2020-01-23

Developing multiplexed genome-edit platforms in yeast (Saccharomyces cerevisiae) using a single gRNA-mediated CRISPR-Cas9

Baek, Sihyun

UNIVERSITY OF CALGARY

Developing multiplexed genome-edit platforms in yeast (*Saccharomyces cerevisiae*) using a single gRNA-mediated CRISPR-Cas9

by

Sihyun Baek

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN BIOLOGICAL SCIENCES

CALGARY, ALBERTA

JANUARY, 2020

Abstract

The budding yeast *Saccharomyces cerevisiae* is a versatile microbial platform to build synthetic metabolic pathways for the production of a diverse array of chemicals. To expedite the construction of synthetic pathways in yeast, a novel multiplex genome-editing platform was developed where CRISPR-Cas9 can be applied to simultaneously integrate up to five foreign genes by a single transformation with one gRNA. To choose optimal loci for integration and expression of transgenes, eight desirable intergenic loci, located adjacent to highly expressed genes, were identified. The eight intergenic loci were fully characterized for different parameters after integrating a green fluorescent protein (*GFP*) cassette – CRISPR-mediated *GFP* integration efficiency, degree of *GFP* expression, growth rates of *GFP*-integrated strains, and genomic stability of the *GFP* integration. From these analyses, five loci were selected to build the multiplex platform where a common 23-bp DNA comprised of 20-bp synthetic DNA and 3-bp protospacer adjacent motif (PAM) was seamlessly placed in the five loci in a sequential manner. This process resulted in five distinct yeast strains harboring one to five copies of the synthetic gRNA-binding site in the genome. Using these pre-engineered yeast strains, simultaneous double, triple, quadruple, and quintuple gene integrations were demonstrated at between 84.8% and 98.0% integration efficiencies using a 3-gene betalain biosynthetic pathway and geneticin and hygromycin B resistance markers. The quadruple and quintuple gene integration platforms were applied to successfully generate yeast strains synthesizing a representative sesquiterpene lactone, costunolide, and its precursor, germacrene A acid. This work demonstrates the utility of the single gRNA-mediated CRISPR-Cas9 platform to build complex metabolic pathways in yeast.

Preface

This research project was conceived by Dae-Kyun Ro and Sihyun Baek. Sihyun Baek designed, conducted and analyzed all experiments in this thesis with guidance from his supervisory committee members. The results for the R1-R6 and P1-P2 site characterizations, flow cytometry analysis and yeast growth assay described in Chapter 3 were collected partly in collaboration with Ji Young Lee (Mitacs visiting PhD student from Ewha Womans University, Republic of Korea). Kunal Dalal contributed to the work involving construction and analysis of the sesquiterpenoid-producing yeast strains described in sub-Chapter 4.3.

Acknowledgments

I would first and foremost like to thank my supervisor, Dr. Dae-Kyun Ro, for his invaluable mentorship and full support throughout my degree. It has truly been an honor to learn and grow as a scientist under Dr. Ro's guidance. I would also like to thank my committee members, Dr. Gordon Chua and Dr. Vanina Zaremberg, who have provided me with helpful suggestions and critical questions that were crucial in steering my project in the right direction. The first part of my research project would have been difficult without the help from Ji Young Lee, who offered valuable helping hands in the loci characterization experiments. I am also incredibly grateful to Kunal Dalal, who assisted with many of the labor-intensive experiments of the project. Thanks are also due to all past and current members of the Ro lab for supporting my research with insightful suggestions and feedback. Special thanks go to Dr. Moonhyuk Kwon, Adam Lakusta, Connor Hodgins, Joseph Utomo, Iris Ngo, Irene Cheung and Kunal Dalal for being friendly and encouraging colleagues. I would also like to thank Parbir Grewal at UC Berkeley for providing us with plasmids encoding betalain biosynthetic genes, Rebecca Leong-Quong at the Molecular and Cellular Core Facility for her help with flow cytometry run, and Ralph Hindle at Vogon Laboratory Services for his help with LC-MS run and analysis. Lastly, I would like to thank my Mom and Dad, my sister Hanji, and my cousin Yeounwoo for providing me with endless support and for being there for me and believing in me.

Table of contents

Abstract ii

Preface iii

Acknowledgments iv

Table of contents v

List of Tables viii

List of Figures ix

List of Abbreviations xi

CHAPTER 1: INTRODUCTION 1

1.1 Introduction to *Saccharomyces cerevisiae* 1

1.1.1 Yeast as a model organism 1

1.1.2 Yeast as a cell factory 2

1.1.3 Yeast cell factories for production of plant secondary metabolites 3

1.2 Double-strand break repair mechanisms in yeast 5

1.2.1 Homologous recombination 5

1.2.2 Non-homologous end joining 8

1.3 Genetic modification of yeast 10

1.3.1 Plasmid-based gene expression 10

1.3.2 Genomic modification *via* homologous recombination 12

1.3.3 CRISPR-Cas9 genome editing 15

1.3.3.1 CRISPR-Cas9: origin and mechanism 15

1.3.3.2 Yeast genome editing using CRISPR-Cas9: overview 19

1.3.3.3 Multiplexed CRISPR-Cas9 strategies in yeast 21

1.4 Research rationale, hypothesis, and objectives 24
CHAPTER 2: MATERIALS AND METHODS 28
2.1 Strains, media and growth conditions 28
2.2 DNA design and construction 29
2.2.1 gRNA target design 29
2.2.2 Cas9/gRNA vectors 30
2.2.3 CRISPR integration vectors for R integration sites 31
2.2.4 CRISPR integration vectors for H integration sites 31
2.3 Yeast transformation 32
2.4 Genomic DNA extraction and genotyping 33
2.5 Fluorescence analysis 34
2.6 Growth assay and gene stability assay 34
2.7 Evaluation of synthetic targets 35
2.8 Strain building 36
2.9 Yeast metabolite extraction 36
2.10 LC-MS analysis of yeast metabolites 37
CHAPTER 3: CHARACTERIZATION OF CRISPR GENOMIC INTEGRATION SITES IN YEAST 38
3.1 Evaluation of R intergenic sites for CRISPR-mediated genomic integration 38
3.2 Evaluation of H intergenic sites and different Cas9/gRNA expression platforms 43
3.3 Further characterization of H intergenic sites for CRISPR-mediated genomic integration 49

CHAPTER 4: DEVELOPMENT AND APPLICATION OF THE SGM-CRISPR MULTIPLEX PLATFORM 57

4.1 Synthetic gRNA target testing and strain construction 57

4.2 Evaluation of the SGM-CRISPR platform using marker genes 62

4.3 Proof of concept: construction of plant sesquiterpenoid biosynthetic pathways 71

CHAPTER 5: DISCUSSION 76

5.1 New set of intergenic loci for CRISPR-mediated gene integration and expression 76

5.2 SGM-CRISPR platform: significance and future directions 79

5.3 Conclusion 83

References 84

Appendix 94

List of Tables

Table 1. Description of the R intergenic sites used for CRISPR-mediated gene integration ... 39

Table 2. Efficiency of *LacZ* and *GFP* cassette integration for the R intergenic sites..... 42

Table 3. Efficiency of single gene integration: different integration sites and different Cas9/gRNA expression platforms..... 45

Table 4. Efficiency of single gene integration: marker-free or marker-dependent gene integrations..... 48

Table 5. Description of the H intergenic sites used for CRISPR-mediated gene integration ... 50

Table 6. Summary of CRISPR-Cas9 multiplex strategies for integration of exogenous gene cassettes..... 80

Table 7. Yeast strains used in this work 94

Table 8. Primers used in this work..... 96

Table 9. Plasmids used in this work..... 102

List of Figures

Figure 1. Overview of the eukaryotic HR mechanism. 7

Figure 2. Mode of action of CRISPR-Cas9 18

Figure 3. SGM-CRISPR yeast platforms developed in this project 25

Figure 4. Overview of the SGM-CRISPR platform 27

Figure 5. Hypothesized mode of action of the SGM-CRISPR platform 27

Figure 6. Plasmid map of the pCut plasmid carrying a gRNA cassette..... 30

Figure 7. Plasmid map of the pTargetH-GFP plasmid with unique enzyme sites. 32

Figure 8. Different Cas9/gRNA plasmid systems used for the single gene integration experiment..... 44

Figure 9. Different plasmid systems used for the single gene integration control experiment 46

Figure 10. Efficiency of *GFP* cassette integration for the H integration sites..... 51

Figure 11. Flow cytometry analysis of genome-based and plasmid-based *GFP* expression.... 53

Figure 12. Flow cytometry analysis of the *GFP*-harboring yeast strains 54

Figure 13. Genomic stability of the *GFP* integration in the H integration sites..... 55

Figure 14. Growth rates of the *GFP*-harboring yeast strains..... 56

Figure 15. CRISPR editing efficiencies of synthetic gRNA targets..... 59

Figure 16. PCR-Genotyping of sTarget-mediated single gene integration..... 60

Figure 17. Workflow of the strain construction for the development of the SGM-CRISPR platforms 61

Figure 18. Overview of the betalain biosynthetic pathway 63

Figure 19. Simultaneous 2- and 3- gene integrations demonstrated by the SGM-CRISPR platform..... 64

Figure 20. Simultaneous 4- and 5- gene integrations demonstrated by the SGM-CRISPR platform 66

Figure 21. Summary of the 2-, 3-, 4- and 5- gene integration efficiencies 67

Figure 22. Genotyping of yeast colonies for confirmation of the five marker gene integrations 69

Figure 23. Genomic stability of the five marker genes in the selected integration sites 70

Figure 24. Overview of the costunolide biosynthetic pathway 72

Figure 25. Overview of the gene cassettes used for the construction of sesquiterpenoid biosynthetic pathways in yeast 73

Figure 26. LC-MS analysis of selected transformants for germacrene A acid production 74

Figure 27. LC-MS analysis of selected transformants for costunolide production 75

List of Abbreviations

Abbreviation	Definition
5-FOA	5-Fluoroorotic Acid
Aa	*Artemisia annua*
ARS	Autonomously replicating sequence
bp	base pair
Bv	*Beta vulgaris*
Cas	CRISPR associated protein
Cas9	CRISPR-associated protein 9 from *Streptococcus pyogenes*
CEN	Yeast centromeric sequence
COS	Costunolide synthase
CPR	Cytochrome P450 reductase
Cre	Cre recombinase
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
CYP76AD1	Tyrosine hydroxylase
DOD	DOPA dioxygenase
cDOPA5GT	*cyclo*-DOPA 5-O-glucosyl transferase
DSB	Double-strand break
DSBR	Double-strand break repair
Flp	Flippase recombinase
FPP	Farnesyl diphosphate

FRT	Flippase recognition target
G418	Geneticin
GAA	Germacrene A acid
GAO	Germacrene A oxidase
GAS	Germacrene A synthase
GFP	Green fluorescent protein
gRNA	Guide RNA
kb	kilobase
HDR	Homology-directed repair
HDV	Hepatitis delta virus ribozyme
HEPES	4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid
HH	Hammerhead ribozyme
HR	Homologous recombination
HygB	Hygromycin B
LacZ	β-galactosidase
LC-MS	Liquid chromatography–mass spectrometry
Ls	*Lactuca sativa*
lox	Locus of X-over P1
Mj	*Mirabilis jalapa*
NHEJ	Non-homologous end joining
P450	Cytochrome P450-dependent monooxygenase
PAM	Protospacer adjacent motif

PCR	Polymerase chain reaction
SC	Synthetic complete
SDSA	Synthesis-dependent strand annealing
SGM-CRISPR	Single gRNA-mediated CRISPR
sTarget	Synthetic target
STL	Sesquiterpene lactone
TAP	Tandem affinity purification
tHMGR	Truncated 3-hydroxy-3-methylglutaryl-coenzyme A reductase
tracrRNA	Trans-activating crRNA
yeGFP	Yeast-enhanced green fluorescent protein
YPDA	Yeast extract-peptone-dextrose-adenine

CHAPTER 1: INTRODUCTION

1.1 Introduction to *Saccharomyces cerevisiae*

1.1.1 Yeast as a model organism

The budding yeast, *Saccharomyces cerevisiae*, is one of the most extensively used and studied eukaryotic organisms throughout the history of mankind. As one of the main industrial microorganisms, yeast has been frequently employed to supply a diverse array of useful commodities (e.g. bread, wine and beer) to humans for thousands of years, with the earliest human use of *S. cerevisiae* dating back to 7,000 years ago (Mortimer, 2000). Yeast has been a predominant industrial microorganism for thousands of years, but the beginning of modern yeast research was relatively recent, as the first use of yeast in the scientific community took place in the 1930s. Shortly after, genetic crosses were used in 1957 to construct a stable haploid S288C strain, a reference isolate strain that would later be used for the yeast genome sequencing project and become one of the most widely used strains in the field of yeast research (Mortimer and Johnston, 1986). The further establishment of yeast as a model organism arose from the development of genetic modification tools that gave yeast geneticists an ability to freely move genes in and out of yeast with ease. The advances in the transformation and genome modification techniques allowed scientists to precisely manipulate the yeast genome with commonly used selectable marker genes and develop designer deletion strains (Brachmann et al., 1998; Hinnen et al., 1978). Furthermore, another highlight of yeast genetics research was the completion of the yeast genome project that made *S. cerevisiae* to be the first ever eukaryotic organism to have its whole genome sequenced, allowing dissection of the anatomy of a eukaryotic genome (Goffeau et al., 1996). The availability of the genetic modification tools and sequencing information

enabled generation of different yeast strain libraries, such as the Yeast Knock-Out deletion collection (Giaever et al., 2002), yeast overexpression library (Sopko et al., 2006), and the libraries for GFP- and TAP- tagged yeast strains for protein localization and quantitative proteome studies (Ghaemmaghami et al., 2003; Huh et al., 2003). With the ability to investigate the functional biology of nearly all genes present in the yeast genome and introduce modifications in any desired location in the genome, researchers have been able to efficiently carry out advanced studies on diverse topics including cell cycle, protein interactions, gene interactions, regulatory networks, and various mammalian diseases including cancer (Botstein and Fink, 2011; Miller-Fleming et al., 2008; Steffen et al., 2009).

1.1.2 Yeast as a cell factory

S. cerevisiae is a main industrial organism that is well-suited for large-scale production of various consumable products, due to the robustness in industrial fermentation conditions, the status as GRAS (generally regarded as safe) organism, and the availability of well-characterized genetics database and powerful recombinant DNA technologies (Borodina and Nielsen, 2014; Da Silva and Srikrishnan, 2012). In the last couple of decades, the yeast scientific community saw incredible advancements in biomolecular engineering and genome engineering technologies which greatly increased the production capacity of yeast by means of metabolic engineering. Metabolic engineering is a discipline that is dedicated to the construction of novel organisms with modulated metabolic pathways for mass production of valuable chemicals. One of the main foundations of metabolic engineering lies in the ability to introduce and stably express foreign genes in a host organism to enable the construction of synthetic pathways (Ostergaard et al., 2000). Other metabolic engineering techniques employing enzyme engineering, modulated gene

expression, gene copy number alteration and metabolic flux analysis also play important roles in further optimizing the pathway introduced in a host organism in order to achieve a desirable yield (Kulkarni, 2016). By allowing the re-constitution of heterologous pathways with optimized conditions for flux and yield, such tools play important roles in transforming yeast into "cell factories" that are capable of producing biochemicals that are heterologous to the native yeast metabolism. The resulting yeast cell factories can utilize cheap, renewable feedstocks (e.g. simples sugars such as glucose) to *de novo* synthesize a diverse array of chemicals in a cost-efficient fashion, thereby offering sustainable and economically attractive alternatives to the conventional methods of fuel and biochemical production from petroleum (Kavšček et al., 2015; Nielsen et al., 2013). As such, a vast number of 'proof of concept' studies as well as several examples of successful commercial applications during the last ten years have illustrated the capability of yeast cell factory to produce a variety of high-value products, including biofuels (e.g. ethanol and isobutanol), pharmaceuticals (e.g. human insulin and anti-malarial drug artemisinin), and nutraceuticals (e.g. resveratrol and squalene) (Gustavsson and Lee, 2016; Hong and Nielsen, 2012).

1.1.3 Yeast cell factories for production of plant secondary metabolites

Secondary metabolites synthesized by plants are a specialized type of metabolites that are not essential for the growth and reproduction of the organism. Although secondary metabolites do not directly take part in the central metabolic processes of plants, these compounds play crucial roles as defense compounds and regulatory molecules, greatly increasing the evolutionary fitness of the plants (Wink, 2019). The plant kingdom is known to possess the greatest variety of secondary metabolites, and different secondary metabolites extracted from plants have been used

by humans as medicines, recreational drugs, and nutrients (Guerriero et al., 2018). Despite the high relevance of plant secondary metabolites to human health, a sufficient supply of plant secondary metabolites is not easy to achieve as very low quantities of these compounds accumulate in plants and chemical synthesis is often not feasible due to the complex structure and stereochemistry of most of the secondary metabolites (Siddiqui et al., 2012). This not only presents a problem to various industries in society that depend on the availability of valuable plant products, but also hampers the scientific efforts to elucidate the biochemistry of plant secondary metabolism by causing the absence of authentic standards and substrates (Nguyen et al., 2010). Microbial biosynthesis strategy overcomes the obstacles associated with *in planta* synthesis and chemical synthesis, with fast growth, controllable yield and an excellent capacity to accommodate various types of exogenous biosynthetic pathways. Among many microbial platforms, yeast is a preferred host for this strategy as it has an endoplasmic reticulum to accommodate membrane-bound enzymes originating from plants. The most notable application of yeast for the production of high-value plant secondary metabolites is the semi-synthetic production of anti-malarial drug artemisinin at the yield of 25 g/L (Paddon et al., 2013; Ro et al., 2006). More recent examples of yeast-synthesized plant products that are of high-relevance include biosynthesis of opioid compounds including thebaine and hydrocodone, and major cannabinoids such as Δ9-tetrahydrocannabinolic acid (Galanie et al., 2015; Luo et al., 2019). As evident from these examples, bioproduction of plant natural products in yeast is a growing field of research expanding in scope to accommodate more variety of plant natural products with increased control over achievable yields, benefitting the scientific community as well as the society as a whole.

1.2 Double-strand break repair mechanisms in yeast

Double-strand breaks (DSBs) are one of the most lethal DNA damages that is caused by either failure of endogenous processes, including DNA replication and repair, or exposure to ionizing radiations and various types of toxic chemicals (Rodgers and McVey, 2016). DSBs are considered to be the most detrimental of DNA lesions as unrepaired damage from DSB can easily lead to cell death due to accumulation of mutations and genomic rearrangements (Rodgers and McVey, 2016). In eukaryotes, two evolutionarily conserved pathways exist to repair DSBs: homologous recombination (HR) and non-homologous end joining (NHEJ). HR mode of DNA repair is based on sequence homology as the repair process involves a search for stretches of similar base sequences in the genome, followed by a synthesis of a new strand at the DSB site based on the homologous sequence. NHEJ, on the other hand, carries out a rejoining of broken ends without the need for a homologous repair template. In budding yeast, HR functions as a default mechanism to counteract DSB while NHEJ occurs at a much less frequency and will only act as a dominant repair mechanism if the HR pathway is blocked or if a homologous repair template is not present (Clikeman et al., 2001). Interestingly, strong preference of NHEJ over HR is observed in mammalian species and most other yeast species, such as *Yarrowia lipolytica* and *Kluyveromyces lactis* (Featherstone and Jackson, 1999; Löbs et al., 2017).

1.2.1 Homologous recombination

Currently, there are two primary models for HR mode of DNA repair: double-strand break repair (DSBR) and synthesis-dependent strand annealing (SDSA) (Figure 1). It is believed that SDSA is a major pathway for DSB repair in mitotically proliferating cells while DSBR model is commonly used to describe meiotic recombination events (Andersen and Sekelsky,

2010). Therefore, the commonly used haploid laboratory strains derived from S288C are likely to utilize SDSA as a major DSB repair pathway. DSBR and SDSA share stages of resection, strand invasion, and DNA synthesis. In the resection step, trimming of the 5' ends of the broken DNA ends generate short 3' overhangs of single-stranded DNA. The key factors involved in this step are Mre11-Rad50-Xrs2 (MRX) complex, Sae2 protein, Sgs1 helicase, and Dna2 endonuclease (Gnügge and Symington, 2017). This is followed by the formation of nucleoprotein filament which searches for homologous DNA repair template sequences similar to the sequences at the 3' overhangs (Aylon and Kupiec, 2004). Strand invasion involves a movement of the filament into a homologous donor DNA to form a displacement loop, a region of heteroduplex DNA (Wright et al., 2018). After strand invasion, a new DNA can be synthesized at the end of the invading 3' end by DNA polymerase. In the DSBR pathway, the displacement loop becomes a Holliday junction which is processed into recombination products by nicking endonucleases, usually resulting in crossover products or non-crossover products depending on where the nick is made (Figure 1) (San Filippo et al., 2008). In the SDSA pathway, DNA synthesis continues through branch migration without forming a stable Holliday junction and the newly-synthesized strand is able to anneal to the opposite broken end in the damaged chromosome, providing a template for ligation and subsequent formation of non-crossover recombination products (Figure 1) (San Filippo et al., 2008).

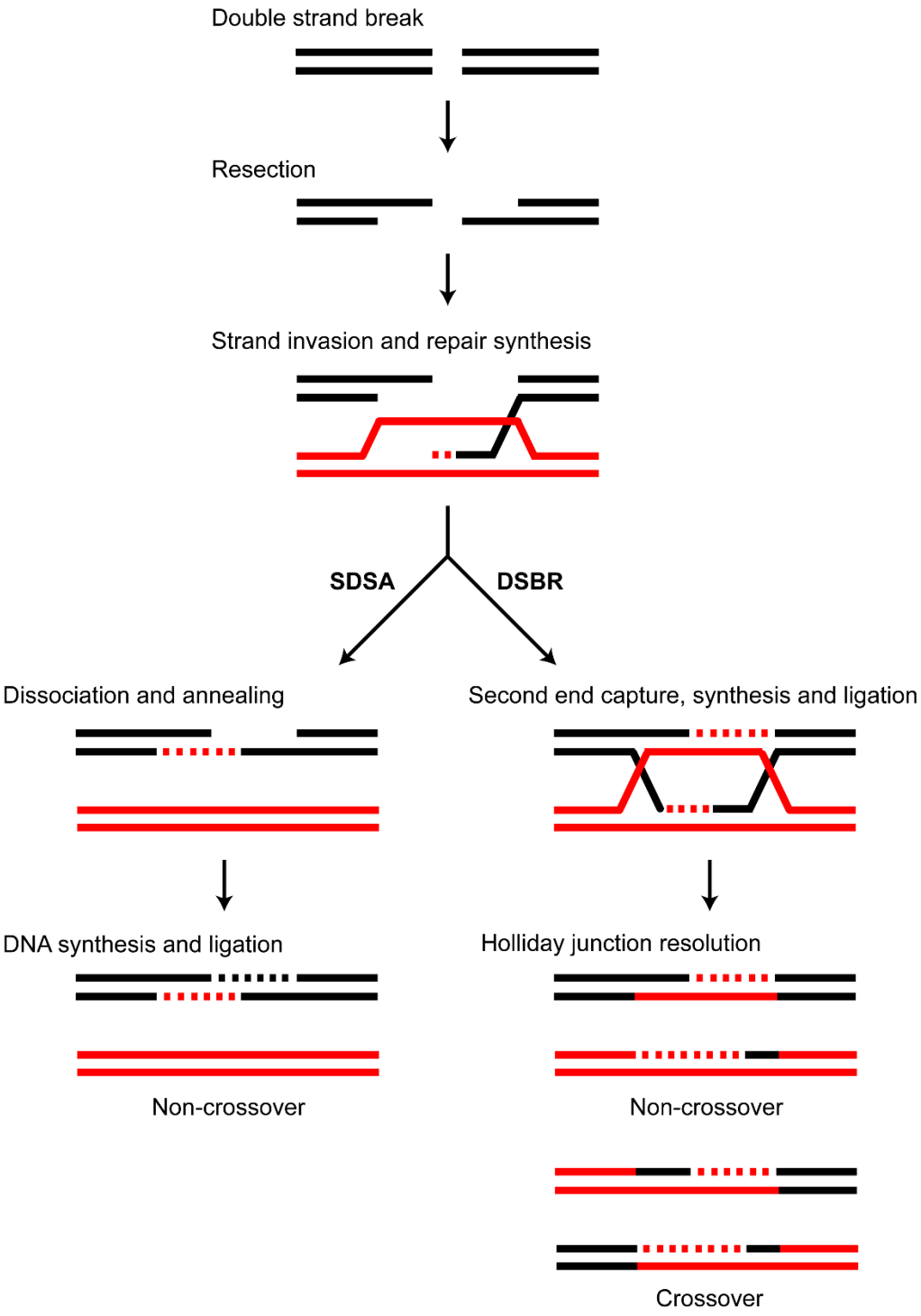


Figure 1. Overview of the eukaryotic HR mechanism.

1.2.2 Non-homologous end joining

In absence of HR-repair donor template or in a mutant strain defective in HR, NHEJ is the dominant DNA repair mechanism of such genomic damage in budding yeast. The three key factors in yeast known to play critical roles in classical NHEJ (C-NHEJ) are as follows: Yku70/80, Dnl4/Lif1/Nej1, and MRX complex (Emerson and Bertuch, 2016). Upon formation of DSB in the genome, Yku and MRX bind to the damaged genomic site and promote protection and bridging of the cut DNA ends. Subsequently, end-processing factors such as Rad27 nuclease and Pol4 polymerase 'polish' the cut DNA ends by deleting mismatched or damaged nucleotides and adding new nucleotides in order to generate compatible ends (Deshpande and Wilson, 2007; Emerson and Bertuch, 2016). Following the end processing step, the Dnl4/Lif1/Nej1 complex, corresponding to DNA ligase IV complex in yeast, is recruited by Yku and is responsible for ligation of the broken ends (Deshpande and Wilson, 2007). Yeast has been found to possess another form of NHEJ which is known as an alternative NHEJ (Alt-NHEJ) pathway that uses 5-25-bp long microhomologous sequences during pairing of broken ends and typically gives rise to base deletions at the repair site (Chiruvella et al., 2013; Daley et al., 2005). The Alt-NHEJ mechanism in yeast, however, still requires further studies as the exact mechanism and the components involved in the pathway need to be identified.

Most DSBs cause complex DNA lesions, often generating incompatible ends with no complementary overhangs that cannot be repaired by simple ligation. NHEJ is able to correctly restore the original DNA sequence if the DNA ends can directly be used for ligation, but the process of concise genome repairing becomes very difficult with non-complementary broken ends. NHEJ is conservative in that it is designed to maximize the possibility of retaining the original sequence of the incompatible repair junction with minimal genetic modifications,

however NHEJ is still prone to the risk of indel errors (Bétermier et al., 2014). Therefore, NHEJ is commonly considered to be error-prone as insertions or deletions of nucleotides with various sizes at the break site are often observed (Aylon and Kupiec, 2004; Pawelczak et al., 2018). The indel errors generated by NHEJ are extremely heterogeneous, differing greatly in terms of length and DNA content (Tao et al., 2018).

1.3 Genetic modification of yeast

In contemporary metabolic engineering, reliable and speedy manipulations of the native and foreign genes involved in metabolic pathways and regulatory networks are essential to achieve desirable metabolic outcomes. The following sub-sections provide descriptions of genetic modification techniques exercised to date that have contributed to the developments of various yeast cell factories in modern research.

1.3.1 Plasmid-based gene expression

In yeast research, various expression plasmid systems have been used to introduce heterologous genes with different levels of plasmid copy numbers. The majority of yeast expression plasmids are 'shuttle vectors', meaning they can propagate both in *E. coli* and *S. cerevisiae*. Shuttle vectors are usually constructed and replicated in *E. coli* and then introduced to *S. cerevisiae* for expression of desired genes. In addition to the bacterial origin of replication and antibiotic selection marker for propagation in bacteria, yeast expression plasmids contain a yeast-specific origin of replication and a selection marker that allows for the maintenance and selection of the plasmid in yeast. Commonly used selection markers include auxotrophic markers and dominant markers.

Auxotrophic markers complement a mutation in essential metabolic enzymes in yeast, meaning only the plasmid-carrying cells can survive in an environment where the corresponding end metabolic product is absent. In yeast, the most commonly used auxotrophic markers are wild-type alleles of genes that encode for key enzymes in metabolic pathways, such as *ADE2*, *LEU2*, *LYS2*, *HIS3*, *TRP1*, *MET15* and *URA3* (Gnügge and Rudolf, 2017). As the use of auxotrophic markers requires availability of yeast strains that possess mutations in the

corresponding metabolic pathways, collections of various types of auxotrophic yeast strains have been developed. Several auxotrophic markers allow counterselection and subsequent removal of the plasmid from yeast cells. For example, *URA3* can be counter-selected by 5-fluoroorotic acid and *TRP1* can be counter-selected by 5-fluoroanthranilic acid, since the use of these chemicals as substrates results in formation of 5-fluorouracil and 5-fluorotryptophan, respectively, which are toxic metabolites (Boeke et al., 1984; Toyn et al., 2000).

Antibiotic markers provide yeast with resistance to certain chemicals that would otherwise be toxic to yeast. As antibiotic resistance markers grant yeast cells a special genotype to grow in presence of certain toxic chemicals, the use of auxotrophic strains is not necessary. The commonly used antibiotic resistance markers in yeast include Kan, Hph, Nat, Ble and Pat, with each conferring resistance to geneticin, hygromycin B, nourseothricin, phleomycin, and bialaphos (in media lacking glucose), respectively (Gnügge and Rudolf, 2017).

The two major types of yeast plasmids used are centromeric plasmids and episomal plasmids. Yeast centromeric plasmids bear two sequences originating from the yeast chromosome: autonomously replicating sequence (ARS) and centromeric sequence (CEN). Due to the presence of ARS and CEN sequences, yeast centromeric plasmids act as though they are small independent chromosomes and are maintained and replicated during cell divisions. Yeast centromeric plasmids typically exist in about single copy per haploid genome, although there are slight variations of copy numbers in a population of cells, with some cells having up to 5 copies and a small subset of cells having none (Gnügge et al., 2016; Mumberg et al., 1995). Episomal plasmids, on the other hand, contain sequences from the 2µ circle, a natural yeast plasmid, and are therefore commonly referred to as 2µ plasmids. 2µ plasmids are typically present in 10-40 copies per haploid genome (Karim et al., 2013), allowing for ‘high-copy’ number propagation of

the plasmids and over-expression of the genes present on the plasmid. Over the years, a series of yeast expression plasmids has been developed and used in yeast research. Some of the commonly used vector series include: pRS plasmid shuttle vector series and p4xx vector series which offer a set of highly-characterized centromeric and episomal plasmids (Christianson et al., 1992; Mumberg et al., 1995).

Although plasmid-based gene expression is an excellent tool for expression of heterologous genes, the variability in plasmid copy number among cells gives rise to heterozygous populations expressing different levels of proteins, and the need for constant selection pressure leads to high maintenance cost (Da Silva and Srikrishnan, 2012; Zhang et al., 1996). Therefore, the unstable nature of plasmid-based gene expression reduces its applicability in a wide range of settings, including that of industrial fermentation. For this reason, HR-mediated genomic integration approach is often preferred as an engineering strategy as it offers homogenous protein expression levels with defined copy numbers of genes among a population of cells.

1.3.2 Genomic modification *via* homologous recombination

For several decades, the vast majority of genome modifications in yeast have been carried out using HR-mediated genomic integration of heterologous DNA. This genomic modification method is a powerful technique that has greatly advanced the processes of knock-in, knock-out, gene tagging, and gene replacement in yeast. The introduction of exogenous genes *via* homologous recombination involves tagging of a gene cassette of interest with a selectable marker, typically an auxotrophic or antibiotic marker, followed by the flanking of the gene cassette-selection marker construct with homology arms which possess DNA sequences specific

to a target site in the genome (Wach et al., 1994). Upon transformation of yeast with the constructed exogenous DNA, yeast cells displaying the desired gene expression and selection marker expression are selected by growing the transformants on selective media. The delivery of the exogenous DNA can be achieved *via* PCR-based approaches, with the use of as little as ~30-bp of homology arms (Kowalczykowski, 2015; Wach et al., 1994). Alternatively, the heterologous DNA can be delivered in an integrative vector carrying one homologous sequence or two homologous sequences for HR-mediated integration *via* single crossover or double crossover, respectively (Gnügge and Rudolf, 2017).

As a unique selection marker is required for each desired genome edit, strain constructions involving addition or deletion of multiple DNA parts are restricted from the small number of dominant selection markers available for use. Therefore, it is well-acknowledged that there is a limit to the number of sequential genetic modifications that can be made with HR-mediated genomic integrations (Chee and Haase, 2012; Solis-Escalante et al., 2012). To overcome this limitation, several studies aimed to identify unique marker genes or to develop novel strategies for recycling of markers. The representative examples of the marker recycling method are Cre/lox method and Flp/FRT method (Gueldener et al., 2002; Storici et al., 1999). In the marker recycling approaches utilizing the Cre/lox and Flp/FRT recombination techniques, selection markers are flanked by recombinase recognition targets (lox or FRT). Upon integration of the selection marker, the sequence-specific activity of Cre or Flp recombinases excises the selection marker, allowing the re-use of the marker for subsequent gene integrations. One drawback of the Cre/lox and Flp/FRT-based methods is that a copy of lox or FRT site remains in the genome after a marker recycling, which could possibly cause genome instability and internal recombinations after multiple rounds of marker rescue (Solis-Escalante et al., 2015). On the

other hand, a novel marker recycling method known as the *delitto perfetto* method enables scar-less removal of markers (Storici et al., 2001). This method requires insertion of a counter selectable marker and reporter gene (CORE) cassette at a target genomic locus *via* HR before the integration of a desired gene can take place. The yeast cells harboring CORE in the genome are transformed with oligos containing a desired DNA sequence, which is designed to replace the CORE cassette. The transformants are selected for the integration of the oligo DNA sequence using counterselection for the loss of the CORE cassette. A subsequent improvement on the *delitto perfetto* method employs a modified CORE cassette including a I-SceI endonuclease cassette under galactose-inducible promoter and I-SceI cut site (Solis-Escalante et al., 2014). With the aid of I-SceI endonuclease, this approach increases the frequency of targeted HR and allows concurrent, scar-less removal of multiple markers.

The HR-mediated genome manipulation techniques have provided powerful means of introducing and expressing exogenous DNA materials in yeast, allowing for subsequent development of yeast cell factory prototypes. However, the procedure for multiple genome modification using different HR-assisted gene integration toolkits remains a time-consuming and laborious process due to the necessary use of selection markers and marker recovery.

1.3.3 CRISPR-Cas9 genome editing

With the rapid advancements in the field of synthetic biology in the last decade, many genome engineering tools based on the CRISPR-Cas9 technology have come into existence to improve on the traditional approaches for the genetic manipulations. The following sub-chapters provide overview on the origin of the CRISPR-Cas9 technology and its use in yeast genome editing.

1.3.3.1 CRISPR-Cas9: origin and mechanism

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas (CRISPR-associated protein) systems were first discovered in *E. coli* and have been proven to be an adaptive bacterial and archaea immune system against invading plasmids and bacteriophages based on RNA-guided DNA target cleavage (Brouns et al., 2008; Garneau et al., 2010). CRISPR systems are categorized into two major classes based on modes of action. The RNA-guided target cleavage in class 1 systems requires a coordinated action of multiple RNA-guided Cas endonucleases, while only one Cas protein is needed to execute target cleavage in class 2 systems (Makarova et al., 2015). Cas9 originating from *Streptococcus pyogenes* is a representative example of a single RNA-guided endonuclease used in class 2 systems and by far the most characterized Cas enzyme to date. In brief, class 2 CRISPR-Cas9 systems work in three stages (the interference step is illustrated in Figure 2A):

1) Adaptation: CRISPR-utilizing organism obtains DNA fragments from invading phages or plasmids and incorporates the obtained DNA in the CRISPR array as spacers that are located in between crRNA repeats (Deltcheva et al., 2011)
2) Biogenesis: pre-crRNA is transcribed from the CRISPR array and processed into mature crRNAs which possess an approximately 20-bp long spacer sequence specifying the

target, as well as a crRNA repeat sequence which associates with an independently transcribed tracrRNA to form a crRNA:tracrRNA complex (Deltcheva et al., 2011).

3) Interference: with the guidance of the crRNA:tracrRNA duplex, Cas9 is able to recognize a target. Between crRNA and target DNA, an RNA:DNA hybrid forms and subsequently, Cas9 is able to induce a cleavage in the invading DNA, preventing the host cell from infection (Jinek et al., 2012).

The presence of a protospacer adjacent motifs (PAM) in a target sequence is a critical requirement for the success of the CRISPR-Cas9 system. Essentially, PAM is a 3-bp long sequence (5'-NGG-3' in *S. pyogenes*) located 3' immediately downstream of the crRNA recognition sequence to which Cas9 recognizes and binds to (Gasiunas et al., 2012). The cleavage site by Cas9 is located 3-bp upstream of the PAM and the DNA break induced by Cas9 is often a blunt-end cleavage.

The CRISPR-Cas9's ability to introduce cleavages in specific nucleotide sequences means that Cas9 can be repurposed to induce programmed DSBs at specific locations in a genome of interest. The programmed DSBs are followed by the responses of DNA repair mechanisms which can be harnessed to introduce various genetic modifications. Specifically, indel mutations generated by NHEJ will enable gene knock-outs and mutagenesis at a targeted genomic site whereas HR mechanism can be exploited to facilitate gene knock-ins and replacements by providing a user-designed repair donor template. Therefore, many efforts to utilize CRISPR-Cas9 as means of eukaryotic genome editing tool have followed after the initial discovery and characterizations of the CRISPR-Cas9 system. The basic design of the CRISPR-Cas9 genome editing proposed in 2012 involved use of *S. pyogenes* Cas9 and a chimeric single guide RNA construct (gRNA) which is a result of fusing crRNA and tracrRNA into a single

RNA molecule (Figure 2B) (Jinek et al., 2012). The combined expression of *Sp*Cas9 with a gRNA enabled target genome cleavages in human cells in 2013, demonstrating the ability of CRISPR-Cas9 as a genome editing tool (Cong et al., 2013; Mali et al., 2013). Over the years, the applicability of CRISPR-Cas9 expanded as the use of CRISPR-Cas9 was demonstrated in model organisms such as zebrafish, budding yeast, fruit fly and mouse, and is now further extending into the genome editing of non-model organisms (Bassett et al., 2013; DiCarlo et al., 2013; Hwang et al., 2013; Komor et al., 2017; Wang et al., 2013)

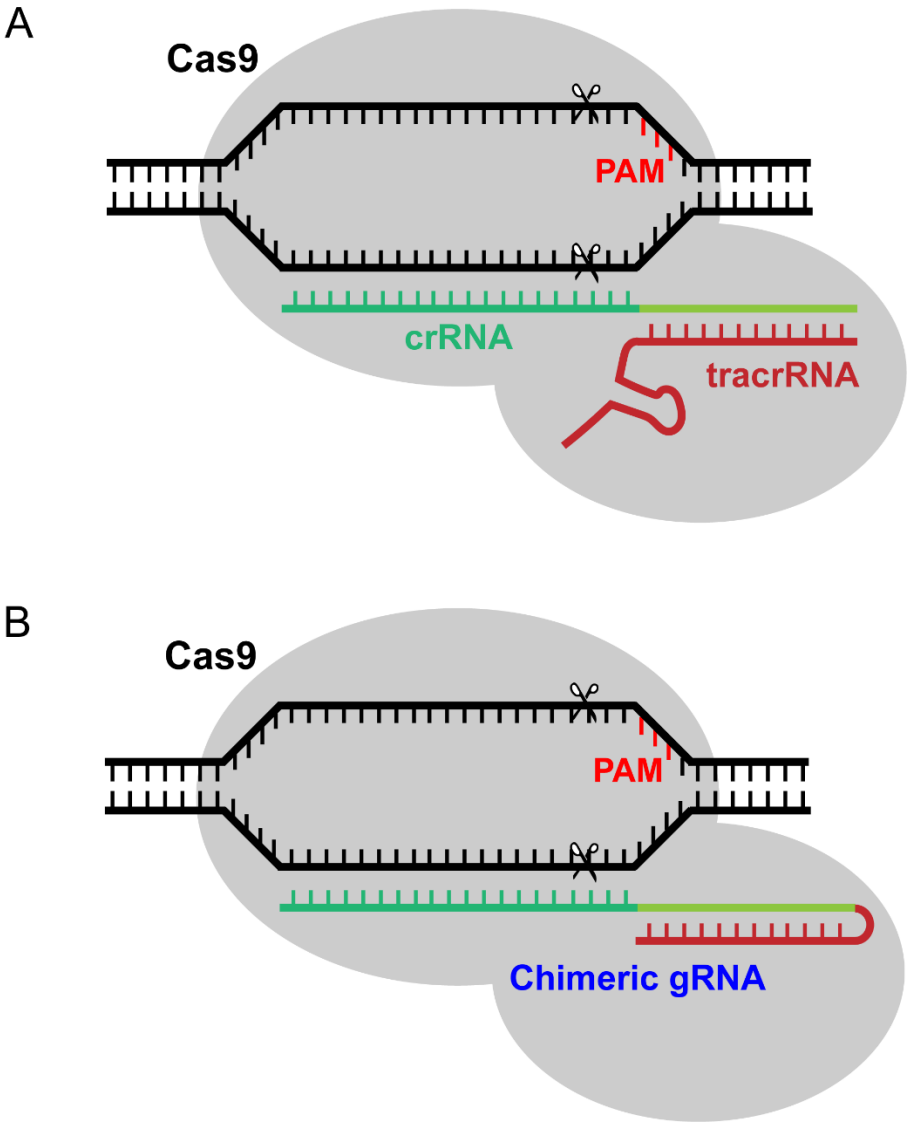


Figure 2. Mode of action of CRISPR-Cas9

(A) Target cleavage by a complex of Cas9 and a naturally occurring crRNA:tracrRNA duplex.
(B) Target cleavage by a complex of Cas9 and a chimeric gRNA that possesses a linker loop.

1.3.3.2 Yeast genome editing using CRISPR-Cas9: overview

The breakthrough of the CRISPR-Cas9 technology was no exception to the yeast research community; since the first example of CRISPR-mediated genome editing yeast in 2013, the yeast research community has seen an increasing use of CRISPR-Cas9 for various genome and metabolic engineering research objectives. The first demonstration of CRISPR-Cas9 in *S. cerevisiae* showed that applying CRISPR-Cas9 in yeast without a repair donor template caused severe toxicity as the NHEJ mechanism does not efficiently operate to repair DSB in yeast (DiCarlo et al., 2013). Furthermore, previous studies showed that the efficiency of HR-mediated gene integration in yeast can increase substantially up to 4,000-fold when DSB is introduced into the genome (Storici et al., 2003). These observations indicate that the toxicity resulting from Cas9-induced DSB can be used as a driving force to significantly boost the process of HR-assisted genomic integration without the need for selection marker use, avoiding the complications associated with selection marker use in genome engineering. This marker-less genomic integration process displays high efficiency, as NHEJ response markedly decreases the number of surviving cells and as a result, the majority of the yeast cells that survive from the Cas9 toxicity in presence of repair templates will be positive for the CRISPR-mediated genomic integration *via* HR (DiCarlo et al., 2013; Mans et al., 2015). Furthermore, the observed toxicity by CRISPR-Cas9 creates a unique opportunity for multiplexing genome edits in yeast by programming DSBs to take place in multiple genomic loci and providing multiple donors for successful scar-less repair of multiple DSBs. These features of CRISPR-Cas9 indicate that researchers now possess a new generation of genome editing tool that greatly advances the capability and efficiency of yeast metabolic engineering.

The foundations of CRISPR-Cas9 genome editing in *S. cerevisiae* were first established in the study by DiCarlo et al. (2013) which modeled their system after the single Cas and single

chimeric gRNA system design by Jinek et al. (2012). In the study by DiCarlo et al. (2013), the delivery of two CRISPR components into yeast was achieved by a plasmid expression of a human-codon-optimized *Sp*Cas9 with SV40 nuclear localization signal and a gRNA construct harboring a 20-bp long target specific sequence and gRNA structural scaffold. This Cas9/gRNA expression strategy accomplished efficient cleavage in the yeast genome and enabled 100% recombination efficiency when a repair template with ~90-bp long homology arms was provided for homology-directed repair (HDR). In the following studies, several different Cas9/gRNA expression strategies have been tested and there are now over two dozen CRISPR-Cas9 toolkits developed for yeast, each employing different methods of Cas9/gRNA expression and design of donor DNA.

Cas9 expression is typically facilitated under the control of a strong yeast promoter (e.g. p*TEF1* and p*TDH3*) from a plasmid or from a Cas9 cassette integrated in the genome, although the platform of Cas9 expression does not hold high importance as the degree of Cas9 expression is not considered to be a critical parameter of CRISPR-Cas9 success (Generoso et al., 2016; Horwitz et al., 2015; Vanegas et al., 2017). With regards to gRNA expression, there has been multiple strategies tested as the optimal transcription of gRNA is important for interaction with Cas9. Polymerase III promoters, such as *SNR52* or tRNA promoters, have been shown to work well for gRNA expression while another gRNA strategy involves flanking of gRNA sequence with HDV and HH ribozymes, allowing the expression of the combined construct under the control of polymerase II promoter (DiCarlo et al., 2013; Gao and Zhao, 2014; Lee et al., 2015). Furthermore, the expression of gRNA cassette is typically carried out on a high-copy 2μ plasmid to ensure the maximum transcript expression.

In general, donor DNAs are provided as single-stranded or double-stranded DNA molecules which typically possess between 40-bp and 500-bp long overhangs that are homologous to a target locus in the genome (Generoso et al., 2016; Horwitz et al., 2015). The PAM sequence is always removed from the donor DNA sequence to prevent subsequent Cas9 cleavage after a genome editing has taken place. The content of the donor DNA depends on the type of application of the CRISPR-Cas9 genome editing. For gene disruption purposes, donor DNAs either contain an indel mutation, a frameshift mutation, a nonsense mutation, or heterologous disrupting sequences such as a barcode sequence, to introduce different changes to the target locus (Bao et al., 2015; Jakočiūnas et al., 2015a; Ryan et al., 2014; Zhang et al., 2014). If the purpose of a genome edit is to introduce a heterologous gene cassette, the gene cassette of interest can be pre-assembled to harbor homology arms prior to being transformed into yeast cells, or *in vivo* assembly approach can be used to combine several overlapping DNA parts to make a complete donor DNA after transformation (Jakočiūnas et al., 2015b; Stovicek et al., 2015). The *in vivo* assembly technique combined with CRISPR-Cas9 allows construction of large pathways and omitting of cloning steps, however the *in vivo* assembly of donor DNAs tends to result in decreased integration efficiencies and requires extensive genotyping (Jakočiūnas et al., 2015b; Mans et al., 2015).

1.3.3.3 Multiplexed CRISPR-Cas9 strategies in yeast

CRISPR-Cas9 can be engineered to direct Cas9 cleavages at multiple loci in the genome, by designing gRNA(s) that target multiple genomic sites. The multiple DSBs induced by Cas9 can be repaired by HDR in presence of multiple donors, allowing knock-outs of multiple endogenous yeast genes or knock-ins of multiple heterologous genes to take place. To execute

multiplex genome editing using CRISPR-Cas9, three major approaches have been practiced in yeast.

The most widely used multiplex strategy is achieved by the expression of multiple gRNAs by individual gRNA expression cassettes. Horwitz et al. (2015) reported introduction of the entire muconic acid biosynthetic pathway by integration of DNA fragments containing multiple genes into three different genomic loci, mediated by the expression of three gRNAs in individual expression cassettes. A very low integration efficiency of 4.2% was observed for this study, likely due to the large sizes of DNA fragments being integrated into the genome (a total size of 24-kb). This approach of using individual gRNA cassettes was further applied in the following independent studies. The CrEdit system (Ronda et al., 2015) demonstrated a concurrent genomic integration of the three carotenogenic pathway genes at the efficiency of 84%. Simultaneous triple gene integrations were also demonstrated by the CasEMBLR system (Jakočiūnas et al., 2015b), which combined the efficacy of CRISPR-mediated targeted integration and *in vivo* assembly of donor DNA parts (a total of 15 parts) to facilitate the re-construction of the carotenoid pathway at 30.6% efficiency.

Another notable multiplex strategy is to express multiple gRNAs by using gRNA constructs flanked by splicing motifs. This approach allows a single promoter to drive the transcription of multiple gRNAs, as all gRNAs are present in one single transcript which is processed to release individual gRNAs. Different types of RNA cleavable sequences demonstrated for use in multiplex settings in yeast include direct repeat elements originating from *S. pyogenes* CRISPR system, self-cleavable RNA moieties (e.g. HDV ribozyme), endogenous RNA processing sequences (e.g., tRNA) or exogenous cleavage-factor recognition sequences (e.g., Csy4) (Bao et al., 2015; Ferreira et al., 2018; Ryan et al., 2014; Zhang et al.,

2019). This multiplexing approach utilizing synthetic gRNA arrays have been exercised for gene disruption purposes, where up to 8 genes were simultaneously knocked out with varying efficiencies.

It was previously found that the performance of a gRNA target sequence is a crucial factor influencing the efficiency of CRISPR-Cas9, as the ability to form a complex with Cas9 and recognize genomic targets is different for every given 20-bp long gRNA target sequence, suggesting that the 'on-target' activities associated with each gRNA play a big role in determining the efficiency of CRISPR-Cas9 (Smith et al., 2016; Thyme et al., 2016). This piece of information indicates that the overall efficiencies of multi-gene integrations of the two aforementioned multiplex strategies are at best limited by the least efficient gRNA. A specificity of a gRNA target sequence is another important property of gRNA to consider, as possible off-target effects can arise from binding of Cas9/gRNA complex at genomic sites other than the intended target genomic locus (Zhang et al., 2015). This possibility of off-target mutations is one of the biggest concerns associated with CRISPR-Cas9 currently. In this respect, the minimization of possible off-target effects is more difficult to accomplish with the multiplex strategies employing multiple gRNA constructs since the presence of more than one gRNA target makes CRISPR-Cas9 genome editing more prone to increased off-target effects. In addition, the two aforementioned multiplex approaches require cumbersome gRNA cloning, albeit the second approach simplifies cloning procedures to some extent.

In order to increase the practicality and simplicity of genome editing, four groups have developed a third approach of multiplex genome editing – targeting of numerous genomic loci simply with one gRNA. In one study, naturally occurring repeat-sequences of Ty1 and Ty2 transposons, present in >100 copies in the yeast genome, were targeted by a single gRNA to

integrate up to 18 copies of donor DNA at 60-80% efficiency (Shi et al., 2016). In other three studies, multiplex genome editing with a single gRNA was achieved by preinstalling synthetic gRNA-binding sites in various genomic sites. The study by Finnigan and Thorner (2016) carried out integration of three foreign genes into *CDC11*, *SHS1*, and *HIS3* loci flanked by identical gRNA-binding sites. This report primarily aimed at providing the proof-of-concept of a single gRNA-mediated, multiplex genome editing by inserting foreign genes in *HIS3*, *CDC11,* and *SHS1* loci. Further investigating the use of synthetic gRNA binding sites for multiplex genome editing, the other two groups employed a 'synthetic DNA landing pad' comprised of a central gRNA-binding site flanked by left and right arms (50-bp – 500-bp) that are pre-engineered in the yeast genome. These synthetic landing pad platforms enabled integrations of foreign genes in pre-determined loci in different numbers, thus benefitting the process of pathway optimization and simplifying the enzyme library screening process by evaluating the dosage effects of a single gene or multiple genes.

1.4 Research rationale, hypothesis, and objectives

The ability of CRISPR-Cas9 to introduce multiple exogenous genes into the yeast genome by a single round of genome-edit greatly benefits the field of yeast metabolic engineering as it allows a speedy construction of synthetic pathways in yeast without selection marker use. Among the different CRISPR-Cas9 multiplex strategies, the multiplex approach utilizing synthetic gRNA binding sites offers a highly-efficient, multi-gene integration platform for pathway construction with the use of a well-characterized synthetic gRNA target that can be thoroughly designed and tested to support optimal gRNA binding parameters. In this regard, Finnigan and Thorner (2016) provided a refined proof-of-concept study of a genome editing

platform based on synthetic target design, however it is not a feasible platform for expression of heterologous genes as *CDC11* is an essential gene required for the survival of yeast and *shs1* mutants show abnormal physiology. The two synthetic landing pad platforms are excellent tools for modulation of gene copy numbers and pathway optimizations (Bourgeois et al., 2018; Hou et al., 2018). However, the synthetic landing pad platforms will integrate foreign genes in genomic loci in a random manner as there is no discrimination of the target loci due to the identical homology arms flanking the gRNA-binding site.

To expand the scope of the yeast multiplex CRISPR toolkits based on synthetic gRNA target design, this project aimed at developing a versatile multiplex gene integration system where a series of yeast strains are engineered to harbour a 23-bp synthetic gRNA binding site from 2- to 5- copies seamlessly in highly characterized intergenic loci without any common homology arms (Figure 3). As each target locus does not share any identical homology arms with other target loci, heterologous genes can be introduced in a locus-specific manner. We refer to this platform as a single gRNA mediated CRISPR (SGM-CRISPR) multiplex gene integration platform.

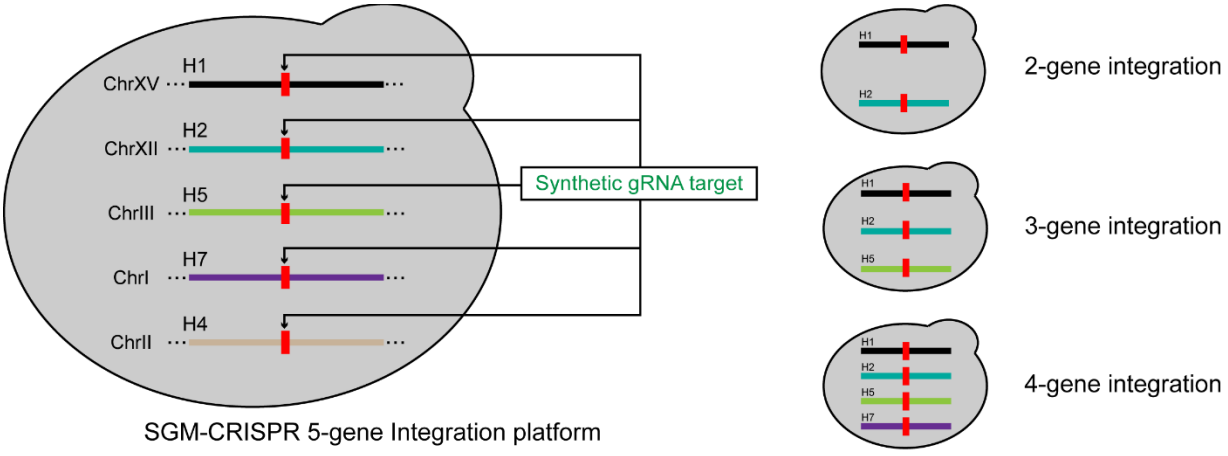


Figure 3. SGM-CRISPR yeast platforms developed in this project

The SGM-CRISPR platform is also equipped with a corresponding Cas9/gRNA plasmid and up to five operating plasmids including ~500-bp long homology arms specifically targeting 2-5 genomic loci (Figure 4). An underlying hypothesis behind the development and use of this multiplex platform is that synthetic gRNA binding sites can be effectively recognized by Cas9/gRNA complex to introduce DSBs and that multiple foreign genes can be simultaneously integrated in the specified genomic loci with known copy numbers in a single transformation, due to the power of simultaneous HR driven by Cas9-induced toxicity (Figure 5 presents an illustration of the hypothesis). Specific objectives of this project are as follows:

1) Identification of intergenic sites in the yeast genome suitable for CRISPR-mediated gene integration.

2) Characterization of the selected intergenic loci by examination of *GFP* transgene integration efficiency, *GFP* expression level, genomic stability of the transgene integration and the effect of transgene insertion on yeast growth.

3) Design of synthetic gRNA targets and assessment for integration efficiency and cell toxicity.

4) Construction of the SGM-CRISPR yeast platforms *via* insertion of a selected synthetic gRNA target into the characterized intergenic loci in a sequential manner.

5) Evaluation of the SGM-CRISPR platform by simultaneous integrations of the three biosynthetic genes involved in the betalain biosynthetic pathway, hygromycin B resistance marker gene, and geneticin resistance marker gene.

6) Application of the SGM-CRISPR platform to synthesize costunolide, a representative sesquiterpene lactone, and its precursor, germacrene A acid.

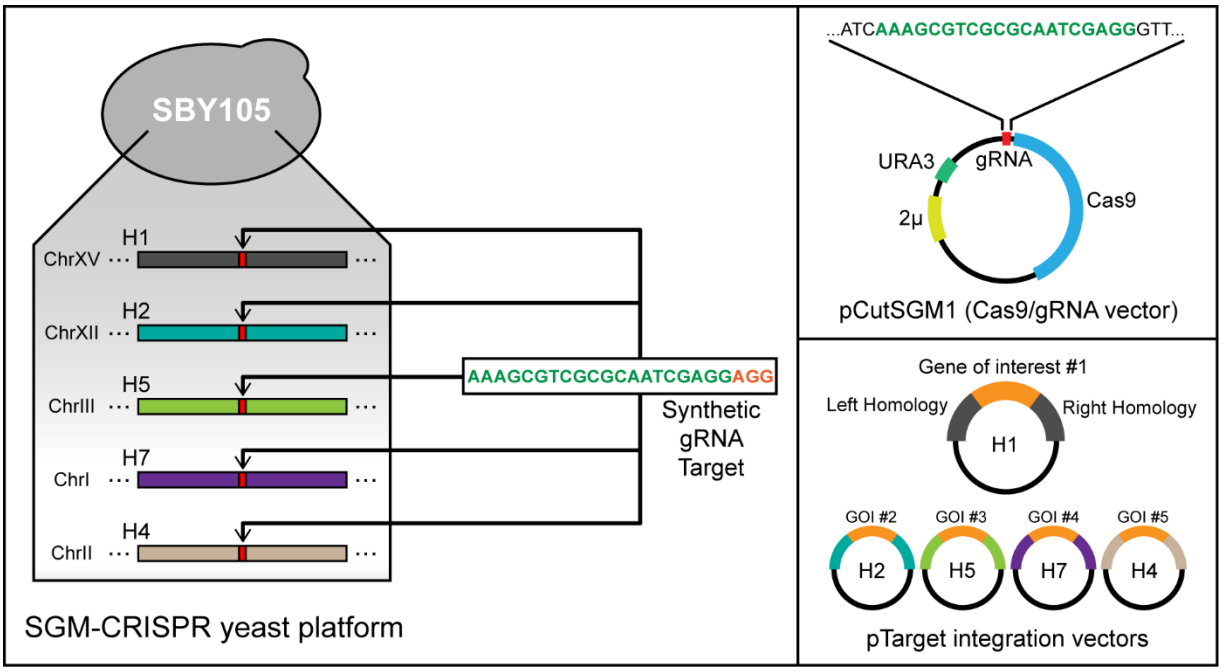


Figure 4. Overview of the SGM-CRISPR platform

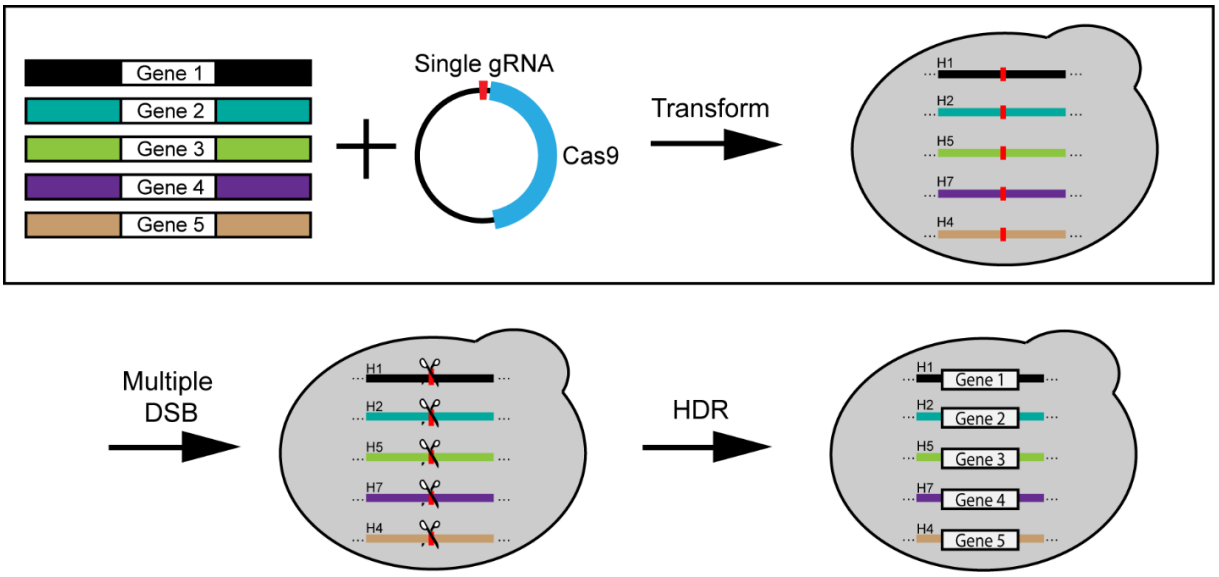


Figure 5. Hypothesized mode of action of the SGM-CRISPR platform

CHAPTER 2: MATERIALS AND METHODS

2.1 Strains, media and growth conditions

All *S. cerevisiae* strains used in this thesis are derivatives of BY4742 and are listed in Table 7. To cultivate yeast, two main types of media were used: synthetic complete (SC) medium (6.7 g/L yeast nitrogen base without amino acids (Bioshop), 1.4 g/L appropriate synthetic dropout mix for SC, SC-URA, SC-URA-LEU, and SC-URA-HIS (Sigma-Aldrich), 20 g/L glucose(Bioshop)) or YPDA medium (10 g/L yeast extract (Bioshop), 20 g/L peptone (VWR Lifesciences), 40 mg/L adenine hemisulfate (Sigma-Aldrich), 20 g/L glucose (Bioshop)). SC-URA medium was used for the selection of the Cas9/gRNA expression plasmids. SC with 5-FOA medium (6.7 g/L yeast nitrogen base without amino acids, 1.4 g/L synthetic compete amino acid mix, 20 g/L glucose, 1 g/L 5-fluoroortic acid (Abcam)) was used for the removal of the Cas9/gRNA expression plasmid from yeast. YPDA medium was supplied with 200 mg/L geneticin (VWR Lifesciences) or 300 mg/L hygromycin B (Thermo scientific) to screen yeast cells harboring antibiotic resistant gene(s). SC-URA and YPDA were supplied with 3 mM L-tyrosine (Sigma-Aldrich) for an increased precursor supplement for production of betalains. SC-URA was supplied with 80 mg/L X-gal (Bioshop) and 10% (w/v) 10X BU salts solution for blue-white colony screening. Yeast inoculations and cultures were carried out at 30 °C and shaking at 200 rpm.

TOP10 *Escherichia coli* strains were cultured in Lysogeny Broth (LB) medium (VWR Lifesciences) supplied with 100 mg/L ampicillin (VWR Lifesciences) at 37°C and shaking at 200 rpm.

2.2 DNA design and construction

All plasmids used in this thesis are described in Table 9. Unless otherwise noted, Gibson Assembly cloning was used for the construction of the plasmids. The oligonucleotides used for DNA amplification and cloning are listed in Table 8. DNA constructs were amplified by Q5® High-Fidelity DNA Polymerase (New England Biolabs) or Taq DNA Polymerase (Applied Biological Materials Inc). PCR products and restriction digest fragments were directly purified by PCR-purification protocol utilizing PB buffer (5 M Gu-HCl and 30% isopropanol) or ran on 0.8% agarose gel containing ethidium bromide and purified with EZ-10 spin column DNA gel extraction kit (Bio Basic).

2.2.1 gRNA target design

CHOPCHOP (www.chopchop.cbu.uib.no/)(Labun et al., 2019) was used to select the endogenous N_{20} gRNA target sequence in the yeast genome for targeting of each CRISPR integration site identified in this thesis. The synthetic gRNA targets used in this thesis were generated using a random DNA generator (www.faculty.ucr.edu/~mmaduro/random.htm) and validated by CHOPCHOP and experimental results. The gRNA targets used in this work have an efficiency score more than 50 on CHOPCHOP and possess zero number of off-targets with 0, 1, 2, and 3 mismatches in the yeast genome. The gRNA expression cassettes were created by overlap extension PCR to stitch Left (p*SNR52*) and Right (gRNA scaffold-t*SUP4*) fragments that had been amplified with 20-bp long overhangs specifying the N_{20} target sequence. Primers #7 to #34 and #103 to #106 were used to amplify the Left and Right fragments from p426-SNR52p-gRNA.CAN1.Y-SUP4t (Addgene #43803) prior to gRNA cassette construction.

2.2.2 Cas9/gRNA vectors

The three plasmids used for Cas9 and/or gRNA expressions in this thesis are as follows: p416-CAS, pLEU-gRNA, and pCut. To create the p416-CAS plasmid, Cas9 cassette was amplified from p414-TEF1p-Cas9-tCyc1 (Addgene #43802) and cloned into the p416 plasmid that had been linearized by *KpnI*/*SacI* digestion. A gRNA cassette was amplified with primers #1 and #2 and cloned into the *KpnI* site of the p416-CAS plasmid to generate a series of p416-CAS plasmids targeting the endogenous gRNA binding site present in each CRISPR integration locus. The construction of the pLEU-gRNA plasmid involved amplification of a gRNA cassette with primers #3 and #4 and cloning of the gRNA cassette into the *PvuII* site of pESC-LEU. To make the pCut plasmid, Cas9 cassette was amplified by primers #5 and #6 and cloned into the *PvuII* site of pESC-URA. A gRNA cassette was amplified with primers #1 and #4 and cloned into the *KpnI* site of the pCut plasmid to generate a series of pCut plasmids targeting the endogenous gRNA binding site present in each CRISPR integration locus or pCutSGM plasmids targeting the synthetic gRNA binding site installed in the genome of the engineered yeast strains (Figure 6).

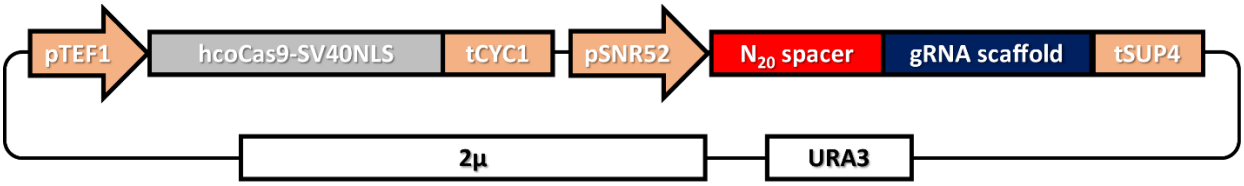


Figure 6. Plasmid map of the pCut plasmid carrying a gRNA cassette

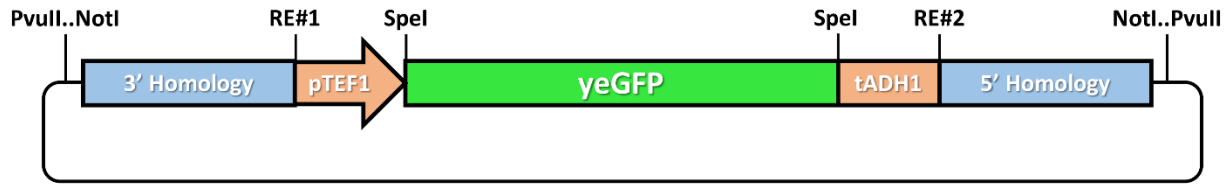
2.2.3 CRISPR integration vectors for R integration sites

To create pTarget integration plasmids for the R1 to R6 integration sites, the intergenic DNA fragments of R1-R6 were amplified from BY4742 genomic DNA using primers #43 to #54. These fragments were then cloned into the multiple cloning site (one of *EcoRI/SpeI*, *KpnI/SpeI*, or *KpnI/ApaI* sites) of pBluescript SK to yield pTargetR1 to pTargetR6 plasmids. To create pTargetR plasmids for *GFP* and *LacZ* cassette integrations, *GFP* and *LacZ* cassettes were amplified using primers #55 to #60, digested either with *KpnI* or *BamHI*, and ligated to the *KpnI* or *BamHI* site of the pTargetR plasmids using T4 DNA ligase (New England Biolabs).

2.2.4 CRISPR integration vectors for H integration sites

To create integration plasmids for H1 to H8 integration sites, 5'- and 3'- homology arms, each approximately 500-bp long, surrounding the PAM site of the gRNA target of each integration site were amplified from BY4742 genomic DNA using primers #61 to #92 with ~25-bp long flanking homology to the *GFP* cassette. Using four-piece Gibson assembly, a 5'-homology arm, *GFP* cassette, 3'-homology arm were cloned into the multicloning site (one of *HindIII/SpeI* or *SalI/NotI* sites) of pBluescript SK (Figure 7). For genomic integration of other genes of interest, pTargetH-GFP plasmids were digested with *BamHI/PstI*, *BamHI* alone, *AvrII/EcoRI*, or *AvrII/HindIII* restriction enzymes to remove the *GFP* cassette and linearize the plasmid (Figure 7). The betalain biosynthetic gene cassettes were amplified from pCMC0756 and pPSG0348 (Grewal et al., 2018) using primers #121 to #126 and cloned into the linearized pTargetH plasmid. The *hphNT1* and *kanMX4* marker gene cassettes were amplified from pYM25 and pYM13 (Janke et al., 2004) using primers #127 to #130 and cloned into the linearized pTargetH plasmid. The sesquiterpenoid biosynthetic gene cassettes were amplified from pESC-

Leu2d::*LsGAS*/*LsGAO*/*AaCPR*/*LsCOS* (Ikezawa et al., 2011) and pESC-HIS::*BTS1*/*tHMGR* using primers #131 to #140 and cloned into the linearized pTargetH plasmid.



pTarget plasmid	RE#1	RE#2
pTargetH1-GFP	PstI	BamHI
pTargetH2-GFP	BamHI	BamHI
pTargetH3-GFP	EcoRI	AvrII
pTargetH4-GFP	EcoRI	AvrII
pTargetH5-GFP	EcoRI	AvrII
pTargetH6-GFP	EcoRI	AvrII
pTargetH7-GFP	HindIII	AvrII
pTargetH8-GFP	EcoRI	AvrII

Figure 7. Plasmid map of the pTargetH-GFP plasmid with unique enzyme sites.
The GFP cassette can be excised out from each pTargetH-GFP plasmid using different sets of restriction enzymes as shown in the table.

2.3 Yeast transformation

The transformation of plasmids and linearized donors into *S. cerevisiae* was carried out according to a modified lithium acetate transformation protocol (Gietz and Robert, 2007). An overnight culture of *S. cerevisiae* in 5 mL of YPDA was diluted 25-fold in 50 mL of YPDA. After 2-4 hours of cultivation, the 50 mL yeast culture was spun down at 3,000 rpm and the resulting pellet was washed once with ddH_2O. The washed yeast was divided into 5-10 aliquots, and each aliquot was used for one transformation. For single gene integration, 1 µg of donor

DNA and 500 ng of Cas9/gRNA expression plasmid was co-transformed into yeast that had been re-suspended in a half-volume of the transformation mastermix (120 µL polyethylene glycol, 18 µL lithium acetate, 25 µL salmon sperm DNA). For double, triple, quadruple and quintuple gene integrations, yeast was re-suspended in a full-volume of the transformation mastermix (240 µL polyethylene glycol, 36 µL lithium acetate, 50 µL salmon sperm DNA) and 1-2 µg of each donor DNA and 0.5-1.0 µg of pCutSGM1 were used for transformation. The yeast transformants were then plated on SC-URA, incubated at 37 °C for one day, and finally incubated at 30 °C for three days. Linearized donors were generated by either linearization of the pTarget integration plasmids by *NotI* or *PvuII* or PCR-amplification of the desired donor sequence from pTarget prior to being transformed into yeast.

2.4 Genomic DNA extraction and genotyping

The extraction of yeast genomic DNAs was carried out using the Yeast DNA Extraction Reagent Kit (Thermo Scientific) according to the manufacturer's protocol. Three mL of overnight yeast culture in YPDA was used for the genomic DNA extraction. Approximately 200 ng of genomic DNA was used for PCR genotyping to confirm the correct integration of donor DNA fragments. When appropriate, the CRISPR target regions showing positive genotyping data were PCR-amplified from the genome and sequenced in order to check the correctness of the genomic integration at a base-pair resolution. Integration efficiency was calculated using the following formula: $\left(100\% \times \frac{Number\ of\ colonies\ positive\ for\ integration}{Total\ number\ of\ colonies\ screened}\right)$.

2.5 Fluorescence analysis

A blue LED transilluminator (IO Rodeo) was used to distinguish between fluorescent and non-fluorescent *S. cerevisiae* colonies on a transformation plate. *GFP* was excited by LED blue light at 470 nm and the fluorescence of individual colonies was observed through an amber viewing filter in dark. When it was required to further visualize individual fluorescing colonies, a Zeiss Axio Imager Z2 upright epifluorescence microscope was used to acquire individual images of each colony. ZEISS Zen blue imaging software and Zeiss plan Apochromat 100 x objective lens was used for image acquisition. Colibri 7 LED light and 90 High Efficiency (HE) filter sets were used for excitation of *GFP*. *GFP* was excited at 470/40 (band width) nm and *GFP* emission was detected at 525/50 nm range.

The expression levels of CRISPR integration sites were assessed by flow cytometry analysis using the FACSCalibur Flow Cytometer (BD Biosciences). The yeast colonies positive for *GFP* cassette integration were first streaked on SC with 5-FOA for removal of the pCut plasmid and consequently on YPDA for further growth. For flow cytometry experiments, the resulting yeast colonies were inoculated in 5 mL of YPDA overnight. The overnight cultures were then diluted to OD_{600} 0.1 prior to being injected to the Flow Cytometer. For fluorescence measurements, a total of 10,000 events were recorded at a flow rate of 60 μL min^{-1}. *GFP* was excited at 488 nm and *GFP* emission was detected at 530 nm. The BD CellQuest™ Pro software was used to analyze the obtained data.

2.6 Growth assay and gene stability assay

S. cerevisiae colonies harboring a *GFP* cassette in the selected CRISPR integration sites were cultured overnight in 5 mL of YPDA. The overnight culture was then diluted 10-fold in 30

mL of YPDA. The 30 mL culture was cultured for a total of 48 hours and the growths of the yeast cultures were recorded by measuring OD_{600} every 1-2 hours until the stationary phase was reached. The growth rate was calculated using the following formula: $\lambda = \frac{\Delta \log_2 OD}{\Delta t}$ (Hall et al., 2014).

S. cerevisiae strains harboring heterologous DNA materials in the selected CRISPR integration sites were tested for stability of the gene integration by repeated passages of growth in YPDA. The yeast cells were first cultured in 400 µL of YPDA in 96 deep-well plates overnight. After 24 hours of cultivation, 20 µL of each culture was transferred over to 400 µL of fresh YPDA. This was repeated for a total of four or five times. At the end of the cultivation after the four or five passages of growth, the yeast cells were plated on YPDA plates and the genomic DNA of each yeast culture was obtained. PCR genotyping reactions were performed to screen the presence of the heterologous DNA at each CRISPR integration locus.

2.7 Evaluation of synthetic targets

To evaluate the integration efficiency and toxicity of each sTarget, *S. cerevisiae* strains harboring sTarget in the H1 integration site (SBY101 and SBY201) were transformed with 500 ng of pCutSGM in absence or presence of 1 µg of *GFP* cassette donor. The integration efficiency of each sTarget was evaluated by assessing the numbers of fluorescing colonies out of the total number of resulting colonies on a transformation plate. Furthermore, eight colonies were randomly picked up from the transformation plate to be genotyped and confirmed for the correct *GFP* integration. The toxicity of each sTarget was evaluated by comparing the numbers of surviving colonies in absence and presence of the *GFP* donor. The wild-type BY4742 strain was included as a negative control in both transformation experiments.

2.8 Strain building

S. cerevisiae strains harboring varying numbers (from 2 to 5) of the sTarget#1 were created by CRISPR-assisted integration of the sTarget#1 into the yeast genome in a sequential manner. The homology arms of a given CRISPR integration site were PCR-amplified with the primers that contain a 23-bp overhang specifying the synthetic gRNA target with 5'-AGG-3' PAM sequence. The two DNA fragments were then stitched together using overlap extension PCR and purified using PCR-purification method. The co-transformation of 250 ng of the resulting donor and 500 ng of pCut plasmid into BY4742 was performed. From the resulting colonies, two to four colonies were screened for the correct integration of the sTarget#1 by genotyping and sequencing. The yeast colonies positive for the sTarget#1 integration were streaked on SC with 5-FOA for removal of the pCut plasmid before the consequent rounds of transformation. The resulting strains utilized in the SGM platform are described in Table 7.

2.9 Yeast metabolite extraction

S. cerevisiae strains being tested for sesquiterpenoid production were cultured overnight in 5 mL of synthetic complete medium with 2% (w/v) glucose. The overnight culture was then diluted 10-fold in 30 mL of the synthetic complete medium with 1.8% (w/v) galactose, 0.2% (w/v) glucose, and 100 mM HEPES buffer. The buffered medium was used to prevent acid-induced cyclization of STL compounds. The yeast was then cultured for a total of 72 hours and the yeast metabolites were extracted with 10 mL of ethyl acetate. The ethyl acetate fractions were evaporated using a rotary evaporator and the metabolites were dissolved in methanol prior to further analysis.

2.10 LC-MS analysis of yeast metabolites

An Agilent 1290 Infinity II LC system (Agilent Technologies) was used for LC-MS analysis of the yeast metabolite samples. 0.1 µL or 1 µL of the samples were injected onto an Agilent Zorbax Eclipse Plus C18 column (2.1 X 50 mm, 1.8 µm particle size; Agilent Technologies). Metabolites were separated using a flow rate of 0.35 ml min^{-1}, and the solvent gradient of 90:10 (A:B) to 0:100 (A:B) for 10 minutes (A: 1 mM ammonium fluoride; B: 100% methanol). The initial 0.5 minute was ran in an isocratic mode with solvent composition of 90:10 (A:B). The high-resolution mass data was obtained using electrospray ionization followed by Q-TOF on the 6545 LC/Q-TOF (Agilent Technologies). The Agilent MassHunter Software was used to analyze the obtained LC-MS data.

CHAPTER 3: CHARACTERIZATION OF CRISPR GENOMIC INTEGRATION SITES IN YEAST

3.1 Evaluation of R intergenic sites for CRISPR-mediated genomic integration

One of the basic requirements for CRISPR-Cas9 genome engineering is the availability of genomic integration sites to which heterologous genes can be placed and expressed. The genomic loci for heterologous gene integrations need to be selected delicately and characterized thoroughly to ensure that the transgenes can be inserted into the loci efficiently and expressed in a reliable manner without affecting the normal physiology of the host organism. In order to systematically identify genomic loci that are suitable for gene integration and expression by CRISPR-Cas9, 6692 intergenic DNA sequences were retrieved from the Yeast Genome Database (yeastgenome.org) and used to identify 'safe harbor sites' for heterologous gene expression. The search for suitable integration site candidates entailed looking for the intergenic DNA sequences that possess a single *BamHI* recognition sequence (5'-GGATCC-3') or *KpnI* recognition sequence (5'-GGTACC-3') flanked by at least 0.5-kb of neighboring DNA. The rationale of this selection criteria is that the presence of the GG and CC motifs in the *BamHI* and *KpnI* recognition sequences automatically generates two protospacer adjacent motifs (PAM), thus simplifying the gRNA target selection process. Additionally, the presence of either of these restriction enzyme sites simplifies the cloning of donor DNA fragments between the left and right homology arms. Furthermore, the requirement for the presence of at least 0.5 kb-long intergenic DNA sequence surrounding the gRNA target minimizes the possibility of disrupting the expression of nearby genes, as the majority of the transcriptional regulator binding sites in the *S. cerevisiae* genome exist within 400-bp upstream from the start codon of open reading

frames (Harbison et al., 2004). The search yielded 43 integration site candidates which were further narrowed down to six candidates that possess a N_{20} gRNA target sequence with excellent on-target activities and minimal numbers of off-targets predicted by the gRNA design tool. These six integration site candidates are referred to as R1-R6 (Table 1).

Table 1. Description of the R intergenic sites used for CRISPR-mediated gene integration

Site name	Chromosomal coordinates	Genes flanking the intergenic region	Size of intergenic region (bp)	N_{20} gRNA target
R1	ChrXIII: 13175-14540	*ERO1* *COX14*	1,366	CATCATTTCAGACACACTGG
R2	ChrX: 717631-719665	*YJR151W-A* *DAL5*	2,035	CATCAGCACAATATCCATGG
R3	ChrX: 194874-196286	*NCA3* *ASF1*	1,413	GCATAATCGGCCCTCACAGA
R4	ChrVI: 259434-263956	*YFR054C* *YFR056C*	4,523	AGTTCGGCAAAGAGCATGGA
R5	ChrXI: 321519-323227	*YKL063C* *MSN4*	1,709	GTAAAGTACGAGCGCGGGGA
R6	ChrXVI: 640959-642207	*ARP7* *GLN1*	1,249	CACTCGCGAGAACCAAAACA

The efficiency of CRISPR-Cas9 mediated gene integration for each of the R1-R6 sites was first examined using *LacZ* expression cassette under *TDH3* promoter as a donor, with 0.5-kb of homology arms. The *LacZ* donor DNA fragments were PCR-amplified from pTarget integration plasmids and transformed into BY4742 yeast strain with p416-CAS, the plasmid with

the CEN6/ARS4 origin of replication expressing Cas9 and gRNA. The blue-white screening of the resulting colonies revealed that less than 20% of the colonies were positive for blue pigment formation for all of the six R intergenic loci tested (Table 2). Additionally, over different number of biological replicates, a transformation plate entirely consisting of white colonies (i.e. no *LacZ* insertion) was observed at least once for each R intergenic site. The genotyping of eight randomly picked colonies showed the integration efficiency of 0% for three loci (R1, R2, and R4) while the R3 site had two out of the eight screened colonies showing positive signal for integration, equating to 25% efficiency. As the *LacZ* cassette is a considerably large reporter gene cassette with the size of ~4-kb, the genomic integration of ~1.5-kb long *yeGFP* cassette under *TEF1* promoter was examined for the R1-R4 sites. It was anticipated that the use of a smaller-sized donor will lead to better genome editing outcomes, as smaller-sized donors are more easily integrated into the genome upon DSB formation and are also more readily transformed into yeast (Horwitz et al., 2015). The *GFP* integration efficiency based on fluorescent colony screening, however, showed similarly low integration efficiencies ranging between 0% and 19.0% (Table 2). Nonetheless, it is notable that the integration efficiency based on random genotyping showed improvements, with the R2, R3 and R4 sites showing 25%, 12.5% and 25% of the screened colonies, respectively, with the desired gene integration (Table 2).

The observed efficacy of the R intergenic sites for heterologous gene integration is too low, even with the use of small-sized donors, to be able to accommodate the downstream application of multiplexed genome editing, especially considering that the gene integration efficiency tends to decrease with increasing number of edits by CRISPR-Cas9 (Bourgeois et al., 2018; Jakočiūnas et al., 2015b). The low integration efficiencies associated with the R

integration sites could be due to many possible factors including the poor chromatin accessibility of the DNA target regions, low cleavage efficiencies of the gRNA targets used, or insufficient expression of Cas9 and/or gRNA. Overall, the results from the preliminary reporter gene integration experiments indicated that an improved design of CRISPR-Cas9 is necessary to facilitate more favorable gene integration outcomes in yeast.

Table 2. Efficiency of *LacZ* and *GFP* cassette integration for the R intergenic sites

The BY4742 yeast strain was co-transformed with Cas9/gRNA plasmid with CEN/ARS origin and linearized donors encoding *LacZ* or *GFP* cassette. A total of 8 colonies were randomly picked from the transformation plates and genotyped to confirm the integration of the reporter gene cassette. N/A : not applicable.

Site name	***LacZ* Integration efficiency**				***GFP* Integration efficiency**		
	Blue-white colony screening			Random Genotyping	Fluorescent colony screening		Random Genotyping
Trial#	1	2	3	1&2&3	1	2	1&2
R1	6.7% (4/60)	10% (7/70)	N/A	0% (0/8)	1.9% (1/52)	9.8% (4/41)	0% (0/8)
R2	0% (0/7)	17.4% (4/23)	N/A	0% (0/8)	19.0% (4/21)	17.4% (8/46)	25% (2/8)
R3	0% (0/3)	18% (3/17)	17% (4/23)	25% (2/8)	0% (0/9)	18.8% (3/16)	12.5% (1/8)
R4	15.8% (3/19)	0% (0/14)	N/A	0% (0/8)	17.5% (7/40)	18.2% (6/33)	25% (2/8)
R5	0% (0/6)	0% (0/12)	7.3% (3/41)	0% (0/8)	N/A		
R6	0% (0/21)	0% (0/2)	5.2% (4/76)	0% (0/8)			

3.2 Evaluation of H intergenic sites and different Cas9/gRNA expression platforms

Previous studies demonstrating CRISPR-Cas9 genome editing in eukaryotic organisms reported that local chromatin accessibility can influence Cas9 cleavage efficiency, and thus genome editing by CRISPR-Cas9 is more efficient in relaxed euchromatic regions than in compact heterochromatic regions (Jensen et al., 2017; Uusi-Makela et al., 2018). Based on this knowledge, it is possible that the selection criteria used to choose the R1-R6 intergenic site candidates could have coincidentally led to selecting loci with lengthy intergenic regions with poor chromatin accessibility, inhibiting efficient target DNA cleavage by Cas9. Therefore, it was reasoned that the efficiency of Cas9-mediated genomic integration can be improved by using genomic sites adjacent to more "open" euchromatic regions that would offer a better accessibility to the genomic target. To test this idea, a total of four transcriptionally active genomic sites were identified. Firstly, the two intergenic loci located adjacent to *ADH1* and *PDC1*, two genes known to belong to the top hundred most highly expressed genes in yeast (Velculescu et al., 1997), were identified and were named H1 and H2 (Table 5). The N_{20} gRNA target sequences for the H1 and H2 sites (H1: CCAATGCTAGTAGAGAAGGG; H2: GAGATAGGTAAATAAACGCG) are located at least 500-bp upstream of the start codons of *ADH1* and *PDC1* coding sequences, respectively. Secondly, the promoter regions of *EFT1* and *APE2*, two non-essential genes that belong to the top two hundred most highly expressed genes in yeast (Velculescu et al., 1997), were identified and named P1 and P2. The N_{20} gRNA target sequences for the P1 and P2 sites (P1: GTCTGGCTTTCTGGCAAATG; P2: TGTATGTTACACGTGAAGCG) are located within 300-bp upstream of the start codons of *EFT1* and *APE2* coding sequences, respectively. Notably, the transcript levels of *ADH1*/*PDC1* and *EFT1*/*APE2* are 100-300 times and 10-50 times higher, respectively, than that of the genes flanking the R1-R6 loci based on the available yeast

transcriptome data (Velculescu et al., 1997). In addition to selecting new loci types, different plasmid systems for improved expression of Cas9 and gRNA were also examined. For this purpose, a high-copy 2µ plasmid expressing both Cas9 and gRNA (referred to as pCut plasmid) and a high-copy 2µ plasmid expressing just gRNA (referred to as pLEU-gRNA plasmid) were developed (Figure 8).

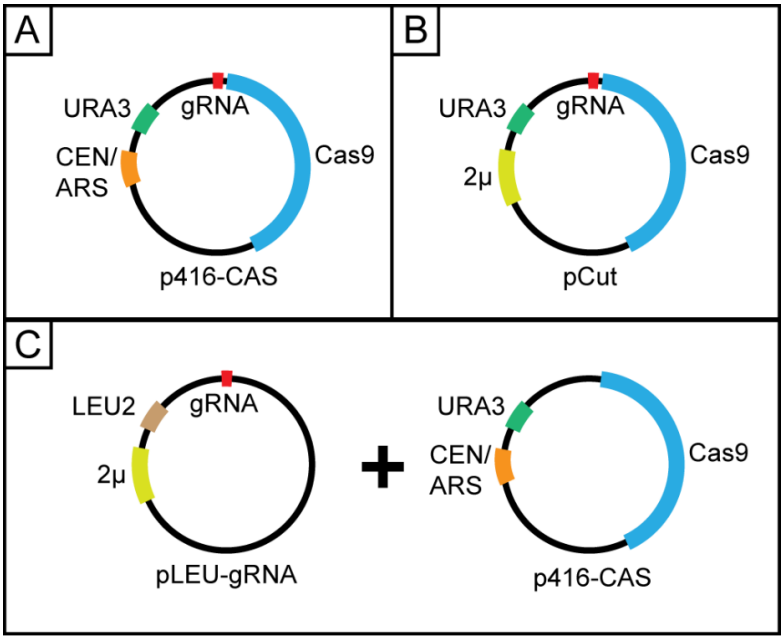


Figure 8. Different Cas9/gRNA plasmid systems used for the single gene integration experiment

Subsequently, the efficiency of *yeGFP* cassette integration for the H1, H2, P1 and P2 sites using different plasmid systems was evaluated, with the R1 and R2 sites as controls. The results from these experiments showed that the selection of CRISPR target loci is critically important for CRISPR-Cas9 efficiency. The overall efficiencies of *GFP* integration for the H1, H2, P1 and P2 integration sites were between 62.5-100%, remarkably higher than 12.5-37.5% observed for the R1 and R2 sites (Table 3). Among the different Cas9/gRNA expression systems tested, the co-

expression of Cas9 and gRNA in a single plasmid (with either CEN/ARS or 2μ origin) resulted in integration efficiencies between 87.5-100% for the H1, H2, P1 and P2 sites (Table 3). On the other hand, expressing gRNA in a high-copy pLEU-gRNA plasmid and Cas9 in a low-copy p416-CAS plasmid resulted in relatively lower efficiencies between 62.5-75%. These results give a strong rationale that the use of transcriptionally active genomic sites for Cas9 targeting provide better outcomes for CRISPR-Cas9. These results also indicate that the co-expression of both gRNA and Cas9 in either low-copy or high-copy plasmid results in similarly high integration efficiencies. Therefore, the H1 and H2 integration sites and a single pCut plasmid system for Cas9 and gRNA expression were determined to be used for further CRISPR-Cas9 experiments moving forward.

Table 3. Efficiency of single gene integration: different integration sites and different Cas9/gRNA expression platforms

The BY4742 strain was transformed with Cas9/gRNA plasmid(s) and *GFP* donor and the *GFP* integration efficiency was determined by the genotyping of 8 randomly selected colonies. CEN/ARS refers to a single-copy plasmid with a CEN6/ARS4 origin and 2μ refers to a high-copy plasmid with a 2μ origin. R: intergenic locus bearing *BamHI* or *KpnI* recognition sites; H: intergenic locus located next to a gene showing a high transcription level; P: promoter region of a non-essential gene showing a high transcription level.

Site Name	*GFP* cassette integration efficiency		
	CEN/ARS: Cas9 & gRNA	2μ: Cas9 & gRNA	CEN/ARS: Cas9 2μ: gRNA
R1	12.5% (1/8)	12.5% (1/8)	12.5% (1/8)
R2	25% (2/8)	37.5% (3/8)	12.5% (1/8)
H1	100% (8/8)	87.5% (7/8)	62.5% (5/8)
H2	87.5% (7/8)	100% (8/8)	62.5% (5/8)
P1	100% (8/8)	100% (8/8)	75% (6/8)
P2	100% (8/8)	100% (8/8)	75% (6/8)

In order to firmly establish that the high integration efficiencies of marker-less gene integrations seen with the newly identified genomic loci are solely due to Cas9-induced toxicity and not by random integration events, a control experiment testing the genomic integration of *His3MX6* donor, a well-characterized selectable marker, into the H2 site was carried out in presence of both Cas9/gRNA expression, in absence of one of Cas9 or gRNA expression or in complete absence of both Cas9/gRNA expression (Figure 9).

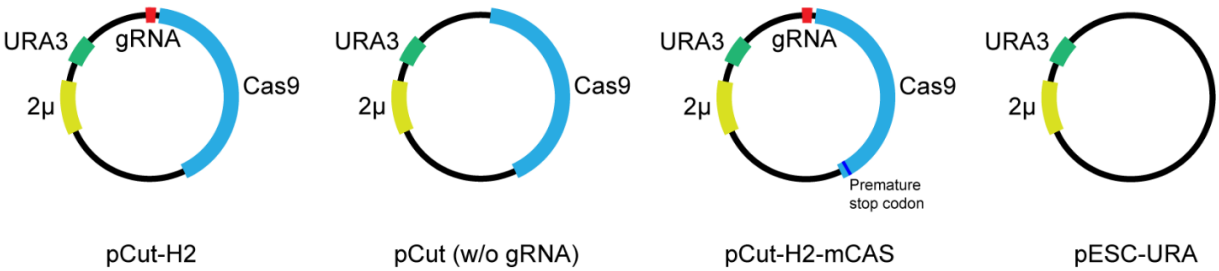


Figure 9. Different plasmid systems used for the single gene integration control experiment

The BY4742 strain co-transformed with *His3MX6* donor and each type of plasmid was plated on SC-URA, selecting for just the plasmid, or SC-URA-HIS, selecting for both the plasmid and the donor. The efficiency of genomic integration was assessed by PCR-genotyping of 8 randomly selected colonies. As expected, when the −HIS selection for the marker donor was absent, the integration of *His3MX6* donor could take place only when both Cas9 and gRNA were present, as evident from the 100% integration efficiency observed in presence of both Cas9 and gRNA expression and 0% integration efficiency resulting from the lack of either one of Cas9 or gRNA expression (Table 4). When donor DNA was not provided to yeast, the cell toxicity by DNA cleavage activity of Cas9 and gRNA was clearly evident as no colony survived in presence of Cas9 and gRNA expression, whereas expressing either Cas9 or gRNA only showed no cell

toxicity and allowed thousands of colonies to form (Table 4). These results, combined together, showed that both Cas9 and gRNA are required for marker-free integration of donor DNA at a desired target locus. Different outcomes were observed when the −HIS selection for the donor was present, as all of the eight screened colonies had the desired integration at the H2 site regardless of Cas9 or gRNA expression (Table 4), indicating that selection pressure for the marker donor is absolutely necessary for successful genomic integration when Cas9-induced DSB is missing. One noteworthy observation is that the number of surviving colonies observed for CRISPR-mediated marker-free genomic integration (1100) is approximately 5-10 fold higher than the number of colonies resulting from marker-dependent HR-mediated gene integration (101-177) (Table 4), consistent with the previous report that DSBs substantially increase the likelihood of HR in yeast (Storici et al., 2003). Overall, the results of the control experiments reinforced the notion that both Cas9 and gRNA are crucial components for the marker-free CRISPR-Cas9-mediated HDR in yeast.

Table 4. Efficiency of single gene integration: marker-free or marker-dependent gene integrations

The BY4742 strain was transformed with plasmids containing or lacking Cas9 and/or gRNA cassettes in presence or absence of *His3MX6* donor. A total of eight colonies from each transformation plate were genotyped to confirm the proper integration of the donor DNA. '+' indicates the delivery of a corresponding CRISPR component into yeast while '−' indicates the lack of delivery of a corresponding CRISPR component into yeast. N/A: Not Applicable.

Selection pressure	Cas9	gRNA	dDNA	Colony Number	Integration Efficiency
No selection for donor marker	+	+	+	1100	100% (8/8)
	−	−	+	2000	0% (0/8)
	+	−	+	1350	0% (0/8)
	−	+	+	1500	0% (0/8)
	+	+	−	0	N/A
	−	−	−	1720	
	+	−	−	1800	
	−	+	−	1650	
Selection for donor marker **(-HIS medium)**	+	+	+	1200	100% (8/8)
	−	−	+	114	100% (8/8)
	+	−	+	101	100% (8/8)
	−	+	+	177	100% (8/8)
	+	+	−	0	N/A
	−	−	−	0	
	+	−	−	0	
	−	+	−	0	

3.3 Further characterization of H intergenic sites for CRISPR-mediated genomic integration

Following the successful demonstration of genome editing using the H1 and H2 integration sites, six additional target intergenic loci were further selected for characterizations to expand the number of genomic integration sites for the CRISPR toolkit. These additional sites were designed following the selection criteria used to design the H1 and H2 sites and are referred to as H3-H8 integration sites (Table 5). The neighboring highly expressed genes used to select these loci are *TDH2* (H3), *TEF2* (H4), *PGK1* (H5), *RPS10B* (H6), *CDC19* (H7), and *YHB1* (H8), which belong to the top hundred most highly expressed genes in yeast (Velculescu et al., 1997). The H1-H8 intergenic sites are each located on different chromosomes to avoid possible chromosomal breakage or translocation events in a multiplexed genome editing. Following the selection of the H intergenic sites, pTarget integration plasmids harboring a *GFP* cassette flanked by at least ~0.5-kb long homology arms were created.

Using the pCut plasmid and *GFP* donor amplified from each pTarget plasmid, the efficacy of heterologous gene integration was tested for the H3-H8 integration sites. A total of eight randomly selected colonies for each integration site were PCR-genotyped to confirm the site-specific integration. The PCR-genotyping strategy employed three primers in a single PCR reaction to unarguably distinguish between colonies positive and negative for *GFP* insertion, based on the presence of amplified DNA fragments with different sizes, rather than the lack of the DNA fragments (Figure 10A). As shown in Figure 10B, the overall integration efficiencies observed for the H1-H8 integration sites were consistently high, ranging between 75% to 100%, reinforcing the interpretation that the poor chromatin accessibility was most likely the cause of low integration efficiencies found for the R1-R6 sites.

Table 5. Description of the H intergenic sites used for CRISPR-mediated gene integration

The bolded genes represent the highly-transcribed genes used to select the integration sites. The numbers indicated in parentheses represent the transcript abundance rankings of the genes according to the available yeast transcriptome data (Velculescu et al., 1997).

Site name	Chromosomal coordinates	Genes flanking the intergenic region	Size of intergenic region (bp)	N_{20} gRNA target
H1	ChrXV: 160595-161578	***ADH1* (6)** *YOL085W-A*	985	CCAATGCTAGTAGAGAAGGG
H2	ChrXII: 234082-235036	***PDC1* (5)** *STU2*	956	GAGATAGGTAAATAAACGCG
H3	ChrX: 455556-456238	***TDH2* (1)** *MET3*	1,558	CGTGATTTATACAAAAGAAG
H4	ChrII: 476438-477670	*TKL2* ***TEF2* (2)**	1,234	CCACCATAACATCAATCATG
H5	ChrIII: 136874-137745	*ADP1* ***PGK1* (10)**	872	TGGCCCTGATAATAGTATGA
H6	ChrXIII: 731124-732413	*RRP5* ***RPS10B* (17)**	1,290	AGGCGAAACACTTCATCCGG
H7	ChrI: 70470-71785	*CYC3* ***CDC19* (23)**	4,266	GTATCACAACCGACGATCCG
H8	ChrVII: 958211-959903	*PHO81* ***YHB1* (60)**	1,694	TTTTCCCAGAGTACCAGCAA

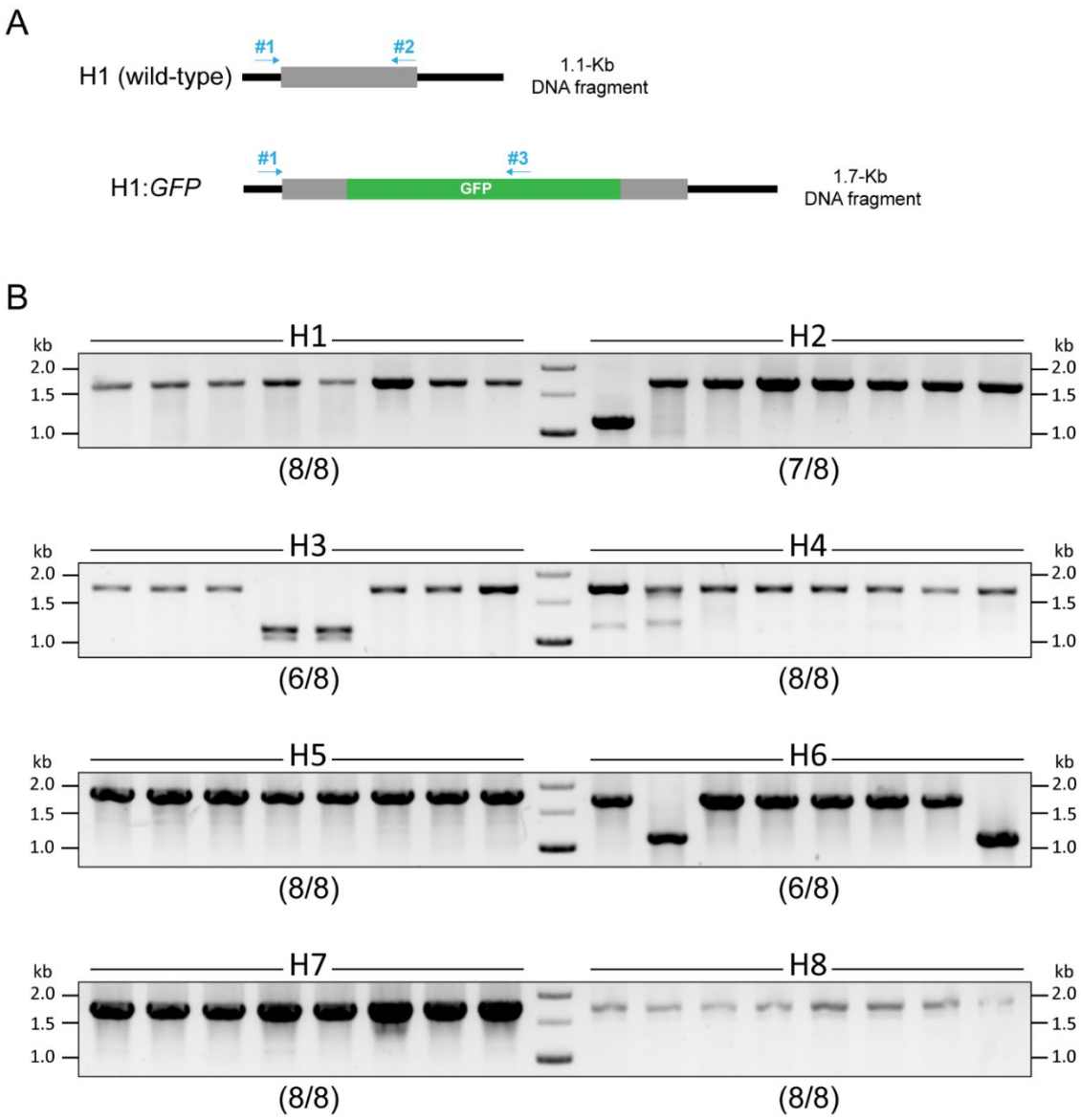


Figure 10. Efficiency of *GFP* cassette integration for the H integration sites

The BY4742 strain was transformed with the pCut plasmid and *GFP* donor and the *GFP* integration efficiency was determined by the genotyping of 8 randomly selected colonies. (A) An overview of the three-primer system utilized to screen the presence or absence of the *GFP* cassette at the integration site. (B) The results of the genotyping for the H1-H8 CRISPR integration sites.

To demonstrate that the selected integration loci can support proper expression of foreign genes, each transformant was subsequently assessed for the degree of *GFP* expression using flow cytometry. To compare the degree of transgene expression for each integration locus with the expression level of native yeast genes, three yeast genes were selected as representatives of high, medium, and low expressers from the published data (Ghaemmaghami et al., 2003) and were tagged with ye*GFP* at the C-terminus for flow cytometry analysis. The three control genes representing endogenous gene expressions are as follows: *FBA1* (rank 3), *RPS28A* (rank 87), and *SPO14* (rank 3866). All of the yeast strains examined, in biological triplicates, produced a clear fluorescence signal when compared to the background fluorescence signals from wild-type BY4742 (for a representative data, see Figure 11A), indicating the manifest of the desired phenotype from the correct integration of the *GFP* cassette. Additionally, it was shown that the fluorescence levels from genome-integrated *GFP* were consistently homogenous compared to the degree of *GFP* expression from high-copy pESC-URA plasmid control, as the plasmid-based expression of *GFP* resulted in a population of cells with extremely heterogeneous levels of *GFP* fluorescence (Figure 11B). This observation showed the suitability of genome-based gene expression for uniform expression of heterologous proteins. Overall, the H1-H8 integration sites harboring the *GFP* cassette exhibited small variations in the fluorescence levels (Figure 12). When compared to the controls representing different endogenous protein expression levels in yeast, the H1-H8 integration sites displayed sufficient levels of *GFP* protein expression (Figure 12).

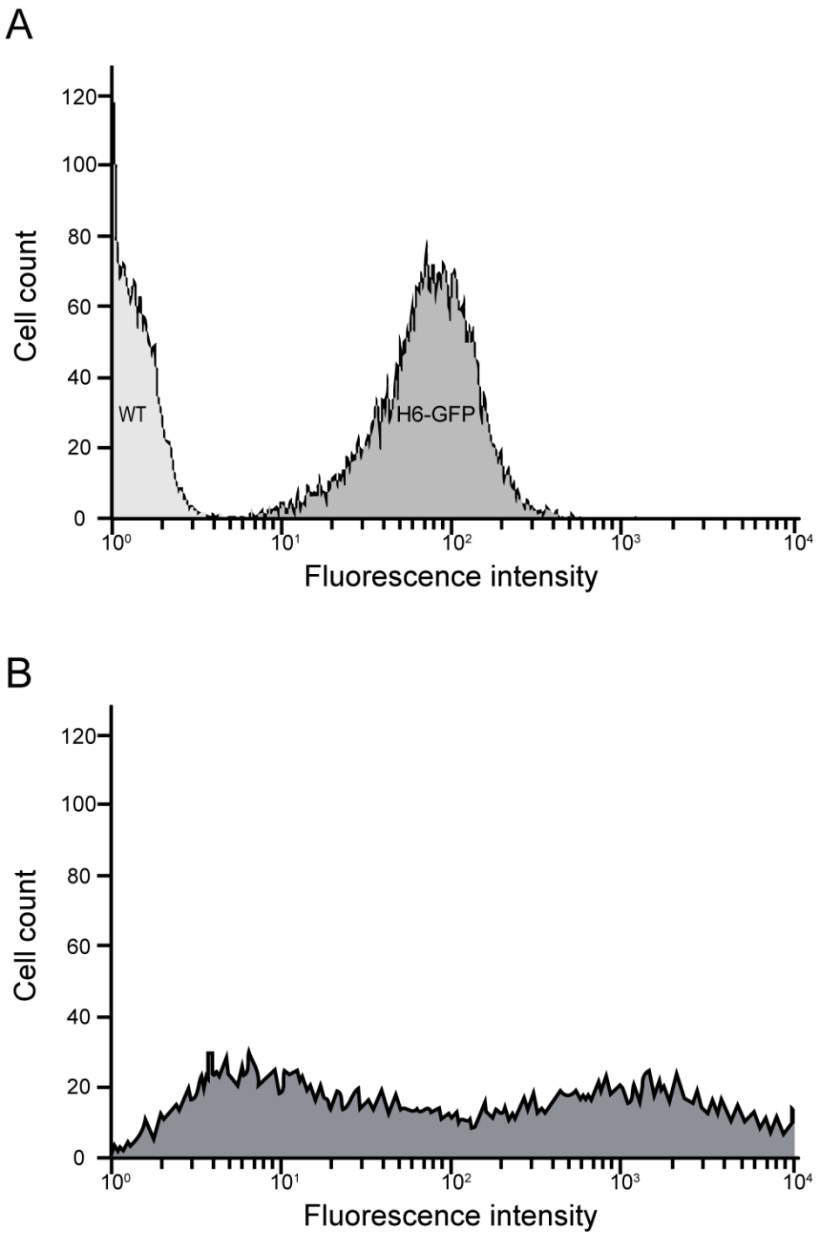


Figure 11. Flow cytometry analysis of genome-based and plasmid-based *GFP* expression

(A) A representative histogram shows the high intensity of *GFP* fluorescence present in a yeast strain harboring a *GFP* cassette at the H6 CRISPR integration site (H6-GFP; dark gray) in comparison to the exceedingly low background fluorescence intensity observed for the wild-type BY4742 (WT; light gray). (B) The variable levels of *GFP* fluorescence observed from yeast harboring a pESC-URA plasmid containing the *GFP* cassette.

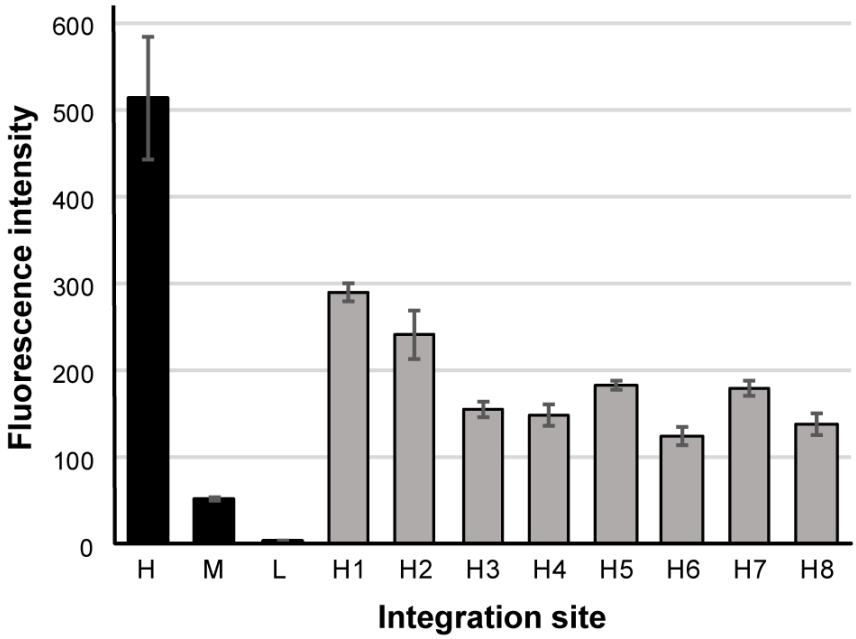


Figure 12. Flow cytometry analysis of the *GFP*-harboring yeast strains

The effect of a genomic locus on the *GFP* reporter gene expression. H: a high level of yeast endogenous protein expression represented by *FBA1-GFP*; M: a medium level of yeast endogenous protein expression represented by *RPS28A-GFP*; L: a low level of yeast endogenous protein expression represented by *SPO14-GFP*. Data are mean ± S.D. from 3 replicates.

Furthermore, the genomic stability of the *GFP* cassette integration was evaluated by carrying out serial cultivations of the *GFP*-harboring yeast strains in YPDA media and screening for the presence of the *GFP* cassette *via* PCR-genotyping after four passages of growth which corresponds to roughly 64 generations. The three-primer system was used for PCR-genotyping in order to clearly distinguish between a locus positive for *GFP* integration and a wild-type locus lacking the *GFP* cassette (Figure 13A). The results of the genotyping indicated that the *GFP* cassette was successfully retained for all eight loci after many generations of growth (Figure 13B).

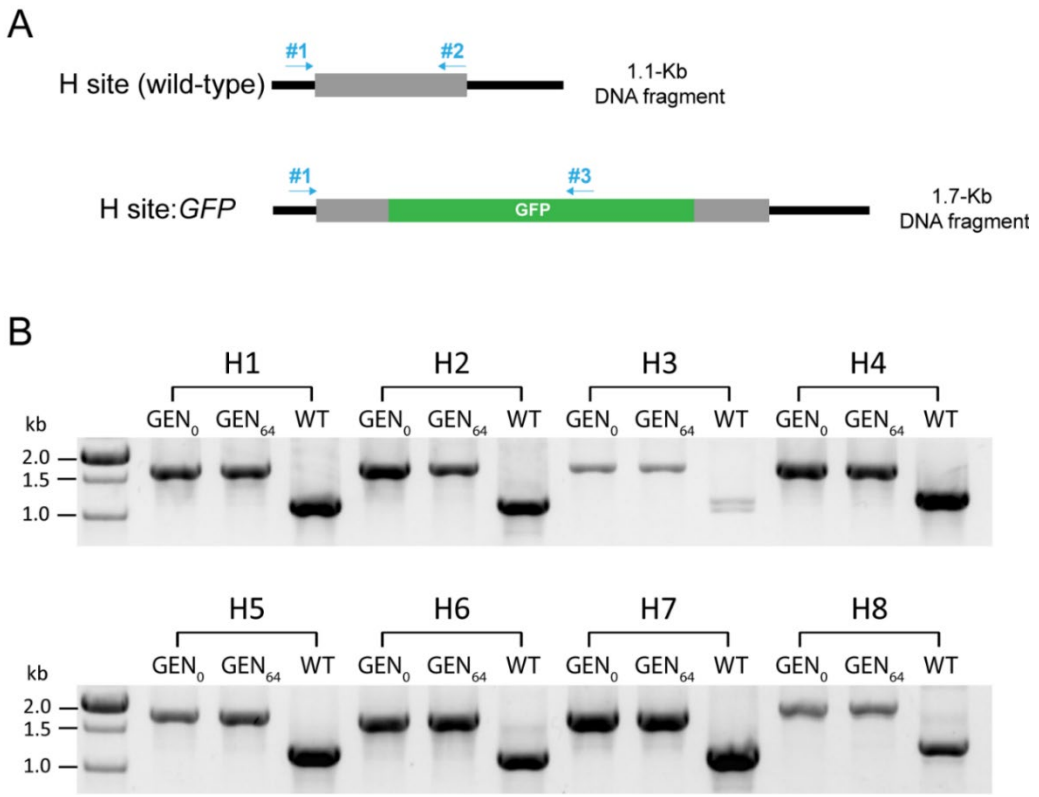


Figure 13. Genomic stability of the *GFP* integration in the H integration sites

(A) An overview of the three-primer system utilized to screen the presence or absence of the *GFP* cassette at each integration site. (B) Screening for the presence of the *GFP* cassette integration in *GFP*-harboring strains prior to serial cultivations in YPDA (GEN_0), *GFP*-harboring strains that have gone through 64 generations of growth in the serial cultivations in YPDA (GEN_{64}) and the wild-type BY4742 strain.

Lastly, an extensive literature search and a growth assay experiment were carried out to ensure that the presence of a heterologous DNA in a target locus does not result in any significant impairment of the growth rate of yeast. First, it was confirmed through literature search that none of the gRNA target cut sites lies in any gene coding sequences, any known regulatory elements associated with the neighboring genes and any 5' or 3' UTRs, minimizing the risk of possible disruption of nearby gene expression. Secondly, the impact of heterologous DNA integration on yeast physiology was examined by measuring the exponential growth rates

of the *GFP*-harboring yeast strains in YPDA media over the course of 48 hours. None of the yeast strains tested showed any significant difference in the mean growth rate compared to the wild-type BY4742 strain (*p*-value >0.05), indicating that the H1-H8 sites are suitable to be used as 'safe harbor' sites for transgene expression (Figure 14).

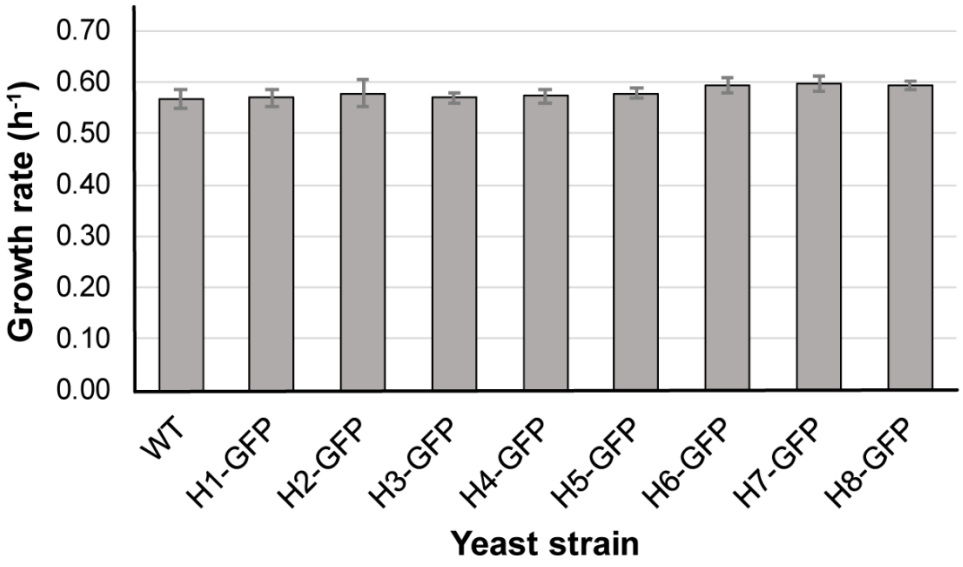


Figure 14. Growth rates of the *GFP*-harboring yeast strains

OD_{600} measurements of yeast cultures in YPDA media were made every 1-2 hours until the stationary phase was reached. Data are mean ± S.D. from 3 replicates.

In conclusion, selecting intergenic loci adjacent to the highly expressed genes remarkably improved the efficiency of foreign gene integration by CRISPR-Cas9. A series of experiments successfully identified the 8 CRISPR integration sites that support high fidelity gene integration and reliable expression of heterologous genes without any significant impairment on the fitness of yeast.

CHAPTER 4: DEVELOPMENT AND APPLICATION OF THE SGM-CRISPR MULTIPLEX PLATFORM

4.1 Synthetic gRNA target testing and strain construction

Of the 8 identified CRISPR integration sites, five sites with high *GFP* expression levels (H1, H2, H4, H5 and H7) were selected and used to integrate a common 23-bp long synthetic gRNA target (sTarget) consisting of a 20-bp long gRNA recognition sequence and a PAM sequence at the 3' end (5'-AGG-3'). Two sTarget candidates were designed using a random DNA sequence generator and the CHOPCHOP gRNA design tool while ensuring maximized parameters for predicted on-target binding efficiency and minimal off-targets in the yeast genome. Specifically, the two sTarget candidates were chosen from a pool of randomly generated 20-bp long DNA sequences displaying the efficiency score more than 70 on CHOPCHOP and possessing a zero number of off-targets with 0, 1, 2 and 3 mismatches in the yeast genome. The sequences of the two sTarget candidates are shown below.

sTarget#1: AAAGCGTCGCGCAATCGAGGAGG

sTarget#2: AAAGCGTCGTACATAACAGGAGG

The two sTarget candidates share the same 9-bp-long sequence starting at the +11 position upstream of PAM; however, they are different by 50% in the nucleotide composition of the 'seed' region (the 10 nucleotides immediately upstream of the PAM) which is important for target recognition by Cas9/gRNA (Jiang et al., 2013).

In order to test the function of the two sTargets, the 23-bp long sTarget DNA fragment was seamlessly placed between ~0.5-kb long left and right arms homologous to the H1-flanking sequences by overlap extension PCR, resulting in a ~1-kb long DNA fragment that was

subsequently integrated into the H1 site in BY4742 by CRISPR-Cas9. This resulted in SBY101 strain with sTarget#1 insertion and SBY102 with sTarget#2 insertion. Following the development of the two yeast strains, a pCut plasmid expressing a gRNA molecule that can bind to the 23-bp synthetic target sequence in the genome was created and named 'pCutSGM'. SBY101 and SBY102 were then transformed with a corresponding *GFP* donor and a corresponding pCutSGM plasmid. Counting of the fluorescent colonies after the transformation showed that 98.9% (±0.6% S.D.; n=4) and 96.3% (±1.7% S.D.; n=4) of the resulting colonies displayed fluorescence from CRISPR-Cas9 genome editing using sTarget#1 and sTarget#2, respectively (Figure 15). When the same batch of yeast cells were transformed with pCutSGM without *GFP* donor fragments, the colony numbers drastically decreased, and only 1.8% (±0.9% S.D.; n=4) and 4.2% (±2.6% S.D.; n=4) of yeast colony numbers, relative to the total number of viable yeast cells in presence of *GFP* donor, were recovered for sTarget#1 and sTarget#2, respectively. These results can be expressed in terms of cell toxicity where percentage toxicity for sTarget#1 and sTarget#2 was 98.2% and 95.8%, respectively (Figure 15). The potent cell toxicity observed from both sTargets indicates that the complex of Cas9 and gRNA can recognize the synthetic target sequence and effectively introduce DNA cleavage to cause severe lethality in yeast. The exceedingly high on-target gene integration efficiencies observed for both sTarget candidates further demonstrate that the DSB introduced on synthetic target sequences by CRISPR-Cas9 can be readily repaired by HR in presence of donor fragments. In addition to these results, the diagnostic PCR-genotyping of randomly picked eight colonies using the three-primer system revealed on-target integration efficiency of 100% and 87.5% for sTarget#1 and sTarget#2, respectively, further confirming the high on-target efficiencies observed from the fluorescence screening (Figure 16). The assessments of the sTarget performance in CRISPR-

Cas9 genome editing led to the conclusion that both of the sTargets possess excellent sequence properties for CRISPR-mediated gene integration. The sTarget#1 was selected for further development of the SGM-CRISPR platform as it offered slightly better integration efficiency and higher toxicity relative to the sTarget#2, although there was no statistically significant difference between them (*p*-value >0.05).

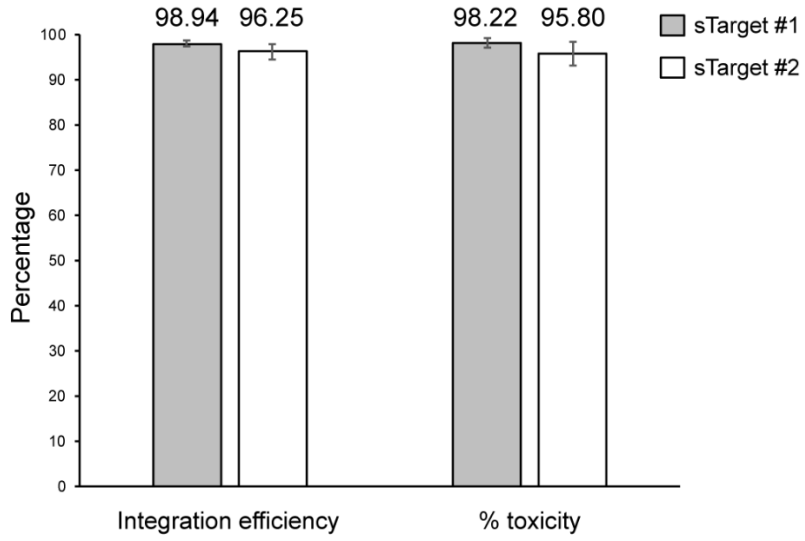


Figure 15. CRISPR editing efficiencies of synthetic gRNA targets

The efficacy of sTarget#1 and #2 in CRISPR-Cas9 genome editing was evaluated by CRISPR-assisted integration of *GFP* cassette at the H1 integration site in the SBY101 and 201 strains. For integration efficiency, the percentages of green fluorescent colonies were calculated. For percentage toxicity, the percentages of lethality were calculated from the numbers of surviving yeast colonies after transforming yeast without donor DNA in comparison to control (with donor DNA). Data are mean ± S.D. from 4 replicates.

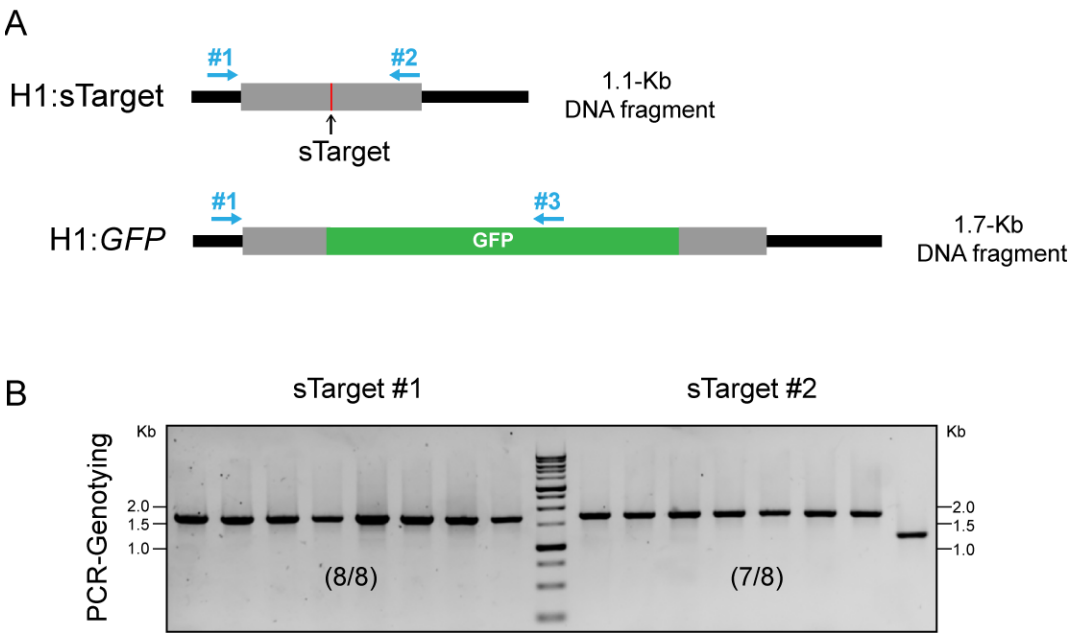


Figure 16. PCR-Genotyping of sTarget-mediated single gene integration

(A) An overview of the three-primer system utilized to screen the presence or absence of the *GFP* cassette at the H1 integration site. (B) Screening of the efficiency of *GFP* integration using the sTargets. A total of 8 randomly selected colonies were genotyped and checked for the success or failure of the *GFP* cassette integration.

Following the evaluation of the sTarget candidates, the sTarget#1 was integrated into the 5 chosen CRISPR integration loci in the order of H1, H2, H5, H7 and H4 in a sequential manner by recycling *URA3* marker through 5-FOA counter-selection (Figure 17). The sequential integration of the 23-bp synthetic target sequence was highly efficient as the yeast strain with the desired integration was able to be obtained from the genotyping of between two to four colonies, often with 100% integration efficiency. This repeated cycle of sTarget#1 integration generated four additional yeast strains, SBY102, SBY103, SBY104, and SBY105, where the last number indicates the number of sTarget#1 integrated in the yeast genome.

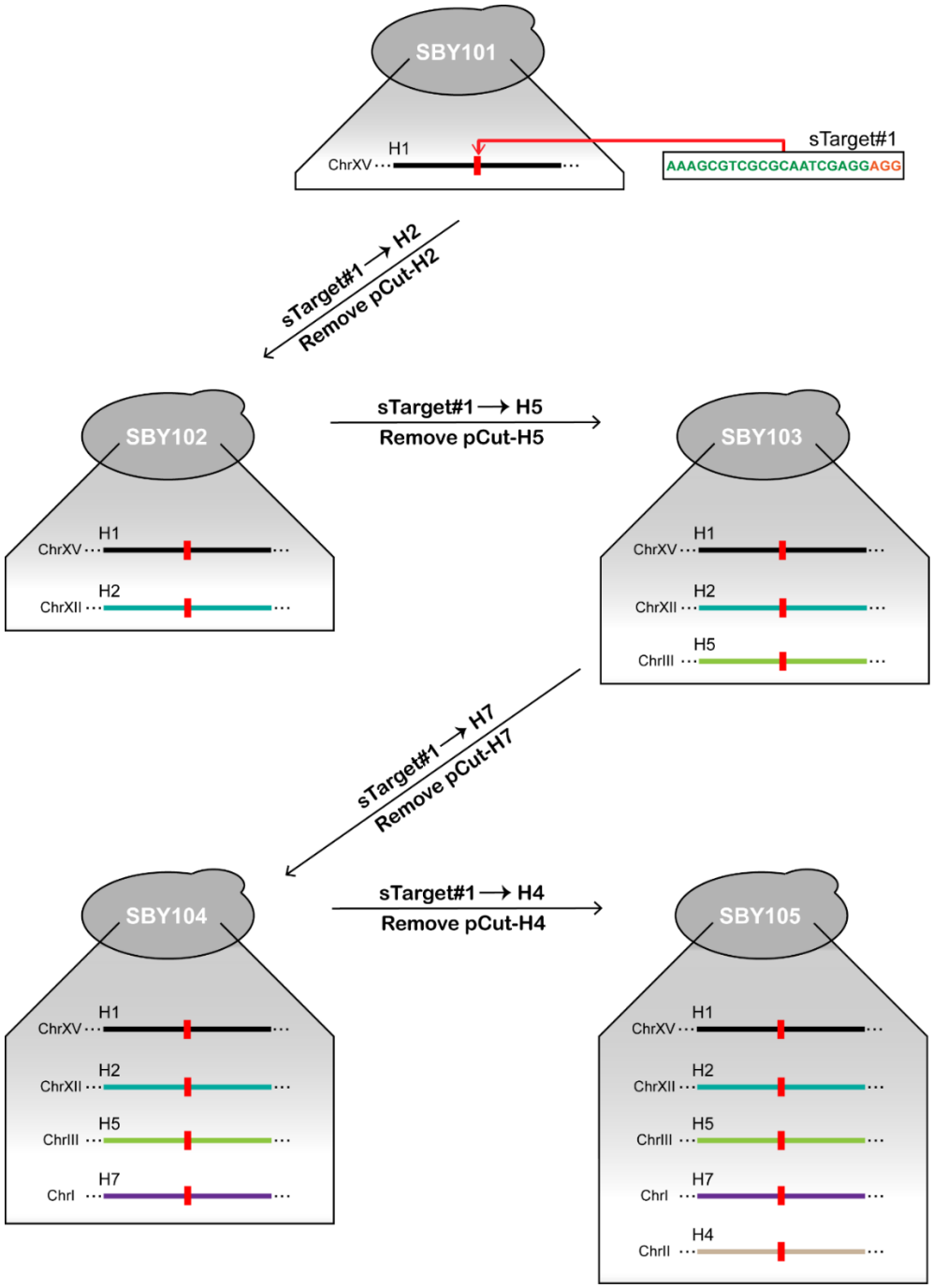


Figure 17. Workflow of the strain construction for the development of the SGM-CRISPR platforms

4.2 Evaluation of the SGM-CRISPR platform using marker genes

Following the successful generation of the yeast platforms for multiplex genome editing, the feasibility of simultaneous 2- and 3- gene integrations was first examined using yeast strains SBY102 and SBY103. To test double and triple gene integrations, the biosynthetic pathway of betalain, a class of yellow and red pigments produced by plants of Caryophyllales, was used as a model pathway. A recent study has successfully re-constituted the biosynthetic pathway of betanin, a red-violet colored betalain pigment, in yeast by the over-expression of three enzymes - a tyrosine hydroxylase from *Beta vulgaris* (*BvCYP76AD1*), a dioxygenase from *Mirabilis jalapa* (*MjDOD*), and a glucosyltransferase from *M. jalapa* (*Mj-cDOPA5GT*) (Grewal et al., 2018). The co-expression of these three genes in yeast results in the formation of yeast colonies with red-violet pigments (Figure 18). On the other hand, the over-expression of the first two enzymes in the pathway – *CYP76AD1* and *DOD* – forms yellow-orange colored betaxanthin pigments, giving yeast colonies a strong yellow color (Figure 18). Taking advantage of the characteristic, visible pigment properties of betaxanthin and betanin compounds, yeast colonies positive for the integration of the betalain biosynthetic genes can be easily screened based on the colony color.

To demonstrate the concurrent integration of two heterologous genes, the SBY102 strain bearing sTarget#1 in the H1 and H2 sites was co-transformed with pCutSGM1 and donors encoding *BvCYP76AD1* and *MjDOD* cassettes with ~0.5-kb long homology arms targeting H1 and H2 sites, respectively. Prior to the transformation, the *BvCYP76AD1* and *MjDOD* genes were put under the control of strong constitutive promoters, p*TDH3* and *pCCW12*, respectively. The transformants were plated on a SC-URA medium supplemented with extra L-tyrosine, a central precursor to the betalain compounds, and were allowed to grow for three days. An

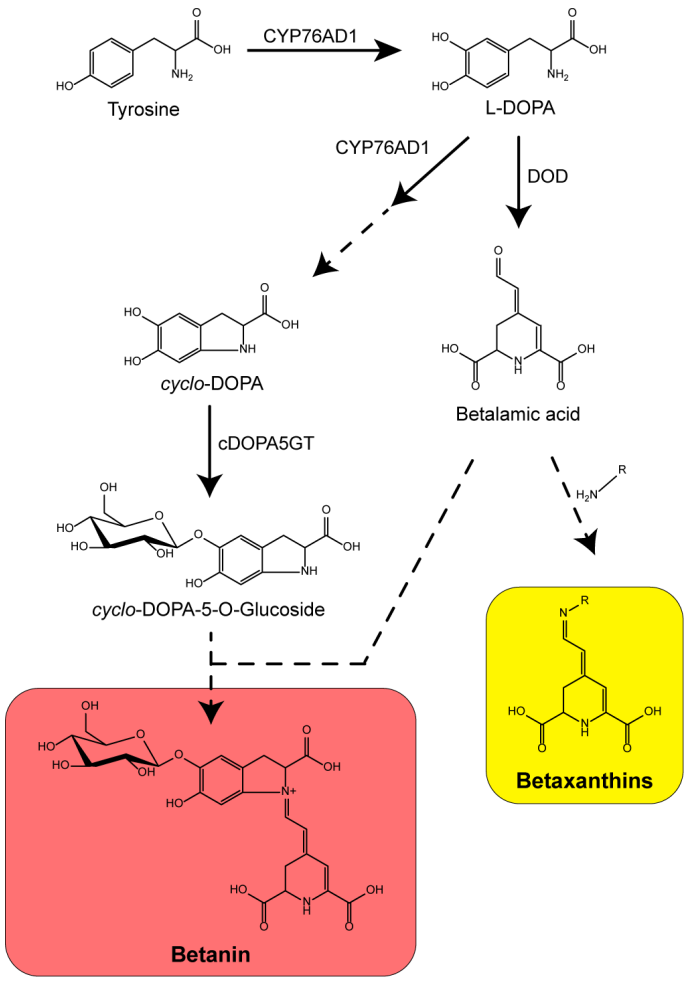


Figure 18. Overview of the betalain biosynthetic pathway

Solid lines indicate enzymatic reactions and dashed lines indicate spontaneous reactions. The overexpression of *CYP76AD1* and *DOD* in *S. cerevisiae* results in the formation of yellow-orange colored betaxanthin pigments. The overexpression of the *CYP76AD1*, *DOD* and *cDOPA5GT* in *S. cerevisiae* results in the formation of red-violet colored betanin pigments.

average of 97.6% (±1.7% S.D.; n=3) of the resulting colonies displayed yellow color, indicating the formation of betaxanthin pigments from the integrations of *CYP76AD1* and *DOD* (Figure 19A). Following the encouraging results from the double gene integrations, simultaneous integrations of *BvCYP76AD1*, *MjDOD* and *Mj-cDOPA5GT* into H1, H2 and H5 sites, respectively, were examined using the SBY103 strain harboring sTarget#1 in the three target

sites. Prior to transformation, the *Mj-cDOPA5GT* gene was put under the control of strong constitutive promoter p*FBA1* and ~0.5-kb long left and right homology arms to the H5 site were attached to the gene cassette. Following the transformation of SBY103 with pCutSGM1 and the three donor DNA components, an average of 98.0% (±2.4% S.D.; n=3) of the resulting colonies displayed red color from the three gene integrations leading to the biosynthesis of betanin (Figure 19B). In contrast, when SBY103 was transformed with donors and pCut without a gRNA cassette, the red pigment formation was not observed from more than 1,000 resulting colonies, illustrating the critical role of targeted DSB in achieving the three-gene pathway construction.

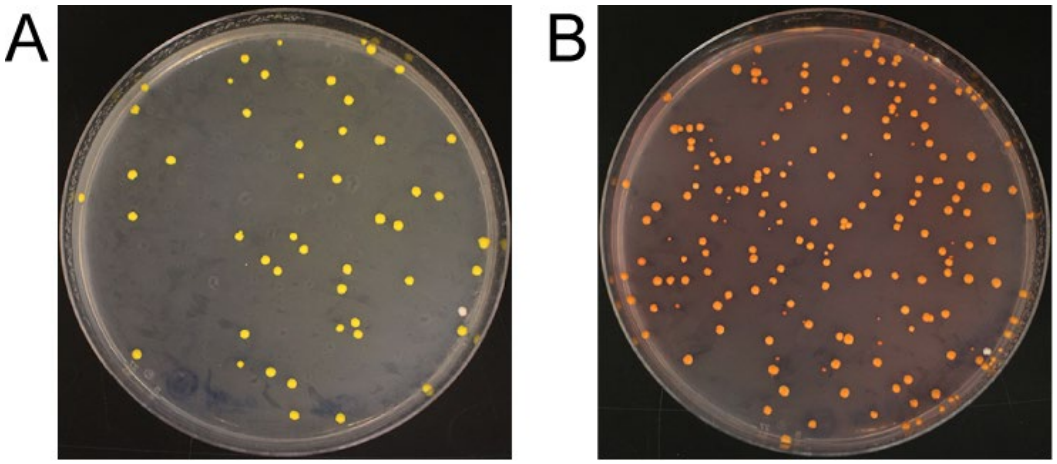


Figure 19. Simultaneous 2- and 3- gene integrations demonstrated by the SGM-CRISPR platform

(A) Simultaneous double gene integration represented by SBY102 transformed with pCutSGM1 and linearized donors encoding *BvCYP76AD1* and *MjDOD*. (B) Simultaneous triple gene integration represented by SBY103 transformed with pCutSGM1 and linearized donors encoding *CYP76AD1*, *DOD* and *cDOPA5GT*.

The integration efficiencies of the double and triple gene integrations were exceptionally high and consistent among the replicates. Intrigued by this, quadruple and quintuple gene

integrations were tested by additional integration of selectable marker cassettes on top of the three betanin biosynthetic gene cassettes. Specifically, quadruple gene integrations were carried out by the co-transformation of SBY104 with pCutSGM1 and donors encoding the three betanin biosynthetic gene cassettes and hygromycin B (HygB) resistance marker gene cassette, *hphNT1*, flanked by homology arms targeting the H7 site. Quintuple gene integrations used SBY105 to perform integrations of the three betanin biosynthetic gene cassettes, *hphNT1* cassette and geneticin (G418) resistance marker gene cassette, *kanMX4,* that possess homology arms to the H4 site. The SBY104 and SBY105 strains transformed with the appropriate DNA components were plated on SC-URA with extra L-tyrosine, and the resulting colonies were re-plated onto YPDA+HygB and YDPA+G418, both of which have been supplemented with extra L-tyrosine. The results of the phenotypic screening for red pigment formation as well as antibiotic resistance showed that an average of 93.7% (±1.5% S.D.; n=3) and 84.8% (±6.3% S.D.; n=3) of the resulting colonies displayed pigment formation and resistance to the corresponding antibiotic(s) from the quadruple and quintuple gene integrations of the marker genes, respectively (Figure 20A/B). The results for the quintuple gene integration demonstrated an introduction of a total of ~9.5-kb long heterologous DNA into the yeast genome. In summary, the series of phenotypic screenings for 2-, 3-, 4-, and 5- gene integrations revealed that double- to quintuple- gene integrations took place at high efficiencies ranging between 84.8% to 98.0% (Figure 21).

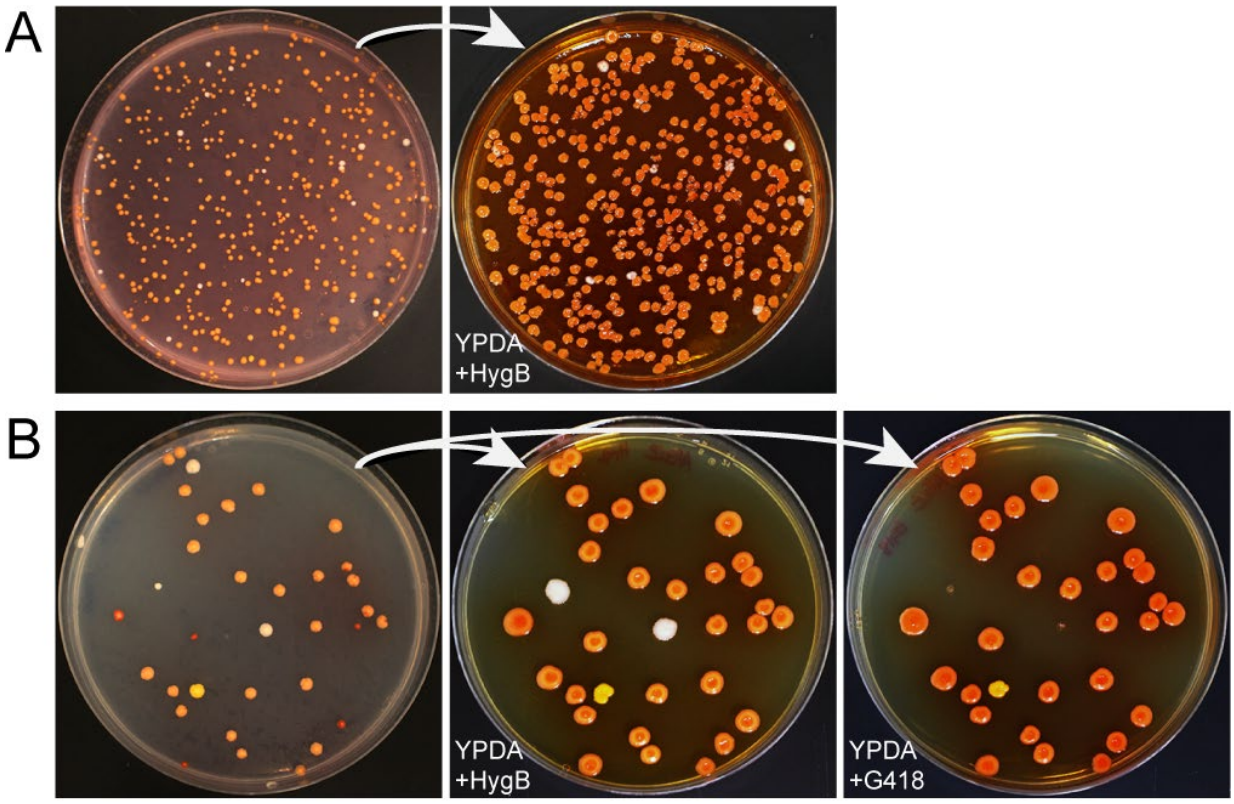


Figure 20. Simultaneous 4- and 5- gene integrations demonstrated by the SGM-CRISPR platform

(A) *S. cerevisiae* colonies showing the simultaneous integration of four gene cassettes. Left: SBY104 transformed with pCutSGM1 and linearized donors encoding *CYP76AD1*, *DOD*, *cDOPA5GT* and *hphNT1*. Right: Re-plating of the yeast colonies on YPDA+HygB. (B) *S. cerevisiae* colonies showing the simultaneous integration of five gene cassettes. Left: SBY105 transformed with pCutSGM1 and linearized donors encoding *CYP76AD1*, *DOD*, *cDOPA5GT*, *hphNT1* and *kanMX4*. Middle and right: re-plating of the yeast colonies on YPDA+HygB and YPDA+G418.

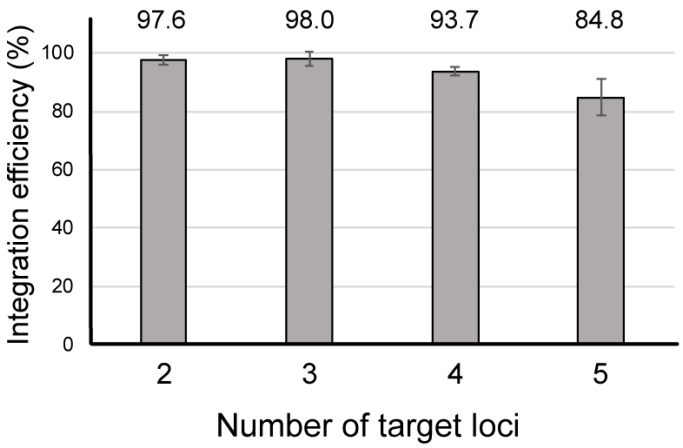


Figure 21. Summary of the 2-, 3-, 4- and 5- gene integration efficiencies

The overall average efficiencies of the simultaneous double, triple, quadruple and quintuple gene integrations achieved by the SGM-CRISPR system. Data are mean ± S.D. from 3 replicates.

One noteworthy observation from the demonstrations of quadruple and quintuple gene integrations is that the overall number of surviving colonies was reduced approximately by 7-fold from the quadruple integration to quintuple integration. The observed decrease in the colony numbers is likely due to the drop in the transformation efficiency from the increasing amount of the donors and pCutSGM1. Another possible reason is that the native HDR machinery in yeast may not be able to cope as efficiently with the elevated Cas9-induced toxicity from the increasing numbers of DSBs. Another noticeable observation from the quadruple and quintuple gene integrations is the presence of partial integrants, exemplified by the white-colored colonies that confer resistance to HygB in Figure 20A and the white- or yellow- colored colonies that confer resistance to HygB or to both HygB and G418 in Figure 20B. The presence of partial integrants suggests that there was either an insufficient Cas9 cleavage leading to intact genomic sites without heterologous gene integrations, or a possible occurrence of NHEJ repair in one or more of the target sites upon Cas9 cleavage, generating indel mutations in the gRNA binding sites that prevent further Cas9 cleavage and subsequent gene integrations.

In order to further strengthen the data for high gene integration efficiencies noted from the phenotypic screenings, a diagnostic PCR-genotyping was performed on eight colonies from each of the 2 – 5 gene-integration experiments that exhibited desired pigment formation as well as resistance to the corresponding antibiotic chemical(s). The resulting genotyping data showed that all of the screened colonies had the correct integrations of the donor genes in the targeted CRISPR integration sites (Figure 22), proving that the pigmented and antibiotic-resistant phenotypes of the transformants were products of targeted gene integrations. Following the genotyping experiments, the genomic stability of the five integrated genes in a multiplexed context was tested by carrying out a serial transfer cultivation experiment on the transformants positive for the five marker gene integrations. After five passages of growth in YPDA, corresponding to roughly 80 generations of growth, the yeast cells were plated on YPDA and the appearing colonies were re-plated on YPDA+HygB and YPDA+G418. All of the resulting colonies from the serial cultivation experiment retained the red color and antibiotic resistance towards HygB and G418 after the 80 generations of growth (Figure 23A). Furthermore, the genotyping data at the end of the five growth passages provided another evidence for the presence of the transgenes at the selected integration sites (Figure 23B). These results show that the five CRISPR integration sites identified in this study can be reliably used to simultaneously contain and express multiple heterologous genes.

In conclusion, the results for the double- to quintuple- gene integration experiments demonstrate that the use of transcriptionally active genomic regions for CRISPR targeting as well as the high-performance single gRNA in the SGM-CRISPR platform enables simultaneous integrations of 2-, 3-, 4-, and 5- genes at a high efficiency.

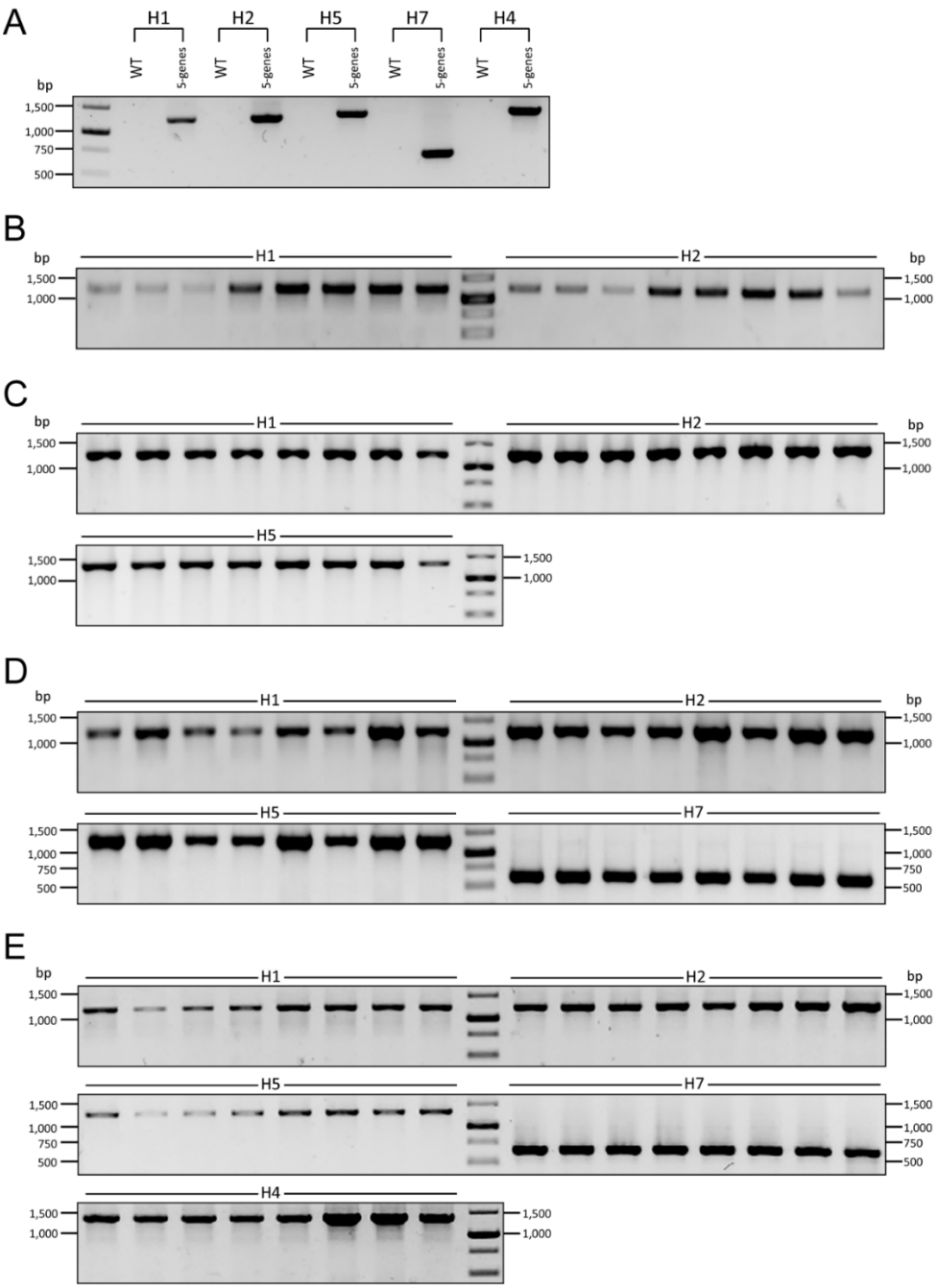


Figure 22. Genotyping of yeast colonies for confirmation of the five marker gene integrations

(A) The comparison of genotyping results between wild-type BY4742 and SBY.B04 with the five marker gene integrations. Expected product size for H1::*BvCYP76AD1* = 1210-bp; expected product size for H2::*MjDOD* = 1215-bp; expected product size for H5::*Mj-cDOPA5GT* = 1299-bp; expected product size for H7::*hphNT1* = 645-bp; expected product size for H4::*kanMX4* = 1378-bp. (B) Genotyping results for double gene integrations (C) Genotyping results for triple gene integrations (D) Genotyping results for quadruple gene integrations (E) Genotyping results for quintuple gene integrations.

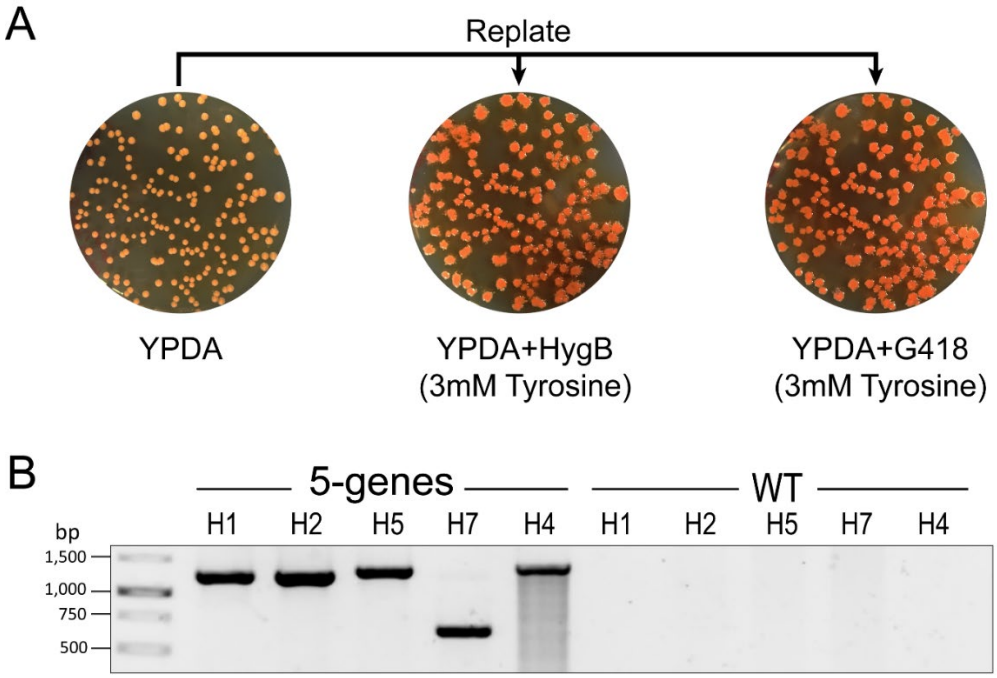


Figure 23. Genomic stability of the five marker genes in the selected integration sites

(A) *S. cerevisiae* transformants positive for the integration of *CYP76AD1*, *DOD*, *cDOPA5GT*, *hphNT1* and *kanMX4* were cultured in YPDA for a total of ~80 generations and plated on YPDA. The resulting colonies were re-plated onto YPDA+HygB and YPDA+G418 media supplemented with extra L-tyrosine. (B) Genotyping of SBY.B04 harboring the five marker genes after ~80 generations of growth and the wild-type BY4742 strain to confirm the retention of the five marker gene cassettes.

4.3 Proof of concept: construction of plant sesquiterpenoid biosynthetic pathways

Sesquiterpene lactones (STLs) are a class of naturally occurring plant secondary metabolites that possess a characteristic γ-lactone moiety on the 15-carbon backbone. There are several STLs that are known to exhibit bioactive activities in the human body, such as anti-cancer thapsigargin, anti-malarial artemisinin, and analgesic lactucopicrin (Nguyen et al., 2010). Among different STLs known, costunolide is the simplest form of STL based on the germacrene A backbone that is believed to serve as a central precursor to diverse bioactive STLs (de Kraker et al., 2002). Costunolide is also a novel compound carrying a therapeutic potential from harnessing various potent biological activities itself, such as anti-inflammatory and anti-cancer properties (Park et al., 2016; Peng et al., 2017). As costunolide accumulates as a metabolic end product only in a small number of species (Ikezawa et al., 2011), an established costunolide producing *S. cerevisiae* strain will serve as a valuable microbial feedstock of costunolide. This would be of industrial relevance for mass production and would also be an effective tool in replacing the *in vitro* analysis for elucidating the biosynthesis of downstream STL derivatives (Xiao et al., 2013).

The biosynthesis of costunolide has been characterized in great detail and the entire pathway is illustrated in Figure 24. It has been shown that farnesyl diphosphate (FPP) is cyclized by germacrene A synthase (*GAS*) to form germacrene A, which goes through three stages of oxidation by germacrene A oxidase (*GAO*) to give germacrene A acid (GAA) (Nguyen et al., 2010). Subsequently, GAA is oxidized by costunolide synthase (*COS*) to yield 6α-hydroxylated germacrene A acid, which goes through a spontaneous lactone ring formation to form costunolide (Ikezawa et al., 2011). The re-constitution of the costunolide biosynthetic pathway in yeast requires the expression of the three biosynthetic genes as well as a cytochrome P450

reductase (*CPR*) which acts as a redox partner for *GAO* and *COS* (Figure 24). Furthermore, the over-expression of N-terminus truncated *HMGR* from yeast (*tHMGR*) can be used to further improve the yield of costunolide by increasing the metabolic flux towards FPP (Figure 24).

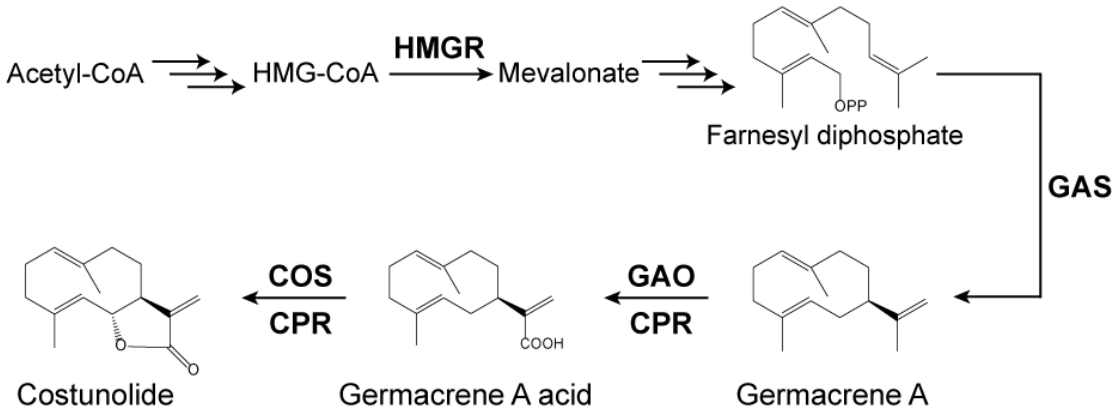
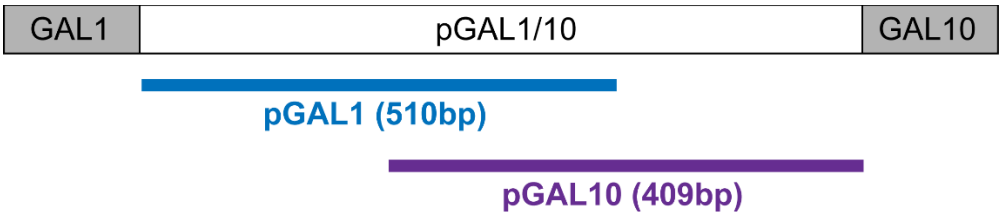


Figure 24. Overview of the costunolide biosynthetic pathway

The mevalonate pathway present in *S. cerevisiae* synthesizes farnesyl diphosphate, a central precursor to various sesquiterpenoid compounds. The coordinated activities of *GAS*, *GAO* and *COS* biosynthesize costunolide.

Using the developed SGM-CRISPR quadruple and quintuple gene integration platforms, the biosynthetic pathways of costunolide and its intermediate GAA were re-constituted in yeast in this work. To facilitate the concurrent integration of the genes required for GAA biosynthesis, the SBY104 strain was co-transformed with pCutSGM1 and donors encoding *Lactuca sativa GAS*, *LsGAO, Artemisia annua CPR* and *tHMGR*, under p*GAL1* or p*GAL10* galactose inducible promoters, with homology arms targeting the H1, H2, H5 and H7 sites, respectively (Figure 25).



Integration site	Gene cassette being integrated
H1	pGAL1-LsGAS-tCYC1
H2	pGAL10-LsGAO-tADH1
H5	pGAL1-AaCPR-tCYC1
H7	pGAL10-tHMGR-tADH1
H4	pGAL10-LsCOS-tADH1

Figure 25. Overview of the gene cassettes used for the construction of sesquiterpenoid biosynthetic pathways in yeast

Different lengths of the *GAL1* and *GAL10* promoters were used to drive the expression of the sesquiterpenoid biosynthetic genes. The biosynthetic gene cassettes for germacrene A acid and costunolide biosynthesis were flanked by ~0.5-kb-long homology arms for each intended CRIPSR integration site.

From the ~50 colonies resulting from the transformation, ten colonies were cultured in 30-mL synthetic media containing galactose for 72 hours and the metabolites were extracted for LC-MS analyses for GAA production. The screening for GAA production using (−)LC-MS analysis detected a $[M\text{-}H]^-$ at m/z = 233.1550 (Δ1.3 ppm from the theoretical mass of GAA) from eight out of the ten screened transformants (Figure 26A/B). Congruent with the previous observation that minor amounts of GAA are modified into ilicic acid in acidic yeast cultures (Nguyen et al., 2010), ilicic acid was also detected at m/z = 251.1656 (Δ1.2 ppm from the theoretical mass of ilicic acid) from the eight GAA-positive colonies.

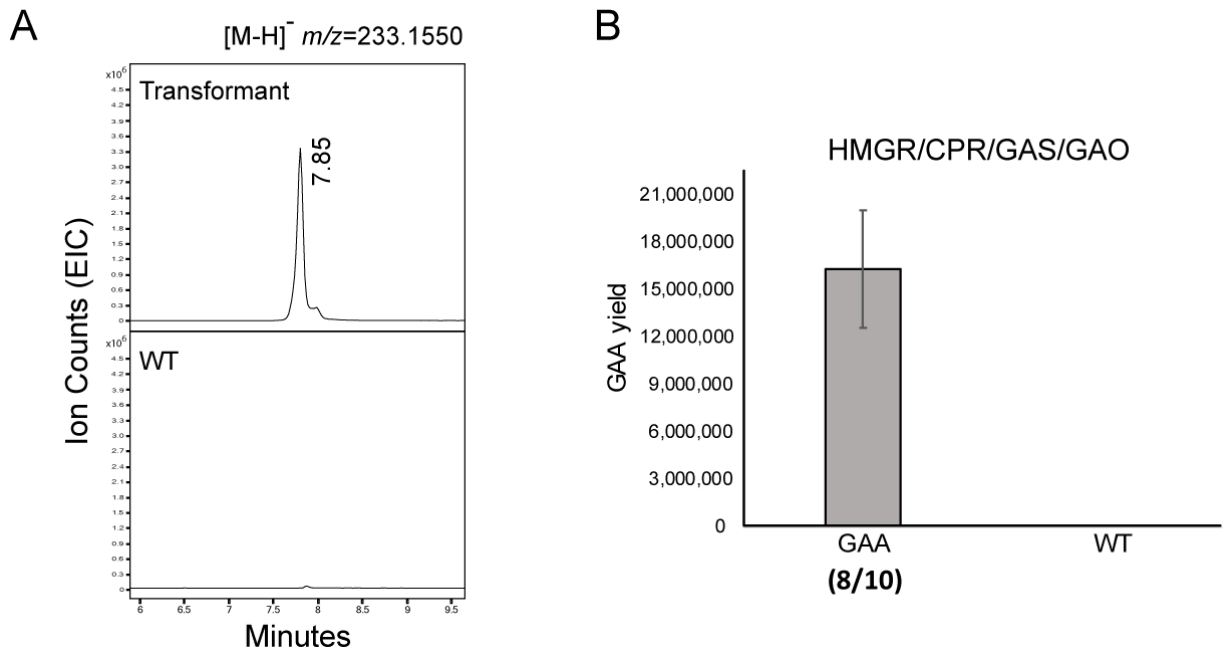


Figure 26. LC-MS analysis of selected transformants for germacrene A acid production

LC-MS analysis of 10 selected transformants resulting from the simultaneous integrations of the four germacrene A acid biosynthetic genes. (A) The metabolite profile of the yeast strain harboring *tHMGR*, *CPR*, *GAS* and *GAO* is shown in the (-)LC-MS scan of $[M-H]^-$ at *m/z* = 233.1550. Germacrene A acid is detected at the retention time of 7.85 minutes. (B) The yield of GAA is expressed in terms of ion counts. The bars for 'GAA' and 'WT' represent the means from eight and three independent colonies, respectively. Data are means ± S.D.

Following the successful re-constitution of the GAA biosynthetic pathway *via* quadruple gene integrations, the costunolide biosynthetic pathway was re-constructed in yeast by simultaneous integrations of *LsGAS*, *LsGAO, AaCPR*, *tHMGR* and *LsCOS* into the H1, H2, H5, H7 and H4 sites of the SBY105 strain, respectively (Figure 25). From the ~25 resulting colonies, 15 randomly picked colonies were subjected to (+)LC-MS analysis for detection of costunolide production. The $[M+H]^+$ ion for costunolide was detected at *m/z* 233.1538 in (+)LC-MS for 9 out of the 15 colonies screened (Δ0.9 ppm from the theoretical mass; Figure 27). Out of the nine colonies positive for the costunolide production, 10.8 mg/L (± 2.0; n=7) of costunolide was synthesized from seven colonies. Interestingly, the remaining two colonies showed much lower

titers of costunolide at 460 µg/L and 30 µg/L (Figure 27B). Genotyping and sequencing of the H5 and H7 integration sites of the two low-costunolide producing strains revealed that the H5 and H7 loci were intact with no sign of indel mutation or foreign gene integration, suggesting that the lack of Cas9 cleavage at these two loci led to the absence of *tHMGR* and *CPR* expression in the two engineered strains.

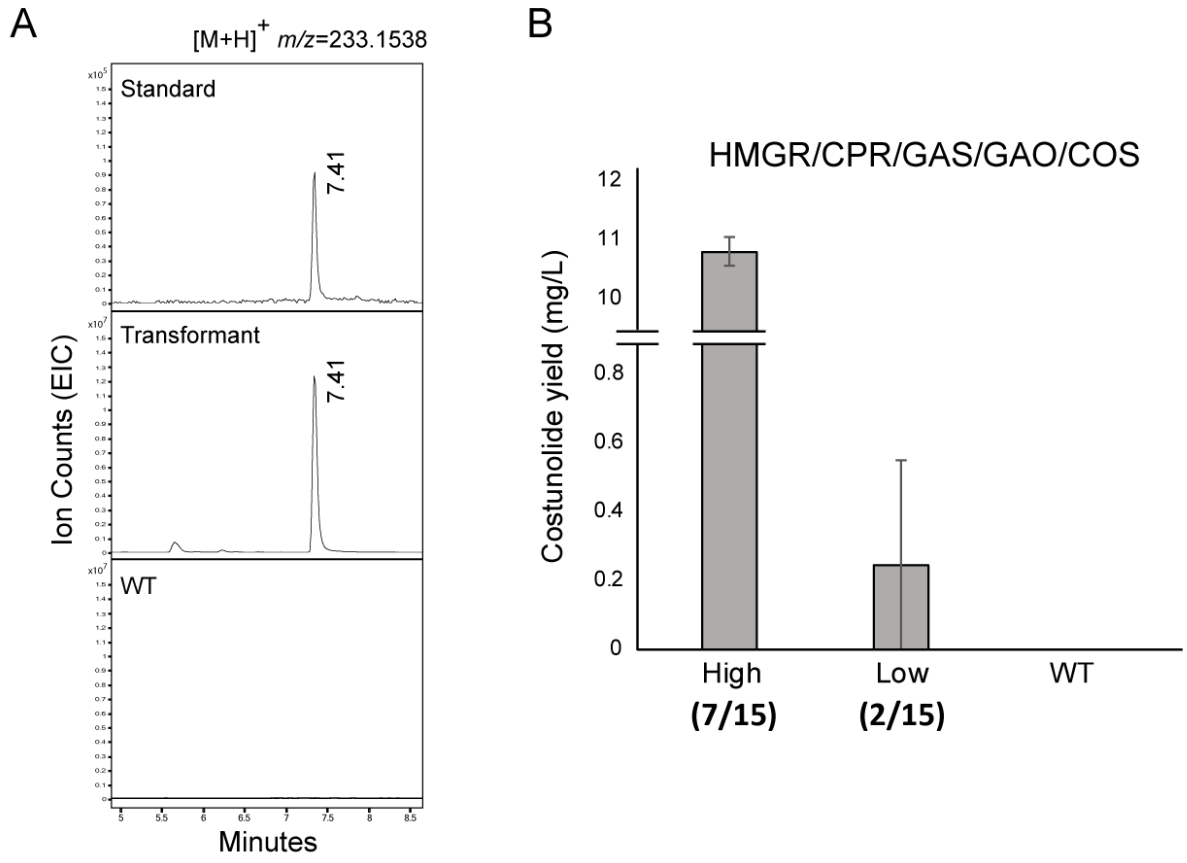


Figure 27. LC-MS analysis of selected transformants for costunolide production

LC-MS analysis of 15 selected transformants resulting from the simultaneous integrations of the five costunolide biosynthetic genes. (A) The metabolite profile of the yeast strain harboring *tHMGR*, *CPR*, *GAS, GAO* and *COS* is shown in the (+)LC-MS scan of $[M+H]^+$ at *m/z* = 233.1538. Costunolide is detected at the retention time of 7.41 minutes. (B) The yield of costunolide is expressed in terms of milligram per liter. The bar for 'High' represents the mean from seven independent colonies, while the bars for 'Low' and 'WT' represent the means from two and three independent colonies, respectively. Data are means ± S.D.

CHAPTER 5: DISCUSSION

5.1 New set of intergenic loci for CRISPR-mediated gene integration and expression

One of the important requirements for successful yeast metabolic engineering is the selection of genomic sites that can support expression of heterologous genes at appropriate expression levels. Moreover, with the increasing use of the CRISPR-Cas9 technology in the field of metabolic engineering, the availability of genomic loci with well-characterized gRNA binding sites for high-fidelity gene integration has also become important to facilitate a fast workflow for construction of synthetic pathways. In this regard, this work provides valuable information on the CRISPR-Cas9 targeting efficiencies associated with different types of genomic loci and a comprehensive report on the set of novel intergenic loci that can accommodate stable expression of heterologous genes.

Currently, the major focus of CRISPR-Cas9 design is the quality of a given gRNA target sequence. As the inherent features of gRNA binding sequences dictate the efficiency of on-target DNA cleavage and possible off-target effects, selection of a proper gRNA target is important to ensure optimal genome editing outcomes. The data from this work suggests that in addition to the gRNA target selection, the selection of integration loci is an important parameter to consider in the context of CRISPR-Cas9 design. This claim is supported by the strikingly different gene integration efficiencies observed between the CRISPR integration sites studied. The R1-R6 intergenic loci, selected based on DNA size and presence of restriction enzyme recognition sequences, exhibited very low gene integration efficiencies ranging between 0% and 37.5%, whereas the P1-P2 promoter sites and the H1-H8 intergenic sites, located near transcriptionally active DNA regions, were able to accomplish gene integrations at high efficiencies ranging

between 75% and 100% for the same Cas9/gRNA plasmid systems used for the R1-R6 sites. From this data, it is clear that the gene integration efficiency can be substantially improved by changing the nature of target genomic loci. As suggested by previous studies reporting the difference in Cas9 cleavage efficiencies between the heterochromatin and euchromatin structures of given target loci in eukaryotic systems (Daer et al., 2017; Uusi-Makela et al., 2018), the high integration efficiencies noted for the P1-P2 sites and H1-H8 sites are likely due to the open, relaxed chromatic nature of the nearby transcriptionally active DNA regions. Shortly after this experiment was carried out, a study looking at the affect of chromatin structure on Cas9 cleavage efficiency in yeast has been published (Yarrington et al., 2018). This study examined the efficiency of Cas9 cleavage on different locations of HO and PHO5 promoters where extensive data on nucleosome-occupied and depleted regions are available. The results of the study revealed that even within narrow sequence windows of ~1,700-bp, strongly positioned nucleosomes reduced or even prevented Cas9 cleavage while nucleosome depleted regions in the promoters were efficiently cut by Cas9. Therefore, from the results of two independent studies in yeast, it is evident that different chromosomal coordinates would exhibit different levels of Cas9 cleavage and that the selection of CRISPR target site is therefore critical for CRISPR-Cas9 efficiency. It should be noted that in both studies, different gRNAs were used to target each genomic region of interest, therefore the local chromatin structure cannot be credited to be the sole factor causing the difference in the observed Cas9 targeting efficiencies between the examined genomic positions. In future studies, a common, synthetic gRNA target can be pre-installed in the genomic loci of interest to better examine the effects of local chromatin structure on Cas9 activity, since the gRNA binding efficiency will be identical for the genomic loci being studied in this approach.

This work identified a total of eight CRISPR integration sites in the *S. cerevisiae* genome that support transgene integration at high efficiencies, allow for reliable expression of transgenes with defined expression levels, and safely maintain the transgene over many generations without affecting the growth of the host organism. This set of highly-characterized intergenic loci is a valuable addition to the available chromosomal integration sites in yeast characterized to date, namely the 17 solo long terminal repeat (LTR) sites (Bai Flagfeldt et al., 2009), the 14 intergenic sites (Mikkelsen et al., 2012), and the 23 intergenic sites (Reider Apel et al., 2017). Interestingly, the chromosomal coordinates of the H7 site identified in this study coincidentally match with the 106a intergenic site examined by Reider Apel et al. (2017), albeit a different gRNA binding sequence was selected to target the locus in their study. The eight intergenic sites identified in this study displayed ~2-fold range in the *GFP* expression which is consistent with the degrees of protein expression levels observed for the previously examined intergenic sites (Mikkelsen et al., 2012; Reider Apel et al., 2017). The integration loci identified by Bai Flagfeldt et al. (2009), however, exhibited a wider range of ~10-fold in the protein expression level based on β-galactosidase assay, likely due to the different method of quantification analysis and different types of loci used. The information on the eight loci and the corresponding gRNA target for each locus broadens the range of possible heterologous gene expression levels and opens up more options for expression of multiple biosynthetic enzymes constituting complex metabolic pathways.

5.2 SGM-CRISPR platform: significance and future directions

The SGM-CRISPR platform developed in this study presents a new single gRNA-mediated multiplex genome editing system that allows precise, concurrent integrations of heterologous DNA into up to five genomic loci in a locus-specific manner. Using carefully selected and characterized 5 intergenic loci (H1, H2, H4, H5 and H7) and the synthetic gRNA target sequence customized for optimal CRISPR editing parameters, the simultaneous marker-free integration of 2-5 genes was achieved at a very high efficiency by a single transformation. The impressive performance of the SGM-CRISPR system shows that multiple copies of the synthetic gRNA binding sites in the genome can be effectively recognized by Cas9/gRNA complex for subsequent DNA cleavage, and the resulting Cas9-induced cell toxicity can be harnessed to integrate multiple foreign DNA fragments *via* simultaneous HDR. The successful re-construction of the biosynthetic pathways of GAA and costunolide further demonstrated the reliable uses of the SGM-CRISPR platform developed in this work. As the exact biosynthetic mechanisms of various guaianolide-type STLs remain unknown, the GAA- and costunolide-producing strains developed in this work will serve as cost-effective sources of biosynthetic substrates and intermediates that will be useful for future biochemical studies. Notably, the gene integration efficiencies achieved by the SGM-CRISPR platform are comparable to or higher than the efficiencies of previous multiplex genome editing strategies which have been discussed in Chapter 1 and summarized in Table 6. To our knowledge, this work is the first study to report the simultaneous genomic integration of five heterologous gene cassettes in a locus-specific manner using CRISPR-Cas9. Overall, the development of the SGM-CRISPR platform expands the scope of multiplex CRISPR-Cas9 toolbox by providing a highly-characterized platform for re-constitution of synthetic biological pathways.

Table 6. Summary of CRISPR-Cas9 multiplex strategies for integration of exogenous gene cassettes

Strategy	Number of target loci	Number of targeting gRNAs	Donor size	Efficiency
Multiplexed integration of metabolic pathways (Horwitz et al., 2015)	3	3	Total size of 24-kb	4.2%
CrEdit (Ronda et al., 2015)	3	3	5.1-kb/5.8-kb/6.6-kb	84%
CasEMBLR (Jakočiūnas et al., 2015a)	2 3	2 3	1.5-kb per site (in-vivo assembled)	58.6% 30.6%
EasyClone-MarkerFree (Jessop-Fabre et al., 2016)	3	3	Unknown	60-70%
Di-CRISPR (Shi et al., 2016)	11 10 10	1	8-kb 16-kb 24-kb	60-80%
mCAL (Finnigan and Thorner, 2016)	3	1	1.3-kb/1.7-kb/1.0-kb	87-100%
Landing Pad (Bourgeois et al., 2018)	2 3 4	1	1.5-kb per site	97% 94% 97%
SGM-CRISPR (This study)	2 3 4 5	1	2.4-kb/1.6-kb 2.4-kb/1.6-kb/2.4-kb 2.4-kb/1.6-kb/2.4-kb/1.6-kb 2.4-kb/1.6-kb/2.4-kb/1.6-kb/1.4-kb	97.6% 98.0% 93.7% 84.8%

Although simultaneous targeting of up to five loci was explored in this study, it is envisioned that the SGM-CRISPR platform can be further optimized to perform up to 8-gene integrations using the other CRISPR integration sites identified in the study. Currently, the upper limit for the number of CRISPR-Cas9 edits that can be accommodated by yeast HDR machinery is not clearly established, and further experiments will be needed to accurately assess the efficacies of 6-, 7- and 8- gene integrations. The integration events requiring more than 5 gene integrations might require improved performance of Cas9/gRNA complex or extended Cas9 exposure to maintain high integration efficiency, as more number of DSBs will need to be introduced to the genome with the increasing number of genome edits. Therefore, strategies for improved expression of Cas9 and gRNA and increased efficiency of Cas9 cleavage will need to be considered. One strategy could include use of an improved variant of Cas9 (iCas9) for increased Cas9 activity. It was previously shown that iCas9, which possesses D147Y and P411T double mutations, function with a higher targeting efficiency compared to the wild-type Cas9, most likely because the double mutations lie in the REC1 domain that plays an important role in DNA cleavage (Bao et al., 2015). Another strategy entails propagation of transformants in a liquid selection medium for an extended time period (2-6 days) to enable prolonged cellular exposure to Cas9 activity. This approach gives more time for Cas9/gRNA complex to generate the necessary DSBs and reduces the possible occurrence of partial integrants resulting from insufficient Cas9 pressure that leaves some integration sites intact with no Cas9 cleavage (Bourgeois et al., 2018). Compared to the typical multiplex strategies that use multiple gRNA expression cassettes, a single gRNA-mediated CRISPR-Cas9 uses a single gRNA expression cassette which might not provide sufficient intracellular levels of gRNA needed for guidance of Cas9 and subsequent DNA cleavage when more than five genomic sites are being targeted.

Therefore, increased expression of gRNA *via* expression on a ultra-high-copy plasmid with a truncated promoter of URA3 selection marker (Bao et al., 2015) or increased stability of a gRNA molecule *via* fusing of gRNA to the 3' end of HDV ribozyme (Ryan et al., 2014) might provide substantially increased amount of gRNA molecules available to associate with Cas9 and cause multiple DSBs in the genome.

Such strategies for improved Cas9/gRNA activity will be also of great interest in improving the current SGM-CRISPR platform for quadruple and quintuple gene integrations. The re-occurring observation from the quadruple and quintuple gene integrations in this work is the presence of partial integrants. One case of partial gene integrations was seen from the two low-costunolide producing strains among the 15 colonies screened for the integration of the five costunolide biosynthetic genes. The sequencing of the two strains revealed that two out of the five integration sites were intact without any sign of Cas9 cleavage or gene integration. Likewise, the partial integrants observed during the demonstrations of the 4- and 5- marker gene integrations were likely the product of insufficient Cas9/gRNA activity that led to the absence of Cas9-induced DSB in some of the target loci. The presence of partial integrants has a negative impact on the efficiency of successful genome editing as it reduces the likelihood of detecting the correct transformants with all of the desired integrations. Considering that partial integrants are likely to arise from insufficient Cas9 cleavage, the aforementioned strategies for improving the activity of Cas9/gRNA complex will be useful in further improving the performance of the current SGM-CRISPR multiplex gene integration platforms.

5.3 Conclusion

The robustness and versatility of the CRISPR-Cas9 technology allow for fast and highly-efficient strain engineering, greatly benefiting the field of yeast metabolic engineering. Among different CRISPR-based cell factory engineering methods, systems that enable a single-step introduction of multi-enzyme pathways in yeast are useful for building cell factories that synthesize high-value biochemicals originating in nature. To further increase the flexibility of targeted gene integration and extend the scope of CRISPR-Cas9 multiplex genome editing toolboxes, this research project provided a systematic characterization of various genomic loci suited for CRISPR-mediated gene integration and described a successful development of a multiplex gene editing platform operated by a single gRNA. This work showed that the selection of intergenic region is critically important for the efficient CRISPR-Cas9 genome-edit, providing the logic framework attesting that intergenic sites beside highly expressed genes are suited well for CRISPR targeting in yeast and likely in other organisms. In addition, this work demonstrated the capability of the SGM-CRISPR platform to achieve precise, single-step multi-gene integrations without the need for markers through a series of experiments. It is envisioned that the current SGM-CRISPR system can be widely used for integrating multiple genes in yeast and can be further advanced to accommodate a greater number of genome-edits in the future.

References

Andersen, S.L., and Sekelsky, J. (2010). Meiotic versus mitotic recombination: Two different routes for double-strand break repair. Bioassays *32*, 1058-1066.

Aylon, Y., and Kupiec, M. (2004). DSB repair: the yeast paradigm, DNA repair *3*, 797-815.

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

Bao, Z., Xiao, H., Liang, J., Zhang, L., Xiong, X., Sun, N., Si, T., and Zhao, H. (2015). Homology-integrated CRISPR-Cas (HI-CRISPR) system for one-step multigene disruption in Saccharomyces cerevisiae. ACS synthetic biology *4*, 585-594.

Bassett, Andrew r., Tibbit, C., Ponting, Chris p., and Liu, J.-L. (2013). Highly Efficient Targeted Mutagenesis of Drosophila with the CRISPR/Cas9 System. Cell Reports *4*, 220-228.

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

Borodina, I., and Nielsen, J. (2014). Advances in metabolic engineering of yeast Saccharomyces cerevisiae for production of chemicals. Biotechnol J *9*, 609-620.

Botstein, D., and Fink, G.R. (2011). Yeast: an experimental organism for 21st Century biology. Genetics *189*, 695-704.

Bourgeois, L., Pyne, M.E., and Martin, V.J.J. (2018). A Highly Characterized Synthetic Landing Pad System for Precise Multicopy Gene Integration in Yeast. ACS synthetic biology *7*, 2675-2685.

Brachmann, C.B., Davies, A., Cost, G.J., Caputo, E., Li, J., Hieter, P., and Boeke, J.D. (1998). 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.

Brouns, S.J.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J.H., Snijders, A.P.L., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van Der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960-964.

Bétermier, M., Bertrand, P., and Lopez, B.S. (2014). Is Non-Homologous End-Joining Really an Inherently Error-Prone Process? PLoS Genet *10*, e1004086.

Chee, M.K., and Haase, S.B. (2012). New and Redesigned pRS Plasmid Shuttle Vectors for Genetic Manipulation of Saccharomyces cerevisiae. G3 *2*, 515-526.

Chiruvella, K.K., Liang, Z., Birkeland, S.R., Basrur, V., and Wilson, T.E. (2013). Saccharomyces cerevisiae DNA Ligase IV Supports Imprecise End Joining Independently of Its Catalytic Activity. PLoS Genet *9*, e1003599.

Christianson, T.W., Sikorski, R.S., Dante, M., Shero, J.H., and Hieter, P. (1992). Multifunctional yeast high-copy-number shuttle vectors. Gene *110*, 119-122.

Clikeman, J., Khalsa, G., Barton, S., and Nickoloff, J. (2001). Homologous recombinational repair of double-strand breaks in yeast is enhanced by MAT heterozygosity through yKU-dependent and -independent mechanisms. Genetics *157*, 579-589.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A.*, et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819.

Da Silva, N.A., and Srikrishnan, S. (2012). Introduction and expression of genes for metabolic engineering applications in Saccharomyces cerevisiae. FEMS Yeast Res *12*, pp. 197-214.

Daer, R.M., Cutts, J.P., Brafman, D.A., and Haynes, K.A. (2017). The Impact of Chromatin Dynamics on Cas9-Mediated Genome Editing in Human Cells. ACS synthetic biology *6*, 428-438.

Daley, J.M., Palmbos, P.L., Wu, D., and Wilson, T.E. (2005). Nonhomologous end joining in yeast. Annu Rev Genet *39*, 431-451.

de Kraker, J.-W., Franssen, M., Joerink, M., de Groot, A., and Bouwmeester, H. (2002). Biosynthesis of costunolide, dihydrocostunolide, and leucodin. Demonstration of cytochrome P450-catalyzed formation of the lactone ring present in sesquiterpene lactones of chicory. Plant Physiology *129*, 257-268.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602-607.

Deshpande, R.A., and Wilson, T.E. (2007). Modes of interaction among yeast Nej1, Lif1 and Dnl4 proteins and comparison to human XLF, XRCC4 and Lig4. DNA Repair *6*, 1507-1516.

DiCarlo, J., Norville, J., Mali, P., Rios Villanueva, X., Aach, J.D., and Church, G.M. (2013). Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Research *41*, 4336-43

Emerson, C.H., and Bertuch, A.A. (2016). Consider the workhorse: Nonhomologous end-joining in budding yeast 1. Biochemistry and Cell Biology *94*, 396-406.

Featherstone, C., and Jackson, S.P. (1999). DNA double-strand break repair. Current Biology *9*, R759-R761.

Ferreira, R., Skrekas, C., Nielsen, J., and David, F. (2018). Multiplexed CRISPR/Cas9 Genome Editing and Gene Regulation Using Csy4 in Saccharomyces cerevisiae. ACS synthetic biology *7*, 10-15.

Finnigan, G.C., and Thorner, J. (2016). mCAL: A new approach for versatile multiplex action of Cas9 using one sgRNA and loci flanked by a programmed target sequence. G3 *6*, 2147-2156

Galanie, S., Thodey, K., Trenchard, I.J., Filsinger Interrante, M., and Smolke, C.D. (2015). Complete biosynthesis of opioids in yeast. Science *349*, 1095-1100.

Gao, Y., and Zhao, Y. (2014). Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. Journal of Integrative Plant Biology *56*, 343-349.

Garneau, J.E., Dupuis, M.-È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature *468*, 67-71.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. PNAS *109*, E2579-E2586.

Generoso, W.C., Gottardi, M., Oreb, M., and Boles, E. (2016). Simplified CRISPR-Cas genome editing for Saccharomyces cerevisiae. Journal of Microbiological Methods *127*, 203-205.

Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R., W., Belle, A., Dephoure, N., O'Shea, E., K., and Weissman, J., S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B.*, et al.* (2002). Functional profiling of the Saccharomyces cerevisiae genome. Nature *418*, 387-391.

Gietz, R.D., and Robert, H.S. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nature Protocols *2*, 31.

Gnügge, R., Liphardt, T., and Rudolf, F. (2016). A shuttle vector series for precise genetic engineering of Saccharomyces cerevisiae. Yeast *33*, 83-98.

Gnügge, R., and Rudolf, F. (2017). Saccharomyces cerevisiae Shuttle vectors. Yeast *34*, 205-221.

Gnügge, R., and Symington, L.S. (2017). Keeping it real: MRX-Sae2 clipping of natural substrates. Genes & development *31*, 2311-2312.

Goffeau, A., Barrell, B., Bussey, H., Davis, R., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J., Jacq, C., Johnston, M.*, et al.* (1996). Life with 6000 genes. Science *274*, 546-573.

Grewal, P.S., Modavi, C., Russ, Z.N., Harris, N.C., and Dueber, J.E. (2018). Bioproduction of a betalain color palette in Saccharomyces cerevisiae. Metab Eng *45*, 180-188.

Gueldener, U., Heinisch, J., Koehler, G., Voss, D., and Hegemann, J. (2002). A second set of loxP marker cassettes for Cremediated multiple gene knockouts in budding yeast. Nucleic Acids Research *30*, e23-e23.

Guerriero, G., Berni, R., Muñoz-Sanchez, J.A., Apone, F., Abdel-Salam, E.M., Qahtan, A.A., Alatar, A.A., Cantini, C., Cai, G., Hausman, J.-F.*, et al.* (2018). Production of Plant Secondary Metabolites: Examples, Tips and Suggestions for Biotechnologists. Genes *9*. E309

Gustavsson, M., and Lee, S.Y. (2016). Prospects of microbial cell factories developed through systems metabolic engineering. Microbial Biotechnology *9*, 610-617.

Hall, B.G., Acar, H., Nandipati, A., and Barlow, M. (2014). Growth rates made easy. Molecular biology and evolution *31*, 232-238.

Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J.-B., Reynolds, D.B., Yoo, J.*, et al.* (2004). Transcriptional regulatory code of a eukaryotic genome. Nature *431*, 99-104.

Hinnen, A., Hicks, J.B., and Fink, G.R. (1978). Transformation of yeast. PNAS *75*, 1929-1933.

Hong, K.-K., and Nielsen, J. (2012). Metabolic engineering of Saccharomyces cerevisiae : a key cell factory platform for future biorefineries. Cell Mol Life Sci *69*, 2671-2690.

Horwitz, A.A., Walter, J.M., Schubert, M.G., Kung, S.H., Hawkins, K., Platt, D.M., Hernday, A.D., Mahatdejkul-Meadows, T., Szeto, W., Chandran, S.S.*, et al.* (2015). Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas. Cell systems *1*, 88-96.

Hou, S., Qin, Q., and Dai, J. (2018). Wicket: A Versatile Tool for the Integration and Optimization of Exogenous Pathways in Saccharomyces cerevisiae. ACS synthetic biology *7*, 782-788.

Huh, W.-K., Falvo, J., V., Gerke, L., C., Carroll, A., S., Howson, R., W., Weissman, J., S., and O'Shea, E., K. (2003). Global analysis of protein localization in budding yeast. Nature *425*, 686.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.R.J., and Joung, J.K. (2013). Efficient In Vivo Genome Editing Using RNA-Guided Nucleases. Nature biotechnology *31*. 227-229

Ikezawa, N., Göpfert, J.C., Nguyen, D.T., Kim, S.U., O'Maille, P.E., Spring, O., and Ro, D.K. (2011). Lettuce costunolide synthase (CYP71BL2) and its homolog (CYP71BL1) from sunflower catalyze distinct regio- and stereoselective hydroxylations in sesquiterpene lactone metabolism. J Biol Chem *286*, 21601-21611.

Jakočiūnas, T., Bonde, I., Herrgård, M., Harrison, S.J., Kristensen, M., Pedersen, L.E., Jensen, M.K., and Keasling, J.D. (2015a). Multiplex metabolic pathway engineering using CRISPR/Cas9 in Saccharomyces cerevisiae. Metabolic Engineering *9*, 213-222.

Jakočiūnas, T., Rajkumar, A.S., Zhang, J., Arsovska, D., Rodriguez, A., Jendresen, C.B., Skjødt, M.L., Nielsen, A.T., Borodina, I., Jensen, M.K.*, et al.* (2015b). CasEMBLR: Cas9-Facilitated Multiloci Genomic Integration of in Vivo Assembled DNA Parts in Saccharomyces cerevisiae. ACS synthetic biology *4*, 1226-1234.

Janke, C., Magiera, M.M., Rathfelder, N., Taxis, C., Reber, S., Maekawa, H., Moreno-Borchart, A., Doenges, G., Schwob, E., Schiebel, E.*, et al.* (2004). A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. Yeast *21*, 947-962.

Jensen, K.T., Fløe, L., Petersen, T.S., Huang, J., Xu, F., Bolund, L., Luo, Y., and Lin, L. (2017). Chromatin accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing efficiency. FEBS letters *591*, 1892-1901.

Jessop-Fabre, M.M., Jakočiūnas, T., Stovicek, V., Dai, Z., Jensen, M.K., Keasling, J.D., and Borodina, I. (2016). EasyClone-MarkerFree: A vector toolkit for marker-less integration of genes intoSaccharomyces cerevisiaevia CRISPR-Cas9. Biotechnology Journal *11*, 1110-1117.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nature Biotechnology *31*, 233-239.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816.

Karim, A.S., Curran, K.A., and Alper, H.S. (2013). Characterization of plasmid burden and copy number in S accharomyces cerevisiae for optimization of metabolic engineering applications. FEMS Yeast Research *13*, 107-116.

Kavšček, M., Stražar, M., Curk, T., Natter, K., and Petrovič, U. (2015). Yeast as a cell factory: current state and perspectives. Microbial cell factories *14*, 94-94.

Komor, A.C., Badran, A.H., and Liu, D.R. (2017). CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell *168*, 20-36.

Kowalczykowski, S.C. (2015). An Overview of the Molecular Mechanisms of Recombinational DNA Repair. Cold Spring Harbor perspectives in biology *7*.

Kulkarni, R. (2016). Metabolic engineering; Biological art of producing useful chemicals. Resonance *21*, 233.

Labun, K., Montague, T.G., Krause, M., Torres Cleuren, Y.N., Tjeldnes, H., and Valen, E. (2019). CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. Nucleic acids research *47*, W171-W174.

Lee, M.E., Deloache, W.C., Cervantes, B., and Dueber, J.E. (2015). A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. ACS Synthetic Biology *4*. 975-986

Luo, X., Reiter, M.A., D'Espaux, L., Wong, J., Denby, C.M., Lechner, A., Zhang, Y., Grzybowski, A.T., Harth, S., Lin, W.*, et al.* (2019). Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. Nature *567*, 123-126

Löbs, A.-K., Schwartz, C., and Wheeldon, I. (2017). Genome and metabolic engineering in non-conventional yeasts: Current advances and applications. Synthetic and Systems Biotechnology *2*, 198-207.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J.J., Charpentier, E., Haft, D.H.*, et al.* (2015). An updated evolutionary classification of CRISPR–Cas systems. Nature Reviews Microbiology *13*, 722-736.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Güell Cargol, M., Dicarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*. 823-826

Mans, R., van Rossum, H.M., Wijsman, M., Backx, A., Kuijpers, N.G.A., van Den Broek, M., Daran-Lapujade, P., Pronk, J.T., van Maris, A.J.A., and Daran, J.-M.G. (2015). CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in Saccharomyces cerevisiae. FEMS yeast research *15*.

Mikkelsen, M.D., Buron, L.D., Salomonsen, B., Olsen, C.E., Hansen, B.G., Mortensen, U.H., and Halkier, B.A. (2012). Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. Metabolic Engineering *14*, 104-111.

Miller-Fleming, L., Giorgini, F., and Outeiro, T.F. (2008). Yeast as a model for studying human neurodegenerative disorders. Biotechnology Journal *3*, 421-421.

Mortimer, R. (2000). Evolution and variation of the yeast (Saccharomyces) genome. Genome Res *10*, 403-409.

Mortimer, R.K., and Johnston, J.R. (1986). Genealogy of principal strains of the Yeast Genetic Stock Center. Genetics *113*, 35-43.

Mumberg, D., Müller, R., and Funk, M. (1995). Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene *156*, 119-122.

Nguyen, D.T., Göpfert, J.C., Ikezawa, N., Macnevin, G., Kathiresan, M., Conrad, J., Spring, O., and Ro, D.-K. (2010). Biochemical conservation and evolution of germacrene A oxidase in asteraceae. The Journal of biological chemistry *285*, 16588-16598.

Nielsen, J., Larsson, C., van Maris, A., and Pronk, J. (2013). Metabolic engineering of yeast for production of fuels and chemicals. Current Opinion in Biotechnology *24*, 398-404.

Ostergaard, S., Olsson, L., and Nielsen, J. (2000). Metabolic Engineering of Saccharomyces cerevisiae. Microbiology and Molecular Biology Reviews *64*, 34-50

Paddon, C.J., Westfall, P.J., Pitera, D.J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M.D., Tai, A., Main, A., Eng, D.*, et al.* (2013). High-level semi-synthetic production of the potent antimalarial artemisinin. Nature *496*, 528-532.

Park, E., Song, J.H., Kim, M.S., Park, S.-H., and Kim, T.S. (2016). Costunolide, a sesquiterpene lactone, inhibits the differentiation of pro-inflammatory CD4+ T cells through the modulation of mitogen-activated protein kinases. International Immunopharmacology *40*, 508-516.

Pawelczak, K.S., Gavande, N.S., Vandervere-Carozza, P.S., and Turchi, J.J. (2018). Modulating DNA Repair Pathways to Improve Precision Genome Engineering. ACS chemical biology *13*, 389-396.

Peng, Z., Wang, Y., Fan, J., Lin, X., Liu, C., Xu, Y., Ji, W., Yan, C., and Su, C. (2017). Costunolide and dehydrocostuslactone combination treatment inhibit breast cancer by inducing cell cycle arrest and apoptosis through c-Myc/p53 and AKT/14-3-3 pathway. Scientific Reports *7*, 41254.

Reider Apel, A., D'Espaux, L., Wehrs, M., Sachs, D., Li, R.A., Tong, G.J., Garber, M., Nnadi, O., Zhuang, W., Hillson, N.J.*, et al.* (2017). A Cas9-based toolkit to program gene expression in Saccharomyces cerevisiae. Nucleic acids research *45*, 496-508.

Ro, D.-K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J.*, et al.* (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature *440*, 940-943.

Rodgers, K., and McVey, M. (2016). Error-Prone Repair of DNA Double-Strand Breaks. Journal of Cellular Physiology *231*, 15-24.

Ronda, C., Maury, J., JakociuInas, T., Baallal Jacobsen, S.A., Germann, S.M., Harrison, S.J., Borodina, I., Keasling, J.D., Jensen, M.K., and Nielsen, A.T. (2015). CrEdit: CRISPR mediated multi-loci gene integration in Saccharomyces cerevisiae. Microbial Cell Factories *14*.

Ryan, O.W., Skerker, J.M., Maurer, M.J., Li, X., Tsai, J.C., Poddar, S., Lee, M.E., Deloache, W., Dueber, J.E., Arkin, A.P.*, et al.* (2014). Selection of chromosomal DNA libraries using a multiplex CRISPR system. eLife *3*.

San Filippo, J., Sung, P., and Klein, H. (2008). Mechanism of Eukaryotic Homologous Recombination. Annu Rev Biochem *77*, 229-257.

Shi, S., Liang, Y., Zhang, M.M., Ang, E.L., and Zhao, H. (2016). A highly efficient single-step, markerless strategy for multi-copy chromosomal integration of large biochemical pathways in Saccharomyces cerevisiae. Metabolic Engineering *33*, 19-27.

Siddiqui, M.S., Thodey, K., Trenchard, I., and Smolke, C.D. (2012). Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools. FEMS Yeast Res *12*, 144-170.

Smith, J.D., Suresh, S., Schlecht, U., Wu, M., Wagih, O., Peltz, G., Davis, R.W., Steinmetz, L.M., Parts, L., and St Onge, R.P. (2016). Quantitative CRISPR interference screens in yeast identify chemical-genetic interactions and new rules for guide RNA design. Genome biology *17*, 45-45.

Solis-Escalante, D., Kuijpers, N.G.A., Bongaerts, N., Bolat, I., Bosman, L., Pronk, J.T., Daran, J.M., and Daran-Lapujade, P.A.S. (2012). amdSYM, a new dominant recyclable marker cassette for Saccharomyces cerevisiae. FEMS Yeast Research, *13*, 1567-1356.

Solis-Escalante, D., Kuijpers, N.G.A., Linden, F.H., Pronk, J.T., Daran, J.M., and Daran-Lapujade, P. (2014). Efficient simultaneous excision of multiple selectable marker cassettes using I-SceI-induced double-strand DNA breaks in Saccharomyces cerevisiae. FEMS Yeast Research *14*, 741-754.

Solis-Escalante, D., van Den Broek, M., Kuijpers, N.G.A., Pronk, J.T., Boles, E., Daran, J.-M., and Daran-Lapujade, P. (2015). The genome sequence of the popular hexose-transport-deficient Saccharomyces cerevisiae strain EBY.VW4000 reveals LoxP/Cre-induced translocations and gene loss. FEMS yeast research *15*.

Sopko, R., Huang, D., Preston, N., Chua, G., Papp, B., Kafadar, K., Snyder, M., Oliver, S.G., Cyert, M., Hughes, T.R.*, et al.* (2006). Mapping Pathways and Phenotypes by Systematic Gene Overexpression. Molecular Cell *21*, 319-330.

Steffen, K.K., Kennedy, B.K., and Kaeberlein, M. (2009). Measuring Replicative Life Span in the Budding Yeast. Journal of Visualized Experiments *28*. 1209

Storici, F., Coglievina, M., and Bruschi, C.V. (1999). A 2-μm DNA-based marker recycling system for multiple gene dispruption in the yeast Saccharomyces cerevisiae. Yeast *15*, 271-283.

Storici, F., Lewis, L.K., and Resnick, M., A. (2001). In vivo site-directed mutagenesis using oligonucleotides. Nature Biotechnology *19*, 773-776

Storici, F., Durham, C., Gordenin, D., and Resnick, M. (2003). Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. PNAS *100*, 14994-14995.

Stovicek, V., Borodina, I., and Forster, J. (2015). CRISPR–Cas system enables fast and simple genome editing of industrial Saccharomyces cerevisiae strains. Metabolic Engineering Communications *2*, 13-22.

Tao, G., Yi-Li, F., Jing-Jing, X., Qian, L., Xiu-Na, S., Ji-Feng, X., Na, K., Si-Cheng, L., Guo-Qiao, C., Yue, W.*, et al.* (2018). Harnessing accurate non-homologous end joining for efficient precise deletion in CRISPR/Cas9-mediated genome editing. Genome Biology *19*, 1-20.

Thyme, S.B., Akhmetova, L., Montague, T.G., Valen, E., and Schier, A.F. (2016). Internal guide RNA interactions interfere with Cas9-mediated cleavage. Nature Communications *7*.

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

Uusi-Makela, M.I.E., Barker, H.R., Bauerlein, C.A., Hakkinen, T., Nykter, M., and Ramet, M. (2018). Chromatin accessibility is associated with CRISPR-Cas9 efficiency in the zebrafish (Danio rerio). PLoS ONE *13*, e0196238.

Vanegas, K.G., Lehka, B.J., and Mortensen, U.H. (2017). SWITCH: a dynamic CRISPR tool for genome engineering and metabolic pathway control for cell factory construction in Saccharomyces cerevisiae. Microbial Cell Factories *16*.

Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E., Hieter, P., Vogelstein, B., and Kinzler, K.W. (1997). Characterization of the Yeast Transcriptome. Cell *88*, 243-251.

Wach, A., Brachat, A., Pöhlmann, R., and Philippsen, P. (1994). New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae. Yeast *10*, 1793-1808.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013). One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. Cell *153*, 910-918.

Wink, M. (2015). Modes of Action of Herbal Medicines and Plant Secondary Metabolites. Meidinces 2, 251-286.

Wright, W., Shah, S.S., and Heyer, W. (2018). Homologous recombination and the repair of DNA double-strand breaks. J Biol Chem *293*, 10524-10535.

Xiao, M., Zhang, Y., Chen, X., Lee, E.-J., Barber, C.J.S., Chakrabarty, R., Desgagné-Penix, I., Haslam, T.M., Kim, Y.-B., Liu, E.*, et al.* (2013). Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. Journal of Biotechnology *166*, 122-134.

Yarrington, R., Verma, S., Schwartz, S., Trautman, J., and Carroll, D. (2018). Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. PNAS *115*, 9351.

Zhang, G.-C., Kong, I., Kim, H., Liu, J.-J., Cate, J., and Jin, Y.-S. (2014). Construction of a Quadruple Auxotrophic Mutant of an Industrial Polyploid Saccharomyces cerevisiae Strain by Using RNA-Guided Cas9 Nuclease. Applied and Environmental Microbiology *80*, 7694-7701

Zhang, X.-H., Tee, L.Y., Wang, X.-G., Huang, Q.-S., and Yang, S.-H. (2015). Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. Molecular Therapy - Nucleic Acids *4*, e264.

Zhang, Y., Wang, J., Wang, Z., Zhang, Y., Shi, S., Nielsen, J., and Liu, Z. (2019). A gRNA-tRNA array for CRISPR-Cas9 based rapid multiplexed genome editing in Saccharomyces cerevisiae. Nature Communications *10*.

Zhang, Z., Moo-Young, M., and Chisti, Y. (1996). Plasmid stability in recombinant Saccharomyces cerevisiae. Biotechnology Advances *14*, 401-435.

Appendix

Table 7. Yeast strains used in this work

Strain Name	Use	Description	Source
BY4742	N/A	MATα, his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0	Brachmann et al. 1998
SBY001	GFP expression analysis	H1::pTEF1-yEGFP-tADH1	This study
SBY002		H2::pTEF1-yEGFP-tADH1	
SBY003		H3::pTEF1-yEGFP-tADH1	
SBY004		H4::pTEF1-yEGFP-tADH1	
SBY005		H5::pTEF1-yEGFP-tADH1	
SBY006		H6::pTEF1-yEGFP-tADH1	
SBY007		H7::pTEF1-yEGFP-tADH1	
SBY008		H8::pTEF1-yEGFP-tADH1	
SBY101	SGM CRISPR platforms	H1::sTarget1	This study
SBY201		H1::sTarget2	
SBY102		H1::sTarget1 H2::sTarget1	
SBY103		H1::sTarget1 H2::sTarget1 H5::sTarget1	
SBY104		H1::sTarget1 H2::sTarget1 H5::sTarget1 H7::sTarget1	
SBY105		H1::sTarget1 H2::sTarget1 H5::sTarget1 H7::sTarget1 H4::sTarget1	
SBY.B01	Betalain production & antibiotic resistant strains	H1:: pTDH3- BvCYP76AD1^{W13L} -tTDH1 H2:: pCCW12-MjDOD-tADH1	This study
SBY.B02		H1:: pTDH3- BvCYP76AD1^{W13L} -tTDH1 H2:: pCCW12-MjDOD-tADH1 H3:: pFBA1-MjcDOPA5GT-tSSA1	
SBY.B03		H1:: pTDH3- BvCYP76AD1^{W13L} -tTDH1 H2:: pCCW12-MjDOD-tADH1 H3:: pFBA1-MjcDOPA5GT-tSSA1 H7:: pAgTEF-HygR-tCYC1	
SBY.B04		H1:: pTDH3- BvCYP76AD1^{W13L} -tTDH1 H2:: pCCW12-MjDOD-tADH1 H3:: pFBA1-MjcDOPA5GT-tSSA1 H7:: pAgTEF-HygR-tCYC1 H4:: pAgTEF-KanR-tAgTEF	
SBY.C01	Sesquiterpenoid	H1:: pGAL1-LsGAS-tCYC1 H2:: pGAL10-LsGAO-tADH1	This study

	production strains	H5:: pGAL1-AaCPR-tCYC1 H7:: pGAL10-tHMGR-tADH1	
SBY.C02		H1:: pGAL1-LsGAS-tCYC1 H2:: pGAL10-LsGAO-tADH1 H5:: pGAL1-AaCPR-tCYC1 H7:: pGAL10-tHMGR-tADH1 H4:: pGAL10-LsCOS-tADH1	

Table 8. Primers used in this work

Underlined sequences indicate homology arms used for Gibson assembly or restriction enzyme sites for traditional ligation-based cloning.

Primer #	Primer sequence	Purpose
1	GCTTTAATTTGCGGCCGGTACTCTTTGAAAAG ATAATGTATGATTATGCTTTC	Gibson assembly of gRNA cassette into p416-CAS, pESC-LEU or pCut
2	CTATAGGGGAATTGGAGACATAAAAAACAA AAAAAGCACCAC	
3	GCCGATTCATTAATGCAGTCTTTGAAAAGAT AATGTATGATTATGCTTTC	
4	TCGCTATTACGCCAGCCTAGGAGACATAAAA AACAAAAAAAGCACCAC	
5	GCCGATTCATTAATGCAGCCTAGGCATAGCT TCAAAATGTTTCTACTCC	Gibson assembly of Cas9 cassette into pESC-URA
6	CTTCGCTATTACGCCAGCCTAGGGTACCGGC CGCAAATTAAAGCC	
7	CCAGTGTGTCTGAAATGATGGATCATTTATCT TTCACTGCGGAG	Assembly of pSNR52-gRNA(R1)-tSUP4
8	CATCATTTCAGACACACTGGGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGG	
9	CCATGGATATTGTGCTGATGGATCATTTATCT TTCACTGCGGAG	Assembly of pSNR52-gRNA(R2)-tSUP4
10	CATCAGCACAATATCCATGGGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGG	
11	TCTGTGAGGGCCGATTATGCGATCATTTATCT TTCACTGCGGAG	Assembly of pSNR52-gRNA(R3)-tSUP4
12	GCATAATCGGCCCTCACAGAGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGG	
13	TCCATGCTCTTTGCCGAACTGATCATTTATCT TTCACTGCGGAG	Assembly of pSNR52-gRNA(R4)-tSUP4
14	AGTTCGGCAAAGAGCATGGAGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGG	
15	TCCCCGCGCTCGTACTTTACGATCATTTATCT TTCACTGCGGAG	Assembly of pSNR52-gRNA(R5)-tSUP4
16	GTAAAGTACGAGCGCGGGGAGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGG	
17	TGTTTTGGTTCTCGCGAGTGGATCATTTATCT TTCACTGCGGAG	Assembly of pSNR52-gRNA(R6)-tSUP4
18	CACTCGCGAGAACCAAAACAGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGG	
19	CCCTTCTCTACTAGCATTGGGATCATTTATCT TTCACTGCGGAG	Assembly of pSNR52-gRNA(H1)-tSUP4
20	CCAATGCTAGTAGAGAAGGGGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGG	
21	CGCGTTTATTTACCTATCTCGATCATTTATCT TTCACTGCGGAG	Assembly of pSNR52-gRNA(H2)-tSUP4
22	GAGATAGGTAAATAAACGCGGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGG	
23	CTTCTTTTGTATAAATCACGGATCATTTATCT TTCACTGCGGAG	Assembly of pSNR52-gRNA(H3)-tSUP4

24	CGTGATTTATACAAAAGAAGGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGG	
25	CATGATTGATGTTATGGTGGGATCATTTATCT TTCACTGCGGAG	Assembly of pSNR52-gRNA(H4)-tSUP4
26	CCACCATAACATCAATCATGGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGG	
27	TCATACTATTATCAGGGCCAGATCATTTATCT TTCACTGCGGAG	Assembly of pSNR52-gRNA(H5)-tSUP4
28	TGGCCCTGATAATAGTATGAGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGG	
29	CCGGATGAAGTGTTTCGCCTGATCATTTATCT TTCACTGCGGAG	Assembly of pSNR52-gRNA(H6)-tSUP4
30	AGGCGAAACACTTCATCCGGGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGG	
31	CGGATCGTCGGTTGTGATACGATCATTTATCT TTCACTGCGGAG	Assembly of pSNR52-gRNA(H7)-tSUP4
32	GTATCACAACCGACGATCCGGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGG	
33	TTGCTGGTACTCTGGGAAAAGATCATTTATCT TTCACTGCGGAG	Assembly of pSNR52-gRNA(H8)-tSUP4
34	TTTTCCCAGAGTACCAGCAAGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGG	
35	CCCTATAGTGAGTCGTATTACGTCGAGTTTAT CATTATCAATACTGCC	Assembly of pESC-URA::LacZ
36	ATCCGTAATCATGGTCATGTTCCACTAGTTCT AGATCCGTCGAAAC	
37	CGGATCTAGAACTAGTGGAACATGACCATGA TTACGGATTCACTG	
38	CTCTGGCGAAGAATTGTTAATTTATTTTTGAC ACCAGACCAACTGG	
39	TCTTAGCTAGCCGCGGTACcGGATCCCATAGC TTCAAAATGTTTCTACTCC	Assembly of pESC-URA::yEGFP
40	GGTACTAGTGGCTCTAGAAAACTTAGATTAG ATTG	
41	AAGTTTTCTAGAGCCACTAGTaccATGTCTAA AGGTGAAGAATTATTCACTGG	
42	GTAATCCATCGATACTAGTGCTTATTTGTACA ATTCATCCATACCATGG	
43	TATCGATAAGCTTGATATCGAATTCTGTCTGT GGTATTCTCGATTTGTG	Assembly of pTargetR1
44	GGCGGCCGCTCTAGAACTAGTCGTAGTGTGC TGGTTCCACC	
45	CACTATAGGGCGAATTGGGTACGCAAACAGA AAGTGCGATTCTC	Assembly of pTargetR2
46	GTGGCGGCCGCTCTAGAACTAGCCCTCATGA TCTCCTAATTTGATG	
47	CACTATAGGGCGAATTGGGTACGACTGTTAG TTGGATATCAGTAATGAGACG	Assembly of pTargetR3
48	GTGGCGGCCGCTCTAGAACTAGCAGCTCTCT CTTAAACACATAGTAACAGC	
49	CTATAGGGCGAATTGGGTACGCGGCCGCCTT CGGCATTAATTTGGCCTAC	Assembly of pTargetR4
50	TACCGTCGACCTCGAGGGGGGGCCCTAATCC AATTGTCAATGAGCAGG	

51	TACCGGGCCCCCCCTCGAGGTCGACCATCTC TTGCAGTATAAGGATGCAG	Assembly of pTargetR5
52	GCTGGAGCTCCACCGCGGTGGCGGCCGCGCC TGTTGACCAGTTGCG	
53	TACCGGGCCCCCCCTCGAGGTCGACGATTGT AGGCGGCTAGAAGGAG	Assembly of pTargetR6
54	GCTGGAGCTCCACCGCGGTGGCGGCCGCAAC AATATGCGTCGGAGTCATC	
55	CACGAGGTACCTCGAGTTTATCATTATCAAT ACTGCC	Cloning of LacZ & GFP cassettes into pTargetR
56	CACGAGGATCCTCGAGTTTATCATTATCAAT ACTGCC	
57	AGGCAGGTACCGAGCGACCTCATGCTATACC TG	
58	AGGCAGGATCCGAGCGACCTCATGCTATACC TG	
59	CACGAGGTACCCATAGCTTCAAAATGTTTCT ACTCC	
60	CACGAGGATCCCATAGCTTCAAAATGTTTCT ACTCC	
61	AGGTCGACGGTATCGATAGCGGCCGCTGACT CTGTGGCGG	Assembly of pTargetH1-GFP
62	GTATAGCATGAGGTCGCTCGGATCCCTTCTCT ACTAGCATTGGAC	
63	AGAAACATTTTGAAGCTATGCTGCAGACACC CCTCCGC	
64	GGCCGCTCTAGAACTAGCTTCGAGATACCCC TCATC	
65	TACCGGGCCCCCCCTCGAGGCGGCCGCGGAG GAGGAAGATGTCAG	Assembly of pTargetH2-GFP
66	GGTATAGCATGAGGTCGCTCGGATCCGTGTG TTTATTACATGAAGAAGAAGTTAG	
67	AGTAGAAACATTTTGAAGCTATGGATCCGCG TTTATTTACCTATCTCTAAACTTC	
68	GCTGGAGCTCCACCGCGGTGGCGGCCGCAGG ACAGCTAGGTTTATGC	
69	TACCGGGCCCCCCCTCGAGGCGGCCGCCAAG CTAGAGATCACAAGC	Assembly of pTargetH3-GFP
70	GGTATAGCATGAGGTCGCTCCTAGGCTTCTTT TGTATAAATCACGGAAGG	
71	AGTAGAAACATTTTGAAGCTATGAATTCAGA ATATGCTGAATACTTGGTATATTAC	
72	GCTGGAGCTCCACCGCGGTGGCGGCCGCTCC TGTAAATGCCGCAGATC	
73	TACCGGGCCCCCCCTCGAGGCGGCCGCAAGC CAATTGAATATACGACTG	Assembly of pTargetH4-GFP
74	GGTATAGCATGAGGTCGCTCCTAGGCATGAT TGATGTTATGGTGG	
75	AGTAGAAACATTTTGAAGCTATGAATTCTTG CTGGTGTATTTACCAATAATG	
76	GCTGGAGCTCCACCGCGGTGGCGGCCGCGGG AATAAACAAGTAGAATAGATAGAG	
77	TACCGGGCCCCCCCTCGAGGCGGCCGCGGTG ATAGACCGCCACAG	Assembly of pTargetH5-GFP

78	GGTATAGCATGAGGTCGCTCCTAGGTGAAGC CAAAATAAAGGATTCG	
79	AGTAGAAACATTTTGAAGCTATGAATTCTCA TACTATTATCAGGGCCAG	
80	GCTGGAGCTCCACCGCGGTGGCGGCCGCGAA AAAGAAAGCATCTAAGAACTTG	
81	TACCGGGCCCCCCCCTCGAGGCGGCCGCTTCG GAATCAGAAACGAC	Assembly of pTargetH6-GFP
82	GGTATAGCATGAGGTCGCTCCTAGGTAAAAA ACTACTTCAGTACATCTCTC	
83	AGTAGAAACATTTTGAAGCTATGAATTCCCG GATGAAGTGTTTCG	
84	GCTGGAGCTCCACCGCGGTGGCGGCCGCGTT TCAGTTGTGCTAGTTCTAC	
85	TACCGGGCCCCCCCCTCGAGGCGGCCGCTTTC TCTCGGGGAAATTTTG	Assembly of pTargetH7-GFP
86	GGTATAGCATGAGGTCGCTCCTAGGGTCATG GCGGCTATTTTTTT	
87	AGTAGAAACATTTTGAAGCTATGAAGCTTGG ATCGTCGGTTGTGATAC	
88	GCTGGAGCTCCACCGCGGTGGCGGCCGCAAG ACATCACAAGCATTCATTTAC	
89	TACCGGGCCCCCCCCTCGAGGCGGCCGCATTT TGTGATAAAGATGGTTAACTGAC	Assembly of pTargetH8-GFP
90	GGTATAGCATGAGGTCGCTCCTAGGGCTGGT ACTCTGGGAAAATG	
91	AGTAGAAACATTTTGAAGCTATGAATTCGCA TCCTTGATGCCTACTC	
92	GCTGGAGCTCCACCGCGGTGGCGGCCGCTAA TGAATAAAGTCTTTGTGTGGTTTG	
93	TAGCAATCTAATCTAAGTTTTCTAGAAATGG ACTAAAGAAGTACTCCATTGGGCTC	Assembly of pCut-mCAS-H2
94	CTGAACCAGTTGGATAAAGAGTTTG	
95	CAAGGCAGAAACTAACTTCTTCTTCATGTAA TAAACACACCGTACGCTGCAGGTCGAC	Assembly of pTargetH2-His3MX6
96	GATATAAGGTGTTGAAGTTTAGAGATAGGTA AATAAACGCATCGATGAATTCGAGCTCG	
97	CACCAAGTCTTTGGAAACTTTCCGTACCACTA ACACTTTACGTACGCTGCAGGTCGAC	FBA1 yEGFP tagging
98	ACTCATTAAAAAACTATATCAATTAATTTGA ATTAACTTAATCGATGAATTCGAGCTCG	
99	GGTTCTAATGGAATCTGAACGTGAAGCTCGT CGTTTGCGT CGTACGCTGCAGGTCGAC	RPS28A yEGFP tagging
100	TTGTATAGCTGCAACCTTCAATCTGCAAATA AGCTTCCTA ATCGATGAATTCGAGCTCG	
101	TTTTAACTCGGATAGACTTTCTCCAATGGAAA TCTACAAT CGTACGCTGCAGGTCGAC	SPO14 yEGFP tagging
102	CAGGTAATGGTGTGTTCCTGGTCGTTTTTATA TTCCCTTA ATCGATGAATTCGAGCTCG	
103	CCTCGATTGCGCGACGCTTTGATCATTTATCT TTCACTGCGGAG	Assembly of pSNR52-syn.gRNA#1-t SUP4
104	AAAGCGTCGCGCAATCGAGGGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGG	

105	CCTGTTATGTACGACGCTTTGATCATTTATCTTTCACTGCGGAG	Assembly of pSNR52-syn.gRNA#2-tSUP4
106	AAAGCGTCGTACATAACAGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG	
107	CCTCCTCGATTGCGCGACGCTTTCTTCTCTACTAGCATTGGAC	Integration of sTarget#1 into H1 integration site
108	AAAGCGTCGCGCAATCGAGGAGGACACCCCTCCGC	
109	CCTCCTGTTATGTACGACGCTTTCTTCTCTACTAGCATTGGAC	Integration of sTarget#2 into H1 integration site
110	AAAGCGTCGTACATAACAGGAGGACACCCCTCCGC	
111	CCTCCTCGATTGCGCGACGCTTTGTGTGTTTATTACATGAAGAAGAAGTTAG	Integration of sTarget#1 into H2 integration site
112	AAAGCGTCGCGCAATCGAGGAGGGCGTTTATTTACCTATCTCTAAACTTC	
113	CCTCCTGTTATGTACGACGCTTTGTGTGTTTATTACATGAAGAAGAAGTTAG	Integration of sTarget#2 into H2 integration site
114	AAAGCGTCGTACATAACAGGAGGGCGTTTATTTACCTATCTCTAAACTTC	
115	CCTCCTCGATTGCGCGACGCTTTTGAAGCCAAAATAAAGGATTCG	Integration of sTarget#1 into H5 integration site
116	AAAGCGTCGCGCAATCGAGGAGGTCATACTATTATCAGGGCCAG	
117	CCTCCTCGATTGCGCGACGCTTTGTCATGGCGGCTATTTTTTT	Integration of sTarget#1 into H7 integration site
118	AAAGCGTCGCGCAATCGAGGAGGGGATCGTCGGTTGTGATAC	
119	CCTCCTCGATTGCGCGACGCTTTCATGATTGATGTTATGGTGG	Integration of sTarget#1 into H4 integration site
120	AAAGCGTCGCGCAATCGAGGAGGTTGCTGGTGTATTTACCAATAATG	
121	TTAATTAAAGTCCAATGCTAGTAGAGAAGGCCATACAGAACCAGCCGTTCAG	Assembly of pTargetH1-*BvCYP76AD1*
122	AGAGCGCGGAGGGGTGTCTGCAGCAGTTCGAGTTTATCATTATCAATACTGC	
123	GCAGAAACTAACTTCTTCTTCATGTAATAAACACACGGATCCGAGCGACCTCATGCTATACCTGAGAAA	Assembly of pTargetH2-*MjDOD*
124	TGAAGTTTAGAGATAGGTAAATAAACGCGCCTAGGATTCCAGATGCGGGAATTAGCTTG	
125	GGCGCGAATCCTTTATTTTGGCTTCACATAAAATTAAAGTAGCAGTACTTCAACCATTAG	Assembly of pTargetH5-*MjcDOPA5GT*
126	CCTTTTTCTGGCCCTGATAATAGTATGAGACGCAAGCCCTAAGAAATGAATAAC	
127	AAAAAAAATAGCCGCCATGACCGACATGGAGGCCCAGAATACC	Assembly of pTargetH7-*hphNT1*
128	ACCGTATCACAACCGACGATCCACTTCGAGCGTCCCAAAACCTTC	
129	CCTATATTCCACCATAACATCAATCATGCGACATGGAGGCCCAGAATACC	Assembly of pTargetH4-*kanMX4*
130	TAAACATTATTGGTAAATACACCAGCAAGCAGTATAGCGACCAGCATTCACATA	

131	TTAATTAAAGTCCAATGCTAGTAGAGAAGGCTTCGAGCGTCCCAAAACCTTC	Assembly of pTargetH1-*LsGAS*
132	AGAGCGCGGAGGGGTGTCTGCAACCTTCTCTTTGGAACTTTCAGTAATACG	
133	GCAGAAACTAACTTCTTCTTCATGTAATAAACACACGGATCCGAGCGACCTCATGCTATACCTGAGAAA	Assembly of pTargetH2-*LsGAO*
134	AGTTTAGAGATAGGTAAATAAACGCGGATCGGTTTGTGGGGCCAGGTTACTG	
135	GGCGCGAATCCTTTATTTTGGCTTCACCTTCGAGCGTCCCAAAACCTTC	Assembly of pTargetH5-*AaCPR*
136	CCTTTTTCTGGCCCTGATAATAGTATGAGACCTTCTCTTTGGAACTTTCAGTAATACG	
137	AAAAAAAATAGCCGCCATGACCGAGCGACCTCATGCTATACCTG	Assembly of pTargetH7-*tHMGR*
138	ACCGTATCACAACCGACGATCCAGGTTTGTGGGGCCAGGTTAC	
139	CCTATATTCCACCATAACATCAATCATGCGAGCGACCTCATGCTATACCTG	Assembly of pTargetH4-*LsCOS*
140	TAAACATTATTGGTAAATACACCAGCAAGGGTTTGTGGGGCCAGGTTAC	
141	GGACATACTTAGCAGGTAGC	Genotyping of R1 integration site
142	CTTTCACATCAACACTACTTTCAAGATG	Genotyping of R2 integration site
143	GGAAAGTAAGGTACCGTCTATGAGG	Genotyping of R3 integration site
144	GCAAGGCAACCTGAATATGC	Genotyping of R4 integration site
145	CTTAAAGCCGCAAAAGATGAAG	Genotyping of R5 integration site
146	AGGCGCAAATATCCAAAAGG	Genotyping of R6 integration site
147	GTAGTGTGCGTGAATGAAGG	Genotyping of H1 integration site
148	GATTGCAAGGAGAGTGAAAGAG	Genotyping of H2 integration site
149	CTTTGTTAGCCGTAGCTGC	Genotyping of H3 integration site
150	ACCGTATATTCTAAAAACAAGTACTCC	Genotyping of H4 integration site
151	AAATCTTGGACAGACAACTTTGAAG	Genotyping of H5 integration site
152	CTTGAACGGTTAGTTATGATCTGTG	Genotyping of H6 integration site
153	GGTACCTAGCATCATATGGGAAG	Genotyping of H7 integration site
154	CGGTTGCTTTGATTATGGAAC	Genotyping of H8 integration site
155	TTATTTGTACAATTCATCCATACCATGG	Genotyping of yEGFP integration
156	GGATGGTGGCGCTGG	Genotyping of LacZ integration
157	CTTGGTGTTTTAAAACTAAAAAAAAGACTAAC	Genotyping of *BvCYP76AD1* integration
158	TATGTTAGGAAAGAATAAAGGAGAATAAG	Genotyping of *MjDOD* integration
159	TTTGAATATGTATTACTTGGTTATGGTTATATATGAC	Genotyping of *MjcDOPA5GT* integration
160	CTGTACAGACGCGTGTACGC	Genotyping of *hphNT1* integration
161	GATTGCGCCTGAGCGAGAC	Genotyping of *kanMX4* integration

Table 9. Plasmids used in this work

Plasmid name	Content	Origin of replication	Selection marker	Source
pESC-URA	N/A	2μ	URA3	Agilent
pESC-LEU	N/A	2μ	LEU2	Agilent
pBluescript SK	N/A	N/A	N/A	Agilent
p416	N/A	CEN/ARS	URA3	Mumberg et al. 1995
p414-TEF1p-Cas9-CYC1t	pTEF1-hcoCas9-tCYC1	CEN/ARS	TRP1	DiCarlo et al. 2013
p426-SNR52p-gRNA.CAN1.Y-SUP4t	pSNR52-gRNA.CAN1.Y-tSUP4	2μ	URA3	DiCarlo et al. 2013
p416-CAS	pTEF1-hcoCas9-tCYC1	CEN/ARS	URA3	Gina Kwon (unpublished)
p416-CAS-gRNA-R1	pTEF1-hcoCas9-tCYC1 pSNR52-gRNA(R1)-tSUP4	CEN/ARS	URA3	This study
p416-CAS-gRNA-R2	pTEF1-hcoCas9-tCYC1 pSNR52-gRNA(R2)-tSUP4	CEN/ARS	URA3	This study
p416-CAS-gRNA-R3	pTEF1-hcoCas9-tCYC1 pSNR52-gRNA(R3)-tSUP4	CEN/ARS	URA3	This study
p416-CAS-gRNA-R4	pTEF1-hcoCas9-tCYC1 pSNR52-gRNA(R4)-tSUP4	CEN/ARS	URA3	This study
p416-CAS-gRNA-R5	pTEF1-hcoCas9-tCYC1 pSNR52-gRNA(R5)-tSUP4	CEN/ARS	URA3	This study
p416-CAS-gRNA-R6	pTEF1-hcoCas9-tCYC1 pSNR52-gRNA(R6)-tSUP4	CEN/ARS	URA3	This study
p416-CAS-gRNA-H1	pTEF1-hcoCas9-tCYC1 pSNR52-gRNA(H1)-tSUP4	CEN/ARS	URA3	This study
p416-CAS-gRNA-H2	pTEF1-hcoCas9-tCYC1 pSNR52-gRNA(H2)-tSUP4	CEN/ARS	URA3	This study
p416-CAS-gRNA-P1	pTEF1-hcoCas9-tCYC1 pSNR52-gRNA(P1)-tSUP4	CEN/ARS	URA3	This study
p416-CAS-gRNA-P2	pTEF1-hcoCas9-tCYC1 pSNR52-gRNA(P2)-tSUP4	CEN/ARS	URA3	This study
pLEU-gRNA-R1	pSNR52-gRNA(R1)-tSUP4	2μ	LEU2	This study
pLEU-gRNA-R2	pSNR52-gRNA(R2)-tSUP4	2μ	LEU2	This study
pLEU-gRNA-H1	pSNR52-gRNA(H1)-tSUP4	2μ	LEU2	This study
pLEU-gRNA-H2	pSNR52-gRNA(H2)-tSUP4	2μ	LEU2	This study
pLEU-gRNA-P1	pSNR52-gRNA(P1)-tSUP4	2μ	LEU2	This study
pLEU-gRNA-P2	pSNR52-gRNA(P2)-tSUP4	2μ	LEU2	This study
pCut	pTEF1-hcoCas9-tCYC1	2μ	URA3	This study
pCut-R1	pTEF1-hcoCas9-tCYC1 pSNR52-gRNA(R1)-tSUP4	2μ	URA3	This study

pCut-R2	pTEF1-hcoCas9-tCYC1 pSNR52-gRNA(R2)-tSUP4	2μ	URA3	This study
pCut-H1	pTEF1-hcoCas9-tCYC1 pSNR52-gRNA(H1)-tSUP4	2μ	URA3	This study
pCut-H2	pTEF1-hcoCas9-tCYC1 pSNR52-gRNA(H2)-tSUP4	2μ	URA3	This study
pCut-P1	pTEF1-hcoCas9-tCYC1 pSNR52-gRNA(P1)-tSUP4	2μ	URA3	This study
pCut-P2	pTEF1-hcoCas9-tCYC1 pSNR52-gRNA(P2)-tSUP4	2μ	URA3	This study
pCut-H3	pTEF1-hcoCas9-tCYC1 pSNR52-gRNA(H3)-tSUP4	2μ	URA3	This study
pCut-H4	pTEF1-hcoCas9-tCYC1 pSNR52-gRNA(H4)-tSUP4	2μ	URA3	This study
pCut-H5	pTEF1-hcoCas9-tCYC1 pSNR52-gRNA(H5)-tSUP4	2μ	URA3	This study
pCut-H6	pTEF1-hcoCas9-tCYC1 pSNR52-gRNA(H6)-tSUP4	2μ	URA3	This study
pCut-H7	pTEF1-hcoCas9-tCYC1 pSNR52-gRNA(H7)-tSUP4	2μ	URA3	This study
pCut-H8	pTEF1-hcoCas9-tCYC1 pSNR52-gRNA(H8)-tSUP4	2μ	URA3	This study
pESC-URA::LacZ	pTDH3-*LacZ*-tADH1	2μ	URA3	This study
pESC-URA:: yEGFP	pTEF1-*yEGFP*-tADH1	2μ	URA3	This study
pTargetR1	R1 integration site fragment	N/A	N/A	This study
pTargetR2	R2 integration site fragment	N/A	N/A	This study
pTargetR3	R3 integration site fragment	N/A	N/A	This study
pTargetR4	R4 integration site fragment	N/A	N/A	This study
pTargetR5	R5 integration site fragment	N/A	N/A	This study
pTargetR6	R6 integration site fragment	N/A	N/A	This study
pTargetR1-LacZ	R1 integration site 5'homology pTDH3-*LacZ*-tADH1 R1 integration site 3'homology	N/A	N/A	This study
pTargetR2-LacZ	R2 integration site 5'homology pTDH3-*LacZ*-tADH1 R2 integration site 3'homology	N/A	N/A	This study
pTargetR3-LacZ	R3 integration site 5'homology pTDH3-*LacZ*-tADH1 R3 integration site 3'homology	N/A	N/A	This study
pTargetR4-LacZ	R4 integration site 5'homology pTDH3-*LacZ*-tADH1 R4 integration site 3'homology	N/A	N/A	This study
pTargetR5-LacZ	R5 integration site 5'homology pTDH3-*LacZ*-tADH1 R5 integration site 3'homology	N/A	N/A	This study
pTargetR6-LacZ	R6 integration site 5'homology pTDH3-*LacZ*-tADH1 R6 integration site 3'homology	N/A	N/A	This study

pTargetR1-yEGFP	R1 integration site 5'homology pTEF1-*yEGFP*-tADH1 R1 integration site 3'homology	N/A	N/A	This study
pTargetR2-yEGFP	R2 integration site 5'homology pTEF1-*yEGFP*-tADH1 R2 integration site 3'homology	N/A	N/A	This study
pTargetR3-yEGFP	R3 integration site 5'homology pTEF1-*yEGFP*-tADH1 R3 integration site 3'homology	N/A	N/A	This study
pTargetR4-yEGFP	R4 integration site 5'homology pTEF1-*yEGFP*-tADH1 R4 integration site 3'homology	N/A	N/A	This study
pTargetH1-GFP	H1 integration site 5'homology pTEF1-*yEGFP*-tADH1 H1 integration site 3'homology	N/A	N/A	This study
pTargetH2-GFP	H2 integration site 5'homology pTEF1-*yEGFP*-tADH1 H2 integration site 3'homology	N/A	N/A	This study
pTargetH3-GFP	H3 integration site 5'homology pTEF1-*yEGFP*-tADH1 H3 integration site 3'homology	N/A	N/A	This study
pTargetH4-GFP	H4 integration site 5'homology pTEF1-*yEGFP*-tADH1 H4 integration site 3'homology	N/A	N/A	This study
pTargetH5-GFP	H5 integration site 5'homology pTEF1-*yEGFP*-tADH1 H5 integration site 3'homology	N/A	N/A	This study
pTargetH6-GFP	H6 integration site 5'homology pTEF1-*yEGFP*-tADH1 H6 integration site 3'homology	N/A	N/A	This study
pTargetH7-GFP	H7 integration site 5'homology pTEF1-*yEGFP*-tADH1 H7 integration site 3'homology	N/A	N/A	This study
pTargetH8-GFP	H8 integration site 5'homology pTEF1-*yEGFP*-tADH1 H8 integration site 3'homology	N/A	N/A	This study
pCut-H2-mCAS	pTEF1-hcoCas9^{K3X}-tCYC1 pSNR52-gRNA(H2)-tSUP4	2μ	URA3	This study
pYM44	yEGFP-pTEF-*His3MX6*-tTEF	N/A	N/A	Janke et al. 2004
pTargetH2-His3MX6	H2 integration site 5'homology pTEF-*His3MX6*-tTEF H2 integration site 3'homology	N/A	N/A	This study
pCutSGM1	pTEF1-hcoCas9-tCYC1 pSNR52-Syn.gRNA1-tSUP4	2μ	URA3	This study
pCutSGM2	pTEF1-hcoCas9-tCYC1 pSNR52-Syn.gRNA2-tSUP4	2μ	URA3	This study
pCMC0756	pCCW12-*MjDOD*-tADH1 pTDH3-*BvCYP76AD1*^{W13L}-tTDH1 pTEF1-*ScARO4*^{K229L}-tENO2	N/A	URA3	Grewal et al. 2018

pPSG348	pTEF2-*MjcDOPA5GT*-tSSA1	N/A	LEU2	Grewal et al. 2018
pTargetH1-*BvCYP76AD1*	H1 integration site 5'homology pTDH3- *BvCYP76AD1*^{W13L} - tTDH1 H1 integration site 3' homology	N/A	N/A	This study
pTargetH2-*MjDOD*	H2 integration site 5'homology pCCW12-*MjDOD*-tADH1 H2 integration site 3' homology	N/A	N/A	This study
pTargetH5-*MjcDOPA5GT*	H5 integration site 5'homology pFBA1-*MjcDOPA5GT*-tSSA1 H5 integration site 3' homology	N/A	N/A	This study
pYM25	*yEGFP*-pAgTEF-*HygR*-tCYC1	N/A	N/A	Janke et al. 2004
pYM13	pAgTEF-*KanR*-tAgTEF	N/A	N/A	Janke et al. 2004
pTargetH7-*hphNT1*	H7 integration site 5'homology pAgTEF-*HygR*-tCYC1 H7 integration site 3' homology	N/A	N/A	This study
pTargetH4-*kanMX4*	H4 integration site 5'homology pAgTEF-*KanR*-tAgTEF H4 integration site 3' homology	N/A	N/A	This study
pESC-Leu2d::*LsGAS*/*LsGAO*/*AaCPR*/*LsCOS*	pGAL1/10-*LsGAS*-tCYC1- *LsCOS*-tADH1 pGAL1/10-*AaCPR*-tCYC1- *LsGAO*-tADH1	2μ	LEU2	Ikezawa et al. 2011
pESC-HIS::*BTS1*/*tHMGR*	pGAL1/10-*BTS1*-tCYC1- *tHMGR*-tADH1	2μ	HIS	Kwon et al. (unpublished)
pTargetH1-*LsGAS*	H1 integration site 5'homology pGAL1-*LsGAS*-tCYC1 H1 integration site 3' homology	N/A	N/A	This study
pTargetH2-*LsGAO*	H2 integration site 5'homology pGAL10-*LsGAO*-tADH1 H2 integration site 3' homology	N/A	N/A	This study
pTargetH5-*AaCPR*	H5 integration site 5'homology pGAL1-*AaCPR*-tCYC1 H5 integration site 3' homology	N/A	N/A	This study
pTargetH7-*tHMGR*	H7 integration site 5'homology pGAL10-*tHMGR*-tADH1 H7 integration site 3' homology	N/A	N/A	This study
pTargetH4-*LsCOS*	H4 integration site 5'homology pGAL10-*LsCOS*-tADH1 H4 integration site 3' homology	N/A	N/A	This study