A Multiplex CRISPR-Mediated Yeast Engineering Platform for the Expression of Complex Biosynthetic Pathways

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Franklin Leyang Gong

December 2020

ABSTRACT

Natural products (NPs) and their derivatives play an important role in modern healthcare as frontline therapeutics for many diseases and as commodity chemicals. Despite their ubiquity, NPs are historically challenging to produce. Many high-value metabolites are produced in nature by organisms that are not ideal for large-scale production. Therefore, interest exists in expressing the biosynthetic pathways that create these compounds in organisms that are more suitable for industrial production, a key example being the yeast *Saccharomyces cerevisiae*. The explosion of genomic sequence data and the significant advancements in synthetic biology have led to developments in both the discovery of new natural products as well as their corresponding biosynthetic pathways. However, challenges remain in the development of metabolic engineering tools that allow for the reconstruction of complex pathways in new host microorganisms. Every natural product is synthesized from the core metabolism by a large network of enzymes, whose coding genes must be integrated and stably expressed within the new host organism. The integration of these complex genetic pathways has traditionally formed a major bottleneck in the creation of NP-producing microbes. We report a biosynthetic pathway reconstruction platform based on the type II Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)–CRISPR associated proteins (Cas) system to generate multiple gene disruptions simultaneously in *S. cerevisiae* to allow for the homology-based recombination of exogenous gene fragments into the yeast chromosome. The results presented here demonstrate a one-step system that can insert specific genes into targeted sites of the yeast genome, allowing for the expression of large biosynthetic pathways.

BIOGRAPHICAL SKETCH

Franklin Leyang Gong was born May 23, 1996 and raised in Cookeville, Tennessee. He and his sister, Lisa Gong, are the only children of Mei Hu and Shaowei Gong. After graduating from Cookeville High School in 2014, Franklin attended Vanderbilt University and in May of 2018, he graduated Magna cum Laude with a Bachelor of Engineering in Chemical Engineering. He began pursing his Ph.D. at Cornell university shortly after, working in the group of Dr. Sijin Li.

This work is dedicated to my family for their support throughout my education

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Sijin Li, for taking me as a student, and for guiding me along this journey. I am deeply grateful for what she has contributed to my scientific development, providing constant support and opportunities. Under her patience and guidance, I have become more mature as a scientist and as a person. She is a great mentor, and none of this would have been possible without her. I would also like to thank my committee members, Dr. Jeff Varner and Dr. Buz Barstow, for following my research and for their continued support and encouragement. I received tremendous support and assistance from the fellow members of the Li Lab. My undergraduate student, Maple Chen, contributed to several key experiments for this project. I wish her a great success in the future. My friends also deserve thanks for helping me with issues scientific and beyond: Austin, Jacob, Dakota, and Kurt to name a few.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Natural Products Overview

Natural products (NPs) are chemicals of great structural complexity and diversity which have found many uses in human medicine, nutrition, animal health, and plant crop protection [1-5]. This large family of diverse chemical entities originate from a myriad of sources including terrestrial plants, animals, marine organisms, microorganisms, and terrestrial vertebrates and invertebrates, and they are typically products of the "secondary metabolism": molecules that are not absolutely required for survival of the host and are not specifically needed by the core metabolism of the cells that produce them, but which provide advantages to the whole organism in its natural environment. Unlike primary metabolites, which are required for growth and are mostly the same across the spectrum of living organisms, secondary metabolites can vary widely from species to species and encompass a diverse array of complex chemical structures. NPs provide key scaffolds in drug development and in the development of other high-value agents. Novel NP discovery has been driven by the findings that they are important in many fields such as pharmaceuticals, herbicides, insecticides, etc. As of 1990, 80% of drugs in use were natural products, their derivatives, or natural product analogs [6]. Since the discovery of penicillin more than 75 years ago, >23,000 NPs have been characterized, with the most abundant sources being bacterial, mainly by the family Actinomycetaceae [7]. Nonetheless, the discovery and manufacturing of plant NPs have lagged that of microbial NPs primarily due to the size and complexity of plant genomes and the difficulty in plant biosynthetic pathway prediction and engineering as well as

the size and complexity of plant biosynthetic pathways, which may include dozens of unique genes extending from the core metabolism [8].

1.2 Microbial Production of Plant Natural Products

Plant natural products (PNPs) are currently of great importance in the pharmaceutical and agricultural fields, but like most secondary metabolites, they typically exhibit low abundance in nature and low production levels in their native plant hosts [9-10]. Due to the long growth time, low yield, and environmental consequences of harvesting large volumes of plant biomass, there is a significant incentive to produce these valuable compounds in microbial hosts. Microbial production of PNPs can overcome these challenges by enabling (1) on-demand production capabilities associated with microbial cells, (2) scalable and controlled production in fermentation facilities, and (3) the capacity to produce PNPs and PNP intermediates at higher purity or yield than those provided by the native plant host. Recent demonstrations of bio-based, industrial-scale production of PNPs in bacterial and fungal hosts have shown that microbial biosynthesis from fermentable building blocks is a promising approach toward scalable production and modification of these high-value compounds [11-14]. However, challenges remain in the effective application of this approach. Many PNPs are derived from very complex and lengthy metabolic pathways [15]. Thus, reconstituting these plant-specialized pathways in microbes is a slow and difficult process. Furthermore, the biosynthetic routes for most PNPs have not been completely discovered, and validation is slowed by the aforementioned difficulties in pathway reconstruction in microbial systems. Modern microbial genome manipulation techniques can integrate up to 10 pathway genes at one time with a typical turnaround time of around six weeks [16-18]. This allows for the

integration of smaller pathways but is insufficient when synthesizing more complex members of the PNP library.

1.2.1 Yeast as a Platform for the Production of Natural Products

Baker's yeast (*Saccharomyces cerevisiae*) is a well-studied production host with a long history of industrial use. It offers many advantages for this role such as robustness in many environments, high genetic tractability, and safety to human health [19]. To reconstruct foreign pathways into yeast, pathway genes with promotors and terminators compatible the yeast biology must be internalized in a functional state. Then the cells must translate these new genes into their respective enzymes which create a metabolic pathway where metabolic flux leads to a desired product. Thus, the first step of any PNP metabolic engineering endeavor is the identification of the pathway genes of interest from the host organism and the construction of these genes into a deliverable and usable form. To identify target genes related to the pathway of interest, the genome of the host is mined using various techniques from bioinformatics, mass spectrometry, proteomics, transcriptomics, metabolomics, and gene expression [20]. Then the genes are synthesized or cloned with the appropriate promotors and integrated into the yeast genome. Most metabolic engineering endeavors in *S. cerevisiae* employ episomal plasmids which are not connected to the microbial chromosome and are independently active and replicative [21]. However, there are inherent problems associated with non-integrative plasmid-based metabolic engineering strategies such as genetic instability due to plasmid loss, the need to continually maintain a selection pressure, and variable gene expression due to different copy numbers within the population [22]. These limitations are especially prevalent during long term or large-scale cultivations such as those seen in industrial settings, or in cultivations occurring in poorly defined media,

limiting the utility of plasmid-based engineering endeavors outside of strict laboratory settings. Due to the limitations of plasmid-based approaches, integrating pathway genes directly into the yeast chromosome is typically preferred, especially for industrial production.

1.2.2 Homologous Recombination in Yeast

In yeast and most other eukaryotes, there are two principal pathways for the repair of cytotoxic DNA lesions and for the rescue of collapsed replication forks, homologous recombination (HR) and nonhomologous end-joining (NHEJ). In HR, a DNA sequence with homology to the chromosome on each side of the lesion is used as a direct repair template, with the chromosome reconstructed to precisely match the template sequence between the regions of homology. In NHEJ based repair of DSBs, the lesion is directly ligated together in an error-prone fashion. Typically, genome engineering at the chromosomal level is achieved by leveraging the cell's own repair machinery. This can come from the error-prone NHEJ pathway that leads to insertion/deletion (indel) mutations, which can be used to create point mutations or knock out genes, or, alternatively, we can supply a repair template to overwrite the site of a double-stranded break (DSB) for more-precise genome engineering via the HR pathway (Figure 1).

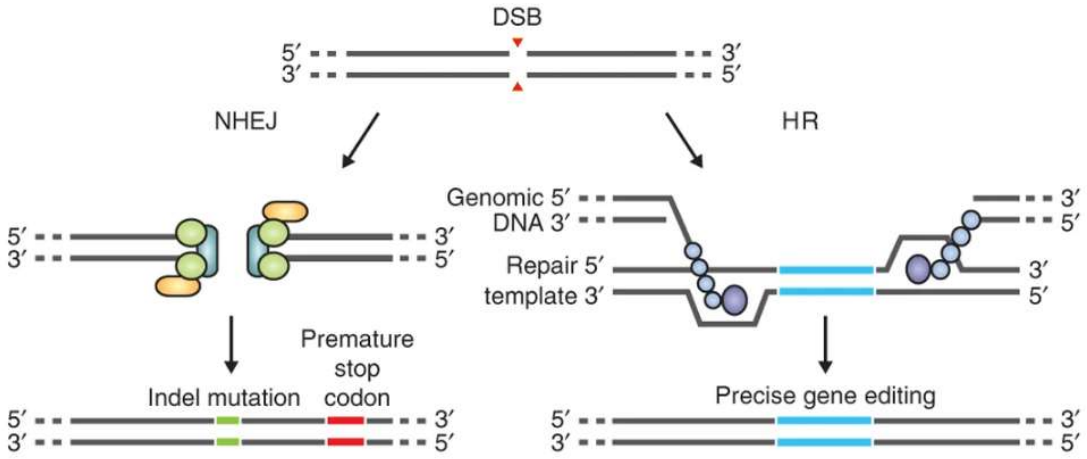


Figure 1. DNA double-stranded breaks facilitate alteration of the genome. [23]

HR is the primary mode of chromosomal repair in *Saccharomyces*, with NHEJ only occurring at a frequency of 0.1% compared to HR in HR-competent haploid cells, and NHEJ only occurring significantly when the primary HR mechanisms, the RAD52 epistasis group of genes, are blocked [24]. Because of the inefficiency and inaccuracy of NHEJ, HR-arrested yeast cells are much worse at surviving a DSB than cells which utilize HR, and this reliance upon HR for the vast majority of DNA repair is what leads to the high genetic tractability of *Saccharomyces cerevisiae*. Any sequence that shares homology to a specific part of the yeast chromosome can be utilized as a repair template in HR, so a DNA fragment containing a foreign pathway gene flanked on each side by yeast homology can be perfectly reconstructed in the chromosome during a DSB event. This forms the basis for most genetic pathway reconstruction techniques in yeast.

1.3.1 Conventional Pathway Integration in Yeast

Traditional yeast integration methods use the endogenous homologous recombination and DNA repair machinery of S. cerevisiae to integrate genes of interest. Here, linear fragments of DNA are flanked by arms homologous to specific integration sites and

delivered into the yeast cells. Spontaneously occurring damage and double strand breaks (DSBs) in the chromosome are typically repaired using homologous chromosomes or native DNA as a template, but with the delivered linear fragments, repair will now be driven to include the new genes of interest [25]. The natural homologous repair mechanisms of *S. Cerevisiae* are extremely efficient, able to repair homologous linear fragments into constructs reaching the size of an entire, episomal genome [26]. However, there are many difficulties associated with the genomic integration of entire pathways into the yeast chromosome itself, and this is especially apparent for long and complex pathways such as those seen in PNP synthesis. Insertions of genes into the chromosome first requires spontaneous and random damage to the chromosome at the specific target locus in order to generate a DSB which allows for homology driven repair to begin. However, the efficiency of this transformation technique is very low for even a single insertion site, and this efficiency declines dramatically for each additional integration site that is targeted, limiting the number of genes that can be integrated at any given time [27]. Another challenge presented by this technique is the lack of selectivity. Any cell which can repair a DSB without the use of the delivered pathway repair template, either through HR with the cell's endogenous templates or through precise NHEJ, will still be able to survive and proliferate. This problem is amplified when targeting multiple sites simultaneously as the cell can accept the foreign pathway in a mosaic fashion, leading to a heterogenous genotype throughout the cell culture. Thus, to integrate an entire large biosynthetic pathway into yeast under these techniques, this homologous recombination driven integration must be performed multiple times with modules of a few genes being inserted into the chromosome at a time, with each module containing its own selective marker, until the entire pathway has been reconstructed. The current inefficiency of conventional pathway integration presents a

major bottleneck in the creation of yeast strains holding complex biosynthetic pathways and forces researchers to focus on smaller, more manageable biosynthetic targets.

1.3.2 Advantages of CRISPR-Mediated Pathway Reconstruction

To address the issues with traditional chromosomal integration strategies in yeast, there has been recent interest in the use of targetable endonucleases such as TALENs [28] or CRISPR-associated systems (Cas) [29] to induce DSBs in the chromosome and drive homologous repair. Here the same linear DNA fragments containing pathway genes flanked by homology to specific integration sites will be delivered to yeast cells; however, a CRISPR-Cas9 plasmid containing guide sequences targeting the homologous integration sites will also be delivered. This will specifically induce DSBs at the genetic points of interest, allowing HR to proceed with much greater efficiency. Using these endonucleases, yeast integrations would no longer have to rely on spontaneous damage happening in the sites of interest for homology-driven integration to occur. Because multiple genome sites can be induced to begin HR simultaneously, much larger gene sets can be integrated in a single-step and with much higher efficiency. Additionally, this CRISPR system has innate selectivity. Any HR or NHEJ which regenerates the wild-type yeast genotype will regenerate the targeted guide sequence and simply induce another DSB from the CRISPR system. The only means the cell has to quench Cas9's activity is to integrate the exogenous pathway gene which will eliminate the protospacer sequence, or to use imprecise NHEJ which will generate a significant enough error to prevent Cas9 binding, something that is rare and inefficient in yeast. Overall, this strategy represents a powerful tool for rapidly integrating large biosynthetic pathways into multiple, specific sites of the yeast chromosome simultaneously.

1.4.1 Research Objectives

As mentioned above, DSB formation is known to induce HR-based gene integration. Therefore, the first aim of this study is to develop a CRISPR-Cas9 system which is functional in yeast and investigate the ability of this system to aid in the integration of exogenous biosynthetic-pathway-gene fragments into specific chromosomal loci. Here we used an ultrahigh copy number receiver plasmid constructed for expressing the CRISPR–Cas9 gene in yeast and harboring a HR disruption guide targeting a single site on the yeast chromosome. This CRISPR plasmid and a linearized DNA insert with homology towards to targeted DSB site were delivered into yeast and the resulting cell pool was analyzed to determine if this CRISPR HR system was functional.

The second objective of this research is to engineer this CRISPR-Cas9 system to target multiple loci within the yeast chromosome and allow for the simultaneous integration of multiple pathway fragments each with homology to separate target sites. Here we transfected a CRISPR-Cas9 expression plasmid containing a crRNA array targeting 5 nonessential loci on the yeast genome and additionally delivered 5 linear DNA inserts, each with homology to these target sites. The resulting cell pool was analyzed to determine if this multilocus CRISPR system was functional.

The ultimate goal of these characterizations is to develop an alternative technology to make pathway integration much more efficient and with greater scalability. Accordingly, the third goal we are working towards is to test its ability to integrate an entire PNP biosynthetic pathway into yeast in a single step and compare its efficiency with that of currently existing methods. To this end we will deliver CRISPR plasmids

targeting 1, 2, and 3 or more sites into yeast alongside increasingly large biosynthetic pathways and comparing the efficiency of transformation against conventional non-CRISPR based methods. The results of this study will provide a single-step method for the markerless integration of large biosynthetic pathways into yeast and will lead to further developments and optimizations of the CRISPR system.

CHAPTER 2

MATERIALS AND METHODS

2.1 Overview

To reach all the objectives of this study, all new CRISPR plasmids and gene fragment holding vectors were ligated *in vitro* and cloned in *E. coli*. CRISPR plasmids and gene fragments amplified from holding vectors were transformed into yeast for Cas9 expression and gene integration. Whole colony PCR and colony formation were used to determine the effectiveness of the CRISPR/Cas9 integration system.

2.2 Strains, Plasmids, and Protospacers

The *E. coli* strain used for all the plasmid cloning and library construction was DH5α. Yeast CENPK2.1D was used for all CRISPR/Cas9 integration studies. Details of these strains are listed in Table 1 below.

Table 1. Strains used in this study

Strains	Relevant genotype	Source
DH5α	F– endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK– mK+), λ–	Laboratory Stock
CENPK2.1D	MATalpha; his3D1; leu2-3_112; ura3-52; trp1-289; MAL2-8c; SUC2	Laboratory Stock

All of the plasmids used in this study are listed in Table 2. Details of the cloning and construction of these plasmids are described afterward.

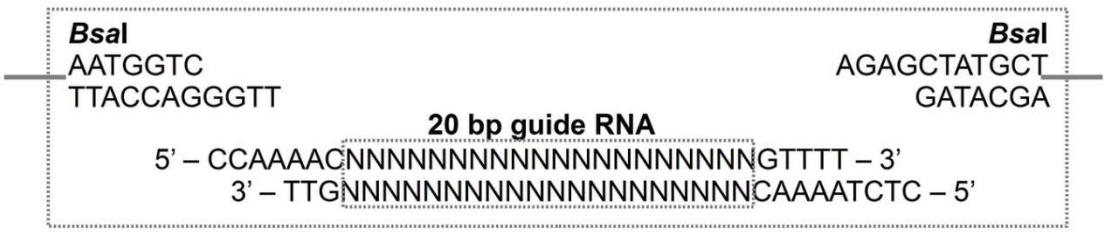
Table 2. Plasmids used in this study

Plasmids	Description	Source
pSL 4	P_{TEF1}, T_{CYC1}, Kanamycin	Laboratory stock
pSL 49	P_{TEF1}, mCherry, T_{TEF1}, Leu2, 2μ	Laboratory stock
pSL 86	P_{TEF1}, mCherry, T_{CYC1}, Kanamycin	This study
pSL 68	P_{TEF1}, CaVAMT, T_{CYC1}, Kanamycin	Laboratory stock
pCRCT	tracrRNA, iCas9, Ampicillin, URA3, 2μ	Bao et al. [30]
pSL 87	pCRCT, ADH6-1 crRNA	This study
pSL 88	pCRCT, ADH6-2 crRNA	This study
pSL 89	pCRCT, Locus 1-5 crRNA array	This study
pSL 90	pCRCT, Locus 1 crRNA	This study
pSL 91	pCRCT, Locus 2 crRNA	This study
pSL 92	pCRCT, Locus 3 crRNA	This study
pSL 93	pCRCT, Locus 4 crRNA	This study
pSL 94	pCRCT, Locus 5 crRNA	This study

Plasmids used in this study can be separated into two main groups: library holding vectors and CRISPR expression plasmids. The main cloning backbone used in this study was derived from pSL 4 and contains the high-expression yeast protomer/terminator pair TEF1/CYC1 inside of a high-copy *E. coli* cloning vector. The plasmid pSL 49 is a high-copy yeast expression cassette housing the mCherry gene. For cloning, this gene was PCR amplified and assembled with the pSL 4 backbone using Gibson assembly to form pSL 86. The plasmid pSL 68 is a holding vector similarly derived from pSL 4

housing the vanillin aminotransferase gene from the pungent pepper *Capsicum annuum*. The plasmid pCRCT is a yeast CRISPR/Cas9 expression plasmid that lacks a functional protospacer guide. These guides can be inserted in between two *Bsa*I restriction sites on pCRCT using Golden Gate assembly. Here pCRCT is digested using the *Bsa*I restriction enzyme. Two ssDNA primers containing the 20 bp guide sequence are annealed together with overhangs complementary to the cut site of *Bsa*I. This annealed guide sequence and the digested pCRCT backbone are ligated together to form the functional CRISPR/Cas9 expression plasmid. A schematic of this assembly is shown in Figure 2 below. This assembly method was used for the construction of pSL 87- 94 varying only in the 20 bp guide sequence delivered.

Figure 2. Guide sequence oligonucleotide assembly into *Bsa*I sites of pCRCT [31]



One of the requirements for what sequence can be utilized as a functional guide RNA is what is called a protospacer adjacent motif (PAM). This is important because a given species may have one or multiple types of CRISPR systems, and each CRISPR system may have a unique PAM. The CRISPR system (Figure 3) used in pCRCT derives from *Streptococcus pyogenes* and the PAM for that species is the NGG trinucleotide motif where N is any nucleotide. This means that the CRISPR system can only target sites adjacent to NGG.

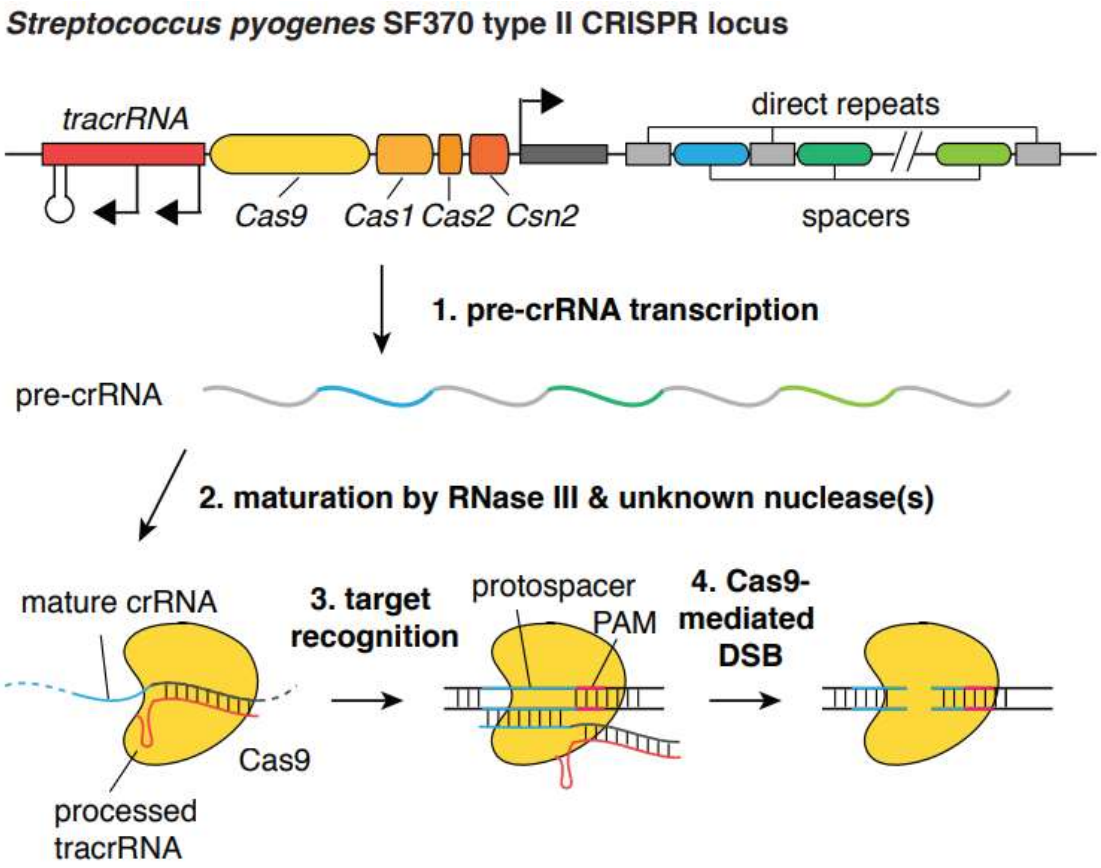


Figure 3. *Streptococcus pyogenes* CRISPR system. [32]

We designed guides adjacent to this motif using an algorithm reported by Doench et al. [33]. This algorithm uses spacer motifs commonly associated with high CRISPR activity to generate an efficiency score and uses alignment throughout the yeast genome to generate a specificity score. It then ranks sequences adjacent to the NGG PAM within the target locus accordingly. This algorithm was used to design all of the guide RNAs used in this study which are listed in Table 3. Details of these guide sequences are described afterwards.

Table 3. Guide sequences used in this study

Target loci	Sequence	Plasmid assembly
ADH6 (gRNA 1)	CTAGGGCCCAAGTCAAACAG	pSL 87
ADH6 (gRNA 2)	GACATTAAGATCGAAGCATG	pSL 88
Locus 1 (YBR197C)	TAAAGATCAAGAACCAGACG	pSL 89, 90
Locus 2 (YDR514C)	GTGCAAGAAAGCATACGAAG	pSL 89, 91
Locus 3 (YMR206W)	TCATGTCTAAAAGACAACAG	pSL 89, 92
Locus 4 (YPL250C)	GATGGGGGAAGATTCCACAG	pSL 89, 93
Locus 5 (YBL059W)	TATTTGGTATGGGCATCACA	pSL 89, 94

Two separate guide sequences were made to target the native yeast gene alcohol dehydrogenase 6. These two guides were assembled with pCRCT to make pSL 87 and 88, respectively. Five other guide sequences were created to target functionless sites within the yeast genome known to be nonessential. These five sites were dubbed Loci 1-5. Each of the Locus 1-5 guides were individually assembled within pCRCT to form pSL 90-94. All 5 of the guides were assembled together to form a CRISPR DNA array and this array was assembled with pCRCT to form pSL 89, a multilocus CRISPR plasmid containing 5 separate functional guides.

2.3 Plasmid Transformation and Gene Integration

After constructing the plasmids using either Gibson or Golden Gate assembly, ligation mixtures were directly transformed into competent DH5α *E. coli* using a 42 °C heat shock for 45 seconds. After outgrowth in Super Optimal broth with Catabolite repression (SOC) at 37 °C for 1 hour, cultures were grown on Luria-Bertani (LB) broth plates at 37 °C containing antibiotics as needed. Concentrations of the antibiotics are

listed as following: 100 μg/ml ampicillin (Amp) and 50 μg/ml kanamycin (Kan). Colonies were then collected and cultured in liquid LB media for plasmid purification and the correct assembly was confirmed using Sanger Sequencing.

Yeast integration was performed using electroporation according to the method reported by Colby et al. [34]. Linear gene inserts which are to be integrated into the yeast genome were PCR amplified from their respective library plasmids such that the fragments were flanked on each side by 40 base pairs containing homology to the target locus surrounding the predicted cut site to facilitate homologous recombination. 500 ng of the relevant CRISPR expression plasmid alongside 200 ng of each linear homologous repair template sharing homology to the CRISPR cut sites were electro-transformed into *S. cerevisiae* CENPK2.1D. Transformants were outgrown in Yeast Extract–Peptone–Dextrose (YPD) Medium for 1 hour before being transferred to Synthetic Defined (SD) Media lacking the amino acid uracil and cultured at 30 °C for 48 hours. This was done to increase the copy number of the URA3 containing CRISPR plasmid, give the cells time to make the necessary cuts and repairs, and reduce to number of untransformed cells in the pool. These cells were then plated on SD -Ura plates and incubated at 30 °C for 4 days. The resulting yeast colonies were then analyzed to determine the functionality of the CRISPR system.

CHAPTER 3

RESULTS AND DISCUSSION

3.1 Single Target Validation – ADH6/VAMT Integration

Before utilizing this system to integrate an entire pathway, it was first necessary to validate the functionality of our CRISPR plasmid and guide sequence design by inserting a single gene into a single target site. Thus, we first integrated the CaVAMT gene held in pSL 68 into the ADH6 gene locus targeted by pSL 87 and 88. These plasmids were electroporated into yeast CENPK2.1D either with or without a VAMT homologous repair template. The resulting colonies formed were then counted to roughly determine the efficiency of the CRISPR system. A positive control was performed by delivering pCRCT, a CRISPR plasmid without a usable guide sequence, alone into the yeast cells. The results of these transformations are shown in Table 4.

Table 4. Colonies counted on plates for each single target ADH6/VAMT experiment

Cas9 Vector	Repair Template	Colonies on Plate
(-) pCRCT	(-) None	848
(+) pSL 87	(+) VAMT	336
(+) pSL 87	(-) None	309
(+) pSL 88	(+) VAMT	721
(+) pSL 88	(-) None	52

When the CRISPR plasmid is introduced without a repair template, it will repeatedly cleave the target locus, causing toxicity. Yeast cells are rescued from DSB lethality when an appropriate repair template is provided. The uracil dropout medium will select

against cells which failed to take up the CRISPR plasmid (which confers uracil prototrophy), but because the CRISPR plasmid is toxic to cells unless a successful replacement occurs (eliminating the CRISPR target locus) only cells which have a replaced locus are expected to survive. Cells containing pSL 87 were equally capable of growing on SD – Ura regardless of the presence of a repair template. This indicates that the guide sequence assembled into pSL 87 was not functional as the plasmid was not able to confer a lethal targeted DSB. Cells containing pSL 88 were not capable of growing nearly as well as the pCRCT control unless rescued by the VAMT repair insert, which conferred an over 13-fold increase in the number of colony forming units. This indicates that the guide sequence of pSL 88 was successful in inducing a lethal DSB that could be repaired using a VAMT insert flanked with homology to the ADH6 gene. Due to NHEJ quenching the protospacer region, mutations in the CRISPR target locus, and cells which manage to survive CRISPR-associated DSBs, there will be a background rate in the form of false transformant colonies which do not carry the correct genomic replacements, as indicated by the surviving colonies containing pSL 88 without the appropriate insert. To ensure that the colonies contained the VAMT gene inserted into the ADH6 locus, we employed whole yeast colony PCR for screening. A forward primer annealing to the original yeast locus and a reverse primer annealing to the new gene coding region were used to amplify a ~300 bp PCR fragment containing the 5' homology site. This reaction was performed using three different colonies containing both pSL 88 and the VAMT insert as well as three different wild type CENPK2.1D colonies. The PCR products were run on a 1% agarose gel and the bands are shown in Figure 4.

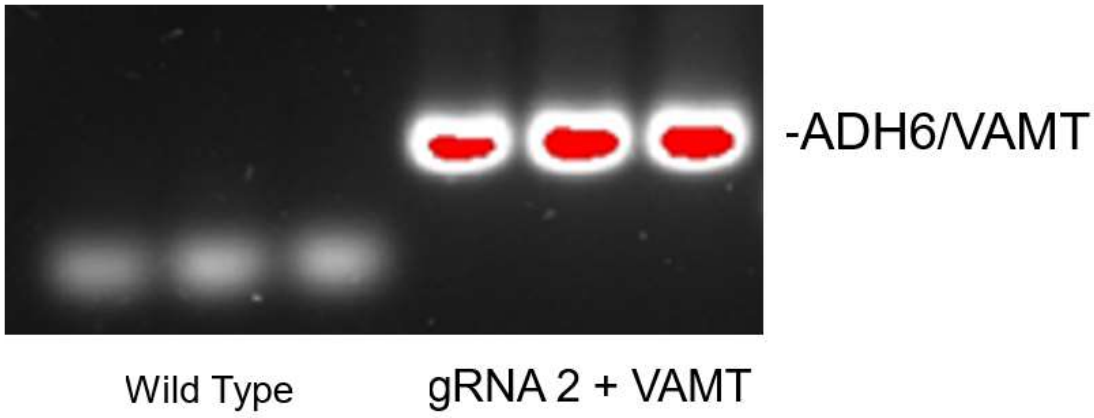


Figure 4. Results of agarose gel electrophoresis on Colony PCR for ADH6/VAMT integration.

We were able to successfully amplify a band of the correct size over the ADH6/VAMT homology region. This is strong evidence of successful integration of the VAMT repair template into the ADH6 locus, showing the CRISPR/Cas9 system is capable to delivering a gene into a specific site on the yeast chromosome.

3.2 Multi Target Validation – Locus1-5/mCherry

To further investigate the capabilities of this CRISPR system, we then integrated 5 separate mCherry inserts derived from pSL 86, each with homology to the 5 loci targeted by pSL 89, into the yeast genome simultaneously. First, we validated each of the 5 guide sequences present in pSL 89 on an individual basis in a manner similar to the ADH6/VAMT single target validation. Plasmids pSL 90-94 were transformed into yeast with or without their respective mCherry repair inserts. The resulting colony formations are shown in Table 5.

Table 5. Colony count validation of single target Locus 1-5 CRISPR plasmids

Cas9 Vector	Repair Template	Colonies on Plate
(+) pSL 90	(+) mCherry/Locus1	67
(+) pSL 90	(-) None	1
(+) pSL 91	(+) mCherry/Locus2	21
(+) pSL 91	(-) None	2
(+) pSL 92	(+) mCherry/Locus3	29
(+) pSL 92	(-) None	2
(+) pSL 93	(+) mCherry/Locus4	93
(+) pSL 93	(-) None	2
(+) pSL 94	(+) mCherry/Locus5	44
(+) pSL 94	(-) None	2

Based on the colony counts from the single target validation of the Locus 1-5 guides, each of the 5 guide sequences appear to be functional as each of the plasmids significantly arrested growth except in the presence of a repair template with appropriate homology.

To characterize the CRISPR system's ability to simultaneously integrate in multiple sites, several electroporations using pSL 89 were performed, each lacking an additional repair template until none remained. The results of these transformations are shown in Table 6.

Table 6. Colony count validation of multi target Locus1-5 CRISPR plasmid experiments after 48 hours in liquid culture

Cas9 Vector	Repair Template	Colonies on Plate
(-) pCRCT	(-) None	142
(+) pSL 89	(+) mCherry/Locus1, (+) mCherry/Locus2, (+) mCherry/Locus3, (+) mCherry/Locus4, (+) mCherry/Locus5	224
(+) pSL 89	(+) mCherry/Locus2, (+) mCherry/Locus3, (+) mCherry/Locus4, (+) mCherry/Locus5	71
(+) pSL 89	(+) mCherry/Locus3, (+) mCherry/Locus4, (+) mCherry/Locus5	44
(+) pSL 89	(+) mCherry/Locus4, (+) mCherry/Locus5	21
(+) pSL 89	(+) mCherry/Locus5	10
(+) pSL 89	(-) None	0

These results indicate that the multilocus CRISPR system is functional as the cells can grow very effectively when all repair templates are fed but is incapable of growth when no repair templates are present, with growth declining with each additional repair template removed. However, the activity of the CRISPR system against any singular target appears to decline when multiple loci are targeted simultaneously as the cells are

still relatively capable of growth even when one or more inserts are not present. This is most likely because the amount of activating tracrRNA and Cas9 protein does not increase even when more sites are targeted, and these CRISPR factors become a bottleneck in DSB formation.

The cultures from the previous experiment were passaged and further cultured in SD – URA media for an additional 4 days and plated every 48 hours to see if the relative colony counts would change over time. The resulting colony counts are shown in Figure 5.

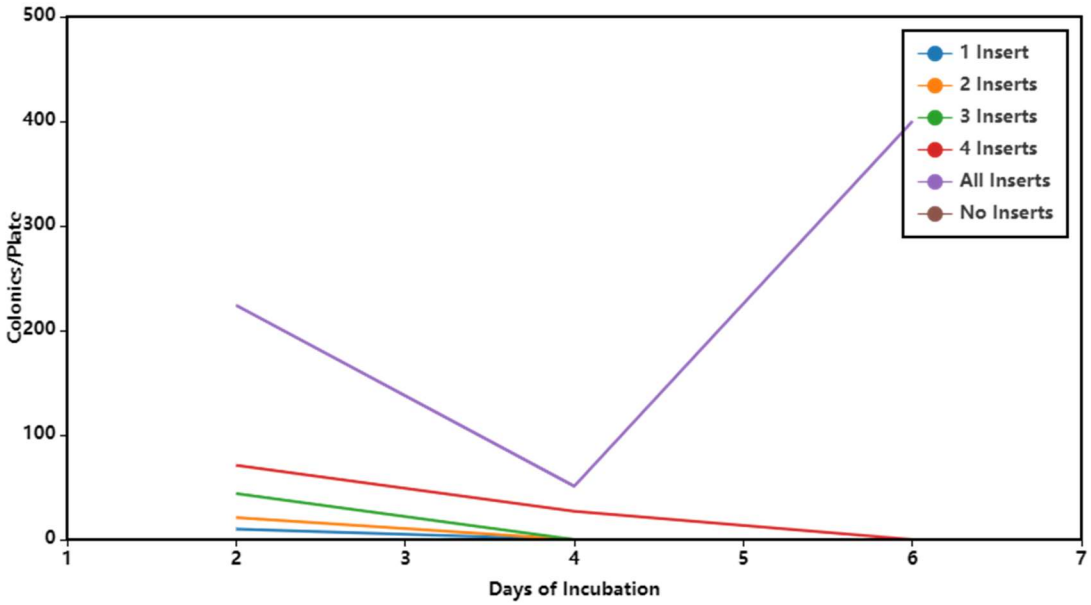


Figure 5. Colony formation for the multilocus CRISPR system over time.

This shows that the survival of the cells decreases over time as the cells lacking the full repair templates are unable to fully quench the activity of Cas9 and eventually succumb. Only cell populations containing all necessary repair templates are able to fully recover.

CHAPTER 4

SUMMARY AND FUTURE WORK

This study investigated the capability of a CRISPR/Cas9 system to induce damage into specific sites of the yeast chromosome in order to facilitate the efficient integration of delivered gene fragments through homologous recombination. Several Cas9 expression plasmids targeting user-defined loci were constructed and delivered into yeast. These plasmids were shown to be able to successfully integrate genes into target sites of the yeast chromosome. The system was demonstrated to be able to integrate genes into up to five separate loci simultaneously, and the markerless selectivity of the CRISPR system through DSB toxicity was thoroughly characterized. In conclusion, this system demonstrates great potential in expanding the yeast genome engineering toolbox and facilitating the rapid integration of large biosynthetic pathways.

The ultimate goal of one-step large pathway integration first requires several intermediate goals. First, the system must be demonstrated with smaller pathways integrated into single loci and directly compared with conventional integration methods to show that it is truly more efficient. Second, additional high expression synthetic promoter and terminator sets need to be constructed. As we are aiming to integrate many genes simultaneously, fragments sharing the same promoter and terminator sets will present a problem as they share significant homology with each other. Finally, the system will need to be fully validated by performing a complete demonstration of the system's ability to integrate an entire biosynthetic pathway into multiple loci simultaneously. These three aims together yield a methodology that enables the rapid integration of large biosynthetic pathways in a single step.

REFERENCES

[1] Bhanot A, Sharma R, Noolvi MN (2011) Natural sources as potential anti-cancer agents: a review. Int J Phytomed 3:9–26

[2] Demain AL (2014) Importance of microbial natural products and the need to revitalize their discovery. J Ind Microbiol Biotechnol 41:185–201

[3] Newman DJ, Cragg GM (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. J Nat Prod 75:311–335

[4] Giddings LA, Newman DJH (2013) Microbial natural products: molecular blueprints for antitumor drugs. J Ind Microbiol Biotechnol 40:1181–1210

[5] Newman DJ, Cragg GM, Snader KM (2000) The influence of natural products upon drug discovery. Nat Prod Rep 17:214–234

[6] Harvey, Alan L. "Natural products in drug discovery." Drug discovery today 13.19-20 (2008): 894-901.

[7] Bérdy J (2012) Thoughts and facts about antibiotics: where we are now and where we are heading. J Antibiot 65:385–395

[8] McChesney, James D., Sylesh K. Venkataraman, and John T. Henri. "Plant natural products: back to the future or into extinction?." Phytochemistry 68.14 (2007): 2015-2022.

[9] Barone, Rebecca P., et al. "The production of plant natural products beneficial to humanity by metabolic engineering." Current Plant Biology (2019): 100121.

[10] Ouyang, L., et al. "Plant natural products: from traditional compounds to new emerging drugs in cancer therapy." Cell proliferation 47.6 (2014): 506-515.

[11] Paddon, C. J. & Keasling, J. D. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. Nat. Rev. Microbiol. 12, 355–367 (2014).

[12] Bomgardner, M. M. Evolva pursues new route to resveratrol. Chem. Eng. News 92, 14 (2014).

[13] Galanie, S., Thodey, K., Trenchard, I. J., Interrante, F. M. & Smolke, C. D. Complete biosynthesis of opioids in yeast. Science349, 1095–1100 (2015).

[14] Trenchard, I. J., Siddiqui, M. S., Thodey, K. & Smolke, C. D. De novoproduction of the key branch point benzylisoquinoline alkaloid reticuline in yeast. Metab. Eng.31, 74–83 (2015).

[15] Brown, S., Clastre, M., Courdavault, V. & O'Connor, S. E. De novo production of the plant-derived alkaloid strictosidine in yeast. Proc. Natl Acad. Sci. USA112, 3205–3210 (2015).

[16] Naseri, Gita, et al. "COMPASS for rapid combinatorial optimization of biochemical pathways based on artificial transcription factors." Nature communications 10.1 (2019): 2615.

[17] Mitchell LA, et al. Versatile genetic assembly system (VEGAS) to assemble pathways for expression in S. cerevisiae. Nucleic Acids Res 43, 6620-6630 (2015).
[18] Lian J, HamediRad M, Hu S, Zhao H. Combinatorial metabolic engineering using an orthogonal tri-functional CRISPR system. Nat Commun 8, 1688 (2017).
[19] Hong, Kuk-Ki, and Jens Nielsen. "Metabolic engineering of Saccharomyces cerevisiae: a key cell factory platform for future biorefineries." Cellular and Molecular Life Sciences 69.16 (2012): 2671-2690.
[20] Rudolf JD, Yan X, Shen B (2015) Genome neighborhood network reveals insights into enediyne biosynthesis and facilitates prediction and prioritization for discovery. J Ind Microbiol Biotechnol. doi:10.1007/s10295-015-1671-0
[21] Karim, Ashty S., Kathleen A. Curran, and Hal S. Alper. "Characterization of plasmid burden and copy number in Saccharomyces cerevisiae for optimization of metabolic engineering applications." FEMS yeast research 13.1 (2013): 107-116.
[22] Da Silva, Nancy A., and Sneha Srikrishnan. "Introduction and expression of genes for metabolic engineering applications in Saccharomyces cerevisiae." FEMS yeast research 12.2 (2012): 197-214.
[23] Ran, F. Ann, et al. "Genome engineering using the CRISPR-Cas9 system." Nature protocols 8.11 (2013): 2281-2308.
[24] Clikeman, Jennifer A., et al. "Homologous recombinational repair of double-strand breaks in yeast is enhanced by MAT heterozygosity through yKU-dependent and-independent mechanisms." Genetics 157.2 (2001): 579-589.
[25] Storici, Francesca, et al. "Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast." Proceedings of the National Academy of Sciences 100.25 (2003): 14994-14999.
[26] Gibson, Daniel G., et al. "One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome." Proceedings of the National Academy of Sciences 105.51 (2008): 20404-20409.
[27] DiCarlo, James E., et al. "Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems." Nucleic acids research 41.7 (2013): 4336-4343.
[28] Yuan, Jifeng, and Chi Bun Ching. "Combinatorial assembly of large biochemical pathways into yeast chromosomes for improved production of value-added compounds." ACS synthetic biology 4.1 (2014): 23-31.
[29] Christian, Michelle, et al. "Targeting DNA double-strand breaks with TAL effector nucleases." Genetics 186.2 (2010): 757-761.
[30] Bao, Zehua, et al. "Homology-integrated CRISPR–Cas (HI-CRISPR) system for one-step multigene disruption in Saccharomyces cerevisiae." ACS synthetic biology 4.5 (2014): 585-594.

[31] Shi, Shuobo, et al. "Delta Integration CRISPR-Cas (Di-CRISPR) in Saccharomyces cerevisiae." Microbial Metabolic Engineering. Humana Press, New York, NY, 2019. 73-91.
[32] Cong, Le, et al. "Multiplex genome engineering using CRISPR/Cas systems." Science 339.6121 (2013): 819-823.
[33] Doench, John G., et al. "Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9." Nature biotechnology 34.2 (2016): 184-191.
[34] Colby, David W., et al. "Engineering antibody affinity by yeast surface display." Methods in enzymology. Vol. 388. Academic Press, 2004. 348-358.