	Construct	Primers	Template
		Lycopene pathway		
1 (pLW2592 donor)	1a	p*GPD*	LMW433/445	LW2647Y gDNA
	1b	*crtE*	LMW446/447	pSC203
	1c	t*ADH*	1) LMW448/449 2) LMW448/367	1) LW2647Y gDNA 2) PCR (1)
2 (pLW2593 donor)	2a	*crtB* (reverse complement)	1) LMW450/451 2) LMW374/451	1) pSC203 2) PCR (1)
	2b	p*PGK* (reverse complement)	1) LMW452/453 2) LMW451/375	1) LW2647Y gDNA 2) PCR (1)

3 (pLW2592 donor)	3a	p*TEF*	1) LMW454/455 2) LMW374/455	1) LW2647Y gDNA 2) PCR (1)
	3b	*crtI*	LMW456/457	pSC203
	3c	t*ACT*	LMW458/459	LW2647Y gDNA
	3d	*TRP1*	1) LMW460/461 2) LMW460/367	1) pRS414GAL 2) PCR (1)
4 (pLW2593 donor)	4a	p*GPD*	1) LMW464/465 2) LMW374/465	1) LW2589Y gDNA 2) PCR (1)
	4b	*tHMG1*	LMW466/467	pUCAD-tHMG1
	4c	t*ADH*	1) LMW468/469 2) LMW468/375	1) LW2589Y gDNA 2) PCR (1)
5 (pLW2592 donor)	5a	*ERG20* (reverse complement)	1) LMW472/473 2) LMW374/473	1) pSC103 2) PCR (1)
	5b	p*TEF* (reverse complement)	1) LMW474/475 2) LMW474/367	1) LW2589Y gDNA 2) PCR (1)
6 (pLW2593 donor)	6a	p*PGK*	1) LMW476/477 2) LMW374/477	1) LW2589Y gDNA 2) PCR (1)
	6b	*ALD6*	LMW478/479	pSC302
	6c	t*ACT*	LMW480/481	LW2589Y gDNA
	6d	*MET15*	1) LMW482/483 2) LMW482/375	1) FY251 gDNA 2) PCR (1)
7 (pLW2592 donor)	7a	p*TEF*	1) LMW484/485 2) LMW374/485	LW2589Y gDNA 2) PCR (1)
	7b	*acs*	LMW486/487	pSC302
	7c	t*ADH*	1) LMW488/489 2) LMW488/367	1) LW2589Y gDNA 2) PCR (1)
8 (pLW2593 donor)	8a	*UPC2* (reverse complement)	1) LMW501/502 2) LMW374/502	1) pSC103 2) PCR (1)
	8b	p*PGK* (reverse complement)	1) LMW503/504 2) LMW503/375	1) LW2589Y gDNA 2) PCR (1)
Lycopene libraries				
2 (pLW2593 donor)	2a	*crtB-stop* (reverse complement)	LMW374/451	pLW2597
	2a	*crtB-silent* (reverse complement)	LMW374/451	pLW2598
	2b	p*PGK* (reverse complement)	LMW452/375	pLW2597
3 (pLW2592 donor)	3abc	p*TEF*-(*crtI-stop*)-t*ACT*	LMW374/459	pLW2599
	3abc	p*TEF*-(*crtI-stop*)-t*ACT*	LMW374/459	pLW2600
	3d	*TRP1*	LMW460/367	pLW2599

***E. coli* carotenoid pathway**				
1 (pLW2594 donor)	1a	p*Lac*	LMW370/371	pAC-crtE-crtB-crtI
	1b	*crtE*	1) LMW372/373 2) LMW372/367	1) pAC-crtE-crtB-crtI 2) PCR (1)
2 (pLW2593 donor)	2	*crtB*	1) LMW376/377 2) LMW374/375	1) pAC-crtE-crtB-crtI 2) PCR (1)
3 (pLW2592 donor)	3	*crtI14*	1) LMW382/383 2) LMW374/367	1) pAC-crtE-crtB-crtI14 2) PCR (1)
4 (pLW2593 donor)	4	KanMX	1)LMW429/430 2)LMW374/375	1) pVC2265 2) PCR (1)

Library construction via Reiterative Recombination. Based on phylogenetic analyses of the *crtB* and *crtI* genes[63,64], we selected two adjacent, fully conserved amino acid residues in each gene to replace with stop codons. Using plasmid gap repair, plasmids were created containing *crtB* alleles with the following sequences at nucleotide positions 514 through 531 (mutated residues in lower case):

***crtB*-stop:** AAC ATt GCG taa tag ATT (stop codons at amino acid residues 175 and 176 and a new BsrDI restriction site, underlined)

***crtB*-silent:** AAC ATC GCG AGg GAT ATc (only silent mutations and a new EcoRV restriction site, underlined)

and *crtI* alleles with the following sequences at nucleotide positions 52 through 66:

***crtI*-stop:** ATA taa tag CAA GCT (stop codons at amino acid residues 19 and 20 and a new Cac8I restriction site, underlined)

***crtI*-silent:** ATt cgt ctc CAA GCT (only silent mutations and a new BsmBI restriction site, underlined)

These alleles were used as PCR templates for amplification of the *crtB* and *crtI* genes for the lycopene libraries, and purified PCR products were mixed in the desired

ratios (0:1, 10:1, or 100:1 stop/silent) during the round 2 and round 3 transformations. “Libraries” containing only the stop or only the silent alleles at each position were constructed in parallel, and the resulting cured colonies are shown in **Figure 5-2, C-F**. Following the third round of Reiterative Recombination, libraries were plated on SC(−Leucine, −Tryptophan, 0.1% 5-FOA) media for curing. Plates were grown at 30°C for three days rather than two days to allow full development of the orange color. The expected percentage of orange colonies was, for example, 100 x (1/101) x (1/11) = 0.90% for the 100:1 *crtB* stop/silent x 10:1 *crtI* stop/silent library. Single colonies from the libraries were analyzed by colony PCR and restriction analysis, following streak purification if necessary.

Analysis of *HIS LEU* mutants. The pool of cured recombinants was lifted from the curing plate for the rounds of Reiterative Recombination shown in **Table 5-1**. Serial dilutions of the resuspended cells were plated on SC(−Leucine) and SC(−Histidine) media. After two days of growth at 30°C, the number of colonies on each type of media was counted. Colonies that grew on the “wrong” media (SC(−Leucine) for even rounds, and SC(−Histidine) for odd rounds) were assayed for growth on the “correct” media (SC(−Histidine) for even rounds, and SC(−Leucine) for odd rounds) to confirm that they were prototrophic for both amino acids, and all tested colonies grew on both types of media. The percentages of colonies with this *HIS LEU* phenotype observed in four different Reiterative Recombination trials are shown in **Table 5-1**.

Shuttling of Carotenoid Pathway to *E. coli.* Strain LW2663Y was transformed with pLW2662. The pool of resulting colonies were grown overnight in SC(−Histidine, −Uracil) media and then plated on SC(−Histidine, −Uracil, 2% galactose, 2% raffinose).

The resulting pool of colonies was miniprepped and retransformed into *E. coli* TG1 cells, giving 6 Kan^R colonies. Colony PCR analysis indicated that one of these colonies had the *crtI* construct. This colony was miniprepped, and the plasmid was transformed into JM109 cells. All but one (of ~30) transformants had the desired orange phenotype, including LW2664E. JM109 cells were not used for the initial transformation of yeast plasmid DNA into *E. coli* only because the available competent JM109 cells had very low transformation efficiencies.

5.5 Strains, plasmids, and oligonucleotides

Table 5-4. Strains used in this study

Name	Genotype	Source/Reference
FY251	*MATa trp1Δ63 his3Δ200 ura3-52 leu2Δ1 GAL+*	M. Carlson
LW2589Y	BY4733 *MATa-inc*	This study
LW2591Y	Reiterative Recombination parental acceptor strain BY4733 *MATa-inc* pLW2590 integrated	This study
LW2647Y	BY4733 *MATΔ::URA3*	This study
LW2656Y	Round 1 recombinant from *E. coli* carotenoid pathway	This study
LW2657Y	Round 2 recombinant from *E. coli* carotenoid pathway	This study
LW2658Y	Round 3 recombinant from *E. coli* carotenoid pathway	This study
LW2663Y	Round 4 recombinant from *E. coli* carotenoid pathway	This study
LW2664E	JM109 *E. coli* with pLW2662 + recovered round 4 carotenoid pathway	This study
LW2669Y	Round 1 recombinant from lycopene pathway	This study
LW2670Y	Round 2 recombinant from lycopene pathway	This study
LW2671Y	Round 4 recombinant from lycopene pathway	This study
LW2674Y	Round 5 recombinant from lycopene pathway	This study
LW2675Y	Round 6 recombinant from lycopene pathway	This study
LW2676Y	Round 7 recombinant from lycopene pathway	This study
LW2677Y	Round 8 recombinant from lycopene pathway	This study
LW2678Y	Round 3 recombinant from lycopene pathway with *crtB-stop, crtI-stop*	This study
LW2679Y	Round 3 recombinant from lycopene pathway with *crtB-stop, crtI-silent*	This study
LW2680Y	Round 3 recombinant from lycopene pathway with *crtB-silent, crtI-stop*	This study
LW2681Y	Round 3 recombinant from lycopene pathway with *crtB-silent, crtI-silent*	This study

Table 5-5. Plasmids used in this study

Name	Details	Source/Reference
pAC-crtE-crtB-crtI	Modified pACYC184 containing the operon pLac-crtE-crtB-crtI, Cm^R	C. Schmidt-Dannert/53
pAC-crtE-crtB-crtI14	Modified pACYC184 containing the operon pLac-crtE-crtB-crtI14, Cm^R	C. Schmidt-Dannert/53
pRS414GAL	*pGAL1 CEN6/ARSH4 TRP1* pBIISK ori amp^R	ATCC #87336
pSC103	*TRP1* integration plasmid with overexpressed *tHMG1*, *ERG20*, and *UPC2*	P. Ajikumar and G. Stephanopoulos
pSC203	*URA3* integration plasmid with overexpressed *crtE*, *crtB*, and *crtI*	P. Ajikumar and G. Stephanopoulos
pSC302	*LEU2* integration plasmid with overexpressed *ALD6* and *acs*	P. Ajikumar and G. Stephanopoulos
pUCAD-tHMG1	*ADE1* integration plasmid with overexpressed *tHMG1*	P. Ajikumar and G. Stephanopoulos
pVC2265	HO-polylinker-KanMX4-HO	D. Stillman/65
pLW2592	Universal odd donor plasmid for Reiterative Recombination *pGAL1-HO-tCYC GFP-LEU2-SceI cleavage site CEN6/ARSH4 URA3* pBIISK ori amp^R	This study
pLW2593	Universal even donor plasmid for Reiterative Recombination *pGAL1-SceI-tCYC GFP-HIS3-HO cleavage site CEN6/ARSH4 URA3* pBIISK ori amp^R	This study
pLW2594	Round 1 donor plasmid for Reiterative Recombination *pGAL1-HO-tCYC GFP-LEU2-SceI cleavage site-KanMX-HO(L) CEN6/ARSH4 URA3* pBIISK ori amp^R	This study
pLW2597	pLW2593 with lycopene 2 fragment with *crtB-stop*	This study
pLW2598	pLW2593 with lycopene 2 fragment with *crtB-silent*	This study
pLW2599	pLW2592 with lycopene 3 fragment with *crtI-stop*	This study
pLW2600	pLW2592 with lycopene 3 fragment with *crtI-silent*	This study
pLW2654	pRS416 with *HO(R)-pPYK-GFP-LEU2-SceI site-HO(L)*	This study
pLW2662	pLW2654 with *pGAL-SceI-tCYC*	This study

Table 5-6. Oligonucleotides used in this study

Name	Sequence (5'-3')
LMW300	CGATTAAGTTGGGTAACGCCAGGGTTTTCCACGGATTAGAAGCCGCC
LMW301	CTGCAGGAATTCGATATCAAGCTTATCGATACCAAAGCCTTCGAGCGTCC
LMW367	TCAGTACAATCTTAGGGATAACAGGGTAAT
LMW370	AAAATTGTGCCTTTGGACTTAAAATGGCGTCCGACTGGAAAGCGGG
LMW371	TTTTTTGCGCAGACCGTCATTTTGTAATCCTCCTTCTAGA
LMW372	TCTAGAAGGAGGATTACAAAATGACGGTCTGCGCAAAAA
LMW373	CTTAGGGATAACAGGGTAATGCCATTTTGTAATCCTCCTGCATCCTTAACTGACGGC
LMW374	TGAGAAGGTTTTGGGACGCTCGAAGGCTTT
LMW375	GCACAGTTATACTGTTGCGGAAAGCTGAAA
LMW376	GGACGCTCGAAGGCTTTGCCGTCAGTTAAGGATGCAGGAGGATTACAAAATGGCA

LMW377	CTGTTGCGGAAAGCTGAAATTTGTAATCCTCCTCTCGAGCTAGAGCGGGC GCTGCC
LMW382	GGACGCTCGAAGGCTTTAGCGCCCGCTCTAGCTCGAGAGGAGGATTACAA AATGAAA
LMW383	CTTAGGGATAACAGGGTAATTTTGTAATCCTCCTGAATTCTCAAATCAGATC CTCCAGC
LMW429	GGACGCTCGAAGGCTTTGATGCTGGAGGATCTGATTTGAGAATTAGCTTG CCTCGTCCCC
LMW430	CTGTTGCGGAAAGCTGAAATCGACACTGGATGGCGG
LMW433	AAAATTGTGCCTTTGGACTTAAAATGGCGTAGTTTATCATTATCAATACTCG
LMW445	GCTTTCGAACCAGAAACCATTTTGTTTGTTTATGTGTGTTTAT
LMW446	AACACACATAAACAAACAAAATGGTTTCTGGTTCGAAAG
LMW447	CATAAATCATAAGAAATTCGCTTAGGCGATCTTCATCACT
LMW448	CAGTGATGAAGATCGCCTAAGCGAATTTCTTATGATTTATG
LMW449	CTTAGGGATAACAGGGTAATTATGGCAAAGACCTGTCTAAGAGCGACCTCA TGCTATA
LMW450	GGACGCTCGAAGGCTTTGGTATAGCATGAGGTCGCTCTTAGACAGGTCTT TGCCATA
LMW451	TTTACAACAAATATAAAACAATGAGTCAACCACCTTTGTT
LMW452	AACAAAGGTGGTTGACTCATTGTTTTATATTTGTTGTAAAAAGT
LMW453	CTGTTGCGGAAAGCTGAAATAGAAACATTTTGAAGCTATAGACGCGAATTT TTCGAAGA
LMW454	GGACGCTCGAAGGCTTTTCTTCGAAAAATTCGCGTCTATAGCTTCAAAATG TTTCTACT
LMW455	ATCACTACGGTTTTCTTCATTTTGTAATTAAAACTTAGATTAGA
LMW456	ATCTAAGTTTTAATTACAAAATGAAGAAAACCGTAGTGATT
LMW457	TACGCGCACAAAAGCAGAGATTACTGCAAATCCTCGATCA
LMW458	TGATCGAGGATTTGCAGTAATCTCTGCTTTTGTGCGCG
LMW459	ATATATATAGTAATGTCGTTACACTATGATATATAAATATAATAG
LMW460	ATATTTATATATCATAGTGTAACGACATTACTATATATATAATA
LMW461	CTTAGGGATAACAGGGTAATAGTATTGATAATGATAAACTAGGCAAGTGCA CAAACAATA
LMW464	GGACGCTCGAAGGCTTTTATTGTTTGTGCACTTGCCTAGTTTATCATTATCA ATACTCG
LMW465	GTTTTCACCAATTGGTCCATTTTGTTTGTTTATGTGTGTTTAT
LMW466	AACACACATAAACAAACAAAATGGACCAATTGGTGAAAAC
LMW467	CATAAATCATAAGAAATTCGCTTAGGATTTAATGCAGGTGA
LMW468	TCACCTGCATTAAATCCTAAGCGAATTTCTTATGATTTATG
LMW469	CTGTTGCGGAAAGCTGAAATTTACAAGAGAAGCAAATAGGAGCGACCTCAT GCTATA
LMW472	GGACGCTCGAAGGCTTTGGTATAGCATGAGGTCGCTCCTATTTGCTTCTCT TGTAAAC
LMW473	ATCTAAGTTTTAATTACAAAATGGCTTCAGAAAAAGAAATTA
LMW474	ATTTCTTTTTCTGAAGCCATTTTGTAATTAAAACTTAGATTAGA
LMW475	CTTAGGGATAACAGGGTAATTTACAACAAATATAAAACAATAGCTTCAAAAT GTTTCTAC
LMW476	GGACGCTCGAAGGCTTTGGAGTAGAAACATTTTGAAGCTATTGTTTTGCAA GTACCACTG
LMW477	TCAAAGTGTAGCTTAGTCATTGTTTTATATTTGTTGTAAAAAGT
LMW478	TTTACAACAAATATAAAACAATGACTAAGCTACACTTTGA

LMW479	TACGCGCACAAAAGCAGAGATTACAACTTAATTCTGACAGC
LMW480	CTGTCAGAATTAAGTTGTAATCTCTGCTTTTGTGCGCG
LMW481	GATTCGAACCCTTGCATCCGACACTATGATATATAAATATAATAG
LMW482	ATATTTATATATCATAGTGTCGGATGCAAGGGTTCGAA
LMW483	CTGTTGCGGAAAGCTGAAATAGAAACATTTTGAAGCTATTAGATGAGTCTTGTTCACC
LMW484	GGACGCTCGAAGGCTTTGGTGAACAAGACTCATCTAATAGCTTCAAAATGTTTCTACT
LMW485	TGTTTATGTGTTTGGCTCATTTTGTAATTAAAACTTAGATTAGA
LMW486	ATCTAAGTTTTAATTACAAAATGAGCCAAACACATAAACA
LMW487	CATAAATCATAAGAAATTCGCTTATGACGGCATCGCGAT
LMW488	CCATCGCGATGCCGTCATAAGCGAATTTCTTATGATTTATGA
LMW489	CTTAGGGATAACAGGGTAATCTCTGATTTTTCGTTATGAGAGCGACCTCATGCTATA
LMW491	CGCGATGTTTGTCAGCTG
LMW501	GGACGCTCGAAGGCTTTGGTATAGCATGAGGTCGCTCTCATAACGAAAAATCAGAGAA
LMW502	TTTACAACAAATATAAAACAATGAGCGAAGTCGGTATAC
LMW503	TGTATACCGACTTCGCTCATTGTTTTATATTTGTTGTAAAAAG
LMW504	CTGTTGCGGAAAGCTGAAAAGACGCGAATTTTTCGAAG
LMW529	GGTTTGGCTTTTCAGCTGACAAACATTGCGTAATAGATTATCGACG
LMW530	GGTTTGGCTTTTCAGCTGACAAACATCGCGAGGGATATCATCGACG
LMW531	CAATAGCACTGTAGGAATACCTGCAGCTTGCTATTATATAGCCAAAG
LMW532	CAATAGCACTGTAGGAATACCTGCAGCTTGGAGACGAATAGCCAAAG

5.6 References

1. Newman, D.J. & Cragg, G.M. Natural products as sources of new drugs over the last 25 years. *J. Nat. Prod.* **70**, 461-477 (2007).

2. Cragg, G.M., Grothaus, P.G. & Newman, D.J. Impact of natural products on developing new anti-cancer agents. *Chem. Rev.* **109**, 3012-3043 (2009).

3. Nicolaou, K.C., Vourloumis, D., Winssinger, N. & Baran, P.S. The art and science of total synthesis at the dawn of the twenty-first century. *Angew. Chem. Int. Ed. Engl.* **39**, 44-122 (2000).

4. Wilson, R.M. & Danishefsky, S.J. Small molecule natural products in the discovery of therapeutic agents: The synthesis connection. *J. Org. Chem.* **71**, 8329-8351 (2006).

5. Rodriguez, E., Menzella, H.G. & Gramajo, H. Heterologous production of polyketides in bacteria. *Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates* **459**, 339-365 (2009).

6. Ro, D.K. et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940-943 (2006).

7. Ajikumar, P.K. et al. Isoprenoid pathway optimization for taxol precursor overproduction in Escherichia coli. *Science* **330**, 70-74 (2010).

8. Hawkins, K.M. & Smolke, C.D. Production of benzylisoquinoline alkaloids in Saccharomyces cerevisiae. *Nat. Chem. Biol.* **4**, 564-573 (2008).

9. Leonard, E. et al. Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 13654-13659 (2010).

10. Wenzel, S.C. & Muller, R. Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways. *Curr. Opin. Biotechnol.* **16**, 594-606 (2005).

11. Galm, U. & Shen, B. Expression of biosynthetic gene clusters in heterologous hosts for natural product production and combinatorial biosynthesis. *Expert Opin. Drug Discovery* **1**, 409-437 (2006).

12. Olano, C., Lombo, F., Mendez, C. & Salas, J.A. Improving production of bioactive secondary metabolites in actinomycetes by metabolic engineering. *Metab. Eng.* **10**, 281-292 (2008).

13. Mutka, S.C., Bondi, S.M., Carney, J.R., Da Silva, N.A. & Kealey, J.T. Metabolic pathway engineering for complex polyketide biosynthesis in Saccharomyces cerevisiae. *FEMS Yeast Res.* **6**, 40-47 (2006).

14. Kittendorf, J.D. & Sherman, D.H. Developing tools for engineering hybrid polyketide synthetic pathways. *Curr. Opin. Biotechnol.* **17**, 597-605 (2006).

15. Zhou, H., Xie, X.K. & Tang, Y. Engineering natural products using combinatorial biosynthesis and biocatalysis. *Curr. Opin. Biotechnol.* **19**, 590-596 (2008).

16. Runguphan, W., Qu, X. & O'Connor, S.E. Integrating carbon-halogen bond formation into medicinal plant metabolism. *Nature* **468**, 461-464 (2010).

17. Eustaquio, A.S., O'Hagan, D. & Moore, B.S. Engineering fluorometabolite production: Fluorinase expression in *Salinispora tropica* yields fluorosalinosporamide. *J. Nat. Prod.* **73**, 378-382 (2010).

18. Menzella, H.G. et al. Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* **23**, 1171-1176 (2005).

19. Xue, Q., Ashley, G., Hutchinson, C.R. & Santi, D.V. A multiplasmid approach to preparing large libraries of polyketides. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 11740-11745 (1999).

20. Kodumal, S.J. et al. Total synthesis of long DNA sequences: Synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 15573-15578 (2004).

21. Shao, Z., Luo, Y. & Zhao, H. Rapid characterization and engineering of natural product biosynthetic pathways via DNA assembler. *Mol. Biosyst.* **7**, 1056-1059 (2011).

22. Shao, Z.Y., Zhao, H. & Zhao, H.M. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.* **37**, e16 (2009).

23. Alper, H., Miyaoku, K. & Stephanopoulos, G. Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. *Nat. Biotechnol.* **23**, 612-616 (2005).

24. Tyo, K.E., Kocharin, K. & Nielsen, J. Toward design-based engineering of industrial microbes. *Curr. Opin. Microbiol.* **13**, 255-262 (2010).

25. Steen, E.J. et al. Metabolic engineering of Saccharomyces cerevisiae for the production of n-butanol. *Microb. Cell Fact.* **7**, 36 (2008).

26. Fischbach, M.A., Lai, J.R., Roche, E.D., Walsh, C.T. & Liu, D.R. Directed evolution can rapidly improve the activity of chimeric assembly-line enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 11951-11956 (2007).

27. Alper, H., Fischer, C., Nevoigt, E. & Stephanopoulos, G. Tuning genetic control through promoter engineering. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 12678-12683 (2005).

28. Pfleger, B.F., Pitera, D.J., Smolke, C.D. & Keasling, J.D. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat. Biotechnol.* **24**, 1027-1032 (2006).

29. Ramon, A. & Smith, H.O. Single-step linker-based combinatorial assembly of promoter and gene cassettes for pathway engineering. *Biotechnol. Lett.* **33**, 549-555 (2011).

30. Gibson, D.G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343-345 (2009).

31. Gibson, D.G., Smith, H.O., Hutchison, C.A., 3rd, Venter, J.C. & Merryman, C. Chemical synthesis of the mouse mitochondrial genome. *Nat. Methods* **7**, 901-903 (2010).

32. Gibson, D.G. et al. Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science* **319**, 1215-1220 (2008).

33. Gibson, D.G. et al. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20404-20409 (2008).

34. Itaya, M., Fujita, K., Kuroki, A. & Tsuge, K. Bottom-up genome assembly using the Bacillus subtilis genome vector. *Nat. Methods* **5**, 41-43 (2008).

35. Itaya, M., Tsuge, K., Koizumi, M. & Fujita, K. Combining two genomes in one cell: Stable cloning of the Synechocystis PCC6803 genome in the Bacillus subtilis 168 genome. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15971-15976 (2005).

36. Yamano, S., Ishii, T., Nakagawa, M., Ikenaga, H. & Misawa, N. Metabolic engineering for production of beta-carotene and lycopene in Saccharomyces cerevisiae. *Biosci. Biotechnol. Biochem.* **58**, 1112-1114 (1994).

37. Ajikumar, P.K. et al. Terpenoids: Opportunities for biosynthesis of natural product drugs using engineered microorganisms. *Mol. Pharmaceutics* **5**, 167-190 (2008).

38. Roberts, S.C. Production and engineering of terpenoids in plant cell culture. *Nat. Chem. Biol.* **3**, 387-395 (2007).

39. Chang, M.C. & Keasling, J.D. Production of isoprenoid pharmaceuticals by engineered microbes. *Nat. Chem. Biol.* **2**, 674-681 (2006).

40. Chang, M.C., Eachus, R.A., Trieu, W., Ro, D.K. & Keasling, J.D. Engineering Escherichia coli for production of functionalized terpenoids using plant P450s. *Nat. Chem. Biol.* **3**, 274-277 (2007).

41. Van Noorden, R. Demand for malaria drug soars. *Nature* **466**, 672-673 (2010).

42. Shiba, Y., Paradise, E.M., Kirby, J., Ro, D.K. & Keasling, J.D. Engineering of the pyruvate dehydrogenase bypass in Saccharomyces cerevisiae for high-level production of isoprenoids. *Metab. Eng.* **9**, 160-168 (2007).

43. Shimada, H. et al. Increased carotenoid production by the food yeast Candida utilis through metabolic engineering of the isoprenoid pathway. *Appl. Environ. Microbiol.* **64**, 2676-2680 (1998).

44. Verwaal, R. et al. High-level production of beta-carotene in Saccharomyces cerevisiae by successive transformation with carotenogenic genes from Xanthophyllomyces dendrorhous. *Appl. Environ. Microbiol.* **73**, 4342-4350 (2007).

45. Donald, K.A., Hampton, R.Y. & Fritz, I.B. Effects of overproduction of the catalytic domain of 3-hydroxy-3-methylglutaryl coenzyme A reductase on squalene synthesis in Saccharomyces cerevisiae. *Appl. Environ. Microbiol.* **63**, 3341-3344 (1997).

46. Engels, B., Dahm, P. & Jennewein, S. Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards taxol (paclitaxel) production. *Metab. Eng.* **10**, 201-206 (2008).

47. Davies, B.S., Wang, H.S. & Rine, J. Dual activators of the sterol biosynthetic pathway of Saccharomyces cerevisiae: Similar activation/regulatory domains but different response mechanisms. *Mol. Cell. Biol.* **25**, 7375-7385 (2005).

48. Lutke-Eversloh, T. & Stephanopoulos, G. Combinatorial pathway analysis for improved L-tyrosine production in Escherichia coli: Identification of enzymatic bottlenecks by systematic gene overexpression. *Metab. Eng.* **10**, 69-77 (2008).

49. Peralta-Yahya, P., Carter, B.T., Lin, H., Tao, H. & Cornish, V.W. High-throughput selection for cellulase catalysts using chemical complementation. *J. Am. Chem. Soc.* **130**, 17446-17452 (2008).

50. Kealey, J.T., Liu, L., Santi, D.V., Betlach, M.C. & Barr, P.J. Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 505-509 (1998).

51. Wattanachaisaereekul, S., Lantz, A.E., Nielsen, M.L. & Nielsen, J. Production of the polyketide 6-MSA in yeast engineered for increased malonyl-CoA supply. *Metab. Eng.* **10**, 246-254 (2008).

52. Siewers, V., Chen, X., Huang, L., Zhang, J. & Nielsen, J. Heterologous production of non-ribosomal peptide LLD-ACV in Saccharomyces cerevisiae. *Metab. Eng.* **11**, 391-397 (2009).

53. Schmidt-Dannert, C., Umeno, D. & Arnold, F.H. Molecular breeding of carotenoid biosynthetic pathways. *Nat. Biotechnol.* **18**, 750-753 (2000).

54. Orr-Weaver, T.L. & Szostak, J.W. Yeast recombination - the association between double-strand gap repair and crossing-over. *Proc. Natl. Acad. Sci. U. S. A.* **80**, 4417-4421 (1983).

55. Rothstein, R. Targeting, disruption, replacement, and allele rescue - integrative DNA transformation in yeast. *Methods Enzymol.* **194**, 281-301 (1991).

56. Yuan, L.W. & Keil, R.L. Distance-independence of mitotic intrachromosomal recombination in *Saccharomyces cerevisiae*. *Genetics* **124**, 263-273 (1990).

57. Toyn, J.H., Gunyuzlu, P.L., White, W.H., Thompson, L.A. & Hollis, G.F. A counterselection for the tryptophan pathway in yeast: 5-fluoroanthranilic acid resistance. *Yeast* **16**, 553-560 (2000).

58. Chattoo, B.B. et al. Selection of *lys2* mutants of the yeast *Saccharomyces cerevisiae* by the utilization of α-aminoadipate. *Genetics* **93**, 51-65 (1979).

59. Gibson, D.G. et al. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52-56 (2010).

60. Carr, P.A. & Church, G.M. Genome engineering. *Nat. Biotechnol.* **27**, 1151-1162 (2009).

61. Sauer, U. Evolutionary engineering of industrially important microbial phenotypes. *Adv. Biochem. Eng. Biotechnol.* **73**, 129-169 (2001).

62. Wang, H.H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-898 (2009).

63. Robinson, G.W., Tsay, Y.H., Kienzle, B.K., Smithmonroy, C.A. & Bishop, R.W. Conservation between human and fungal squalene synthetases - similarities in structure, function, and regulation. *Mol. Cell. Biol.* **13**, 2706-2717 (1993).

64. Xu, Z.J., Tian, B., Sun, Z.T., Lin, J. & Hua, Y.J. Identification and functional analysis of a phytoene desaturase gene from the extremely radioresistant bacterium Deinococcus radiodurans. *Microbiology* **153**, 1642-1652 (2007).

65. Voth, W.P., Richards, J.D., Shaw, J.M. & Stillman, D.J. Yeast vectors for integration at the HO locus. *Nucleic Acids Res.* **29**, e59 (2001).

Appendix

Sequences of Pathways Constructed by Reiterative Recombination

A.1 Sequence of the reporter gene pathway

Sequence of the completed reporter gene pathway, which is shown schematically in **Figure 4-7**. To show the context of the sequence in the chromosome, the last 30 bp of the *HO(L)* region on the 5' end and the first 30 bp of the acceptor module (containing the SceI cleavage site and the *LEU2* terminator) on the 3' end are shown.

```
   1 . . 30   HO(L)
  31 . . 4145 Fragment 1 (pADH-lacZ-tADH)
4146 . . 5957 Subfragment 2a (gusA, reverse complement)
5958 . . 6452 Subfragment 2b (pCYC, reverse complement)
6453 . . 8086 Subfragment 3a (tMET-MET15, reverse complement)
8087 . . 8498 Subfragment 3b (pTEF, reverse complement)
8499 . . 8528 Acceptor module
```

```
   1  AAAATTGTGC CTTTGGACTT AAAATGGCGT CAACTTCTTT TCTTTTTTTT
  51  TCTTTTCTCT CTCCCCCGTT GTTGTCTCAC CATATCCGCA ATGACAAAAA
 101  AATGATGGAA GACACTAAAG GAAAAAATTA ACGACAAAGA CAGCACCAAC
 151  AGATGTCGCT GTTCCAGAGC TGATGAGGGG TATCTCGAAG CACACGAAAC
 201  TTTTTCCTTC CTTCATTCAC GCACGCTACT CTCTAATGAG CAACGGTATA
 251  CGGCCTTCCT TCCAGTTACT TGAATTTGAA ATAAAAAAAA GTTTGCTGTC
 301  TTGCTATCAA GTATAAATAG ACCTGCAATT ATTAATCTTT TGTTTCCTCG
 351  TCATTGTTCT CGTTCCCTTT CTTCCTTGTT TCTTTTTCTG CACAATATTT
 401  CAAGCTATAC CAAGCATACA ATCAACTCCA AGCTTGAATT AATACCGGGC
 451  GGAATGACTA AATCTCATTC AGAAGAAGTG ATTGTACCTG AGTTCAATTC
 501  TAGCGCAAAG GAATTACCAA GACCATTGGC CGAAAAGTGC GGAATTCCAA
 551  GCTTGGCCAA GCCCGGATCC GGAGCTTGGC TGTTGCCCGT CTCACTGGTG
 601  AAAAGAAAAA CCACCCTGGC GCCCAATACG CAAACCGCCT CTCCCCGCGC
 651  GTTGGCCGAT TCATTAATGC AGCTGGCACG ACAGGTTTCC CGACTTAATC
 701  GCCTTGCAGC ACATCCCCCT TTCGCCAGCT GGCGTAATAG CGAAGAGGCC
```

751 CGCACCGATC GCCCTTCCCA ACAGTTGCGC AGCCTGAATG GCGAATGGCG

801 CTTTGCCTGG TTTCCGGCAC CAGAAGCGGT GCCGGAAAGC TGGCTGGAGT

851 GCGATCTTCC TGAGGCCGAT ACTGTCGTCG TCCCCTCAAA CTGGCAGATG

901 CACGGTTACG ATGCGCCCAT CTACACCAAC GTAACCTATC CCATTACGGT

951 CAATCCGCCG TTTGTTCCCA CGGAGAATCC GACGGGTTGT TACTCGCTCA

1001 CATTTAATGT TGATGAAAGC TGGCTACAGG AAGGCCAGAC GCGAATTATT

1051 TTTGATGGCG TTAACTCGGC GTTTCATCTG TGGTGCAACG GGCGCTGGGT

1101 CGGTTACGGC CAGGACAGTC GTTTGCCGTC TGAATTTGAC CTGAGCGCAT

1151 TTTTACGCGC CGGAGAAAAC CGCCTCGCGG TGATGGTGCT GCGTTGGAGT

1201 GACGGCAGTT ATCTGGAAGA TCAGGATATG TGGCGGATGA GCGGCATTTT

1251 CCGTGACGTC TCGTTGCTGC ATAAACCGAC TACACAAATC AGCGATTTCC

1301 ATGTTGCCAC TCGCTTTAAT GATGATTTCA GCCGCGCTGT ACTGGAGGCT

1351 GAAGTTCAGA TGTGCGGCGA GTTGCGTGAC TACCTACGGG TAACAGTTTC

1401 TTTATGGCAG GGTGAAACGC AGGTCGCCAG CGGCACCGCG CCTTTCGGCG

1451 GTGAAATTAT CGATGAGCGT GGTGGTTATG CCGATCGCGT CACACTACGT

1501 CTGAACGTCG AAAACCCGAA ACTGTGGAGC GCCGAAATCC CGAATCTCTA

1551 TCGTGCGGTG GTTGAACTGC ACACCGCCGA CGGCACGCTG ATTGAAGCAG

1601 AAGCCTGCGA TGTCGGTTTC CGCGAGGTGC GGATTGAAAA TGGTCTGCTG

1651 CTGCTGAACG GCAAGCCGTT GCTGATTCGA GGCGTTAACC GTCACGAGCA

1701 TCATCCTCTG CATGGTCAGG TCATGGATGA GCAGACGATG GTGCAGGATA

1751 TCCTGCTGAT GAAGCAGAAC AACTTTAACG CCGTGCGCTG TTCGCATTAT

1801 CCGAACCATC CGCTGTGGTA CACGCTGTGC GACCGCTACG GCCTGTATGT

1851 GGTGGATGAA GCCAATATTG AAACCCACGG CATGGTGCCA ATGAATCGTC

1901 TGACCGATGA TCCGCGCTGG CTACCGGCGA TGAGCGAACG CGTAACGCGA

1951 ATGGTGCAGC GCGATCGTAA TCACCCGAGT GTGATCATCT GGTCGCTGGG

2001 GAATGAATCA GGCCACGGCG CTAATCACGA CGCGCTGTAT CGCTGGATCA

2051 AATCTGTCGA TCCTTCCCGC CCGGTGCAGT ATGAAGGCGG CGGAGCCGAC

2101 ACCACGGCCA CCGATATTAT TTGCCCGATG TACGCGCGCG TGGATGAAGA

2151 CCAGCCCTTC CCGGCTGTGC CGAAATGGTC CATCAAAAAA TGGCTTTCGC

2201 TACCTGGAGA GACGCGCCCG CTGATCCTTT GCGAATACGC CCACGCGATG

2251 GGTAACAGTC TTGGCGGTTT CGCTAAATAC TGGCAGGCGT TTCGTCAGTA

2301 TCCCCGTTTA CAGGGCGGCT TCGTCTGGGA CTGGGTGGAT CAGTCGCTGA

2351 TTAAATATGA TGAAAACGGC AACCCGTGGT CGGCTTACGG CGGTGATTTT

2401 GGCGATACGC CGAACGATCG CCAGTTCTGT ATGAACGGTC TGGTCTTTGC

2451 CGACCGCACG CCGCATCCAG CGCTGACGGA AGCAAAACAC CAGCAGCAGT

2501 TTTTCCAGTT CCGTTTATCC GGGCAAACCA TCGAAGTGAC CAGCGAATAC

2551 CTGTTCCGTC ATAGCGATAA CGAGCTCCTG CACTGGATGG TGGCGCTGGA

2601 TGGTAAGCCG CTGGCAAGCG GTGAAGTGCC TCTGGATGTC GCTCCACAAG

2651 GTAAACAGTT GATTGAACTG CCTGAACTAC CGCAGCCGGA GAGCGCCGGG

2701 CAACTCTGGC TCACAGTACG CGTAGTGCAA CCGAACGCGA CCGCATGGTC

2751 AGAAGCCGGG CACATCAGCG CCTGGCAGCA GTGGCGTCTG GCGGAAAACC

2801 TCAGTGTGAC GCTCCCCGCC GCGTCCCACG CCATCCCGCA TCTGACCACC

2851 AGCGAAATGG ATTTTTGCAT CGAGCTGGGT AATAAGCGTT GGCAATTTAA

2901 CCGCCAGTCA GGCTTTCTTT CACAGATGTG GATTGGCGAT AAAAAACAAC

2951 TGCTGACGCC GCTGCGCGAT CAGTTCACCC GTGCACCGCT GGATAACGAC

3001 ATTGGCGTAA GTGAAGCGAC CCGCATTGAC CCTAACGCCT GGGTCGAACG

3051 CTGGAAGGCG GCGGGCCATT ACCAGGCCGA AGCAGCGTTG TTGCAGTGCA

3101 CGGCAGATAC ACTTGCTGAT GCGGTGCTGA TTACGACCGC TCACGCGTGG

3151 CAGCATCAGG GGAAAACCTT ATTTATCAGC CGGAAAACCT ACCGGATTGA

3201 TGGTAGTGGT CAAATGGCGA TTACCGTTGA TGTTGAAGTG GCGAGCGATA

3251 CACCGCATCC GGCGCGGATT GGCCTGAACT GCCAGCTGGC GCAGGTAGCA

3301 GAGCGGGTAA ACTGGCTCGG ATTAGGGCCG CAAGAAAACT ATCCCGACCG

3351 CCTTACTGCC GCCTGTTTTG ACCGCTGGGA TCTGCCATTG TCAGACATGT

3401 ATACCCCGTA CGTCTTCCCG AGCGAAAACG GTCTGCGCTG CGGGACGCGC

3451 GAATTGAATT ATGGCCCACA CCAGTGGCGC GGCGACTTCC AGTTCAACAT

3501 CAGCCGCTAC AGTCAACAGC AACTGATGGA AACCAGCCAT CGCCATCTGC

3551 TGCACGCGGA AGAAGGCACA TGGCTGAATA TCGACGGTTT CCATATGGGG

3601 ATTGGTGGCG ACGACTCCTG GAGCCCGTCA GTATCGGCGG AATTCCAGCT

3651 GAGCGCCGGT CGCTACCATT ACCAGTTGGT CTGGTGTCAA AAATAATTAC

3701 AACAGGTGTT GTCCTCTGAG GACATAAAAT ACACACCGAG ATTCATCAAC

3751 TCATTGCTGG AGTTAGCATA TCTACAATTG GGTGAAATGG GGAGCGATTT

3801 GCAGGCATTT GCTCGGCATG CCGGTAGAGG TGTGGTCAAT AAGAGCGACC

3851 TCATGCTATA CCTGAGAAAG CAACCTGACC TACAGGAAAG AGTTACTCAA

3901 GAACAAGAAT TTTCGTTTTA AAACCTAAGA GTCACTTTAA AATTTGTATA

3951 CACTTATTTT TTTTATAACT TATTTAATAA TAAAAATCAT AAATCATAAG

4001 AAATTCGCTT ATTTAGAAGT GTCAACAACG TATCTACCAA CGATTTGACC

4051 CTTTTCCATC TTTTCGTAAA TTTCTGGCAA GGTAGACAAG CCGACAACCT

4101 TGATTGGAGA CTTGACCAAA CCTCTGGCGA AGAAGTCCAA AGCTTTCATT

4151 GTTTGCCTCC CTGCTGCGGT TTTTCACCGA AGTTCATGCC AGTCCAGCGT

4201 TTTTGCAGCA GAAAAGCCGC CGACTTCGGT TTGCGGTCGC GAGTGAAGAT

4251 CCCTTTCTTG TTACCGCCAA CGCGCAATAT GCCTTGCGAG GTCGCAAAAT

4301 CGGCGAAATT CCATACCTGT TCACCGACGA CGGCGCTGAC GCGATCAAAG

4351 ACGCGGTGAT ACATATCCAG CCATGCACAC TGATACTCTT CACTCCACAT

4401 GTCGGTGTAC ATTGAGTGCA GCCCGGCTAA CGTATCCACG CCGTATTCGG

4451 TGATGATAAT CGGCTGATGC AGTTTCTCCT GCCAGGCCAG AAGTTCTTTT

4501 TCCAGTACCT TCTCTGCCGT TTCCAAATCG CCGCTTTGGA CATACCATCC

4551 GTAATAACGG TTCAGGCACA GCACATCAAA GAGATCGCTG ATGGTATCGG

4601 TGTGAGCGTC GCAGAACATT ACATTGACGC AGGTGATCGG ACGCGTCGGG

4651 TCGAGTTTAC GCGTTGCTTC CGCCAGTGGC GCGAAATATT CCCGTGCACC

4701 TTGCGGACGG GTATCCGGTT CGTTGGCAAT ACTCCACATC ACCACGCTTG

4751 GGTGGTTTTT GTCACGCGCT ATCAGCTCTT TAATCGCCTG TAAGTGCGCT

4801 TGCTGAGTTT CCCCGTTGAC TGCCTCTTCG CTGTACAGTT CTTTCGGCTT

4851 GTTGCCCGCT TCGAAACCAA TGCCTAAAGA GAGGTTAAAG CCGACAGCAG

4901 CAGTTTCATC AATCACCACG ATGCCATGTT CATCTGCCCA GTCGAGCATC

4951 TCTTCAGCGT AAGGGTAATG CGAGGTACGG TAGGAGTTGG CCCCAATCCA

5001 GTCCATTAAT GCGTGGTCGT GCACCATCAG CACGTTATCG AATCCTTTGC

5051 CACGTAAGTC CGCATCTTCA TGACGACCAA AGCCAGTAAA GTAGAACGGT

5101 TTGTGGTTAA TCAGGAACTG TTCGCCCTTC ACTGCCACTG ACCGGATGCC

5151 GACGCGAAGC GGGTAGATAT CACACTCTGT CTGGCTTTTG GCTGTGACGC

5201 ACAGTTCATA GAGATAACCT TCACCCGGTT GCCAGAGGTG CGGATTCACC

5251 ACTTGCAAAG TCCCGCTAGT GCCTTGTCCA GTTGCAACCA CCTGTTGATC

5301 CGCATCACGC AGTTCAACGC TGACATCACC ATTGGCCACC ACCTGCCAGT

5351 CAACAGACGC GTGGTTACAG TCTTGCGCGA CATGCGTCAC CACGGTGATA

5401 TCGTCCACCC AGGTGTTCGG CGTGGTGTAG AGCATTACGC TGCGATGGAT

5451 TCCGGCATAG TTAAAGAAAT CATGGAAGTA AGACTGCTTT TTCTTGCCGT

5501 TTTCGTCGGT AATCACCATT CCCGGCGGGA TAGTCTGCCA GTTCAGTTCG

5551 TTGTTCACAC AAACGGTGAT ACGTACACTT TTCCCGGCAA TAACATACGG

5601 CGTGACATCG GCTTCAAATG GCGTATAGCC GCCCTGATGC TCCATCACTT

5651 CCTGATTATT GACCCACACT TTGCCGTAAT GAGTGACCGC ATCGAAACGC

5701 AGCACGATAC GCTGGCCTGC CCAACCTTTC GGTATAAAGA CTTCGCGCTG

5751 ATACCAGACG TTGCCCGCAT AATTACGAAT ATCTGCATCG GCGAACTGAT

5801 CGTTAAAACT GCCTGGCACA GCAATTGCCC GGCTTTCTTG TAACGCGCTT

5851 TCCCACCAAC GCTGATCAAT TCCACAGTTT TCGCGATCCA GACTGAATGC

5901 CCACAGGCCG TCGAGTTTTT TGATTTCACG GGTTGGGGTT TCTACAGGAC

5951 GTAACATTAT TAATTTAGTG TGTGTATTTG TGTTTGTGTG TCTATAGAAG

6001 TATAGTAATT TATGCTGCAA AGGTCCTAAT GTATAAGGAA AGAATATTTA

6051 GAGAAAAGAA GAAAACAAGA GTTTTATATA CATACAGAGC ACATGCATGC

6101 CATATGATCA TGTGTCGTCG CACACATATA TATATGCCTG TATGTGTCAG

6151 CACTAAAGTT GCCTGGCCAT CCACGCTATA TATACACGCC TGGCGGATCT

6201 GCTCGAGGAT TGCCTACGCG TGGGCTTGAT CCACCAACCA ACGCTCGCCA

6251 AATGAACTGG CGCTTTGGTC TTCTGCCATC GTCCGTAAAC CCCGGCCAAA

6301 GAGACCGGAA AGATCGGTGA AAACATCTTG ATCTTGCTCC CGGGAATTTT

6351 AGATTCAGGT AGGAAATTGA TTACATCAAT ACTGTTACCC TGAATCATAT

6401 TCGACGATGT CGTCTCACAC GGAAATATAA TTCATTTCTT GGTTTTCCAA

6451 AGAAACCTCC ATCATCCTCT TTTGTAACTT GGTCCTACAA TAAATTTATC

6501 CAGTGTGACA GCTTTATAGG AGGCGTAAAG TAGTCTCATG AAGTAGATGA

6551 GTCTTGTTCA CCATATTTTT CTTCTCGACT GCGAATTAAC ACTGTTCTTT

6601 GATGTTAGAA CAATTTAGGT TCAAAGTACG AGTCACGACA TGTAACAAGG

6651 GAAAAAAAGG ATATTCATTT CAATAAAGTT CGTTTTATAA AAGTATAGTA

6701 CTTGTGAGAG AAAGTAGGTT TATACATAAT TTTACAACTC ATTACGCACA

6751 CTCATGGTTT TTGGCCAGCG AAAACAGTTT CAAAAGATTG CTGGAAGTCT

6801 GCAATAATGT CATCAATAAA TTCGATACCA ACAGAGACAC GAATTAAGTC

6851 CTTGGTAACA CCAGATGCCA ACTTTTCTTT GTCATTTAAT TGTTTGTGGG

6901 TAGTGAAGTA TGGAGCAATG ACTAAGGTCT TGGCATCACC AACATTGGCC

6951 AAGTTAGAGG CAAGCTTTAA ATTGTCAACA ACTTGAGCAC CAGAAAGTTT

7001 GAATGGGTCA GTTTCCTTGT CGGCATTTGG TAAGTCTTTT ACACCGAAAG

7051 ATAAGACACC ACCGAAACCG TTAGATAGAT ACTTCTTAGC ATTTTCATGA

7101 TGAGAATGAG ATGCTAAACC AGGGTATGAA ACCCAAGATA CGTATGGGGA

7151 TTGTTCTAAC CATTTGGCTA ACTTCAATGC ATTTTCACCG TGTCTTTCAG

7201 CTCTCAAAGA TAATGTTTCA ACACCTTGTA GTAGCAAGAA AGAGGCAAAT

7251 GGGTTCATCA ATGGACCCAA ATCTCTTAAT AGTTCAGTTC TAACATGAAC

7301 GATGTATGCC AAGTTACCGT AGGCTTCATT GTAGATAGTA CCGTGATATC

7351 CTTCGGCAGG TTGAGAGAAT TGAGGGAACT TTTCTGGGTA GTCCTTCCAT

7401 GGGAACTTAC CAGAGTCAAC AATAATACCA CCGATAGTAG TACCATGACC

7451 ACCAATCCAT TTGGTAGCAG AATGTGTTAC AATATCAGCA CCGTATTTAA

7501 TTGGCTGACA GAAGTAACCA CCGGCACCAA ATGTGTTGTC AACGACAACT

7551 GGAATACCGT GTTTGTGAGC AATTGCAACA ATTTTTTCAA AATCCGGAAC

7601 ATTGTACTTT GGATTACCAA TGGTTTCCAA ATAAACAGCC TTGGTTCTTT

7651 CATCAAAGAC CTTTTCGAAT TCTTCTGGAT TGTCACCTTC AACAAATCTA

7701 GCCTCGATAC CAAATCTTTT GAACGAGATT TTGAACTGGT TATAAGTACC

7751 ACCGTATAAG TAAGAAGTGG AAACGATGTT GTCACCAGTG TGTGCCAAAC

7801 CTTGGATGGC AAGGGTTTGA GCGGCTTGAC CGGAGGAAAC AGCCAAAGCA

7851 GCAGCACCAC CTTCTAAAGC AGCAATTCTT TCTTCCAAAA CATTACTGGT

7901 TGGGTTTTGG AAACGGGAAT AGACGTAACC TGGAACTTCT AGACCAAACA

7951 ATTGCGAACC ATGCTTAGAG TTTTCGAAAA CATAAGAAGT GGTGGCGTAA

```
8001  ATTGGTACAG CTCTGGATCT GTGAGCATTG TCACCAGGGT TCTCTTGGCC
8051  GGCGTGTAGT TGAACAGTAT CGAAATGAGA TGGCATTTTG TAATTAAAAC
8101  TTAGATTAGA TTGCTATGCT TTCTTTCTAA TGAGCAAGAA GTAAAAAAAG
8151  TTGTAATAGA ACAAGAAAAA TGAAACTGAA ACTTGAGAAA TTGAAGACCG
8201  TTTATTAACT TAAATATCAA TGGGAGGTCA TCGAAAGAGA AAAAAATCAA
8251  AAAAAAAAAT TTTCAAGAAA AAGAAACGTG ATAAAAATTT TTATTGCCTT
8301  TTTCGACGAA GAAAAAGAAA CGAGGCGGTC TCTTTTTTCT TTTCCAAACC
8351  TTTAGTACGG GTAATTAACG ACACCCTAGA GGAAGAAAGA GGGGAAATTT
8401  AGTATGCTGT GCTTGGGTGT TTTGAAGTGG TACGGCGATG CGCGGAGTCC
8451  GAGAAAATCT GGAAGAGTAA AAAAGGAGTA GAAACATTTT GAAGCTATAT
8501  TACCCTGTTA TCCCTAagat tgtactga
```

A.2 Sequence of the lycopene biosynthetic pathway

Sequence of the completed 8-round lycopene biosynthetic pathway, which is shown schematically in **Figure 5-3**. The minimal lycopene biosynthetic pathway (**Fig. 5-2**) consists of subfragments 1a through 3d. To show the context of the sequence in the chromosome, the last 30 bp of the *HO(L)* region on the 5' end and the first 30 bp of the acceptor module (containing the HO cleavage site and the *HIS3* terminator) on the 3' end are shown.

```
    1 . . 30  HO(L)
   31 . . 703 Subfragment 1a (pGPD)
  704 . . 1627 Subfragment 1b (crtE)
 1628 . . 1793 Subfragment 1c (tADH)
 1794 . . 2723 Subfragment 2a (crtB, reverse complement)
 2724 . . 3723 Subfragment 2b (pPGK, reverse complement)
 3724 . . 4135 Subfragment 3a (pGPD)
 4136 . . 5614 Subfragment 3b (crtI)
```

5615 . . 5869 Subfragment 3c (tACT)

5870 . . 6872 Subfragment 3d (TRP1)

6873 . . 7545 Subfragment 4a (pGPD)

7546 . . 9123 Subfragment 4b (tHMG1)

9124 . . 9289 Subfragment 4c (tADH)

9290 . . 10358 Subfragment 5a (ERG20, reverse complement)

10359 . . 10760 Subfragment 5b (pTEF, reverse complement)

10761 . . 11712 Subfragment 6a (pPGK)

11713 . . 13215 Subfragment 6b (ALD6)

13216 . . 13470 Subfragment 6c (tACT)

13471 . . 15396 Subfragment 6d (MET15)

15397 . . 15808 Subfragment 7a (pTEF)

15809 . . 17767 Subfragment 7b (acs)

17768 . . 17933 Subfragment 7c (tADH)

17934 . . 20675 Subfragment 8a (UPC2, reverse complement)

20676 . . 21675 Subfragment 8b (pPGK, reverse complement)

21676 . . 22705 Acceptor module

1 AAAATTGTGC CTTTGGACTT AAAATGGCGT AGTTTATCAT TATCAATACT CGCCATTTCA

61 AAGAATACGT AAATAATTAA TAGTAGTGAT TTTCCTAACT TTATTTAGTC AAAAAATTAG

121 CCTTTTAATT CTGCTGTAAC CCGTACATGC CCAAAATAGG GGGCGGGTTA CACAGAATAT

181 ATAACATCGT AGGTGTCTGG GTGAACAGTT TATTCCTGGC ATCCACTAAA TATAATGGAG

241 CCCGCTTTTT AAGCTGGCAT CCAGAAAAAA AAAGAATCCC AGCACCAAAA TATTGTTTTC

301 TTCACCAACC ATCAGTTCAT AGGTCCATTC TCTTAGCGCA ACTACAGAGA ACAGGGGCAC

361 AAACAGGCAA AAAACGGGCA CAACCTCAAT GGAGTGATGC AACCTGCCTG GAGTAAATGA

421 TGACACAAGG CAATTGACCC ACGCATGTAT CTATCTCATT TTCTTACACC TTCTATTACC

481 TTCTGCTCTC TCTGATTTGG AAAAAGCTGA AAAAAAAGGT TGAAACCAGT TCCCTGAAAT

541 TATTCCCCTA CTTGACTAAT AAGTATATAA AGACGGTAGG TATTGATTGT AATTCTGTAA

601 ATCTATTTCT TAAACTTCTT AAATTCTACT TTTATAGTTA GTCTTTTTTT TAGTTTTAAA

661 ACACCAGAAC TTAGTTTCGA ATAAACACAC ATAAACAAAC AAAATGGTTT CTGGTTCGAA

721 AGCAGGAGTA TCACCTCATA GGGAAATCGA AGTCATGAGA CAGTCCATTG ATGACCACTT

781 AGCAGGATTG TTGCCAGAAA CAGATTCCCA GGATATCGTT AGCCTTGCTA TGAGAGAAGG

841 TGTTATGGCA CCTGGTAAAC GTATCAGACC TTTGCTGATG TTACTTGCTG CAAGAGACCT

901 GAGATATCAG GGTTCTATGC CTACACTACT GGATCTAGCT TGTGCTGTTG AACTGACACA

961 TACTGCTTCC TTGATGCTGG ATGACATGCC TTGTATGGAC AATGCGGAAC TTAGAAGAGG

1021 TCAACCAACA ACCCACAAGA AATTCGGAGA ATCTGTTGCC ATTTTGGCTT CTGTAGGTCT

1081 GTTGTCGAAA GCTTTTGGCT TGATTGCTGC AACTGGTGAT CTTCCAGGTG AAAGGAGAGC

1141 ACAAGCTGTA AACGAGCTAT CTACTGCAGT TGGTGTTCAA GGTCTAGTCT TAGGACAGTT

1201 CAGAGATTTG AATGACGCAG CTTTGGACAG AACTCCTGAT GCTATCCTGT CTACGAACCA

1261 TCTGAAGACT GGCATCTTGT TCTCAGCTAT GTTGCAAATC GTAGCCATTG CTTCTGCTTC

1321 TTCACCATCT ACTAGGGAAA CGTTACACGC ATTCGCATTG GACTTTGGTC AAGCCTTTCA

1381 ACTGCTAGAC GATTTGAGGG ATGATCATCC AGAGACAGGT AAAGACCGTA ACAAAGACGC

1441 TGGTAAAAGC ACTCTAGTCA ACAGATTGGG TGCTGATGCA GCTAGACAGA AACTGAGAGA

1501 GCACATTGAC TCTGCTGACA AACACCTGAC ATTTGCATGT CCACAAGGAG GTGCTATAAG

1561 GCAGTTTATG CACCTATGGT TTGGACACCA TCTTGCTGAT TGGTCTCCAG TGATGAAGAT

1621 CGCCTAAGCG AATTTCTTAT GATTTATGAT TTTTATTATT AAATAAGTTA TAAAAAAAAT

1681 AAGTGTATAC AAATTTTAAA GTGACTCTTA GGTTTTAAAA CGAAAATTCT TATTCTTGAG

1741 TAACTCTTTC CTGTAGGTCA GGTTGCTTTC TCAGGTATAG CATGAGGTCG CTCTTAGACA

1801 GGTCTTTGCC ATAAACCAGC AGGTCTTGGT GTAACTCTGG TTGTTTTGGC ACGAATGACT

1861 TGTCCTGGTG CAGCCATAAG CATCGCAATC TTCTCTCCTT TGCTAGTGTG TTGACGTCTA

1921 TCCCAAGCAG AACCACCTGC AGCTTTTACC TTGATGCCAA TCTCTCTGTA GACAGATCTT

1981 GCAGTAGCTA TAGCCCAAGC ACATCTAGGT GGTAGATCAT GCAATCCAGC TTGACTTGAG

2041 ATGTAATAGG GTTCAGCAGC GTCTATCAGC CTTTCAGCAA CTCTTGCTAA TGCAGCTCTG

2101 TTCTCTCTTG CAGCGTAATT CTCAGGAGTT AGACCAGCAT CTTGCAACCA TTCAGCAGGT

2161 AGATAGCATC TGTCAATAGC TGCATCGTCG ATAATATCTC TCGCGATGTT TGTCAGCTGA

2221 AAAGCCAAAC CTAGATCACA AGCTCTGTCC AAAACCCTTT CGTCTCTTAC ACCCATTACT

2281 CTTGCCATCA TCAAACCAAC TACTCCAGCA ACATGGTAGC AGTATCTCAA GGTGTCTTCA

2341 AAGGTCACGT AACGAGTTTG AGCAACATCC ATTGCGAAAC CATCCAAGTG ATCAAGTGCC

2401 ATTCTTGGCG TAATACCGTG TGTTAGTGCA ACTTCTTGGA ATGCAGCAAA AGCAGGATCT

2461 TGCATCTCAG CACCTTCAAA AGCTGCAAGT GTAAGCGTTC TCAATCTAGC CAATCTCTGA

2521 GTAGCCTCTT CTTCTGCAGC AGCTTCAGAT GCGAAACCAT GTGTCTGGTC ATCTATAACG

2581 TCATCACAGT GTCTACACCA AGTGTACAGC ATAAGCACTG ATCTTCTAGT AGCTGGGTCA

2641 AACAGTTTAG CTGCTGTAGC AAAGGACTTG GAACCATTAG CCATCGTTTG AGTAGCATGA

2701 TCCAACAAAG GTGGTTGACT CATTGTTTTA TATTTGTTGT AAAAAGTAGA TAATTACTTC

2761 CTTGATGATC TGTAAAAAAG AGAAAAAGAA AGCATCTAAG AACTTGAAAA ACTACGAATT

2821 AGAAAAGACC AAATATGTAT TTCTTGCATT GACCAATTTA TGCAAGTTTA TATATATGTA

2881 AATGTAAGTT TCACGAGGTT CTACTAAACT AAACCACCCC CTTGGTTAGA AGAAAAGAGT

2941 GTGTGAGAAC AGGCTGTTGT TGTCACACGA TTCGGACAAT TCTGTTTGAA AGAGAGAGAG

3001 TAACAGTACG ATCGAACGAA CTTTGCTCTG GAGATCACAG TGGGCATCAT AGCATGTGGT

3061 ACTAAACCCT TTCCCGCCAT TCCAGAACCT TCGATTGCTT GTTACAAAAC CTGTGAGCCG

3121 TCGCTAGGAC CTTGTTGTGT GACGAAATTG GAAGCTGCAA TCAATAGGAA GACAGGAAGT

3181 CGAGCGTGTC TGGGTTTTTT CAGTTTTGTT CTTTTTGCAA ACAAATCACG AGCGACGGTA

3241 ATTTCTTTCT CGATAAGAGG CCACGTGCTT TATGAGGGTA ACATCAATTC AAGAAGGAGG

3301 GAAACACTTC CTTTTTCTGG CCCTGATAAT AGTATGAGGG TGAAGCCAAA ATAAAGGATT

3361 CGCGCCCAAA TCGGCATCTT TAAATGCAGG TATGCGATAG TTCCTCACTC TTTCCTTACT

3421 CACGAGTAAT TCTTGCAAAT GCCTATTATG CAGATGTTAT AATATCTGTG CGTCTTGAGT

3481 TGAAGTCAGG AATCTAAAAT AAAAATTAAG GTTAATAAAA AGAGGAAAGA AAAAAAAATT

3541 AATCGATTTA CAGAAACTTG CACACTAAAA ATACACAACT AAAAGCAATT ACAGTATGGG

3601 AAGTCATCGA CGTTATCTCT ACTATAGTAT ATTATCATTT CTATTATTAT CCTGCTCAGT

3661 GGTACTTGCA AAACAAGATA AGACCCCATT CTTTGAAGGT ACTTCTTCGA AAAATTCGCG

3721 TCTATAGCTT CAAAATGTTT CTACTCCTTT TTTACTCTTC CAGATTTTCT CGGACTCCGC

3781 GCATCGCCGT ACCACTTCAA AACACCCAAG CACAGCATAC TAAATTTCCC CTCTTTCTTC

3841 CTCTAGGGTG TCGTTAATTA CCCGTACTAA AGGTTTGGAA AAGAAAAAAG AGACCGCCTC

3901 GTTTCTTTTT CTTCGTCGAA AAAGGCAATA AAAATTTTTA TCACGTTTCT TTTTCTTGAA

3961 AATTTTTTTT TTTGATTTTT TTCTCTTTCG ATGACCTCCC ATTGATATTT AAGTTAATAA

4021 ACGGTCTTCA ATTTCTCAAG TTTCAGTTTC ATTTTTCTTG TTCTATTACA ACTTTTTTTA

4081 CTTCTTGCTC ATTAGAAAGA AAGCATAGCA ATCTAATCTA AGTTTTAATT ACAAAATGAA

4141 GAAAACCGTA GTGATTGGTG CAGGTTTTGG TGGTTTAGCT TTGGCTATAC GTCTACAAGC

4201 TGCAGGTATT CCTACAGTGC TATTGGAGCA AAGAGACAAA CCAGGAGGAA GAGCTTATGT

4261 TTGGCACGAT CAAGGCTTTA CCTTTGATGC TGGTCCTACA GTCATCACTG ATCCTACTGC

4321 ATTGGAAGCT TTGTTCACCT TAGCTGGTAG AAGAATGGAA GATTATGTCC GTCTATTGCC

4381 TGTCAAGCCG TTTTACAGAT TGTGTTGGGA ATCTGGTAAA ACCCTAGATT ACGCCAATGA

4441 CAGTGCTGAA CTAGAAGCTC AGATTACGCA GTTTAATCCC AGAGATGTCG AAGGTTACAG

4501 GAGATTCCTT GCCTATTCCC AAGCTGTTTT CCAAGAGGGT TATCTTCGTT TGGGTTCAGT

4561 TCCATTCCTG TCCTTTAGGG ATATGCTTAG AGCAGGTCCT CAGTTGTTGA AGCTACAAGC

4621 ATGGCAAAGT GTGTATCAGT CTGTTTCGAG ATTTATCGAG GATGAACATC TGAGACAAGC

4681 ATTCTCATTC CACAGTCTTC TAGTTGGAGG TAATCCCTTT ACCACATCGA GCATATATAC

4741 GTTGATTCAC GCTTTGGAAA GAGAATGGGG AGTTTGGTTT CCTGAAGGTG GAACAGGTGC

4801 TTTGGTTAAT GGTATGGTGA AGCTATTCAC GGATTTGGGT GGAGAAATAG AGCTGAATGC

4861 AAGAGTGGAA GAACTTGTTG TAGCAGACAA CAGAGTCTCA CAAGTTAGAC TTGCTGATGG

4921 TAGGATCTTC GATACAGATG CTGTAGCTTC AAACGCAGAT GTAGTGAACA CTTATAAAAA

4981 GTTGTTGGGA CATCATCCTG TTGGACAAAA GAGAGCAGCT GCTTTGGAGA GGAAATCTAT

5041 GAGCAACTCG TTGTTTGTCC TTTACTTTGG GCTGAATCAA CCACACTCAC AACTAGCTCA

5101 TCACACAATC TGCTTTGGTC CTAGATACAG AGAGCTGATA GATGAAATTT TCACTGGATC

5161 TGCTTTAGCA GACGATTTTT CCCTGTACTT GCATTCACCA TGTGTTACTG ATCCCTCTTT

5221 AGCACCACCT GGTTGTGCTA GCTTCTATGT ACTAGCACCT GTACCACATT TGGGTAATGC

5281 TCCATTAGAT TGGGCACAAG AAGGACCGAA ATTGAGGGAT AGGATCTTCG ACTATTTGGA

5341 AGAACGTTAC ATGCCAGGTT TGAGATCTCA GTTGGTTACA CAGAGGATAT TCACACCAGC

5401 TGATTTTCAT GATACTCTAG ATGCGCATTT AGGTAGCGCT TTTTCCATTG AGCCACTTTT

5461 GACGCAAAGT GCTTGGTTTA GACCACACAA CAGAGATTCT GACATTGCCA ATCTGTACCT

5521 AGTAGGTGCA GGAACTCATC CAGGAGCTGG TATTCCTGGA GTTGTAGCTT CTGCTAAAGC

5581 TACTGCTAGT CTGATGATCG AGGATTTGCA GTAATCTCTG CTTTTGTGCG CGTATGTTTA

5641 TGTATGTACC TCTCTCTCTA TTTCTATTTT TAAACCACCC TCTCAATAAA ATAAAAATAA

5701 TAAAGTATTT TTAAGGAAAA GACGTGTTTA AGCACTGACT TTATCTACTT TTTGTACGTT

5761 TTCATTGATA TAATGTGTTT TGTCTCTCCC TTTTCTACGA AAATTTCAAA AATTGACCAA

5821 AAAAAGGAAT ATATATACGA AAAACTATTA TATTTATATA TCATAGTGTA ACGACATTAC

5881 TATATATATA ATATAGGAAG CATTTAATAG ACAGCATCGT AATATATGTG TACTTTGCAG

5941 TTATGACGCC AGATGGCAGT AGTGGAAGAT ATTCTTTATT GAAAAATAGC TTGTCACCTT

6001 ACGTACAATC TTGATCCGGA GCTTTTCTTT TTTTGCCGAT TAAGAATTAA TTCGGTCGAA

6061 AAAAGAAAAG GAGAGGGCCA AGAGGGAGGG CATTGGTGAC TATTGAGCAC GTGAGTATAC

6121 GTGATTAAGC ACACAAAGGC AGCTTGGAGT ATGTCTGTTA TTAATTTCAC AGGTAGTTCT

6181 GGTCCATTGG TGAAAGTTTG CGGCTTGCAG AGCACAGAGG CCGCAGAATG TGCTCTAGAT

6241 TCCGATGCTG ACTTGCTGGG TATTATATGT GTGCCCAATA GAAAGAGAAC AATTGACCCG

6301 GTTATTGCAA GGAAAATTTC AAGTCTTGTA AAAGCATATA AAAATAGTTC AGGCACTCCG

6361 AAATACTTGG TTGGCGTGTT TCGTAATCAA CCTAAGGAGG ATGTTTTGGC TCTGGTCAAT

6421 GATTACGGCA TTGATATCGT CCAACTGCAT GGAGATGAGT CGTGGCAAGA ATACCAAGAG

6481 TTCCTCGGTT TGCCAGTTAT TAAAAGACTC GTATTTCCAA AAGACTGCAA CATACTACTC

6541 AGTGCAGCTT CACAGAAACC TCATTCGTTT ATTCCCTTGT TTGATTCAGA AGCAGGTGGG

6601 ACAGGTGAAC TTTTGGATTG GAACTCGATT TCTGACTGGG TTGGAAGGCA AGAGAGCCCC

6661 GAAAGCTTAC ATTTTATGTT AGCTGGTGGA CTGACGCCAG AAAATGTTGG TGATGCGCTT

6721 AGATTAAATG GCGTTATTGG TGTTGATGTA AGCGGAGGTG TGGAGACAAA TGGTGTAAAA

6781 GACTCTAACA AAATAGCAAA TTTCGTCAAA AATGCTAAGA AATAGGTTAT TACTGAGTAG

6841 TATTTATTTA AGTATTGTTT GTGCACTTGC CTAGTTTATC ATTATCAATA CTCGCCATTT

6901 CAAAGAATAC GTAAATAATT AATAGTAGTG ATTTTCCTAA CTTTATTTAG TCAAAAAATT

6961 AGCCTTTTAA TTCTGCTGTA ACCCGTACAT GCCCAAAATA GGGGGCGGGT TACACAGAAT

7021 ATATAACATC GTAGGTGTCT GGGTGAACAG TTTATTCCTG GCATCCACTA AATATAATGG

7081 AGCCCGCTTT TTAAGCTGGC ATCCAGAAAA AAAAAGAATC CCAGCACCAA AATATTGTTT

7141 TCTTCACCAA CCATCAGTTC ATAGGTCCAT TCTCTTAGCG CAACTACAGA GAACAGGGGC

7201 ACAAACAGGC AAAAAACGGG CACAACCTCA ATGGAGTGAT GCAACCTGCC TGGAGTAAAT

7261 GATGACACAA GGCAATTGAC CCACGCATGT ATCTATCTCA TTTTCTTACA CCTTCTATTA

7321 CCTTCTGCTC TCTCTGATTT GGAAAAAGCT GAAAAAAAAG GTTGAAACCA GTTCCCTGAA

7381 ATTATTCCCC TACTTGACTA ATAAGTATAT AAAGACGGTA GGTATTGATT GTAATTCTGT

7441 AAATCTATTT CTTAAACTTC TTAAATTCTA CTTTTATAGT TAGTCTTTTT TTTAGTTTTA

7501 AAACACCAGA ACTTAGTTTC GAATAAACAC ACATAAACAA ACAAAATGGA CCAATTGGTG

7561 AAAACTGAAG TCACCAAGAA GTCTTTTACT GCTCCTGTAC AAAAGGCTTC TACACCAGTT

7621 TTAACCAATA AAACAGTCAT TTCTGGATCG AAAGTCAAAA GTTTATCATC TGCGCAATCG

7681 AGCTCATCAG GACCTTCATC ATCTAGTGAG GAAGATGATT CCCGCGATAT TGAAAGCTTG

7741 GATAAGAAAA TACGTCCTTT AGAAGAATTA GAAGCATTAT TAAGTAGTGG AAATACAAAA

7801 CAATTGAAGA ACAAAGAGGT CGCTGCCTTG GTTATTCACG GTAAGTTACC TTTGTACGCT

7861 TTGGAGAAAA AATTAGGTGA TACTACGAGA GCGGTTGCGG TACGTAGGAA GGCTCTTTCA

7921 ATTTTGGCAG AAGCTCCTGT ATTAGCATCT GATCGTTTAC CATATAAAAA TTATGACTAC

7981 GACCGCGTAT TTGGCGCTTG TTGTGAAAAT GTTATAGGTT ACATGCCTTT GCCCGTTGGT

8041 GTTATAGGCC CCTTGGTTAT CGATGGTACA TCTTATCATA TACCAATGGC AACTACAGAG

8101 GGTTGTTTGG TAGCTTCTGC CATGCGTGGC TGTAAGGCAA TCAATGCTGG CGGTGGTGCA

8161 ACAACTGTTT TAACTAAGGA TGGTATGACA AGAGGCCCAG TAGTCCGTTT CCCAACTTTG

8221 AAAAGATCTG GTGCCTGTAA GATATGGTTA GACTCAGAAG AGGGACAAAA CGCAATTAAA

8281 AAAGCTTTTA ACTCTACATC AAGATTTGCA CGTCTGCAAC ATATTCAAAC TTGTCTAGCA

8341 GGAGATTTAC TCTTCATGAG ATTTAGAACA ACTACTGGTG ACGCAATGGG TATGAATATG

8401 ATTTCTAAAG GTGTCGAATA CTCATTAAAG CAAATGGTAG AAGAGTATGG CTGGGAAGAT

8461 ATGGAGGTTG TCTCCGTTTC TGGTAACTAC TGTACCGACA AAAAACCAGC TGCCATCAAC

8521 TGGATCGAAG GTCGTGGTAA GAGTGTCGTC GCAGAAGCTA CTATTCCTGG TGATGTTGTC

8581 AGAAAAGTGT TAAAAAGTGA TGTTTCCGCA TTGGTTGAGT TGAACATTGC TAAGAATTTG

8641 GTTGGATCTG CAATGGCTGG GTCTGTTGGT GGATTTAACG CACATGCAGC TAATTTAGTG

8701 ACAGCTGTTT TCTTGGCATT AGGACAAGAT CCTGCACAAA ATGTTGAAAG TTCCAACTGT

8761 ATAACATTGA TGAAAGAAGT GGACGGTGAT TTGAGAATTT CCGTATCCAT GCCATCCATC

8821 GAAGTAGGTA CCATCGGTGG TGGTACTGTT CTAGAACCCC AAGGTGCCAT GTTGGACTTA

8881 TTAGGTGTAA GAGGCCCGCA TGCTACCGCT CCTGGTACCA ACGCACGTCA ATTAGCAAGA

8941 ATAGTTGCCT GTGCCGTCTT GGCAGGTGAA TTATCCTTAT GTGCTGCCCT AGCAGCCGGC

9001 CATTTGGTTC AAAGTCATAT GACCCACAAC AGGAAACCTG CTGAACCAAC AAAACCTAAC

9061 AATTTGGACG CCACTGATAT AAATCGTTTG AAAGATGGGT CCGTCACCTG CATTAAATCC

9121 TAAGCGAATT TCTTATGATT TATGATTTTT ATTATTAAAT AAGTTATAAA AAAAATAAGT

9181 GTATACAAAT TTTAAAGTGA CTCTTAGGTT TTAAAACGAA AATTCTTATT CTTGAGTAAC

9241 TCTTTCCTGT AGGTCAGGTT GCTTTCTCAG GTATAGCATG AGGTCGCTCC TATTTGCTTC

9301 TCTTGTAAAC TTTGTTCAAG AACGCAGTTA AGACATCAGC TTTGAAGCCA CGAGACTCAT

9361 CGACCTGAGA AATTTTGGCC TTCAAATCCT TGGCAATAGA CTCTTCATAT TCGTGGTATA

9421 GCTGTTCAAT TTTCAAGTCA TTGAAAATCT TTTTGCATTT GGCTTCTGCG ACTGAGTCCT

9481 TCTTACCGTA ATTTTCGTCT AAAGTCTTTC TTTGTTCTGC GGAAGCAAGT TCCAATGCCT

9541 TGTTGATTAC CCAAGAACAT TTGTTATCTT GGATATCTGT ACCGATCTTA CCGATCTGTT

9601 CTGGGGTACC GAAGCAGTCT AAGTAGTCAT CTTGAATTTG GAAGTATTCA CCCAATGGAA

9661 TCAAGACATC TCTGGCTTGT TTCAAATCCT TTTCATCCGT GATACCGGCA ACGTACATGG

9721 CCAATGCGAC AGGCAAGTAG AAAGAATAGT AAGCAGTCTT GAAAGTAACT ATGAAGGAGT

9781 GCTTCTTTAG GGAGAACTTA CTCAAGTCGA CTTTGTCTTC AGGTGCAGTG ATTAAGTCCA

9841 TCAATTGGCC CAATTCGGTT TGGAAGGTGA CCTCATGGAA CAATTCGGTG ATATCTATGT

9901 AGTATTTTTC GTTTCTGAAG TGAGATTTCA AAAGCTTGTA GATAGCAGCC TCTAACATGA

9961 ATGCGTCATT GATGGCAATT TCCCCAACTT CAGGAACCTT GTACCAACAT GGTTGGCCTC

10021 TTCTGGTAAT GGACTTGTCC ATCATATCAT CGGCGACCAA GAAGTAAGCC TGCAACAACT

10081 CAATGCACCA ACCTAGAATG GCAACCTTTT CGTATTCTTC TTGCCCCAAT TGTTCAACGG

10141 TCTTGTTGGA GAGAATAGCA TACGTGTCCA CAACGGACAA ACCTCTATTT AGCTTACCGC

10201 CTGGAGTGTT GTAGTTCAAT GAGTGGGCAT ACCAGTCACA TGCTTCCTTA GGCATACCGT

10261 AAGCCAAAAG CGATGCGTTC AATTCCTCTA CTAATTTAGG GAAAACGTTC AAGAATCTCT

10321 CTCTCCTAAT TTCTTTTTCT GAAGCCATTT TGTAATTAAA ACTTAGATTA GATTGCTATG

10381 CTTTCTTTCT AATGAGCAAG AAGTAAAAAA AGTTGTAATA GAACAAGAAA AATGAAACTG

10441 AAACTTGAGA AATTGAAGAC CGTTTATTAA CTTAAATATC AATGGGAGGT CATCGAAAGA

10501 GAAAAAAATC AAAAAAAAAA ATTTTCAAGA AAAAGAAACG TGATAAAAAT TTTTATTGCC

10561 TTTTTCGACG AAGAAAAAGA AACGAGGCGG TCTCTTTTTT CTTTTCCAAA CCTTTAGTAC

10621 GGGTAATTAA CGACACCCTA GAGGAAGAAA GAGGGGAAAT TTAGTATGCT GTGCTTGGGT

10681 GTTTTGAAGT GGTACGGCGA TGCGCGGAGT CCGAGAAAAT CTGGAAGAGT AAAAAAGGAG

10741 TAGAAACATT TTGAAGCTAT TGTTTTGCAA GTACCACTGA GCAGGATAAT AATAGAAATG

10801 ATAATATACT ATAGTAGAGA TAACGTCGAT GACTTCCCAT ACTGTAATTG CTTTTAGTTG

10861 TGTATTTTTA GTGTGCAAGT TTCTGTAAAT CGATTAATTT TTTTTTCTTT CCTCTTTTTA

10921 TTAACCTTAA TTTTTATTTT AGATTCCTGA CTTCAACTCA AGACGCACAG ATATTATAAC

10981 ATCTGCATAA TAGGCATTTG CAAGAATTAC TCGTGAGTAA GGAAAGAGTG AGGAACTATC

11041 GCATACCTGC ATTTAAAGAT GCCGATTTGG GCGCGAATCC TTTATTTTGG CTTCACCCTC

11101 ATACTATTAT CAGGGCCAGA AAAAGGAAGT GTTTCCCTCC TTCTTGAATT GATGTTACCC

11161 TCATAAAGCA CGTGGCCTCT TATCGAGAAA GAAATTACCG TCGCTCGTGA TTTGTTTGCA

11221 AAAAGAACAA AACTGAAAAA ACCCAGACAC GCTCGACTTC CTGTCTTCCT ATTGATTGCA

11281 GCTTCCAATT TCGTCACACA ACAAGGTCCT AGCGACGGCT CACAGGTTTT GTAACAAGCA

11341 ATCGAAGGTT CTGGAATGGC GGGAAAGGGT TTAGTACCAC ATGCTATGAT GCCCACTGTG

11401 ATCTCCAGAG CAAAGTTCGT TCGATCGTAC TGTTACTCTC TCTCTTTCAA ACAGAATTGT

11461 CCGAATCGTG TGACAACAAC AGCCTGTTCT CACACACTCT TTTCTTCTAA CCAAGGGGGT

11521 GGTTTAGTTT AGTAGAACCT CGTGAAACTT ACATTTACAT ATATATAAAC TTGCATAAAT

11581 TGGTCAATGC AAGAAATACA TATTTGGTCT TTTCTAATTC GTAGTTTTTC AAGTTCTTAG

11641 ATGCTTTCTT TTTCTCTTTT TTACAGATCA TCAAGGAAGT AATTATCTAC TTTTTACAAC

11701 AAATATAAAA CAATGACTAA GCTACACTTT GACACTGCTG AACCAGTCAA GATCACACTT

11761 CCAAATGGTT TGACATACGA GCAACCAACC GGTCTATTCA TTAACAACAA GTTTATGAAA

11821 GCTCAAGACG GTAAGACCTA TCCCGTCGAA GATCCTTCCA CTGAAAACAC CGTTTGTGAG

11881 GTCTCTTCTG CCACCACTGA AGATGTTGAA TATGCTATCG AATGTGCCGA CCGTGCTTTC

11941 CACGACACTG AATGGGCTAC CCAAGACCCA AGAGAAAGAG GCCGTCTACT AAGTAAGTTG

12001 GCTGACGAAT TGGAAAGCCA AATTGACTTG GTTTCTTCCA TTGAAGCTTT GGACAATGGT

12061 AAAACTTTGG CCTTAGCCCG TGGGGATGTT ACCATTGCAA TCAACTGTCT AAGAGATGCT

12121 GCTGCCTATG CCGACAAAGT CAACGGTAGA ACAATCAACA CCGGTGACGG CTACATGAAC

12181 TTCACCACCT TAGAGCCAAT CGGTGTCTGT GGTCAAATTA TTCCATGGAA CTTTCCAATA

12241 ATGATGTTGG CTTGGAAGAT CGCCCCAGCA TTGGCCATGG GTAACGTCTG TATCTTGAAA

12301 CCCGCTGCTG TCACACCTTT AAATGCCCTA TACTTTGCTT CTTTATGTAA GAAGGTTGGT

12361 ATTCCAGCTG GTGTCGTCAA CATCGTTCCA GGTCCTGGTA GAACTGTTGG TGCTGCTTTG

12421 ACCAACGACC CAAGAATCAG AAAGCTGGCT TTTACCGGTT CTACAGAAGT CGGTAAGAGT

12481 GTTGCTGTCG ACTCTTCTGA ATCTAACTTG AAGAAAATCA CTTTGGAACT AGGTGGTAAG

12541 TCCGCCCATT TGGTCTTTGA CGATGCTAAC ATTAAGAAGA CTTTACCAAA TCTAGTAAAC

12601 GGTATTTTCA AGAACGCTGG TCAAATTTGT TCCTCTGGTT CTAGAATTTA CGTTCAAGAA

12661 GGTATTTACG ACGAACTATT GGCTGCTTTC AAGGCTTACT TGGAAACCGA AATCAAAGTT

12721 GGTAATCCAT TTGACAAGGC TAACTTCCAA GGTGCTATCA CTAACCGTCA ACAATTCGAC

12781 ACAATTATGA ACTACATCGA TATCGGTAAG AAAGAAGGCG CCAAGATCTT AACTGGTGGC

12841 GAAAAAGTTG GTGACAAGGG TTACTTCATC AGACCAACCG TTTTCTACGA TGTTAATGAA

12901 GACATGAGAA TTGTTAAGGA AGAAATTTTT GGACCAGTTG TCACTGTCGC AAAGTTCAAG

12961 ACTTTAGAAG AAGGTGTCGA AATGGCTAAC AGCTCTGAAT TCGGTCTAGG TTCTGGTATC

13021 GAAACAGAAT CTTTGAGCAC AGGTTTGAAG GTGGCCAAGA TGTTGAAGGC CGGTACCGTC

13081 TGGATCAACA CATACAACGA TTTTGACTCC AGAGTTCCAT TCGGTGGTGT TAAGCAATCT

13141 GGTTACGGTA GAGAAATGGG TGAAGAAGTC TACCATGCAT ACACTGAAGT AAAAGCTGTC

13201 AGAATTAAGT TGTAATCTCT GCTTTTGTGC GCGTATGTTT ATGTATGTAC CTCTCTCTCT

13261 ATTTCTATTT TTAAACCACC CTCTCAATAA AATAAAAATA ATAAAGTATT TTTAAGGAAA

13321 AGACGTGTTT AAGCACTGAC TTTATCTACT TTTTGTACGT TTTCATTGAT ATAATGTGTT

13381 TTGTCTCTCC CTTTTCTACG AAAATTTCAA AAATTGACCA AAAAAAGGAA TATATATACG

13441 AAAAACTATT ATATTTATAT ATCATAGTGT CGGATGCAAG GGTTCGAATC CCTTAGCTCT

13501 CATTATTTTT TGCTTTTTCT CTTGAGGTCA CATGATCGCA AAATGGCAAA TGGCACGTGA

13561 AGCTGTCGAT ATTGGGGAAC TGTGGTGGTT GGCAAATGAC TAATTAAGTT AGTCAAGGCG

13621 CCATCCTCAT GAAAACTGTG TAACATAATA ACCGAAGTGT CGAAAAGGTG GCACCTTGTC

13681 CAATTGAACA CGCTCGATGA AAAAAATAAG ATATATATAA GGTTAAGTAA AGCGTCTGTT

13741 AGAAAGGAAG TTTTTCCTTT TTCTTGCTCT CTTGTCTTTT CATCTACTAT TTCCTTCGTG

13801 TAATACAGGG TCGTCAGATA CATAGATACA ATTCTATTAC CCCCATCCAT ACAATGCCAT

13861 CTCATTTCGA TACTGTTCAA CTACACGCCG GCCAAGAGAA CCCTGGTGAC AATGCTCACA

13921 GATCCAGAGC TGTACCAATT TACGCCACCA CTTCTTATGT TTTCGAAAAC TCTAAGCATG

13981 GTTCGCAATT GTTTGGTCTA GAAGTTCCAG GTTACGTCTA TTCCCGTTTC CAAAACCCAA

14041 CCAGTAATGT TTTGGAAGAA AGAATTGCTG CTTTAGAAGG TGGTGCTGCT GCTTTGGCTG

14101 TTTCCTCCGG TCAAGCCGCT CAAACCCTTG CCATCCAAGG TTTGGCACAC ACTGGTGACA

14161 ACATCGTTTC CACTTCTTAC TTATACGGTG GTACTTATAA CCAGTTCAAA ATCTCGTTCA

14221 AAAGATTTGG TATCGAGGCT AGATTTGTTG AAGGTGACAA TCCAGAAGAA TTCGAAAAGG

14281 TCTTTGATGA AAGAACCAAG GCTGTTTATT TGGAAACCAT TGGTAATCCA AAGTACAATG

14341 TTCCGGATTT TGAAAAAATT GTTGCAATTG CTCACAAACA CGGTATTCCA GTTGTCGTTG

14401 ACAACACATT TGGTGCCGGT GGTTACTTCT GTCAGCCAAT TAAATACGGT GCTGATATTG

14461 TAACACATTC TGCTACCAAA TGGATTGGTG GTCATGGTAC TACTATCGGT GGTATTATTG

14521 TTGACTCTGG TAAGTTCCCA TGGAAGGACT ACCCAGAAAA GTTCCCTCAA TTCTCTCAAC

14581 CTGCCGAAGG ATATCACGGT ACTATCTACA ATGAAGCCTA CGGTAACTTG GCATACATCG

14641 TTCATGTTAG AACTGAACTA TTAAGAGATT TGGGTCCATT GATGAACCCA TTTGCCTCTT

14701 TCTTGCTACT ACAAGGTGTT GAAACATTAT CTTTGAGAGC TGAAAGACAC GGTGAAAATG

14761 CATTGAAGTT AGCCAAATGG TTAGAACAAT CCCCATACGT ATCTTGGGTT TCATACCCTG

14821 GTTTAGCATC TCATTCTCAT CATGAAAATG CTAAGAAGTA TCTATCTAAC GGTTTCGGTG

14881 GTGTCTTATC TTTCGGTGTA AAAGACTTAC CAAATGCCGA CAAGGAAACT GACCCATTCA

14941 AACTTTCTGG TGCTCAAGTT GTTGACAATT TAAAGCTTGC CTCTAACTTG GCCAATGTTG

15001 GTGATGCCAA GACCTTAGTC ATTGCTCCAT ACTTCACTAC CCACAAACAA TTAAATGACA

15061 AAGAAAAGTT GGCATCTGGT GTTACCAAGG ACTTAATTCG TGTCTCTGTT GGTATCGAAT

15121 TTATTGATGA CATTATTGCA GACTTCCAGC AATCTTTTGA AACTGTTTTC GCTGGCCAAA

15181 AACCATGAGT GTGCGTAATG AGTTGTAAAA TTATGTATAA ACCTACTTTC TCTCACAAGT

15241 ACTATACTTT TATAAAACGA ACTTTATTGA AATGAATATC CTTTTTTTCC CTTGTTACAT

15301 GTCGTGACTC GTACTTTGAA CCTAAATTGT TCTAACATCA AAGAACAGTG TTAATTCGCA

15361 GTCGAGAAGA AAAATATGGT GAACAAGACT CATCTAATAG CTTCAAAATG TTTCTACTCC

15421 TTTTTTACTC TTCCAGATTT TCTCGGACTC CGCGCATCGC CGTACCACTT CAAAACACCC

15481 AAGCACAGCA TACTAAATTT CCCCTCTTTC TTCCTCTAGG GTGTCGTTAA TTACCCGTAC

15541 TAAAGGTTTG GAAAAGAAAA AAGAGACCGC CTCGTTTCTT TTTCTTCGTC GAAAAAGGCA

15601 ATAAAAATTT TTATCACGTT TCTTTTTCTT GAAAATTTTT TTTTTTGATT TTTTTCTCTT

15661 TCGATGACCT CCCATTGATA TTTAAGTTAA TAAACGGTCT TCAATTTCTC AAGTTTCAGT

15721 TTCATTTTTC TTGTTCTATT ACAACTTTTT TTACTTCTTG CTCATTAGAA AGAAAGCATA

15781 GCAATCTAAT CTAAGTTTTA ATTACAAAAT GAGCCAAACA CATAAACACG CCATTCCCGC

15841 CAACATTGCG GATCGTTGCC TGATAAATCC AGAGCAGTAT GAGACTAAAT ATAAACAGTC

15901 TATTAACGAC CCCGATACGT TTTGGGGCGA ACAGGGAAAA ATTCTCGATT GGATCACGCC

15961 GTACCAAAAA GTGAAAAACA CCTCCTTTGC GCCAGGCAAT GTGTCGATTA AATGGTACGA

16021 GGACGGCACG CTGAATCTGG CGGCGAACTG TCTTGACCGC CATTTGCAGG AAAATGGCGA

16081 TCGCACCGCC ATTATCTGGG AAGGCGATGA CACGTCGCAG AGTAAACATA TCTCTTATCG

16141 CGAACTGCAT CGCGATGTCT GCCGTTTCGC GAATACGCTG CTGGATCTGG GCATTAAAAA

16201 AGGCGATGTG GTAGCGATTT ATATGCCGAT GGTGCCGGAA GCGGCGGTGG CAATGTTGGC

16261 CTGCGCCCGC ATCGGTGCGG TGCATTCGGT GATCTTCGGG GGCTTCTCGC CGGAAGCCGT

16321 CGCCGGACGC ATTATCGACT CCAGCTCGCG GCTGGTGATC ACCGCTGACG AAGGTGTACG

16381 TGCCGGACGC AGTATCCCGC TGAAAAAGAA TGTCGATGAC GCGCTGAAAA ACCCGAATGT

16441 CACTAGCGTT GAGCATGTGA TCGTCCTGAA GCGCACCGGC AGCGACATTG ACTGGCAAGA

16501 AGGCCGCGAC CTGTGGTGGC GCGATTTGAT TGAAAAAGCC AGCCCTGAGC ACCAGCCTGA

16561 AGCGATGAAT GCCGAAGATC CGCTGTTTAT CCTTTATACC TCCGGCTCCA CCGGCAAGCC

16621 GAAAGGCGTG CTGCACACCA CCGGCGGCTA TCTGGTCTAC GCCGCGACCA CCTTTAAGTA

16681 TGTCTTTGAT TATCACCCTG GCGATATTTA CTGGTGTACC GCCGATGTGG GTTGGGTGAC

16741 GGGGCACAGC TATCTGTTGT ATGGCCCGCT GGCCTGCGGC GCGACCACCT TAATGTTTGA

16801 AGGCGTGCCG AATTGGCCAA CGCCCGCTCG CATGTGCCAG GTGGTCGACA AACACCAGGT

16861 CAACATTCTC TATACCGCCC CGACGGCCAT CCGCGCGCTG ATGGCGGAAG GCGATAAAGC

16921 CATTGAAGGC ACCGACCGTT CTTCACTGCG CATTCTGGGT TCCGTCGGCG AGCCGATCAA

16981 TCCCGAAGCG TGGGAATGGT ACTGGAAGAA GATCGGCAAG GAAAAATGTC CGGTCGTCGA

17041 CACCTGGTGG CAGACTGAAA CAGGCGGTTT TATGATCACG CCGCTACCAG GCGCTATCGA

17101 ACTGAAAGCC GGTTCCGCCA CCCGTCCTTT CTTTGGCGTA CAGCCTGCGC TGGTGGATAA

17161 CGAAGGCCAT CCGCAAGAAG GCGCGACGGA AGGCAATCTG GTCATCACCG ATTCCTGGCC

17221 GGGCCAGGCG CGCACTCTGT TCGGCGATCA TGAACGTTTT GAGCAGACCT ATTTCTCTAC

17281 CTTTAAGAAT ATGTATTTCA GCGGCGACGG CGCGCGTCGC GATGAGGACG GCTATTACTG

17341 GATCACCGGT CGCGTGGACG ACGTGTTAAA CGTCTCCGGC CACCGTCTGG GTACGGCGGA

17401 AATCGAGTCA GCGCTGGTGG CGCATCCGAA GATCGCCGAA GCGGCGGTGG TGGGTATTCC

17461 ACACGCTATC AAAGGCCAGG CGATTTACGC TTATGTGACG CTCAACCACG GCGAGGAGCC

17521 GTCGCCAGAA CTGTACGCGG AGGTGCGCAA CTGGGTACGT AAAGAGATTG GCCCACTGGC

17581 GACGCCGGAC GTGCTGCACT GGACCGACTC ACTGCCAAAA ACCCGTTCCG GCAAAATTAT

17641 GCGCCGCATT TTGCGCAAAA TCGCGGCGGG CGATACCAGC AATCTGGGCG ATACCTCGAC

17701 TCTCGCCGAT CCTGGCGTGG TGGAGAAACC GCTCGAAGAG AAGCAGGCCA TCGCGATGCC

17761 GTCATAAGCG AATTTCTTAT GATTTATGAT TTTTATTATT AAATAAGTTA TAAAAAAAAT

17821 AAGTGTATAC AAATTTTAAA GTGACTCTTA GGTTTTAAAA CGAAAATTCT TATTCTTGAG

17881 TAACTCTTTC CTGTAGGTCA GGTTGCTTTC TCAGGTATAG CATGAGGTCG CTCTCATAAC

17941 GAAAAATCAG AGAAATTTGT TGTTGTCATC GATGGTAATC CGCCACCGAG GAAATCTAGC

18001 ATCATATGCA TATCACCACC TCCACTGTAT TCGTCAACAT CTTGAGGCAG CACCTGCGTG

18061 ACTCCTTCGA GAAACCAGAC TTTATCCTTG ACCTCTGTGG CAAATCCTCG AAGTAGTTTA

18121 TAATATGATC TCATAATTCT CATTGCACCT AAATCACCTG TCATCAGTAA TGCCAGGAAT

18181 GTCTTGTCTA GCAATGCTGG AAATGTAAAT ACCCGCAGAA TAAAATCACC CTGGTTTTTT

18241 TCACGGTGCA ATTTATCTAA ATAAGCTAGT GTTATCAAAT ATGGCGAATC TAAGCCGACA

18301 GGATACAAAT CGGCAATACT TTCATCAAAA CATACCAATT CAGTAATTGT TCCAACATCA

18361 GGGTTAATGA CATCGCCTAA ATCGCTAAGA TCAACAGATA TAATGTTATG AAATTTAGAT

18421 CTTTCACTCA AAGGCCACAC AGCGGTTAAA ATTGTTGCAG CACCTTTGAC ATGAAAGATC

18481 CAAGCGCTTG GTGACATGCT ATTCAAACTT TGGTTTCCTA CAGTGCCGTT ACCACTAGCA

18541 TTTGCTAACG AGTCCATGAT TAGTATCAGG GCGCTGGCAA CTAGCGCATC GGTGTTATTC

18601 TCAGATATTT CTAAAACAGC TTCTCTTAAT AATCTCAGAG CGTCTAGGCG GTGAGATGAA

18661 ACGTATTGCT CCAGTCCAGT TTCAGTCCTC GAAAGATGGG TGGCACTGAA TGCCAACAAA

18721 GCGTGCATTA AAAATGGATA GTCAAATGCT AACTCCGGTA TGTAGTCCCT CCATATTTCA

18781 GGCCCAGAAA CTTTGGCCGC TGTAATCGTA GGCCAGACCT TTGTGCAATA ATGATGAAAC

18841 AGTTTCATGT CTATCAGGTT CAGATTGCCT TTGGTAGAAA GATCTAAAAG CTTAGCGATG

18901 TTACTGGTAC CAGCAGAAAC AGAACTTCGG TCAGCTTCGT TATTAATTTT TGAGGTCATG

18961 TCTTGCCGCG ATGCATCGTC TTCATTCTCT GTCTGTGATG TTTGTAAACT TATCGTTGGT

19021 TCGTTGTTCT CCAAGGCTGC TTTTGTCAGA TTTCCTAGAG AACTTATAGA ATCTTGCATC

19081 TCTTGCCTAT CATTCAACAC TGCAGAGGGC GTTGATGGAG AAATGTTGCT GTTTCTGTTC

19141 ATGTTTCTTT CAATCATTGA TATAGGCGCC ATTGTCGAAA CCGCATCAGC CGCCTGAGGT

19201 TCCATAATAC TAGCACTTCC CCCTTGGGCG TTATTCGATG GCGCTGGACT CTTCGTATCA

19261 GAGAGCTGAC CATCATTCTC CTTTACAGAA GCAGCTGCAC GTTCATGATG CTCTTGCATA

19321 TTGGCCAAAG CTTCCTCTGC GTTGCTGGCC TTAGTAGCTC TTGAATTGCC GTTATTTATT

19381 GGAAACTTAA TACCTAACAG TTCCTGAAAG TCATAGGAAA CACCACCAAT ACCAGCTGTT

19441 GGAAAGGTAC TTAACGACTT TAGGCTTAGC CCCATCTTAC TTAACTGAGA TAATGCATCA

19501 GCTTGTAATG TAGCAGCATC GCTGTTTAAA GTAGTTTCGC TTGTTTGTGA TTGTAAAGGC

19561 CCTTCCTTTA CCTTTTTGCT GTTTTCTTCC TGAACGCCAC CTTCCTCATC GTTAGGTTGT

19621 TGCTCTTGCT GAAGCAGTAC TTGTTGCTGC TGTTGCAAAA AATGTTGTTG GTATTGCTGG

19681 AGTTGTTCCT GTTGTTGTTG CTGAACTTGT TGGTGTTGTT CCTGACGTAA TTGTTGATAC

19741 TGCTGCAGCT GAACTTGCTG ATGCTGTTGC AATTGCAGCT GCTGGTGTAA TTGATATTGT

19801 AATTGTTGTT GAGTGTAGTT TTGGCCGGAA GGAATCCTGT TACATAGTAA CATCCCGCTT

19861 AGACTATTTG ATAAGCTCAT TCCCGTGGAT TGCAATGCAG GACTACCCGC AAGATCGAAT

19921 GAACCATATC TCTCGTTTGG AGTTCCACTT GCTTGCTGCG CTTGGGTCTG TGGTTGAGAT

19981 TGCTGCTGCA CTTGCTGCTG TTGCTGAAGT TGGAACGCAC CCATATTTCC AGAAGGCAAT

20041 GGCGGTAATA GTAAGCCATC GATACCTGAA TTCAATCTTT CAGGAGAACC AAAACCAATA

20101 GTGCCATTTT GAGAGAAAAA GTTATTCATC ATATCTTTCT TATTGATGGT ACCCTGGAGG

20161 CCCATACTTA GAGGAGTCAT ATTGTTTTTG TCCTCTATAG GGGCCGATGA CTTTTCCTCG

20221 CTCTCACTCT TAGTTGTAGA GGCGCTTGAT GGAAAGGAAT CATTGGATGA GCCTGAAGCT

20281 TTTACCGCTG ATTGGATATC ATTGAACGGT GTCTGCTCCT TCTTGATCGT AGGAGGTAAA

20341 TCTACCGATG AATCAGACTC CACGCTACCG TCTGCCTTTC TCGTGACATA CTTCACTACT

20401 GTTGCTCCTC TACCTTTCCT TAAATGGATT GGTGTATACT GACATTCCAA CTTCATATTT

20461 GTGCACTTCC TACAGGCTGG CTTCCCTTCA TCACACTTAA CTCTTCTTCT TTTACAGTTA

20521 TCGCACCCAT TCTTTGATTT GTTATGGAAT TTACGTTTAC CGGTTGAAGT CGTACTCACC

20581 TTTTTGCCGT CCACTTCAAT TAGCTCGATG ACTTTTTCTC TTCTTCTGGG TTTTGTCACC

20641 GCTTTCTTGT GATTCTGTAT ACCGACTTCG CTCATTGTTT TATATTTGTT GTAAAAAGTA

20701 GATAATTACT TCCTTGATGA TCTGTAAAAA AGAGAAAAAG AAAGCATCTA AGAACTTGAA

20761 AAACTACGAA TTAGAAAAGA CCAAATATGT ATTTCTTGCA TTGACCAATT TATGCAAGTT

20821 TATATATATG TAAATGTAAG TTTCACGAGG TTCTACTAAA CTAAACCACC CCCTTGGTTA

20881 GAAGAAAAGA GTGTGTGAGA ACAGGCTGTT GTTGTCACAC GATTCGGACA ATTCTGTTTG

20941 AAAGAGAGAG AGTAACAGTA CGATCGAACG AACTTTGCTC TGGAGATCAC AGTGGGCATC

21001 ATAGCATGTG GTACTAAACC CTTTCCCGCC ATTCCAGAAC CTTCGATTGC TTGTTACAAA

21061 ACCTGTGAGC CGTCGCTAGG ACCTTGTTGT GTGACGAAAT TGGAAGCTGC AATCAATAGG

21121 AAGACAGGAA GTCGAGCGTG TCTGGGTTTT TTCAGTTTTG TTCTTTTTGC AAACAAATCA

21181 CGAGCGACGG TAATTTCTTT CTCGATAAGA GGCCACGTGC TTTATGAGGG TAACATCAAT

21241 TCAAGAAGGA GGGAAACACT TCCTTTTTCT GGCCCTGATA ATAGTATGAG GGTGAAGCCA

21301 AAATAAAGGA TTCGCGCCCA AATCGGCATC TTTAAATGCA GGTATGCGAT AGTTCCTCAC

21361 TCTTTCCTTA CTCACGAGTA ATTCTTGCAA ATGCCTATTA TGCAGATGTT ATAATATCTG

21421 TGCGTCTTGA GTTGAAGTCA GGAATCTAAA ATAAAAATTA AGGTTAATAA AAAGAGGAAA

```
21481 GAAAAAAAAA TTAATCGATT TACAGAAACT TGCACACTAA AAATACACAA CTAAAAGCAA
21541 TTACAGTATG GGAAGTCATC GACGTTATCT CTACTATAGT ATATTATCAT TTCTATTATT
21601 ATCCTGCTCA GTGGTACTTG CAAAACAAGA TAAGACCCCA TTCTTTGAAG GTACTTCTTC
21661 GAAAAATTCG CGTCTTTTCA GCTTTCCGCA ACAGTATAAC TGTGC
```

A.3 Sequence of the *E. coli* carotenoid biosynthetic pathway

Sequence of the completed pathway for the biosynthesis of tetradehydrolycopene in *E. coli*, which is shown schematically in **Figure 5-6**. To show the context of the sequence in the chromosome, the last 30 bp of the *HO(L)* region on the 5' end and the first 30 bp of the acceptor module (containing the HO cleavage site and the *HIS3* terminator) on the 3' end are shown.

```
   1 . . 30   HO(L)
  31 . . 167  Subfragment 1a (pLac)
 168 . . 1082 Subfragment 1b (crtE)
1083 . . 1993 Fragment 2 (crtB)
1994 . . 3491 Fragment 3 (crtI14)
3492 . . 4920 Fragment 4 (KanMX)
4921 . . 4950 Acceptor module
```

```
   1 aaaattgtgc ctttggactt aaaatggcgt ccgactggaa agcgggcagt gagcgcaacg
  61 caattaatgt gagttagctc actcattagg caccccaggc tttacacttt atgcttccgg
 121 ctcgtatgtt gtgtggaatt gtgagcgtct agaaggagga ttacaaaatg acggtctgcg
 181 caaaaaaaca cgttcatctc actcgcgatg ctgcggagca gttactggct gatattgatc
 241 gacgccttga tcagttattg cccgtggagg gagaacggga tgttgtgggt gccgcgatgc
 301 gtgaaggtgc gctggcaccg ggaaaacgta ttcgccccat gttgctgttg ctgaccgccc
 361 gcgatctggg ttgcgctgtc agccatgacg gattactgga tttggcctgt gcggtggaaa
 421 tggtccacgc ggcttcgctg atccttgacg atatgccctg catggacgat gcgaagctgc
```

```
 481 ggcgcggacg ccctaccatt cattctcatt acggagagca tgtggcaata ctggcggcgg
 541 ttgccttgct gagtaaagcc tttggcgtaa ttgccgatgc agatggcctc acgccgctgg
 601 caaaaaatcg ggcggtttct gaactgtcaa acgccatcgg catgcaagga ttggttcagg
 661 gtcagttcaa ggatctgtct gaaggggata agccgcgcag cgctgaagct attttgatga
 721 cgaatcactt taaaaccagc acgctgtttt gtgcctccat gcagatggcc tcgattgttg
 781 cgaatgcctc cagcgaagcg cgtgattgcc tgcatcgttt ttcacttgat cttggtcagg
 841 catttcaact gctggacgat ttgaccgatg gcatgaccga caccggtaag gatagcaatc
 901 aggacgccgg taaatcgacg ctggtcaatc tgttaggccc gagggcggtt gaagaacgtc
 961 tgagacaaca tcttcagctt gccagtgagc atctctctgc ggcctgccaa cacgggcacg
1021 ccactcaaca ttttattcag gcctggtttg acaaaaaact cgctgccgtc agttaaggat
1081 gcaggaggat tacaaaatgg cagttggctc gaaaagtttt gcgacagcct caaagttatt
1141 tgatgcaaaa acccggcgca gcgtactgat gctctacgcc tggtgccgcc attgtgacga
1201 tgttattgac gatcagacgc tgggctttca ggcccggcag cctgccttac aaacgcccga
1261 acaacgtctg atgcaacttg agatgaaaac gcgccaggcc tatgcaggat cgcagatgca
1321 cgaaccggcg tttgcggctt ttcaggaagt ggctatggct catgatatcg ccccggctta
1381 cgcgtttgat catctggaag gcttcgccat ggatgtacgc gaagcgcaat acagccaact
1441 ggatgatacg ctgcgctatt gctatcacgt tgcaggcgtt gtcggcttga tgatggcgca
1501 aatcatgggc gtgcgggata acgccacgct ggaccgcgcc tgtgaccttg ggctggcatt
1561 tcagttgacc aatattgctc gcgatattgt ggacgatgcg catgcgggcc gctgttatct
1621 gccggcaagc tggctggagc atgaaggtct gaacaaagag aattatgcgg cacctgaaaa
1681 ccgtcaggcg ctgagccgta tcgcccgtcg tttggtgcag gaagcagaac cttactattt
1741 gtctgccaca gccggcctgg cagggttgcc cctgcgttcc gcctgggcaa tcgctacggc
1801 gaagcaggtt taccggaaaa taggtgtcaa agttgaacag gccggtcagc aagcctggga
1861 tcagcggcag tcaacgacca cgcccgaaaa attaacgctg ctgctggccg cctctggtca
1921 ggcccttact tcccggatgc gggctcatcc tccccgccct gcgcatctct ggcagcgccc
1981 gctctagctc gagaggagga ttacaaaatg aaaccaacta cggtaattgg tgcaggcttc
2041 ggtggcctgg cactggcaat tcgtctacaa gctgcgggga tccccgtctt actgcttgaa
2101 caacgtgata aacccggcgg tcgggcttat gtctacgagg atcaggggtt tacctttgat
2161 gcaggcccga cggttatcac cgatcccagt gccattgaag aactgtttgc actggcagga
```

2221 aaacagttaa aagagtatgt cgaactgctg ccggttacgc cgttttaccg cctgtgttgg

2281 gagtcaggga aggtctttaa ttacgataac gatcaaaccc ggctcgaagc gcagattcag

2341 cagtttaatc cccgcgatgt cgaaggttat cgtcagtttc tggactattc acgcgcggtg

2401 tttaaagaag gctatctaaa gctcggtact gtccctttt tatcgttcag agacatgctt

2461 cgcgccgcac ctcaactggc gaaactgcag gcatggagaa gcgtttacag taaggttgcc

2521 agttacatcg aagatgaaca tctgcgccag gcgttttctt tccactcgct gttggtgggc

2581 ggcaatccct tcgccacctc atccatttat acgttgatac acgcgctgga gcgtgagtgg

2641 ggcgtctggt ttccgcgtgg cggcaccggc gcattagttc aggggatgat aaagctgttt

2701 caggatctgg gtggcgaagt cgtgttaaac gccagagtca gccatatgga aacgacagga

2761 aacaagattg aagccgtgca tttagaggac ggtcgcaggt tcctgacgca agccgtcgcg

2821 tcaaatgcag atgtggttca tacctatcgc gacctgttaa gccagcaccc tgccgcggtt

2881 aagcagtcca acaaactgca gactaagcgc atgagtaact ctctgtttgt gctctatttt

2941 ggtttgaatc accatcatga tcagctcgcg catcacacgg tttgtttcgg cccgcgttac

3001 cgcgagctga ttgacgaaat ttttaatcat gatggcctcg cagaggactt ctcactttat

3061 ctgcacgcgc cctgtgtcac ggattcgtca ctggcgcctg aaggttgcgg cagttactat

3121 gtgttggcgc cggtgccaca tttaggcacc gcgaacctcg actggacggt tgaggggcca

3181 aaactacgcg accgtatttt tgcgtacctt gagcagcatt acatgcctgg cttacggagt

3241 cagctggtca cgcaccggat gtttacgccg tttgattttc gcgaccagct taatgcctat

3301 catggctcag ccttttctgt ggagcccgtt cttacccaga gcgcctggtt tcggccgcat

3361 aaccgcgata aaaccattac taatctctac ctggtcggcg caggcacgca tcccggcgca

3421 ggcattcctg gcgtcatcgg ctcggcaaaa gcgacagcag gtttgatgct ggaggatctg

3481 atttgagaat tagcttgcct cgtccccgcc gggtcacccg gccagcgaca tggaggccca

3541 gaataccctc cttgacagtc ttgacgtgcg cagctcaggg gcatgatgtg actgtcgccc

3601 gtacatttag cccatacatc cccatgtata atcatttgca tccatacatt ttgatggccg

3661 cacggcgcga agcaaaaatt acggctcctc gctgcagacc tgcgagcagg gaaacgctcc

3721 cctcacagac gcgttgaatt gtccccacgc cgcgcccctg tagagaaata taaaaggtta

3781 ggatttgcca ctgaggttct tctttcatat acttcctttt aaaatcttgc taggatacag

3841 ttctcacatc acatccgaac ataaacaacc atgggtaagg aaaagactca cgtttcgagg

3901 ccgcgattaa attccaacat ggatgctgat ttatatgggt ataaatgggc tcgcgataat

3961 gtcgggcaat caggtgcgac aatctatcga ttgtatggga agcccgatgc gccagagttg

4021 tttctgaaac atggcaaagg tagcgttgcc aatgatgtta cagatgagat ggtcagacta

4081 aactggctga cggaatttat gcctcttccg accatcaagc attttatccg tactcctgat

4141 gatgcatggt tactcaccac tgcgatcccc ggcaaaacag cattccaggt attagaagaa

4201 tatcctgatt caggtgaaaa tattgttgat gcgctggcag tgttcctgcg ccggttgcat

4261 tcgattcctg tttgtaattg tccttttaac agcgatcgcg tatttcgtct cgctcaggcg

4321 caatcacgaa tgaataacgg tttggttgat gcgagtgatt ttgatgacga gcgtaatggc

4381 tggcctgttg aacaagtctg gaaagaaatg cataagcttt tgccattctc accggattca

4441 gtcgtcactc atggtgattt ctcacttgat aaccttattt ttgacgaggg gaaattaata

4501 ggttgtattg atgttggacg agtcggaatc gcagaccgat accaggatct tgccatccta

4561 tggaactgcc tcggtgagtt ttctccttca ttacagaaac ggctttttca aaaatatggt

4621 attgataatc ctgatatgaa taaattgcag tttcatttga tgctcgatga gtttttctaa

4681 tcagtactga caataaaaag attcttgttt tcaagaactt gtcatttgta tagttttttt

4741 atattgtagt tgttctattt taatcaaatg ttagcgtgat ttatattttt tttcgcctcg

4801 acatcatctg cccagatgcg aagttaagtg cgcagaaagt aatatcatgc gtcaatcgta

4861 tgtgaatgct ggtcgctata ctgctgtcga ttcgatacta acgccgccat ccagtgtcga

4921 tttcagcttt ccgcaacagt ataactgtgc