Bioluminescent Reporter Development and Testing for Cre-recombinase and CRISPR/Cas9 Activity for Selectable Marker Recovery in *Scheffersomyces stipitis*

By

Serenah Smith

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Master of Science in Biology

Middle Tennessee State University
May 2021

Thesis Committee:
Dr. Brian Robertson, Chair
Dr. David Nelson
Dr. Rebecca Seipelt-Thiemann

ABSTRACT

Scheffersomyces stipitis is the most efficient xylose fermenting yeast known, though it is not as well-studied as other yeast species. The ability of *S. stipitis* to ferment xylose makes it a potentially useful yeast in the fermenting of lignocellulos biomass into biofuels, such as ethanol. One way to make *S. stipitis* more effective in biofuel production is to inhibit the glucose metabolic pathway, which requires the delition of multiple genes. *S. stipitis* belongs to the "CUG clade," which is a group of yeast species that translate the CUG codon as serine instead of a leucine like other eukaryotic organisms. The difference makes most genetic tools ineffective in *S. stipitis*. Genetic tools need to be optimized with this change before they can be used in *S. stipitis*. Selectable markers are one of the tools that would need to be optimized for *S. stipitis*, especially since they are used during gene deletions. To make sequential gene deletions, different selectable markers are used for each deletion. Therefore, the number of deletions is based on the number of selectable markers available. Having a way to recover selectable markers so they can be reused in other gene deletions is an effective way to increase the number of deletions that can be performed. The aim of this research is to test the ease of usage and the effectiveness of two strategies for recovering selectable markers: 1) Cre-recombinase or 2) CRISPR/Cas9.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER II

CHAPTER I
INTRODUCTION

1.1 Biofuels

With the movement towards more environmentally friendly and replenishable energy sources, people are moving away from fossil fuels, like petroleum, towards biofuels, such as ethanol. The transition to biofuels is moving slowly largely due to the economic feasibility of its production primarily due to production costs being higher than for other energy sources, such as crude oil or electricity (Ilic, 2014; Lu, 2015; Mathew, 2014). Production of biofuels takes plant biomass, which is mainly made up of cellulose, and breaks it down into lignocellulosic biomass, a mixture of sugar monomers (e.g. glucose, sucrose, and xylose) (Cunha, 2018; Mathew, 2014; Pathania, 2017). Lignocellulose is the most abundant renewable bioresource known and has attracted attention not only for its conversion into ethanol, but also for its conversion into chemicals, such as funmaric acid which is used in food and medicines (Jiménez-Quero, 2016; Li 2016). Most of the biofuel used today comes from crops grown for food production; ethanol is primarily made from sugar or starch crops, and biodiesel is made from vegetable oil, which comes from the plant seeds (Ilic, 2014; Mathew, 2014). Whereas glucose is the most prominent sugar in lignocellulosic biomass, xylose and other pentose sugars are estimated to make-up about 20-30% of the mass, most of that mass being xylose (Cunha, 2018; Wei, 2015). Efficient utilization of xylose is one of the biggest obstacles for commercial conversion of lignocellulose to useful compounds, especially since the majority of yeasts used in fermenting lignocellulose biomass do not naturally ferment pentose sugars or only ferment a small portion. The majority of yeasts ferment mainly the hexose sugars, such as

glucose (Cunha, 2018; Wei, 2015). Due to this, the percentage of biomass composed of pentose sugar is discarded instead of being utilized during the fermentation of the biomass, which increases the cost of purifying for the desired chemical. The issue of the low yield compared to cost of production for these chemicals has led to increased research and many new developments in the field of lignocellulose utilization (Pathania, 2017; Lu, 2015; Wei, 2015).

1.2 *Scheffersomyces stipitis*

One of the most efficient fermenters of pentose sugars, especially xylose, is *Scheffersomyces stipitis* (*S. stipitis*) (also known as *Pichia stipitis*) (Feng, 2013; Cao, 2018). *S. stipitis* was discovered and characterized in 1967 and is an ascomycotous yeast species (Kurtzman, 2010; Laplaza, 2006). *S. stipitis* is commonly found in the gut flora of passalaid beetles and helps in the digestion of wood materials ingested by the beetles. Possibly due to this, *S. stipitis* is one of the few yeast species naturally capable of fermenting all of the sugars found in wood (Suh, 2003). This ability, as well as the fact that it is one of the most efficient fermenters of xylose and can live in more extreme environments than most yeasts, makes *S. stipitis* a potentially valuable species in the production of ethanol as well as other chemicals derived from lignocellulose biomass (Nigam, 2001; Wei, 2015). Not only does *S. stipitis* have a high potential for usage in biofuel production, it has been proposed that *S. stipitis* is a better-suited microbial host than *Saccharmyces cerevisiae* (*S. cerevisiae*) for production of chemical products (Cao, 2018). Whereas *S. cerevisiae* is one of the most researched yeast types, *S. stipitis* is becoming more popular due to its more active pentose phosphate pathway making more

erythrose 4-phosphate, which is used in the shikimate pathway (Cao, 2018). The products made from the shikimate pathway include many kinds of flavonoids and alkaloids with important pharmaceutical and nutraceutical properties, so finding an organism that potentially can be used to synthesize large amounts of these high-value products is very important (Cao, 2018).

One of the challenges of using *S. stipitis* for fermenting lignocellulosic biomass is that, whereas *S. stipitis* can ferment multiple types of sugars, especially xylose, it has a metabolic preference for glucose. This means it will ferment any glucose available before it will start fermenting any other sugars present, such as xylose (Kim, 2012). A potential solution to the issue of the sugar preference is to inhibit the glucose metabolic pathway in *S. stipitis*. The current prediction is that this can be accomplished by deleting key genes for the glucose metabolic pathway—usually three or more genes need to be deleted (Cunha, 2018; Maitra, 1983; Slininger, 2011). In order to accomplish this, there needs to be genetic tools for *S. stipitis* (Cao, 2018). One difficulty with *S. stipitis* is that common tools used in the genetic manipulation of other yeast like *S. cerevisiae* do not usually work in *S. stipitis* (Laplaza, 2006), which is mainly due to the difference in the genetic code of *S. stipitis*.

S. stipitis belongs to a group of yeasts known as the "CUG clade," which is a group of yeasts that uses the CUG codon differently during translation than most eukaryotic organisms. In most eukaryotic species CUG codes for leucine, but in the "CUG clade" the codon codes for serine. This difference in the codon translation has been attributed to a Ser-tRNA that has a Leucine anticodon in these yeast species (Massey, 2003). It is largely due to this codon difference that genetic tools commonly used in other yeast

species do not work in *S. stipitis* (Laplaza, 2006; Priest, 2015). Almost all amino acids have multiple codons that code for them, the only two exceptions being tryptophan and methionine, which have one codon each (Xia, 2017). The usage of multiple codons for the same amino acid helps reduce the effect of point mutations since the difference between the codons that code for the same amino acid are usually only the third nucleotide, so if there was a point mutation in the third base pair it most likely would have no effect on the amino acid and, therefore, the resulting protein would be unchanged (Sharp, 1988; Xia, 2017). The different codons that code for the same amino acids are referred to as synonymous codons. These codons are often used at different frequencies depending on the species. This difference in preference of codon usage is referred to as codon bias and exists in both prokaryotic and eukaryotic organisms (Machado, 2020). There are two theories that have been proposed to explain codon bias. The theories are: the neutral theory, that the choice of codons is random; and the selection-mutation-drift balance model, which is the thought that the bias was determined through a combination of mutation pressure, genetic drift, and weak selection (Xia, 2017). Studies have shown that the codon bias of species affects the regulation of several parts of protein synthesis, such as the translation rate and protein folding. Codon optimization has become a normal part of bioengineering because of the effect it has on transcription and protein expression (Machado, 2020).

Codon optimization is a common genetic tool used due to codon biases. The expression of a gene inserted into an organism can be increased by changing any occurrence of a codon for the amino acid in a gene sequence to the organism's preferred codon. This is especially helpful when using reporters and selectable markers which are

not naturally found in the organism being studied. Due to the alternative translation phenomenon observed in the yeast species in the "CUG clade," specifically S. stipitis, tools such as reporter genes, selectable markers that are used in other organisms, or any other protein coding genes, are not translated as in other species when expressed in *S. stipitis*. One solution to this problem is to have the genes tailored by codon optimization to amino acids preferred by in *S. stipitis*, and change any CUG codons to a synonymous codon. In *S. stipitis*, 41.8% of the leucine codons are CUU which makes it a viable choice for using as a substitute leucine codon. By changing any CUG codons to CUU, synthetic genes should not only have the structure and function of the translated proteins restored but also increase the efficiency of translation when expressed in *S. stipitis*.

1.4 Bioluminescent Reporters

Reporter genes are gene sequences that are added to an organism to make it easier to identify what is occurring in a cellular level; such as genetic alterations, measuring/detecting transcription rates, protein stability, mRNA splicing, and many other processes. Reporter genes are usually chosen due to the easily identifiable and measurable characteristics expressed in organisms when they are translated. Commonly used reporter genes that induce visually identifiable phenotypes are fluorescent and bioluminescent proteins, such as Green Fluorescent Protein (GFP) and the bioluminescent luciferase enzyme, or proteins such as beta-galactosidase and Chloramphenicol acetyltransferase (CAT), which catalyze reactions that produce colored products. The use of noninvasive reporters has allowed for better understanding of organisms and gene expression in real time without destroying the biological samples

(Mezzanotte, 2011; Robertson, 2008). One of the common bioluminescent reporters used is the firefly luciferase (*Luc*) gene. The *Luc* gene can be used with an endogenous promoter and it is transcribed into a bioluminescent protein that does not require excitation from an external light source, which avoids the common issues fluorescent reporters have, such as photobleaching, phototoxicity, and high amounts of background due to autofluorescence (Robertson, 2008). Another bioluminescent reporter is click beetle green luciferase, that is encoded by the *CBG99* gene from *Pyrophorus plagiophthalamus*, a click beetle found in Jamaica. When the luciferase reacts with luciferin it emits light at 537 nm (Mezzanotte, 2011), which is a bluer light than the yellowish light emitted by firefly luciferase. The bioluminescent reaction occurs when the luciferase protein interacts with D-luciferin, Mg^{+2}, and ATP, making D-luciferyl adenylate which then releases light as it is oxidized (Figure 1) (Baldwin, 1996; Davis, 2020; Krishnamoorthy, 2015). The observation and quantification of the emitted light can be a useful tool in monitoring gene expression in yeast (Davis, 2020; Krishnamoorthy, 2015). *CBG99* as a bioluminescence reporter has been used in *Saccharomyces cerevisiae* and can be used in *S. stipitis* to understand yeast gene expression (Robertson, 2008; Reichard, 2017).

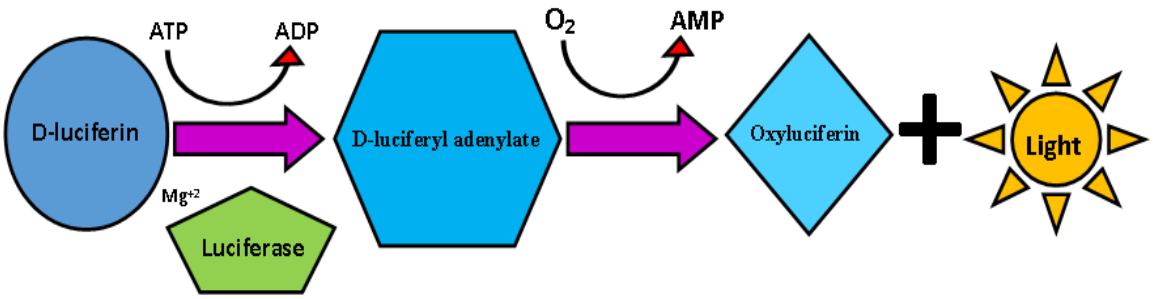

Figure 1: Reaction of luciferase when exposed to D-luciferin Mg^{+2}, and ATP. D-luciferin is converted into D-luciferyl adenylate, which emits light when it is oxidized.

CBG99 can be used to show the success of plasmid transformation by giving a visual result for successful transformations. If *CBG99*, with a constitutive promoter that is always active, is included in a plasmid that has another gene of interest, the successfully transformed cells will emit light when presented with D-luciferin.

Another way to test for success of a transformation as well as monitor gene expression is to include the *CBG99* reporter gene under the control of a native promoter. Once the transformation is done, any time the gene of interest is expressed *CBG99* will be expressed as well since they utilize the same regulatory elements. There are also ways to tell the level of expression by quantifying the amount of bioluminescence relative to cell quantity or another constitutive protein. Observing promoter activity allows for comparisons and speculations to be made about different genes of interest and the effects of different environmental conditions on its expression.

1.5 Homologous Recombination

Genetic engineering is of great interest for multiple types of scientific research. One way to delete genes in a cell is through homologous recombination, which despite being

the first precise gene-editing technique developed, is not very efficient by itself (Loureiro, 2019). Homologous recombination is a multi-step process utilizing a repair mechanism in the cell to exchange a DNA segment with some other DNA sequence (Capecchi, 1989; Cheng, 2000). The homologous recombination strategy for gene deletions involves constructing a linearized DNA sequence with three main components, which are a selectable marker flanked by an upstream and downstream portion of DNA identical to the upstream and downstream regions of the targeted gene. The linearized DNA fragment is then inserted into the cells, and the endogenous repair mechanisms for the DNA in the cell take the linearized DNA fragment and match it to the genome sequence and exchange the portion of the chromosomal sequence between the regions of homology with the selectable marker (Figure 2). The selectable marker then allows for easy identification of successful gene deletion (Tong, 2010). The problems inherent with this strategy are, first, transformations and homologous recombination are rare events, so using this method may take several attempts to actually delete the gene. Second, the sequence could be integrated off-target if random insertion occurs in an unrelated area of the genome, which means the selection may not actually result in all occurrences having the successful gene deletion. Thirdly, once the selectable marker has been used and thereby integrated into the host genome it can no longer be used to verify another gene deletion in the same organism since the organism already contains a working copy of the selectable marker. If multiple gene deletions are required, every deletion will require its own different selectable marker, which limits the number of potential gene deletions to the number of selectable markers available among the genetic tool set for that organism. This could also leads to the issue of a multi-drug resistant yeast strain as well as leaving

few available selectable markers for future genetic manipulation. A solution to the problem with the inability to use the selectable marker for other gene deletions once used is to employ a strategy to remove the integrated selectable marker once it has served its purpose. Two such strategies use CRE/loxP or CRISPR/Cas9 to eliminate the selectable markers.

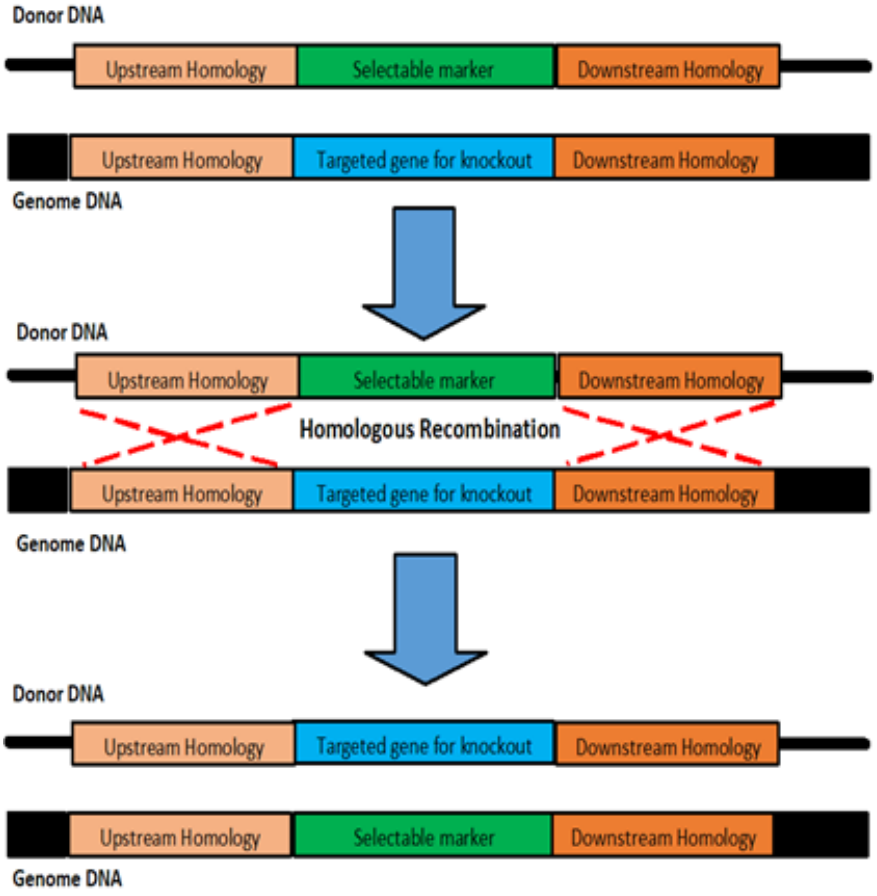

Figure 2: Homologous recombination. Linear donor DNA is introduced into the cell and matched to genome DNA. Recombination occurs and replaces the native gene with the selectable marker from the donor DNA.

1.6 Cre-recombinase

Cre-recombinase belongs to the integrase family of DNA recombinases and was originally isolated from bacteriophage P1 in *E. coli*. Cre-recombinase has been used for

enzyme-facilitated site-directed DNA recombination in several organisms, ranging from bacteria to mammals (Feng, 1999; van Duyne, 2001). Cre-recombinase targets and binds to loxP sites, which are 34 base-pairs (bp) long sequences consisting of two 13 bp palindromic inverted repeats that flank an 8 bp core sequence. The sequence of the 8 bp core determines the directionality of LoxP site (Feng, 1999; van Duyne, 2001). During Cre-recombination, one Cre-recombinase molecule binds to each palindrome half of the loxP site. The Cre molecules then bring the two loxP sites together before cleaving at the first and last bases of the loxP site core and exchanging the sequences connected to the loxP sites (van Duyne, 2001).

There are two types of recombination which can occur between loxP sites and are determined by the orientation of the loxP sites (van Duyne, 2001). The two types of recombination are inter-molecular DNA exchanges and intra-molecular excision or inversion. Inter-molecular DNA exchange occurs when the loxP sites are oriented in the same direction, which leads to the sequence between the sites being circularized and removed during recombination (Figure 3). This process can also be reversed to integrate a circularized piece of DNA into a single loxP site; though the reversed process is rarely used since the sequence would then be flanked to two loxP sites in the same direction, which would result in recombination leading to the sequence being re-circularized and removed. Intra-molecular excision, or inversion, occurs when the loxP sites are oriented in opposite directions, which leads to the sequence between the sites having a 50% chance of being inverted and changing directionality during recombination (Figure 3) (Feng, 1999).

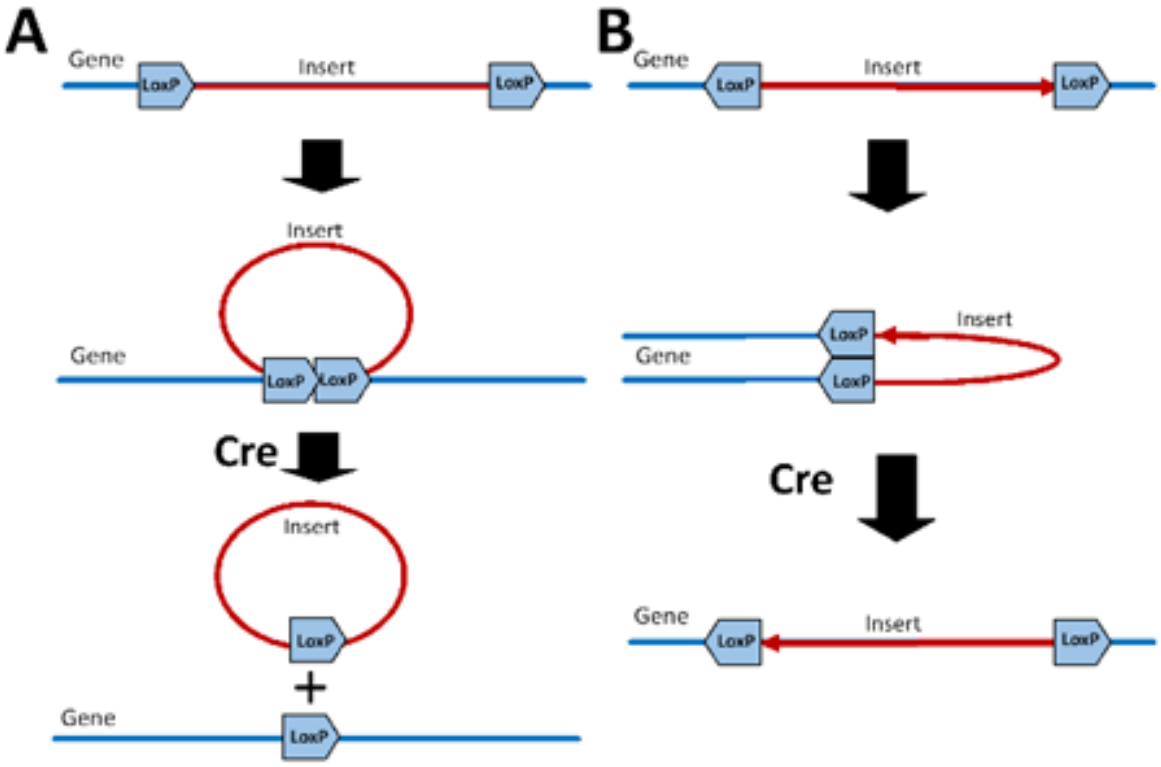

Figure 3: Recombination reactions of Cre-recombinase and paired loxP sites. **A)** Circularization and recombination of the insert due to homologous loxP sites flanking the insert. **B)** 50:50 chance of insert inversion during recombination between inverted loxP sites.

When Cre-recombinase is used to recover a selectable marker, it leaves a loxP scar that still inhibits the transcription of the gene into a functional protein. A solution for removing the loxP scar is to put the selectable marker flanked by loxP sites in an intron. There are two parts to DNA coding sequences in eukaryotes: exons and introns. Exons are parts of the DNA coding sequence that are translated into proteins. Introns are untranslated sequences embedded in the DNA coding sequence that are removed before translation through the process of RNA splicing. Intron splicing works in one of two ways depending on the organism in which it is occurring (Morgan, 2019). The splicing is

either determined by the ends of the introns or by the ends of the exons. In yeast, the two ends of the introns determine where splicing occurs (Parenteau, 2019; Morgan, 2019). The two ends of the intron are brought together by subunits of the spliceosome as it is constructed. The spliceosome then cuts the RNA sequence and removes the intron and connects the two exon pieces together. This process removes the intron leaving a sequence of just the exons from the gene (Figure 4).

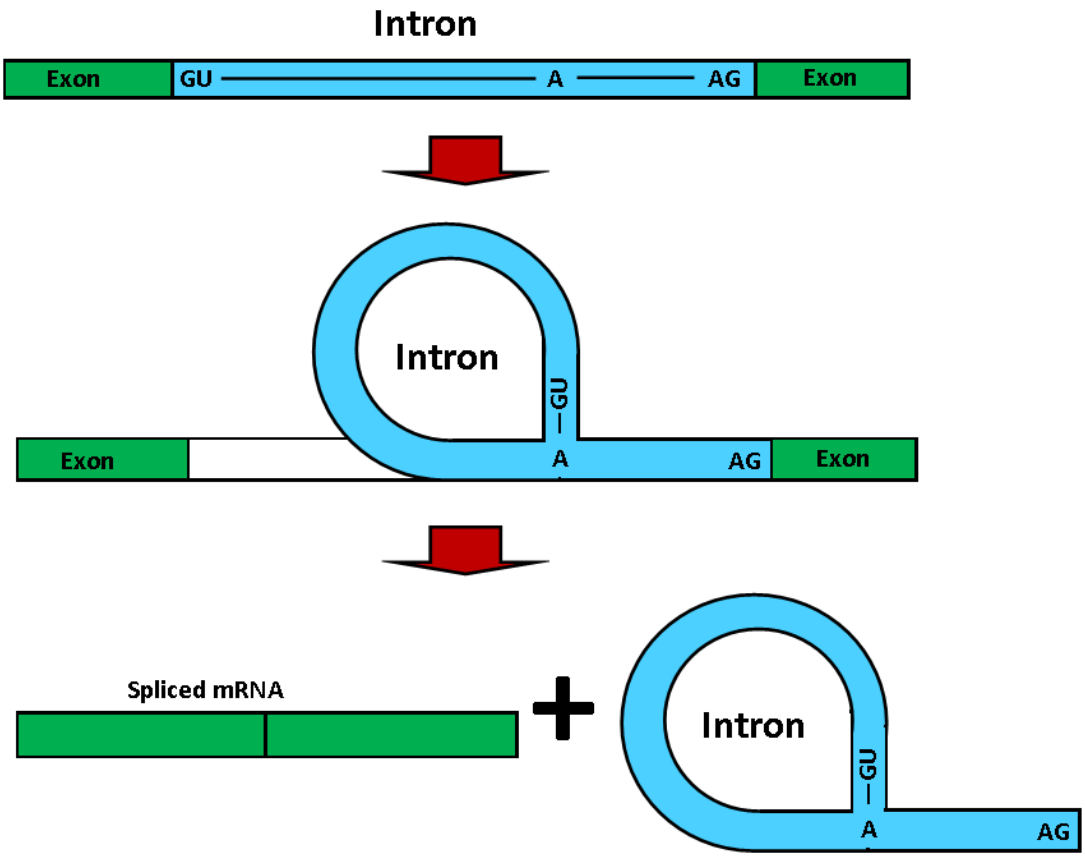

Figure 4: Intron splicing catalyzed by the spliceosome. When the spliceosome is bound, the 5" end of the intron is cut and folded into the middle, then the 3' end is cut removing the intron and the ends of the exons are joined making the complete spliced mRNA.

Cre-recombinase activity can be verified by inserting a selectable marker flanked by loxP sites into an intron and placing the intron into a reporter, such as *CBG99*. *CBG99* is an easily used reporter and, when interrupted by an insert containing the *ShBle* gene

(which codes for zeocin resistance) with a transcription terminator flanked by two loxP sites in an intron, it can be used to monitor Cre-recombinase activity. If Cre-recombinase is not active, the *CBG99* gene will not be fully transcribed due to the presence of the *ShBle* transcription terminator in the insert. When the RNA polymerase reaches the terminator sequence at the end of the *ShBle* gene, transcription will stop, resulting not only in an incomplete RNA that when translated will not make a working protein, but it will also lack the second end to the intron so splicing will not occur. When the truncated RNA is translated a truncated protein is produced that is unable to support bioluminescence (Figure 5.1). If Cre-recombinase is active, then Cre will remove the insert containing the *ShBle* gene with its transcription terminator, allowing for intron splicing to remove the remaining loxP site, which will allow transcription to proceed to the end of the *CBG99* gene. If *CBG99* is fully transcribed, then the RNA will be translated into a functional protein which catalyzes a reaction that emits light in the presence of D-luciferin (Figure 5.2).

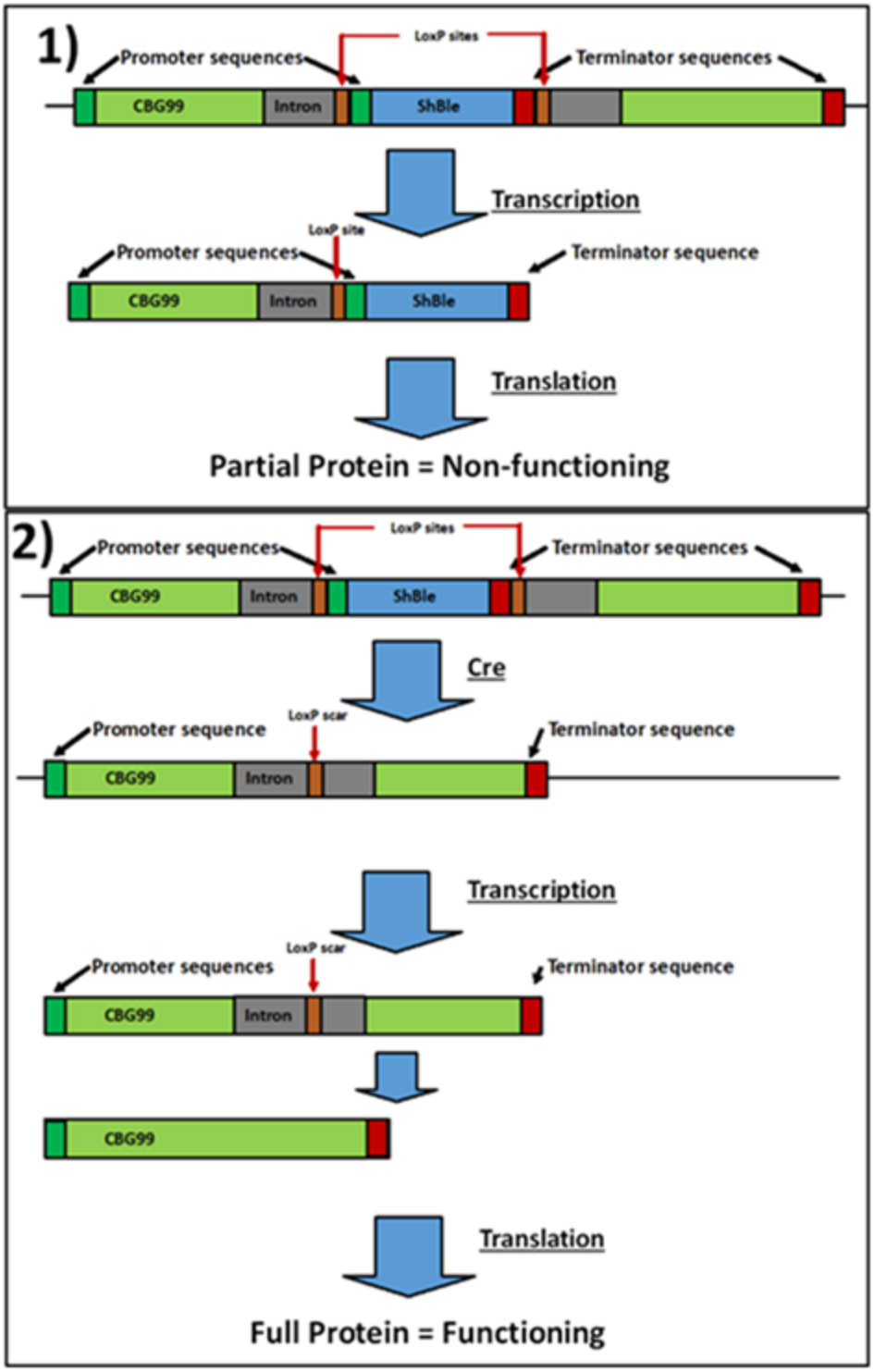

Figure 5: *CBG99* reporter with *ShBle* in intron. **1)** If Cre is not active, then transcription terminates early and the final protein made from the RNA does not function properly. **2)** If Cre is active, the *ShBle* gene with its termination sequence is removed allowing transcription of the full sequence. Intron splicing then removes the intron, resulting in a fully functioning protein after translation.

1.7 CRISPR/Cas9

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated proteins (Cas) systems function as an RNA-mediated immune system in prokaryotes and protects them against bacteriophage and foreign DNA, such as plasmids (Loureiro, 2019). CRISPR was first discovered in *Escherichia coli* and reported in 1987 (Loureiro, 2019; Vijai, 2020). It was thought to be involved in several processes in cells, such as DNA repair and chromosome rearrangement (Vijai, 2020). Currently, depending on the CRISPR motif, CRISPR has been found in 45-50% of bacterial and 87-90% of archaeal genomes (Loureiro, 2019; Vijai, 2020). CRISPR is classified into two classes (1 & 2), six types (I-VI), and 18 subtypes, which all have different signature Cas proteins that target DNA and/or RNA. Cas proteins make up the backbone of CRISPR systems, with some of them being present in most systems, such as Cas1 and Cas2 (Loureiro, 2019). The majority of the Cas protein functions are still unknown, but what is known is that they function with multiple associated proteins for CRISPR to work as an immunity against infections (Vijai, 2020).

Generally, CRISPR-Cas systems work in a three-step process: acquisition, RNA processing, and interference (Singh, 2018; Vijai, 2020). The acquisition phase is where the foreign DNA is recognized, captured, and integrated between the two repeated sequences in the CRISPR gene. These spacers, derived from the foreign DNA, are referred to as protospacers. The next step is the RNA processing phase, where the CRISPR gene with the protospacers is transcribed into a pre-CRISPR RNA which is then cleaved by an endonuclease into CRISPR RNAs. The third step, interference, is where the CRISPR RNAs are incorporated as guide RNAs within a multiprotein complex that

recognizes the sequence that was added at the original protospacer (Loureiro, 2019; Singh, 2018).

CRISPR/Cas9 is a class 2, type II, CRISPR system and is currently the most used CRISPR system for genetic editing due to its simplicity compared to other CRISPR systems. The Cas9 system involves a single protein that makes a double stranded break at a targeted site (Vijai, 2020). CRISPR/Cas9 is currently being used for genetic editing in a variety of fields, from medicine and therapies to chemicals and biofuels. The simplicity of the Cas9 system to only need the Cas9 gene sequence and a guide RNA sequence is what has made it the go-to CRISPR system for doing sequential gene knockouts in many organisms, ranging from bacteria to mammals (Vijai, 2020). The Cas9 complex binds to double stranded DNA, separates the strands, and then moves along the DNA until the single-stranded guide RNA in the Cas9 complex matches to the targeted DNA, then Cas9 cuts the separated strands of DNA (Figure 6).

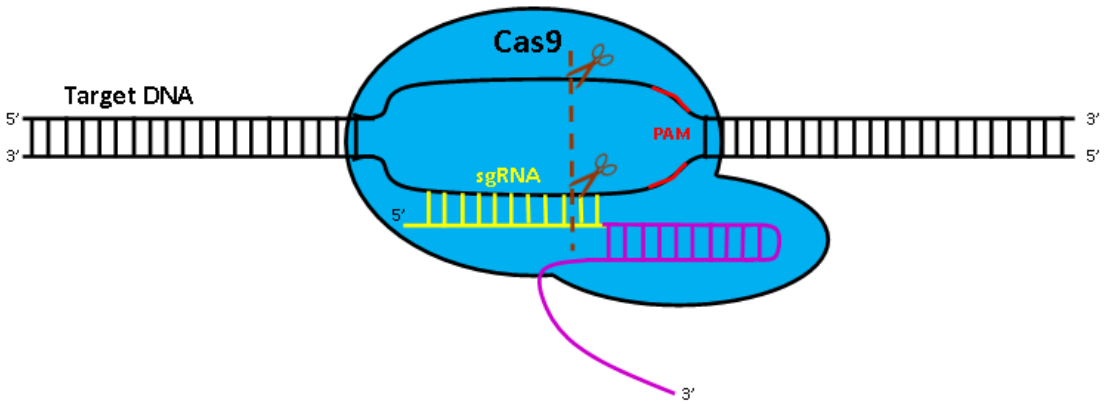

Figure 6: Cas9 binds to DNA and separates the double strands. It then moves down the DNA looking for the sequence that is complementary to the single-stranded guild RNA (sgRNA) the Cas9 is programed with. When the sequence is found, Cas9 cuts both strands of the DNA, leaving a double stranded break in the DNA.

One way to utilize CRISPR/Cas9 for selectable marker recovery during gene knockouts is to construct a plasmid that has a selectable marker next to a bioluminescent reporter flanked by a unique sgRNA sequence and protospacer-adjacent motif (Pam) sites with Cas9-affector-repeats (CAR) sequences and the target genes sequence for integration using homologous recombination. When the plasmid is transformed and integrated into the targeted gene, not only will it delete the gene, it will also be selectable for both by the antibiotic resistance as well as bioluminescence. When Cas9 is then added and expressed, the selectable marker and the bioluminescent reporter will both be removed leaving part of the sgRNA sequences, one Pam site, and the CAR regions on the two ends of the targeted gene. The CAR regions then aligned and homologous recombination occurs, leaving only one CAR region left in the targeted gene (Figure 7). Recovery of the selectable marker can then be screened for by looking for the lack of bioluminescence in the resulting colonies. The same construct can then be used again in multiple knockouts by changing the targeted integration sequence.

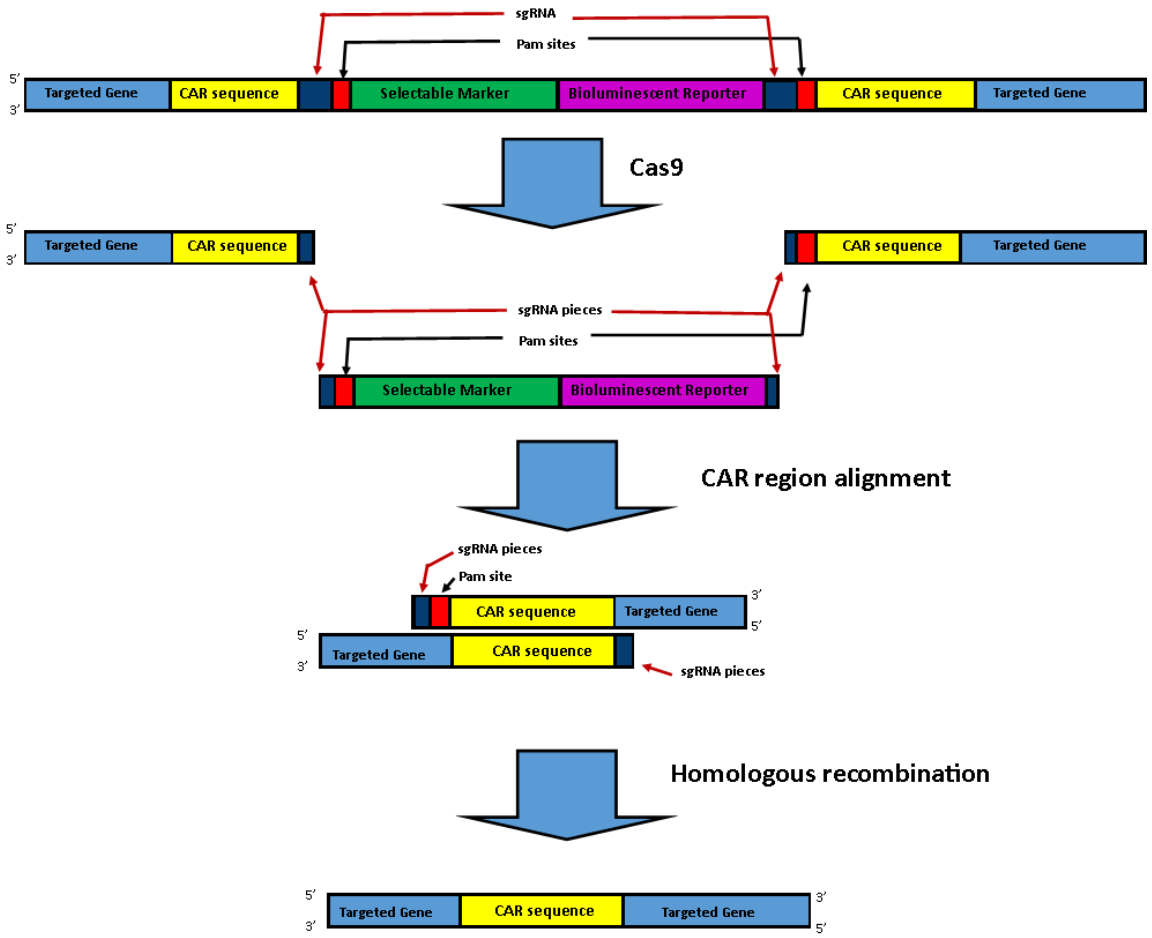

Figure 7: Cas9 reporter with a selectable marker and bioluminescent reporter. Cas9 targets and cuts the sgRNA sequences, removing both the selectable marker and bioluminescent reporter from the DNA sequence, leaving behind part of the sgRNA sequences, one Pam site, and the CAR regions in the targeted gene keeping it form successful expression.

1.8 Significance of the study

With the switch from fossil fuels to renewable fuel sources, biofuels like ethanol are becoming more popular. With this movement towards biofuels there is a need to find ways to increase the efficiency and reduce the cost of the production of biofuels. Bio-ethanol is one of the more prominent biofuels and is made through a multi-step conversion of plant biomass, one of which is the fermentation of the sugars, which is

done by yeast. Increasing the efficiency of the fermentation process will reduce the overall cost of production and make bio-ethanol more practical for large-scale usage as a renewable fuel source.

One of the major issues currently in the fermentation process is the waste of the pentose sugars, specifically xylose, during the fermentation which then have to be purified out. *Scheffersomyces stipitis* (*S. stipitis*) is a strong fermenter of xylose and other pentose sugars. The ability of *S. stipitis* to effectively ferment xylose makes it a valuable species for fermenting lignocellulosic biomass. *S. stipitis*, and the other "CUG clade" yeasts, have not been well-studied and require more genetic tools for more complex studies. One of the tools that would be beneficial is a way to recover selectable markers, which would allow for more genetic manipulation to be done.

1.9 Main Objectives and Hypothesis

1. Construct bioluminescent reporters for Cre-recombinase and CRISPR/Cas9
2. Test and compare effectiveness of selectable marker recovery strategies
3. Identify limitations of the strategies

Hypothesis: Using Cre-recombinase and/or CRISPR/Cas9 will allow for the recovery of selectable markers.

CHAPTER II
DESIGNING BIOLUMINESCENT REPORTERS TO TEST CRE-RECOMBINASE AND CRISPR/CAS9 ACTIVITY IN *S. stipitis*

2.1 Introduction

To achieve sequential gene deletions in *S. stipitis*, there needs to be a method to reuse selectable markers. This can be accomplished by recovering the markers or deleting the genes utilizing a process in which the marker is not permanently added into the genome. Two strategies that have been used in other organisms and were pursued in this study are Cre-recombinase (CRE/loxP) and CRISPR/Cas9 to recover the selectable markers. The testing of both strategies in this study utilized bioluminescent reporters to demonstrate success.

2.2 Evaluating the CRE/loxP strategy

To test Cre-recombinase activity in *S. stipitis*, two plasmids were constructed. The first, called pSES3, had the *ShBle* (zeocin resistance) gene flanked by LoxP sites inserted into the *ACTI* intron from *S. stipitis.* The intron with the selectable marker was then inserted into the *CBG99* gene in a plasmid, which contained a *URA3* element, for linearization and integration into the genome when transformed (Figure 8.A). The second plasmid, called pSES4, was designed to test the ability of *S. stipitis* to splice the *ACTI* intron. pSES4 had the *ACTI* intron inserted into *CBG99*, and it had *ShBle* inserted upstream of the *CBG99* (Figure 8.B). If the intron could be successfully spliced from the *CBG99*, then there would be bioluminescence of the yeast in the presence of luciferin.

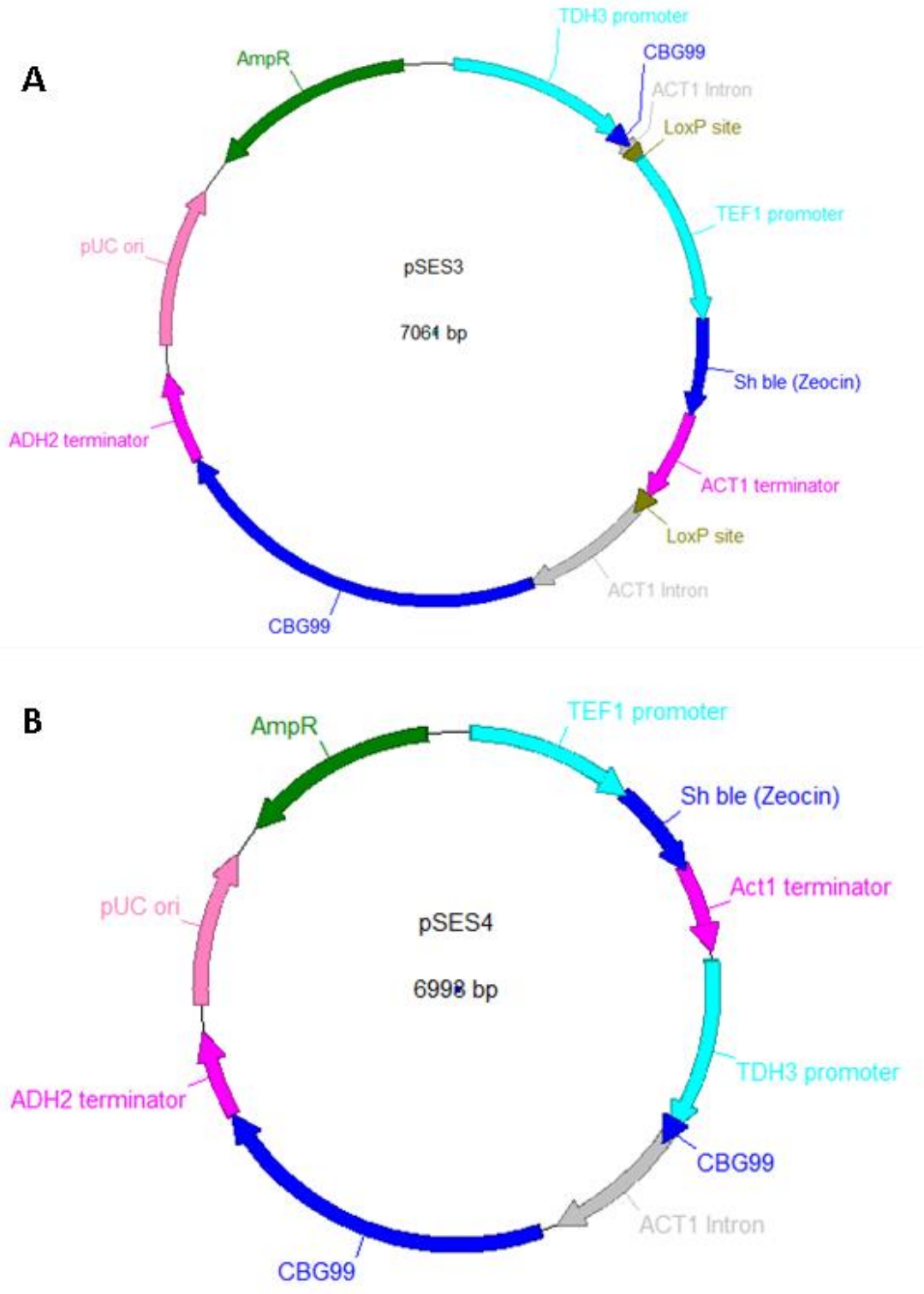

Figure 8: Plasmid layout designs. Both contain pUC origins of replication and ampicillin resistance genes. **A)** pSES3—loxP sites and ShBle selectable marker in intron interrupting CBG99 **B)** pSES4—no LoxP sites and ShBle selectable marker is upstream of CBG99 and Intron.

To construct pSES3 and pSES4, PsCBG99/puc57 was used as the starting plasmid. The first part of the construction was to move the SacI restriction site from outside the *CBG99* sequence to inside the *CBG99* sequence where it could be used as an insertion point for the addition of the intron into the *CBG99* gene. This change was made by cutting with SacI and then using Klenow to fill in and ligating the plasmid back together before using PCR (primers CBGdSac5 and CBGdSac) for inducing a point mutation to construct a SacI site in the *CBG99* sequence. The resulting plasmid was called PsCBG99/puc57-ΔSacI+SacI. The second part of the construction used GenScript to synthesize a DNA sequence that contained the following elements in this order: KpnI restriction site, the first three codons of *CBG99*, an intron from *ACT1* gene in *S. stipitis* (which contains a SalI site), then the remaining twenty-six codons of *CBG99* ending with a SacI site. The synthesized plasmid was transformed into *E. coli* then purified before being moved to the PsCBG99/puc57-ΔSacI+SacI plasmid through a double restriction digest using SacI and BamHI, making PsCBG99-ActInt-puc57. In essence, these modifications effectively added the *ACT1* intron into the *CBG99* coding sequence at the approximate location that the intron is found in the *ACT1* gene.

The next step was to make a plasmid, called pSES1, for the backbone in the constructions of pSES3 and pSES4. pSES1 was constructed using pWR8 and removing the hygromycin (HPH) selectable marker and SalI site by a restriction digest using XhoI and SalI. Once pSES1 was constructed, the *CBG99* intron construct in PsCBG99-ActInt-puc57 plasmid was moved into pSES1 replacing the *CBG99* in pSES1 using restriction digests with KpnI and AflII. The element from PsCBG99-ActInt-puc57 and the pSES1

vector without the *CBG99* element were then ligated together. The completed plasmid was called pSES2.

The P_{TEF1}-ShBle-T_{ACT1} element, for the selectable marker in the reporter, was PCR amplified from pWR10 in two reactions using different primer sets. The first set included loxP sites and XhoI restriction sites as overhangs (Primers PtefLox(Xho)5 and TactLox(Xho)3). The second set included XbaI restriction sites and had no loxP sites (Primers Ptef(Xba)5 and Tact(Xba)3). The PCR product with the loxP sites was cloned into the SalI site of the intron in SES2 and the resulting plasmid was called pSES3. The PCR product without the loxP sites was cloned into the XbaI site upstream of the *CBG99* sequence in pSES2 and the resulting plasmid was named pSES4. Both pSES3 and pSES4 had *URA3* elements added for integration into the genome. The *URA3* element was PCR amplified from pWR16 and cloned into the SacII sites of both pSES3 and pSES4.

Both reporter plasmids were integrated into the chromosomes of separate S. *stipitis* cultures using *URA3* for the integration target. The plasmids were transformed into S. *stipitis* (CBS 6054 strain) and grown on a YPD rich media plate for 24-48 hours. The resulting colonies of the pSES3 and pSES4 transformations were patched on to YPD-Luciferin plates and checked for bioluminescence. The colonies from the original pSES3 transformation were screened for successful integration of pSES3 and then transformed with another plasmid carrying CRE with the antibiotic resistance for hygromycin and plated on YPD+luciferin+hygromycin. The resulting colonies were then checked for bioluminescence.

For the integrated pSES3, when Cre-Recombinase is not active, the terminator sequence on *ShBle* should lead to an incomplete transcription of the *CBG99*, resulting in

no bioluminescence in the presence of luciferin. When Cre-recombinase is added to the *S. stipitis* strain containing the integrated pSES3, then *ShBle* is removed from the intron, leaving a LoxP scar, and a full length transcript is produced. During RNA splicing the intron should then be removed with the LoxP scar leaving a complete *CBG99* RNA for translation, which would then lead to bioluminescence of the yeast in the presence of luciferin. For the integrated pSES4, since the only interruption of *CBG99* is the *ACT1* intron, if the intron is successfully spliced, then there should always be bioluminescence in the presence of luciferin. pSES4 has the Zeocin-selectable marker (*ShBle*) outside of the *CBG99* sequence with the *ACT1* intron and was used to show that *CBG99* expression will still result in bioluminescence, if the intron is not interrupted by the *ShBle* selectable marker.

The transformation of both plasmids resulted in colonies, but showed no bioluminescence—even after several transformation attempts. Both plasmids were then sequenced and found to contain point mutations and deletions ranging from 5-12 bp in length. Both reconstruction of the plasmids and the use of PCR to repair point mutations were attempted to correct the issues, but neither resulted in correct plasmids, with issues arising in other areas of the sequences. The primary mutation location was around the loxP sites and most likely due to the palindromic nature of the loxP sites (Figure 9). This strategy was discarded due to its difficulty of use, especially since the nature of the study was to find easy to use strategies to recover selectable markers.

Figure 9: Sequence alignment between the predicted plasmid sequence and the constructed sequence at the loxP sites in SES3 from the first construction. Pink is the end of the *ACT1* terminator on ShBle, green is the loxP site, lowercase red is the 12 base pair deletion, and blue is the *ACT1* intron.

Table 1: Plasmids used for the Cre-recombinase strategy in this study. Plasmid nomenclature is indicated as "Backbone (promotor gene terminator) and/or (integration element)".

Plasmid	**Description**	**Plasmid Origin**
pUC57-mini	Standard cloning vector from GenScript	Reichard, 2017
PsCBG99/puc57	pUC57 containing CBG99	
PsCBG99/puc57-ΔSacI+SacI	pUC57 containing CBG99 with the SacI restriction site moved into the CBG99 sequence.	
PsCBG99-ActInt-puc57	pUC57 containing CBG99 with ACT1 intron.	
pAllet	Synthesized multiple cloning sites inside pUC57-mini	Reichard, 2017
pWR8	pAllet(P_{TEF1}-coHPH-T_{ACT1})(P_{TDH1}-coCBG99-T_{ADH1})	
pWR16	pAllet(P_{TEF1}-coHPH-T_{ACT1})(P_{TDH1}-coCBG99-T_{ADH1}) (URA3)	
pSES1	pAllet (P_{TDH3}-coCBG99-T_{ADH2})	
pSES2	pAllet (P_{TDH3}-coCBG99-ActInt-T_{ADH2})	
pSES3	pAllet (P_{TDH3}-coCBG99-ActInt (P_{TEF1}-ShBle-T_{ACT})-T_{ADH2}) (URA3)	
pSES4	pAllet (P_{TEF1}-ShBle-T_{ACT}) (P_{TDH3}-coCBG99-ActInt-T_{ADH2}) (URA3)	

Table 2: Primers used for the Cre-recombinase strategy in this study. Numerals at the end of the primer names indicate either 5´ or 3´ direction. Temp is the annealing temperature for PCR in degrees Celsius. Uppercase and underlined sequences indicate the binding sites. Lowercase letters indicate random added base pairs for length off of restriction sites for digestion, unless otherwise indicated in the descriptions.

Primer Name	**Sequence**	**Temp (°C)**	**Description**
CBGdSac5	AAATGTTGTTTCGaGCTCTCCGCAAGC	72	Lowercase red letter is the base pair being changed.
CBGdSac3	GCTTGCGGAGAGCtCGAAACAACATTT	72	Lowercase red letter is the base pair being changed.
PtefLox(Xho)5	actactCTCGAGATAACTTCGTATAATGTATGCTATACGAAGTTATCCGTACACTTATTGTTAACTATGAA	61	loxP sites and XhoI sites
TactLox(Xho)3	agtagtCTCGAGATAACTTCGTATAGCATACATTATACGAAGTTATGAACTTGATTGTAATACAAAATGC	61	loxP sites and XhoI sites
Ptef(XbaI)5	actactTCTAGACCGTACACTTATTGTTAACTATGAA	61	XbaI sites
Tact(XbaI)3	actactTCTAGAGAACTTGATTGTAATACAAAATGC	61	XbaI sites

2.3 Evaluating the CRISPR/Cas9 strategy

To test the CRISPR/Cas9 strategy for selectable marker recovery in *S. stipitis*, two plasmids were constructed. The first plasmid was a reporter plasmid carrying a constitutively expressed *CBG99* gene interrupted by *ShBle* (zeocin resistance), pAllet-ShBle-coCBG99. The purpose of this plasmid was to provide a way to test whether Cas9 was being produced in *S. stipitis* and could be properly targeted by a guide RNA. A portion of the *CBG99* coding sequence (CDS) was duplicated on both sides of a *ShBle* interruption element, so that if Cas9 cut the CDS in either (or both) of these duplicated regions, then homologous recombination would repair the DNA to remove the *ShBle*

interruption element and restore the *CBG99* CDS (Figure 10). The second plasmid provided the functional components for the CRISPR/Cas9 system and included a codon optimized Cas9 gene, a hygromycin resistance gene, and an element that coded for the single-stranded guide RNA (sgRNA). This plasmid was called pWR63-JBRsgRNA. *S. stipitis* was transformed with the first plasmid and transformants grown on plates containing zeocin, creating a stable reporter strain. The resulting reporter strain would then be transformed with pWR63-JBRsgRNA and, if Cas9 was active, then it would remove the selectable marker (*ShBle*) from the middle of the *CBG99* and the resulting colonies would bioluminesce in the presence of luciferin.

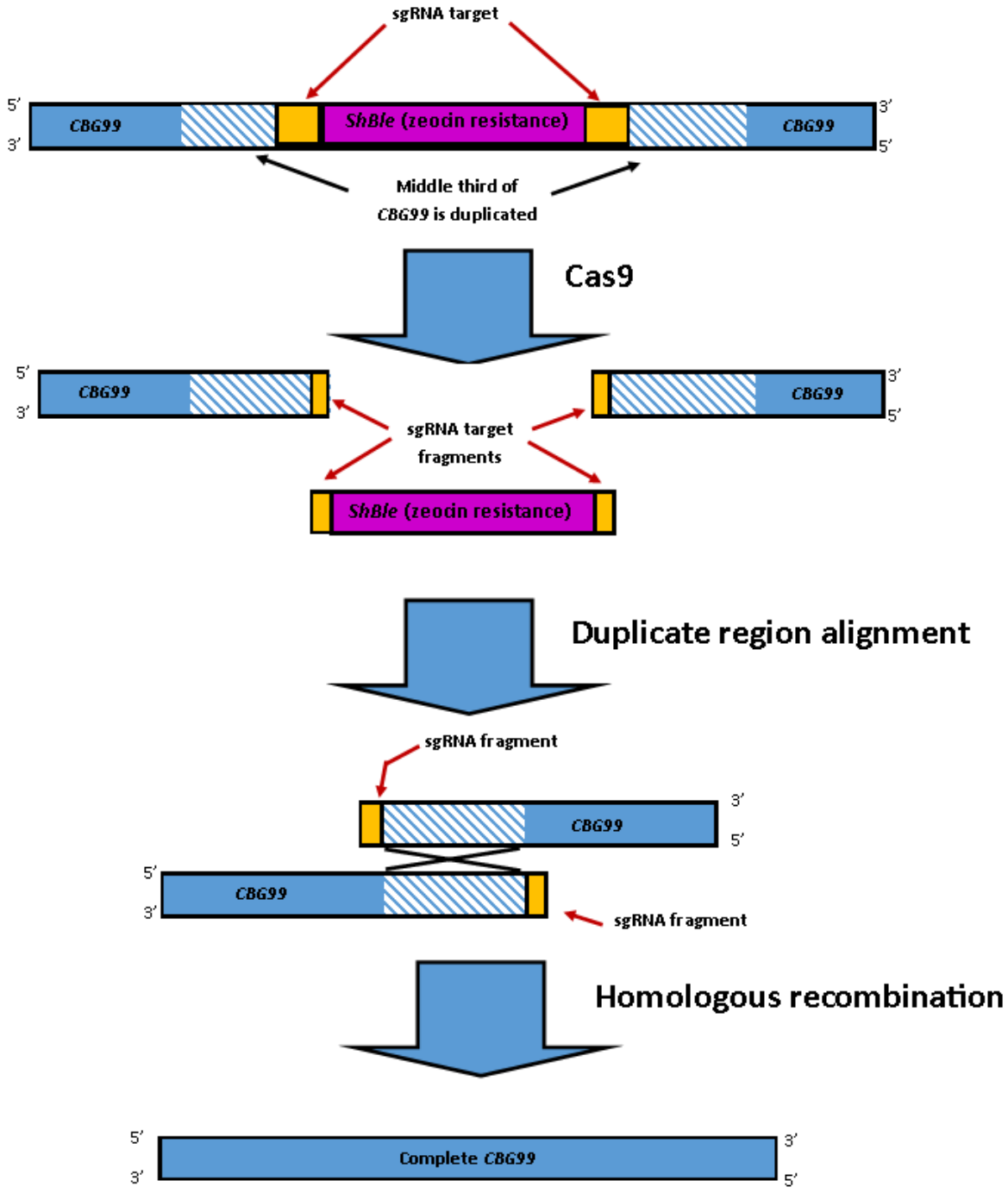

Figure 10: CRISPR/Cas9 activity test strategy. Cas9 targets and cuts at the sgRNA target sequences. *ShBle* is then removed from the sequence and the duplicated middle third of the *CBG99* align. Homologous recombination then occurs between the duplicated middle third resulting in a complete uninterrupted *CBG99*.

Construction for the reporter for CRISPR activity, pAllet-ShBle-coCBG99, used pAllet for the backbone. The first part of the construction was to move the P_{TEF1}-ShBle-T_{ACT1} section from pWR10 with XhoI and XbaI and move it into pAllet to form pAllet-ShBle. The second part of the construction was done by PCR amplifying the *TDH3* promoter and the co*CBG99* upstream two-thirds from pWR9 (primers Ptdh3(Age)5 and CBGint(Xho)3) and then cloning it into pAllet-ShBle with AgeI and XhoI. The third step was to PCR amplify the downstream two-thirds of co*CBG99* and the *ADH2* terminator from pWR9 (primers coCBGint(Sal)5 and Tadh2(Xba)3) and then clone it into the previous plasmid using SalI and XbaI. The final step was to add the *URA3* integration element into the SacI site. The *URA3* element was cut from pWR16 then ligated into the SacI site in the final pAllet-ShBle-coCBG99 plasmid (Figure 11).

The construction of WR63-JBRsgRNA was performed utilizing multiple previously constructed plasmids. First, the CaCas9 element (a codon optimized Cas9 gene from *Candida albicans*, another CUG clade yeast) was PCR amplified from pV1093 using primers that added Kpn and AflII to the ends for cloning (primers CaCas9(Kpn)5 and CaCas9(Afl)3). The PCR product was then cloned into pWR54 (replacing the *CBG99* element in it) and the resulting plasmid was named pWR57. The second step was to cut out the T_{ADH2}-P_{GAL1}-CRE-T_{ADH1} element from pWR33 with AflII and SacII and clone it into pAllet to make pWR58. The third step was to PCR amplify the P_{TDH3} element from pWR33 using primers that added SbfI and FseI site on the end (primers Ptdh3(Sbf)5 and Ptdh3(Fse)3) for cloning into pWR58 (replacing P_{GAL}) to make pWR59. The fourth step was to PCR amplify the *CBG99* element form pWR9 with primers that added FseI and MluI (primers PsCBG(Fse)5 and PsCBG(Mlu)3) to replace the CRE of pWR59 with the

CBG99 to make pWR60. The fifth step was to cut out the P_{TDH3}-coCBG99-T_{ADH1} element from pWR60 with SbfI and SacII and clone it into pWR57 to make pWR61.

The sixth step was to PCR amplify the sgRNA element from the sgRNA-Sap pUC57 plasmid adding Blunt and Mlu sites to the ends (primers sgRNA5 and sgRNA(Mlu)3) and then cloning the piece into WR59 to make WR62. The plasmid sgRNA-Sap pUC57 was built by Genscript and included essential elements for expressing sgRNA under control of standard class II promoters, like P_{TDH3} (design was adapted from Ng and Dean, 2017). The sgRNA-Sap pUC57 plasmid contained four parts: *S. stipitis* tRNA sequence for Alanine; a SapI-SapI tandem cassette for introducing desired sgRNAs; the ctRNA sequence that provides the necessary tandem sequence for the sgRNAs to fit into the Cas9; and the self-splicing viral RNA sequence of HDV virus. The next step was to cut the piece from WR62 with SbfI and SacII and clone it into WR57 to make WR63. The finished WR63 was a working CRISPR/Cas9 plasmid that contained everything but the user's desired sgRNA sequence.

The sgRNA sequence selected for the test for CRISPR activity was a unique sequence that was created (and included in the Cas9 reporter plasmid) and did not naturally occur in *S. stipitis*, so the likelihood of off target cuts was minimized. A pair of 23bp oligos that would dimerize were constructed and clone into the SapI site of pWR63 to make pWR63-JBRsgRNA (Figure 11).

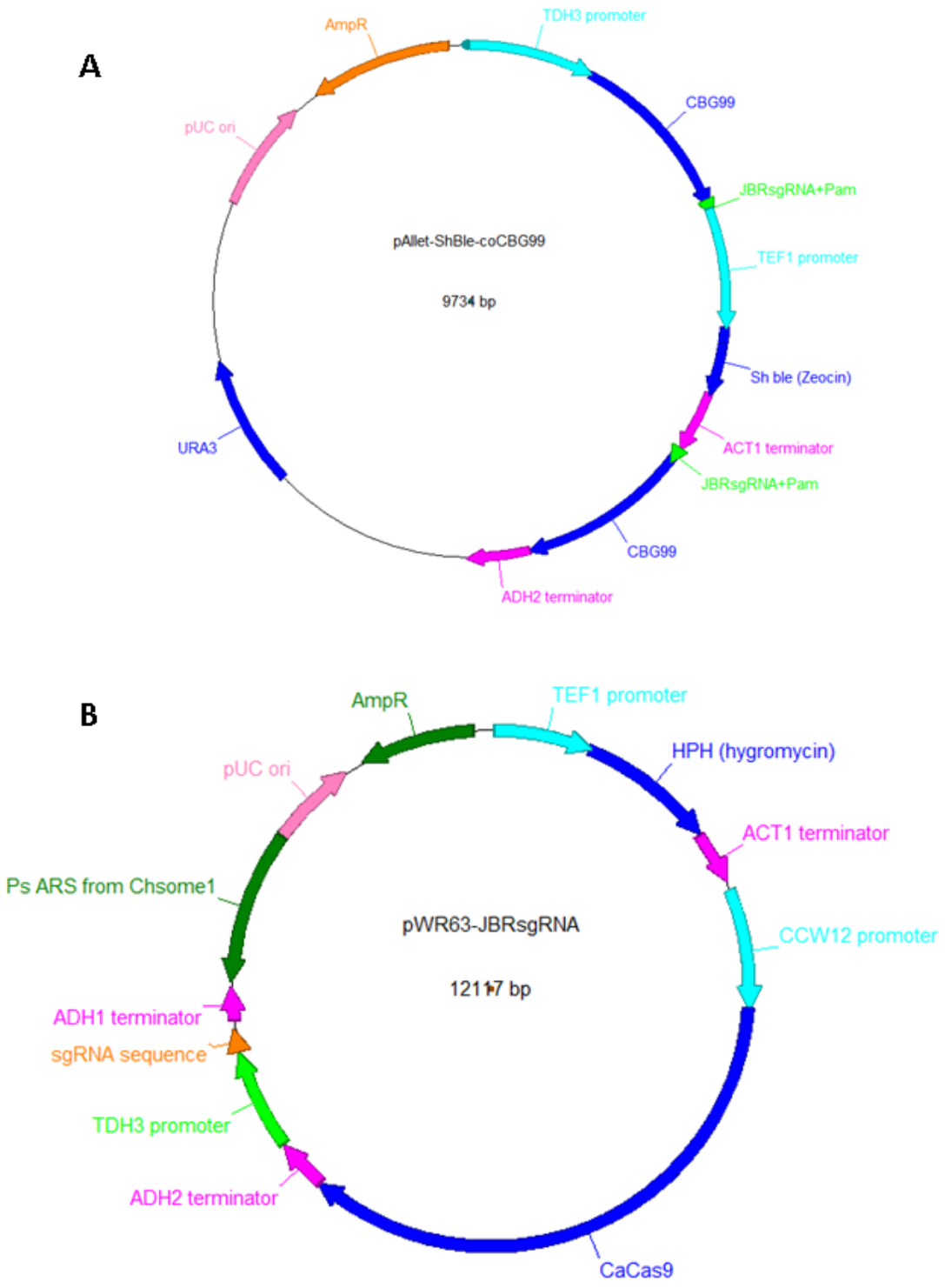

Figure 11: Plasmid layout designs. Both plasmids include a pUC origin or replication and ampicillin resistance gene. **A)** pAllet-ShBle-coCBG99— JBRsgRNA+Pam sites flanking the *ShBle* selectable marker in the middle of *CBG99*, with a *URA3* integration element. **B)** pWR63-JBRsgRNA—Cas9 and the sgRNA sequence, with Hygromycin resistance and ARS from chromosome 1 for replication in *S. stipitis*.

To test the activity of Cas9 in *S. stipitis*, pAllet-ShBle-coCBG99 was transformed into *S. stipitis* and then plated on YPD+Zeocin. Resulting colonies were then patched onto a new YPD plate and incubated at 30°C overnight. Samples from the resulting patches were then screened for successful integration of the reporter by culturing overnight then collecting the DNA and PCR screening for the presence of the reporter. One of the colonies with the integrated plasmid was then cultured again from the patch and transformed with WR63-JBRsgRNA and plated on YPD+luciferin+hygromycin and incubated at 30°C 24-48hrs. The plate was then observed for bioluminescence.

The experiment was conducted twelve times with the resulting numbers of colonies ranging from roughly 50-300. Despite the differences of the number of colonies at the end, the resulting patches showed the same results of roughly 4% non-bioluminescent colonies, 90% of the colonies having a low level of bioluminescence, and 6% with either high bioluminescence or a variance of bioluminescence level through the colony with some areas having high and others low levels. This experiment was also done with plating onto YPD+hygromycin and then patching random colonies onto YPD+luciferin, and then observing for bioluminescence. As with the individual colonies and in the same rough percentages, some colonies had more bioluminescence than others, and some of the patches had variation in bioluminescence where part of the patch displayed bioluminescence and the remainder did not. These variations might be due to a delay in the activity of the CRISPR/Cas9. The experiment was done with controls for both the lack of sgRNA and Cas9, where the second plasmid transformed into the *S. stipitis* strain with the integrated reporter either did not include the sgRNA sequence or did not include the *Cas9* gene. The results showed that both Cas9 and the sgRNA were necessary for the

removal of the *ShBle* gene from the middle of the *CBG99* sequence, and if either was missing, *ShBle* would not be removed and there would not be bioluminescence (Figure 12).

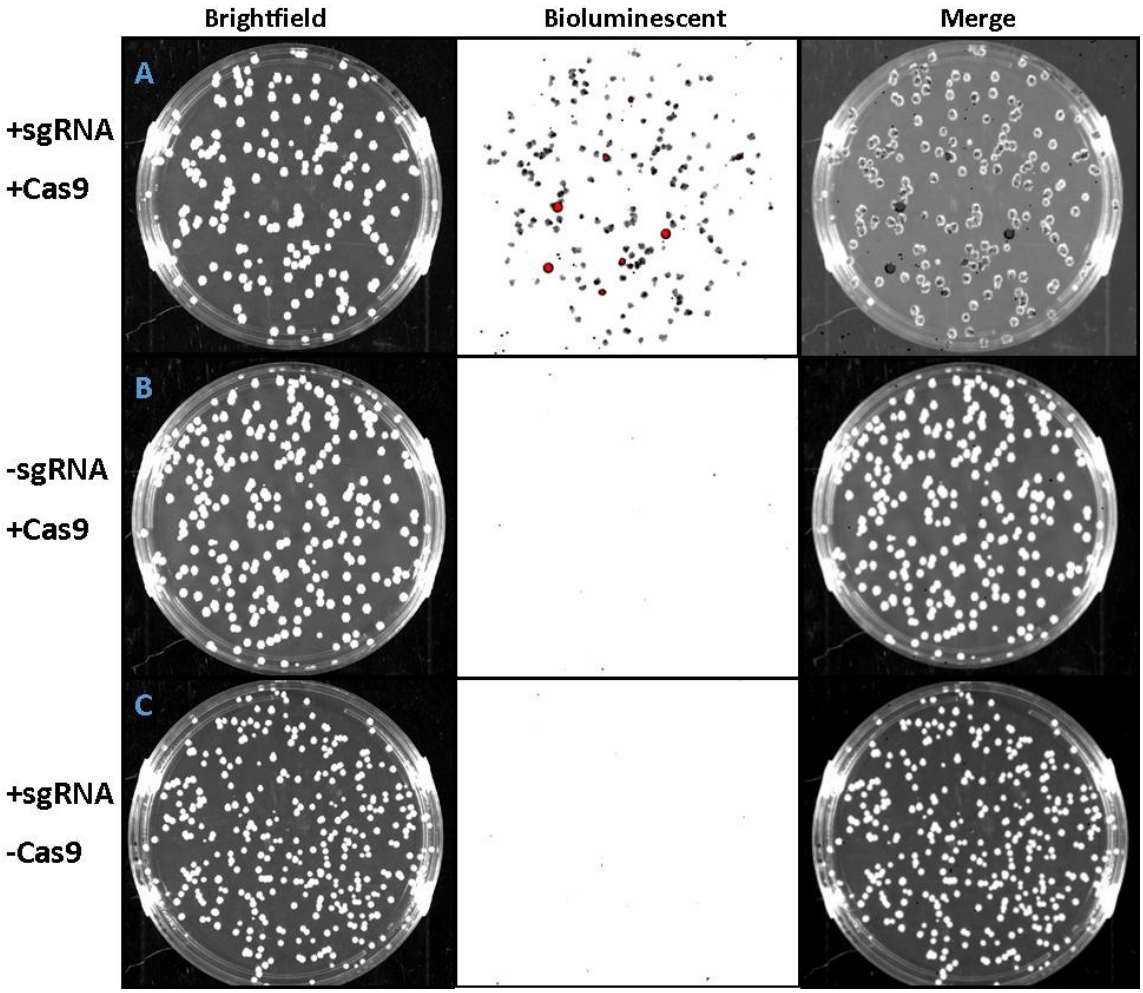

Figure 12: Comparison between the transformation of WR63-JBRsgRNA and the transformations of control plasmids without either Cas9 or the sgRNA. For each of the transformations; brightfield plates on the left, bioluminescent plates in the middle, with the merge on the right. **A)** WR63-JBRsgRNA transformation. **B)** Transformation without the sgRNA. **C)** Transformation without Cas9.

Table 3: Plasmids used for the CRISPR/Cas9 activity strategy in this study. Plasmid nomenclature is indicated as "Backbone (promotor gene terminator) and/or (integration element)".

Plasmid	**Description**	**Plasmid Origin**
pUC57-mini	Standard cloning vector from GenScript	Reichard, 2017
sgRNA-Sap pUC57	Synthesized by GenScript	GenScript
pV1093	Cas9, gifted from Gerald Fink	Vyas, 2015
pAllet	Synthesized multiple cloning sites inside pUC57-mini	Reichard, 2017
pWR9	pAllet (P_{TEF1}-co*HPH -*T_{ACT1}) (P_{TDH1}-co*CBG99-*T_{ADH1}) (*ARS2*)	
pWR10	pAllet (P_{TEF1}-*ShBle*-T_{ACT}) (P_{TDH1}-co*CBG99-*T_{ADH1}) (*ARS2*)	
pWR16	pAllet (P_{TEF1}-co*HPH-*T_{ACT1}) (P_{TDH1}-co*CBG99-*T_{ADH1}) (*URA*)	Reichard, 2017
pWR33	pAllet (P_{TEF1}-Hph-T_{ACT}) (P_{TDH3}-co*CBG99-*T_{ADH1}) (P_{Gal1}-CRE-T_{ADH1}) (*ARS2*)	
pWR54	pAllet (P_{TEF1}-co*HPH-*T_{ACT1}) (P_{CCW12}-co*CBG99-*T_{ADH1}) (*ARS2*)	
pWR57	pAllet (P_{TEF1}-co*HPH-*T_{ACT1}) (P_{CCW12}-*Cas9-*T_{ADH1}) (*ARS2*)	
pWR58	pAllet (T_{ADH1}) (P_{Gal1}-CRE-T_{ADH1})	
pWR59	pAllet (T_{ADH1}) (P_{TDH3}-CRE-T_{ADH1})	
pWR60	pAllet (T_{ADH1}) (P_{TDH3}- co*CBG99*-T_{ADH1})	
pWR61	pAllet (P_{TEF1}-co*HPH-*T_{ACT1}) (P_{CCW12}-*Cas9-*T_{ADH1}) (P_{TDH3}- co*CBG99*-T_{ADH1}) (*ARS2*)	
pWR62	pAllet (T_{ADH1}) (P_{TEF1}-co*HPH-*T_{ACT1}) (P_{TDH3}-*Cas9*-T_{ADH1}) (SapI-SapI tandem cassette)	
pWR63	pAllet (T_{ADH1}) (P_{TEF1}-co*HPH-*T_{ACT1}) (P_{TDH3}-*Cas9*-T_{ADH1}) (sgRNA element)	
pAllet-ShBle	pAllet (P_{TEF1}-ShBle-T_{ACT})	
pAllet-ShBle-coCBG99	pAllet (P_{TDH3}-coCBG99- (P_{TEF1}-ShBle-T_{ACT})-T_{ADH2}) (URA3)	
pWR63-JBRsgRNA	pAllet (T_{ADH1}) (P_{TEF1}-co*HPH-*T_{ACT1}) (P_{TDH3}-*Cas9*-T_{ADH1}) (JBRsgRNA)	

Table 4: Primers used for the CRISPR/Cas9 activity strategy in this study. Numerals at the end of the primer names indicate either 5´ or 3´ direction. Temp is the annealing temperature for PCR in degrees Celsius. Uppercase and underlined sequences indicate the binding sites. Lowercase letters indicate random added base pairs for length off of restriction sites for digestion, unless otherwise indicated in the descriptions.

Primer Name	**Sequence**	**Temp (°C)**	**Description**
Ptdh3(Age)5	tcattaACCGGTCTCGTTAGTATTTCCGTGAAG	59	AgeI sites
CBGint(Xho)3	actatcCTCGAGCCCCGGGGACTAAGTAGAACTTTCAAAGAGCCGCTCTTAAACTC	63	JBRsgRNA+Pam and XboI sites
coCBGint(Sal)5	actacctGTCGACAAAGTTCTACTTAGTCCCCGGGGTTGCAGCCATTTTGTGTAGC	63	JBRsgRNA+Pam and SalI sites
Tadh2(Xba)3	tgccatTCTAGATTCTGCCTTCTGAACGTTT	62	XbaI sites
CaCas9(Kpn)5	actacaGGTACCATGGATAAAAAGTATAGTATTGG	56	KpnI sites
CaCas9(Afl)3	actactCTTAAGTTATCACTTGTCATCGTCATC	56	AflI sites
Ptdh3(Sbf)5	tcattaCCTGCAGGCTCGTTAGTATTTCCGTGAAG	59	SbfI sites
Ptdh3(Fse)3	atgactGGCCGGCCGATGAATTGTTTATAGGGAAGA	58	FseI sites
PsCBG(Fse)5	agggagGGCCGGCCATGGTGAAGCGTGAGAAAAATG	66	FseI sites
PsCBG(Mlu)3	actactACGCGTAATCAATCTGAATAAATAGGCACTTG	66	MluI sites
sgRNA5	AAACGGGCGTGTGGCG	67	
sgRNA(Mlu)3	actactACGCGTGTCCCATTCGCCATGC	64	MluI sites
JBR(gRNA)sap5	CCAAAAGTTCTACTTAGTCCCCG		sgRNA oligos, SapI overhangs
JBR(gRNA)sap3	AACCGGGGACTAAGTAGAACTTT		sgRNA oligos, SapI overhangs

2.4 Recovery of selectable markers

To test the ability of CRISPR to recover selectable markers for sequential gene knockouts in *S. stipitis*, a plasmid, named pWR88, was constructed with the G418 resistance gene and *CBG99* together between two Cas9-Affected-Repeats (CAR regions) (Figure 13). The CAR regions served similar roles as the duplicated portion of the *CBG99* gene in the previously mentioned Cas9 reporter plasmid; they provided locations where homologous recombination could occur. The plasmid was then used to create two modified versions, pWR89 and pWR90, each with a different integration target in the *S. stipitis* genome. The integration sequences chosen were the *EGC3* gene (pWR89) and the *GSC2* gene (pWR90). Both integration genes were chosen because they were suspected to be non-essential genes and would not cause noticeable phenotype changes when removed.

The construction of WR88 started with having Genescript synthesize a codon optimized KanMx coding sequence from pYM-N1 (Janke, 2004) for *S. stipitis* with BsrGI and NotI sites on the ends. The KanMX-PsCO was received in pUC57. The second step was to cut the coKanMX cds out and clone it into pWR1, replacing the coHPH gene. The P_{TEF1}-coKanMX element was then cut out with XhoI and NotI and cloned into pWR9, replacing the P_{TEF1}-HPH element making pWR23 which is selectable on the antibiotic “G-418.” The two-step process of moving the segment first to pWR1 before pWR9 was required to avoid multiple problematic BsrGI sites in pWR9. The third step was to make alterations to pWR23 by removing problematic restriction sites. The first sites removed were BamHI, XbaI, and SalI which was accomplished by cutting with BamHI and SalI then blunting and religating. The AflII site was then removed by the

same process to create pWR23Δ. The fourth step to making pWR88 was to PCR amplify the P_{TEF1}-coKanMx-T_{ACT1}-P_{TDH3}-CBG99-T_{ADH2} element from pWR23Δ and clone it into pAllet using BamHI and SpeI. The next step was to PCR amplify the upstream CAR element with primers that added XbaI and BamHI sites and the 3' primer including the JBR-sgRNA-PAM sequence and add it to the previous plasmid. The actual sequence for the CAR element was a piece of LacZ from pUC8-LacZ (from Altman lab), which was used as the template for the PCR amplification. The sixth step was to PCR amplify the downstream CAR element with primers that added SpeI and AflII sites with the 5' including the JBR-sgRNA-PAM sequence, using pUC8-LacZ as the template. This completed the construction of pWR88. The next step was to remove the EcoRV site from pWR88, because it would be problematic after the integration elements were added, since the plasmids would be linearized with EcoRV for transformation. The EcoRV site was removed by cutting pWR88 with SalI and XhoI, which flanked the EcoRV and had compatible sticky ends. The plasmid was then ligated, making pWR88Δ.

Both pWR89 and pWR90 were made by PCR amplifying their integration elements from *S. stipitis* genomic DNA and then cloning them into the SacI site of pWR88Δ. pWR89 used *EGC3* for the integration element, and pWR90 used *GSC2* as the integration element (Figure 13).

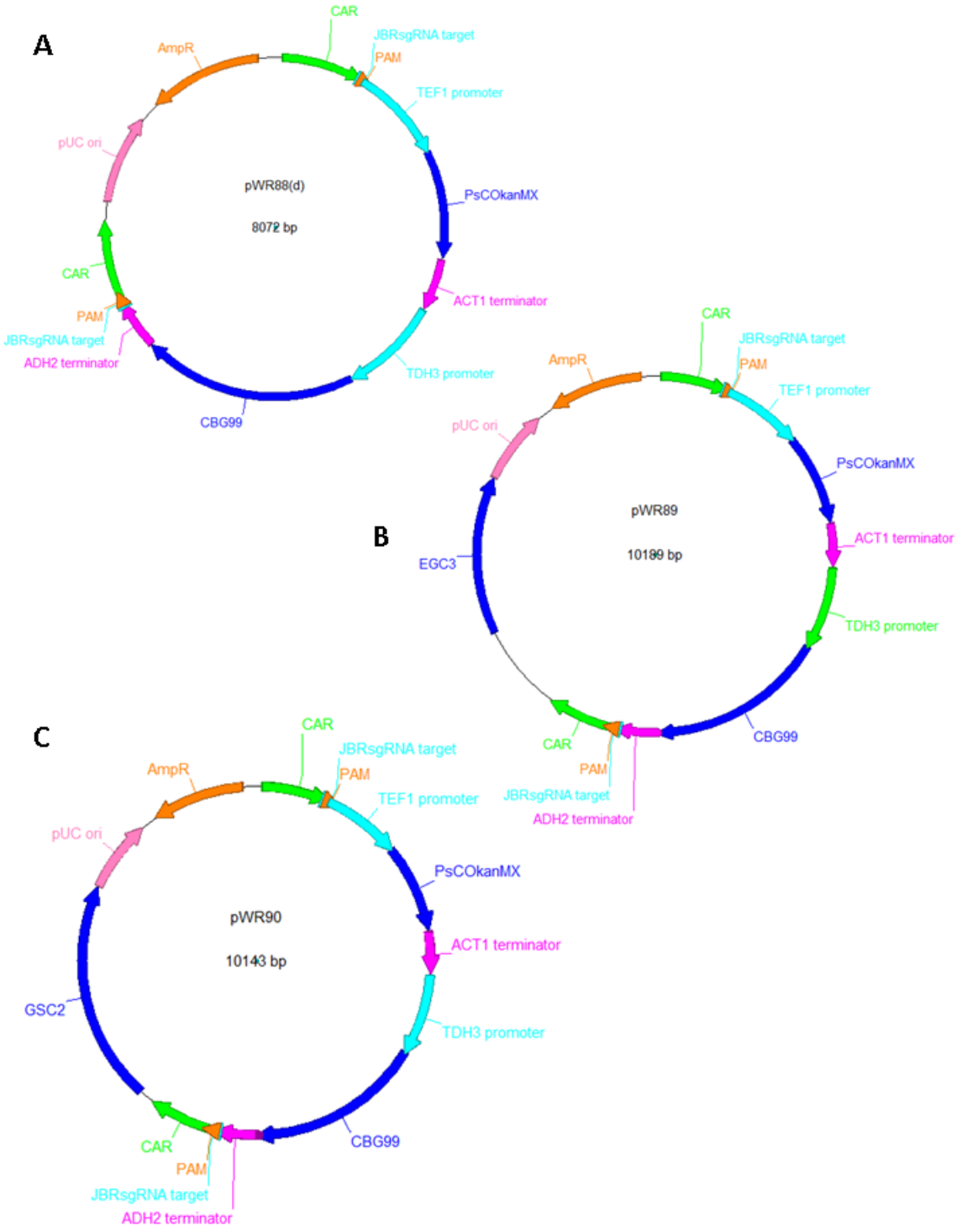

Figure 13: Plasmid layout designs. All plasmids include a pUC origin of replication and ampicillin resistance. **A)** WR88Δ —KanMx with CBG99 and the srRNA+PAM sequences. **B)** WR89— KanMx, CBG99, and srRNA+PAM sequences with EGC3 integration element. **C)** WR90— KanMx, CBG99, and srRNA+PAM sequences with GSC2 integration element.

To test the ability of CRISPR to recover selectable markers in *S. stipitis*, pWR89 was transformed into *S. stipitis* and plated on YPD+luciferin+G418. The plates were observed for bioluminescence (Figure 14.A, B), then one or more of the colonies with bioluminescence were patched onto YPD+luciferin to test the stable integration of the plasmid (Figure 14.C, D). If the plasmid was stably integrated, the colonies would retain their bioluminescence without selection pressure, and all colonies would retain the same level of bioluminescence. If the plasmid was not stably integrated, the plasmid would be lost over time without selective pressure leading to the loss of bioluminescence as the colonies grew, which would result in a variation of bioluminescence levels between colonies and possibly within colonies. Once stability was confirmed, a colony was chosen for further work. The colony was cultured and transformed with WR63-JBRsgRNA and plated on YPD+luciferin+Hygromycin, and observed for bioluminescence (Figure 14.E, F, G). The resulting colonies that did not express bioluminescence were presumed to have had the DNA between the CAR regions removed by Cas9 cutting followed by homologous recombination between the CAR regions. The control for this was to transform the colony with plasmids that were missing either the sgRNA or the Cas9, which would result in all colonies having bioluminescence. One or more of the non-bioluminescent colonies were tested for *G418* resistance/sensitivity by resuspending the candidate colonies in liquid medium, counting the cells by hemocytometer, diluting the suspension to a point where 50 μl of dilution contained 100–200 cells/ml, and then plating 50 μl onto 2 separate plates: YPD+G418 and YPD (Figure 14.H). This method was used to produce single cells spread out so that cell number would not dilute the effect of the antibiotic due to close cells using up the antibiotic, and would make it less likely to

get cells surviving on the antibiotic plate even if they lacked the resistance gene. The results of the plates should show between 100-200 colonies on the plate without antibiotics and no colonies on the plate with antibiotics. These results would confirm the lack *G418* resistance in the cells, meaning the selectable marker had been recovered. The cells were then sequentially diluted, three to four times, by taking 1 µl of culture and adding it to 4 ml of YPD media, and incubating it overnight before repeating. The cultures were grown without antibiotics to cure the strain of any WR63-JBRsgRNA plasmids that remained. The resulting cells were tested for hygromycin resistance in the same manner as used for *G418* resistance. Once sensitivity for both antibiotics was confirmed, the strain was then subjected to another round of the process starting with transforming the *S. stipitis* cells with WR90 and repeating the previous steps (Figure 15).

The results of the experiments show that CRISPR/Cas9 can be used to recover selectable markers, though the approach used did not appear to be highly efficient as it was. Future research can be done to determine ways to improve upon this strategy to increase the effectiveness of recovering selectable markers.

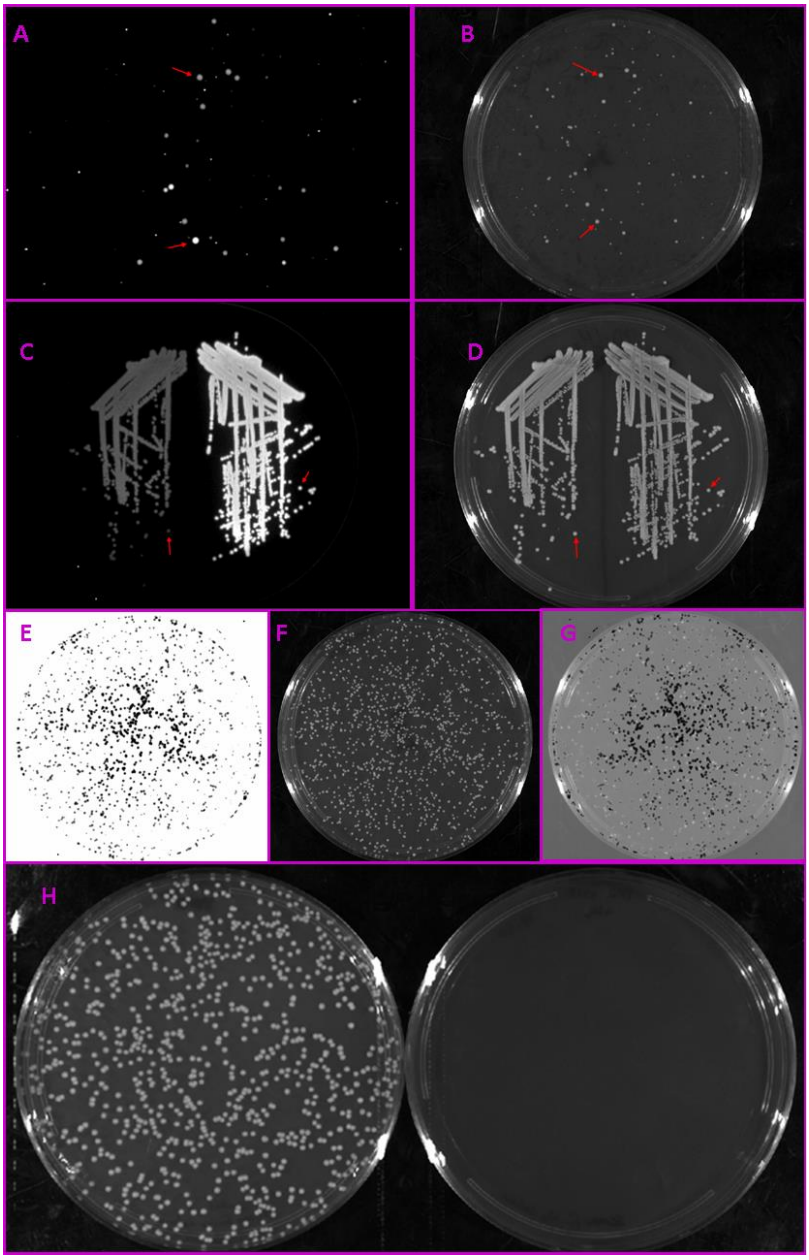

Figure 14: Round 1 of selectable marker recovery. **A, B)** First transformation and colony selection for continuation. Left shows the bioluminescent plate with the brightfield plate on the right. The red arrows indicate colonies that were selected for use in the next step. **C, D)** Test for stable integration of the plasmid and colony selection for continuation. Left shows the bioluminescent plate with the brightfield plate on the right. The red arrows indicate colonies that were selected for use in the next step. **E, F)** Test for selectable marker recovery looking for lack of bioluminescence. Left shows the bioluminescent plate with the brightfield plate on the right. **G)** Merge of selectable marker recovery looking for lack of bioluminescence for determining which colonies lack bioluminescence. **H)** Test for recovery of G418 sensitivity, using colonies that lacked bioluminescence in the previous test. Left plate contains no G418, right plate contains G418.

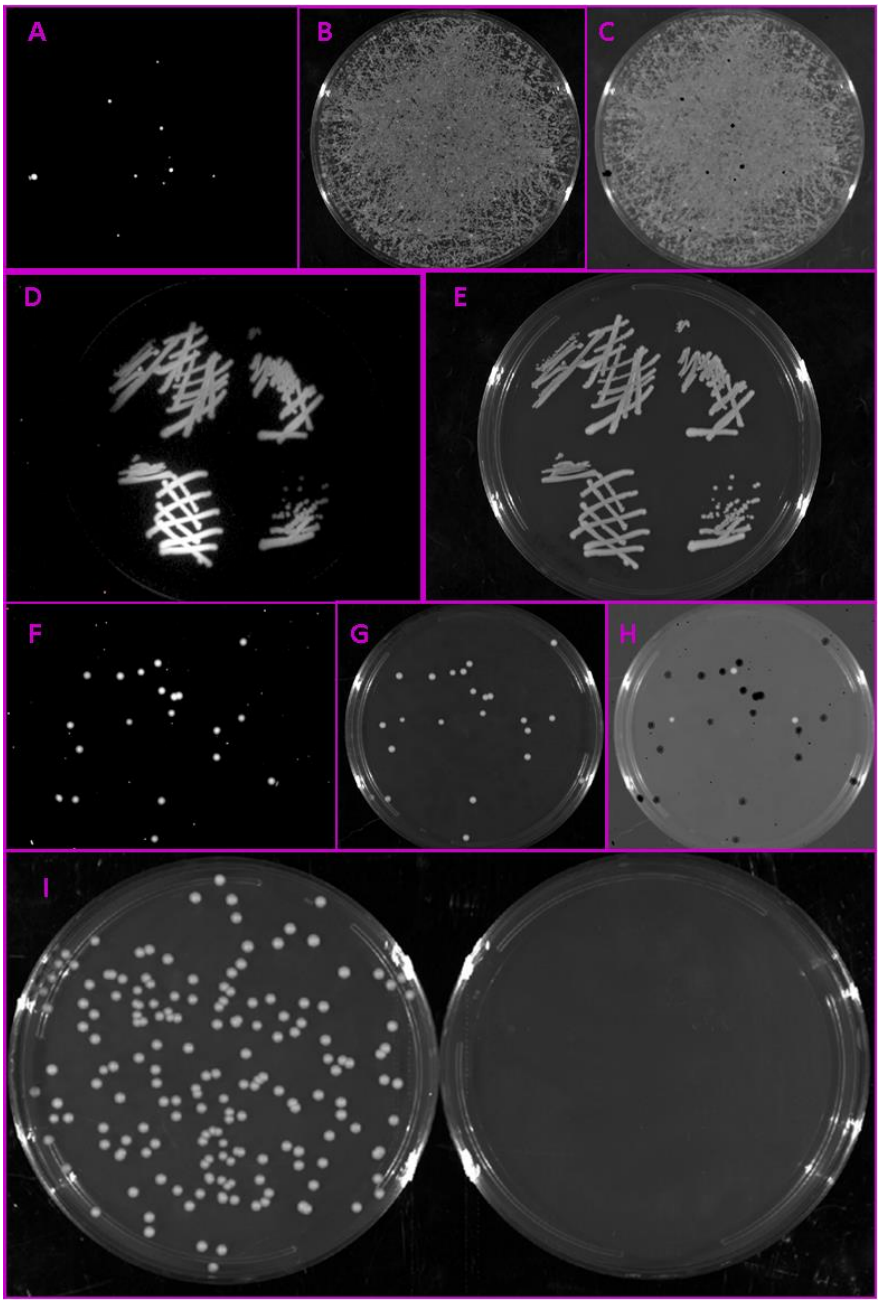

Figure 15: Round 2 of selectable marker recovery. **A, B, C)** First transformation and merge to determine bioluminescent colonies. Left shows the bioluminescent plate, with the brightfield plate in the middle, and the merge of the two on the left. **D, E)** Test for stable integration of the plasmid. Left shows the bioluminescent plate with the brightfield plate on the right. **F, G)** Test for selectable marker recovery looking for lack of bioluminescence. Left shows the bioluminescent plate with the brightfield plate on the right. **H)** Merge of selectable marker recovery looking for lack of bioluminescence for determining which colonies lack bioluminescence. **I)** Test for recovery of G418 sensitivity, using colonies that lacked bioluminescence in the previous test. Left plate contains no G418, right plate contains G418.

Table 5: Plasmids used for the selectable marker recovery using CRISPR/Cas9 in this study. Plasmid nomenclature is indicated as "Backbone (promotor gene terminator) and/or (integration element)".

Plasmid	Description	Plasmid Origin
pYM-N1	(coKanMx)	Janke, 2004
KanMX-PsCO	pUC57(coKanMx)	Genescript
pUC8-LacZ	LacZ	Altman Lab
pWR1	pAllet (P_{TEF1}-co*HPH*-T_{ACT1})	Walter Reichard's Thesis
pWR9	pAllet (P_{TEF1}-co*HPH* -T_{ACT1})(P_{TDH1}-co*CBG99*-T_{ADH1}) (*ARS2*)	
pWR23	pAllet (P_{TEF1}-co*HPH* -T_{ACT1})(P_{TEF1}-coKanMX-T_{ADH1}) (*ARS2*)	
pWR23Δ	pAllet (P_{TEF1}-co*HPH* -T_{ACT1})(P_{TEF1}-coKanMX-T_{ADH1}) (*ARS2*)+ ΔBamHI+ ΔXbaI+ ΔSalI+ ΔAflII	
pWR88	pAllet (P_{TEF1}-coKanMx-T_{ACT1})(P_{TDH3}-CBG99-T_{ADH2})(JBR-sgRNA-PAM)	
pWR88Δ	pAllet (P_{TEF1}-coKanMx-T_{ACT1})(P_{TDH3}-CBG99-T_{ADH2})(JBR-sgRNA-PAM)+ ΔEcoRV	
pWR89	pAllet (P_{TEF1}-coKanMx-T_{ACT1})(P_{TDH3}-CBG99-T_{ADH2})(JBR-sgRNA-PAM)+ ΔEcoRV (EGC3)	
pWR90	pAllet (P_{TEF1}-coKanMx-T_{ACT1})(P_{TDH3}-CBG99-T_{ADH2})(JBR-sgRNA-PAM)+ ΔEcoRV (GSC2)	

Table 6: Primers used for the selectable marker recovery using CRISPR/Cas9 in this study. Numerals at the end of the primer names indicate either 5´ or 3´ direction. Temp is the annealing temperature for PCR in degrees Celsius. Uppercase and underlined sequences indicate the binding sites. Lowercase letters indicate random added base pairs for length off of restriction sites for digestion, unless otherwise indicated in the descriptions.

Primer Name	Sequence	Temp (°C)	Description
Ptef(Bam)5	actactGGATCCGTACACTTATTGTTAACTATG	57	BamI sites
Tadh2(Spe)3	gttcacACTAGTTCTGCCTTCTGAACGTTT	59	SpeI sites
LacZint(Xba)5	actactTCTAGACCATGATTACGGATTCACTGG	63	XbaI sites
LacZint(JBRxBam)3	actataGGATCCCCCGGGGACTAAGTAGAACTTTATCCGCCACATAT	65	Car element and BamI sites

	CCTGATCT		
LacZint(JBRxSpe)5	attacctACTAGTAAAGTTCTACTTAGTCCCCGGGGACCATGATTACGGATTCACTGG	65	Car element and SpeI sites
LacZint(Afl)3	actactCTTAAGATCCGCCACATATCCTGATCT	65	AflI sites
EGC3us(SacI)5	actacaGAGCTCAGCTCAATCCGTACCACATG	62	SacI sites
EGC3(SacI)3	actacaGAGCTCTTTCAATATCTGATTCAATTCC	60	SacI sites
GSC2int(SacI)5	actacaGAGCTCGTTGACATACTTGACAGATTTGG	61	SacI sites
GSC2int(SacI)3	actacaGAGCTCAACTCCTGAACAGCCAAAG	62	SacI sites

2.5 Repairing double-stranded breaks using CRISPR/Cas9

An additional use for CRISPR/Cas9 is in the repair of DNA sequences after double-stranded breaks by using a separate "repair template" to guide the repair process. To test the use of CRISPR/Cas9 in *S. stipitis* to repair double-stranded breaks by a separate repair template, two plasmids were constructed: one with damaged to *CBG99* (pSES-RC) and the other with the correct part for the *CBG99* as a template for the repair (pSES-RT).

The first step in the construction of the pSES-RC plasmid used the pAllet-ShBle plasmid previously constructed during the construction of pAllet-ShBle-coCBG99. The downstream portion of the coCBG+T_{ADH2} from pWR9 was PCR amplified (primers PsCBGdsInt(Spe)5 and Tadh2(EcoRI)3) and cloned into pAllet-ShBle using SpeI and EcoRI. The second step was to PCR amplify the upstream portion of the coCBG+P_{TDH3} from pWR9 (primers Ptdh3(Bam)5 PsCBGusInt(Spe)3) and clone it into the previous plasmid with BamHI and SpeI. The last step was to cut the *URA3* element from pWR16 with SacI and clone it into the previous plasmid resulting in pSES-RC.

The first step in making the repair template pSES-RT was to PCR amplify the coCBG cds from pWR9 using primers with the JRBsgRNA+PAM sequences (primers PsCBG(gRNASal)5 and PsCBG(gRNASal)3) and cloning the product into the SalI site of pWR63-JBRsgRNA. The resulting plasmid was named pSES-RT (Figure 16).

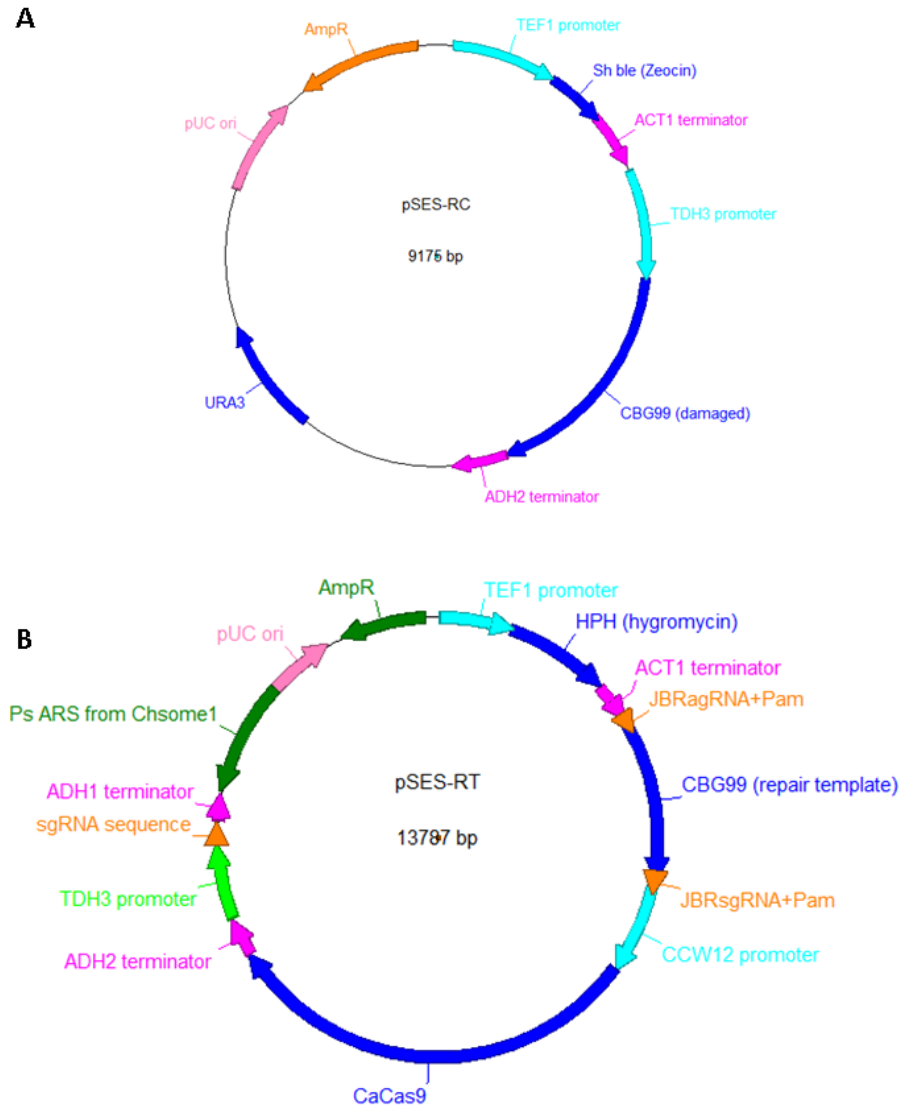

Figure 16: Plasmid layout designs. Both plasmids include a pUC origin of replication and ampicillin resistance. **A**) SES-RC — *CBG99* interrupted with *ShBle* in the middle, with a *URA3* integration element. **B)** SES-RT— Cas9, Hygromycin resistance, and *CBG99* flanked by JBRsgRNA+Pam sequences, with an ARS element from chromosome 1 for replication in *S. stipitis*.

To test the ability of CRISPR to repair double-stranded breaks, pSES-RC was transformed into *S. stipitis* and then plated on YPD+zeocin. Some of the resulting colonies were then patched onto a new YPD plate and screened for successful integration. One of the colonies with the integrated plasmid was then cultured and transformed with pSES-RT and plated on YPD+ hygromycin. Resulting colonies were then patched onto YPD+luciferin plates. The plate was then observed for bioluminescence.

By transforming *S. stipitis* first with the damaged *CBG99* plasmid, which was integrated using *URA3*, and then transforming them with the plasmid carrying Cas9 and the repair template, if Cas9 was active it would cut the damaged *CBG99* and the repair template would be used to fix the cut. The repair of the damaged *CBG99* would then lead to bioluminescence in the colonies in the presence of luciferin. The experiment was conducted five times with the resulting numbers of colonies ranging from 10-300. Despite the differences of the number of colonies at the end, the resulting patches compared to the original CRISPR activity strategy showed the same results in the number of bioluminescent colonies compared to the non-bioluminescent colonies, with roughly 5% having either maximum or varied brightness within the colony, 93% low brightness levels, and 2% with no bioluminescence observed (Figure 17). The result of bioluminescent colonies shows that Cas9 added with a repair template was able to repair the damaged *CBG99* gene. The lower bioluminescent levels indicate that whereas the CBG99 was repaired, there is a low level of expression in the majority of the colonies. This is most likely due to a delay between the colony growth and the success of Cas9 repairing the *CBG99* gene, though with time as cells continue to replicate this decrease may disappear as more of the cells are successfully modified by Cas9. This could also be

due to some colonies ending up with the reporter plasmid integrating in more than one place in the genome and both copies of the gene being repaired, meaning there is twice as much luciferase being translated leading to higher levels of bioluminescence. The test demonstrates that the combination of Cas9 with a repair template should be able to repair any gene sequence in *S. stipitis*.

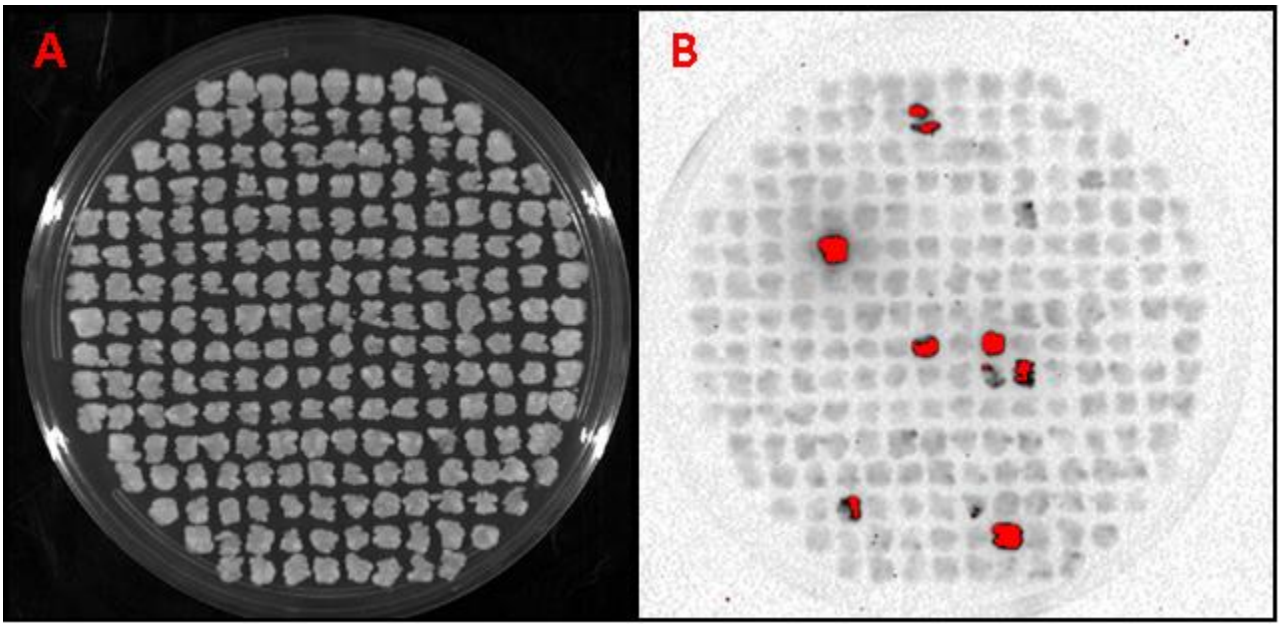

Figure 17: Transformation of SES-RC and SES-RT in S. *stipitis*. **A)** Brightfield plate. **B)** Bioluminescent plate.

Table 7: Plasmids used for the CRISPR/Cas9 repair strategy in this study. Plasmid nomenclature is indicated as "Backbone (promotor gene terminator) and/or (integration element)".

Plasmid	**Description**	**Plasmid Origin**
pYM-N1	(KanMx)	Janke, 2004
KanMX-PsCO	pUC57(coKanMx)	Genescript
pUC8-LacZ	LacZ	Altman Lab
pWR1	pAllet (P_{TEF1}-co*HPH*-T_{ACT1})	Reichard, 2017
pWR9	pAllet (P_{TEF1}-co*HPH* -T_{ACT1}) (P_{TDH1}-co*CBG99*-T_{ADH1}) (*ARS2*)	
pWR16	pAllet (P_{TEF1}-co*HPH*-T_{ACT1})(P_{GAL1}-co*CBG99*fuze-T_{ADH1}) (*ARS2*) (Protein fusion	

	without *CLN2* protein region)	
pAllet-ShBle	pAllet (P_{TEF1}-*ShBle*-T_{ACT})	
pWR63-JBRsgRNA	pAllet (T_{ADH1}) (P_{TEF1}-co*HPH*-T_{ACT1}) (P_{TDH3}-*Cas9*-T_{ADH1}) (JBRsgRNA)	
pSES-RC	pAllet (P_{TDH1}-co*CBG99* (P_{TEF1}-*ShBle*-T_{ACT}) - T_{ADH1}) (*URA3*)	
pSES-RT	pAllet (T_{ADH1}) (P_{TEF1}-co*HPH*-T_{ACT1}) (P_{TDH3}-*Cas9*-T_{ADH1}) (P_{TDH1}-co*CBG99*-T_{ADH1}) (JBRsgRNA+PAM)	

Table 8: Primers used for the CRISPR/Cas9 repair strategy in this study. Numerals at the end of the primer names indicate either 5´ or 3´ direction. Temp is the annealing temperature for PCR in degrees Celsius. Uppercase and underlined sequences indicate the binding sites. Lowercase letters indicate random added base pairs for length off of restriction sites for digestion, unless otherwise indicated in the descriptions.

Primer Name	**Sequence**	**Temp (°C)**	**Description**
PsCBGdsInt(Spe)5	ctaagaACTAGTCATTGCGTGAATTGTGTTG	61	SpeI sites
Tadh2(EcoRI)3	tgccatGAATTCTTCTGCCTTCTGAACGTTTG	62	EcoRI sites
Ptdh3(Bam)5	tcattaGGATCCTCGTTAGTATTTCCGTGAAG	59	BamI sites
PsCBGusInt(Spe)3	ctaagaACTAGTaCCCCGGGGACTAAGTAGAACTTT**cta**GGTGACACCAGGAATCAACTG	63	SpeI sites, JBRsgRNA+PAM, small bold is a stop codon
PsCBG(gRNASal)5	actactGTCGACAAAGTTCTACTTAGTCCCCGGGGATGGTGAAGCGTGAGAAAAATG	66	SalI sites, JBRsgRNA+Pam
PsCBG(gRNASal)3	actactGTCGACCCCCGGGGACTAAGTAGAACTTTCCTTCTCCAACAATTGTTTCAAC	65	SalI sites, JBRsgRNA+Pam

2.6 Improving efficiency of CRISPR/Cas9

Alternate changes to the plasmid construction could increase the efficiency of the CRISPR strategy. Two possible alterations were tested in this study: changing promoters to highly expressed promoters and integrating the Cas9 into the genome.

The first alteration attempted was to change the promoter for both Cas9 and the hygromycin selectable marker to be the *TEF1* promoter that during previous testing showed to have high expression. This was done by utilizing the previously constructed pWR63 plasmid and cutting the *TEF1* promoter from pWR77 using Kpn and AgeI and cloning it into pWR63, replacing the *TDH3* promoter on the *CBG99* with the *TEF1* promoter, so both the Cas9 and *CBG99* had *TEF1* promoters, making pWR63-Tef.

To test the promoter change effect on CRISPR activity, pWR63-Tef was transformed into the *S. stipitis* strain containing pAllet-ShBle-coCBG99 made previously and plated on Hygromycin. At the same time, another transformation was done with pWR63-JRBsgRNA instead of pWR63-Tef. Some of the resulting colonies from both transformations were then patched onto a new YPD-luciferin plate for comparison. The experiment was run multiple times with varying numbers of colonies resulting. Despite the differences of the number of resulting colonies, the resulting patches for pWR63-JBRsgRNA transformation showed the same results matching the other transformations done with it (Figure 18). The colonies from the pWR63-Tef transformation never resulted in bioluminescent colonies (Figure 18). The construction of the plasmid was attempted five times and the transformation was attempted eight times with the same result each time of no bioluminescence, with resulting colony counts ranging from 7-300 colonies. Integrated plasmids carrying duplicate *Tef1* promoters had no problem with expression.

The only problem found was in the non-integrated plasmids carrying the same promoter for two genes. The exact reason for this is unknown, but it is hypothesized that there might be recombination happening between the promoters in the plasmids. Future tests could be performed experimenting with duplicates of different promoters to determine if the same problem persists. Alternately, sequencing the plasmid to see if there were any mutations and/or deletions that would not show in a PCR screening could also be performed to verify that the construction was done correctly. Based on the current results, utilizing the same high transcription promoters more than once in non-integrated plasmids does not appear to be an easy way to improve effectiveness of the CRISPR/Cas9 strategy.

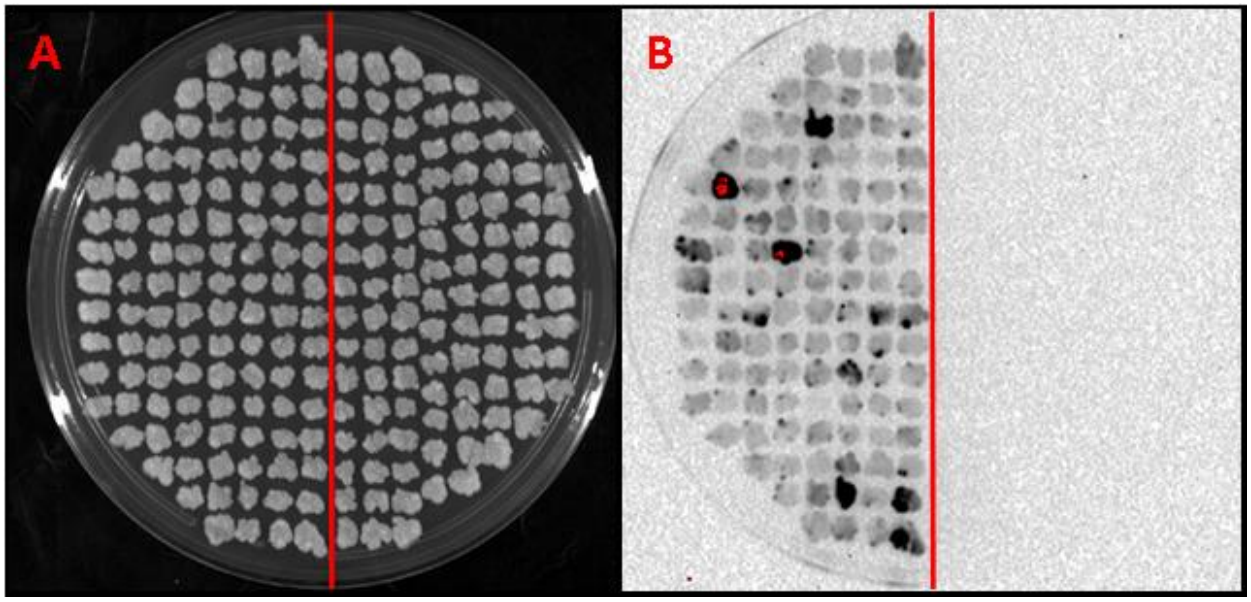

Figure 18: Transformation of pWR68-JBRsgRNA and pWR63-Tef in S. *stipitis*. **A)** Brightfield plate with pWR68-JBRsgRNA on the left and pWR63-Tef on the right. **B)** Bioluminescent plate with pWR68-JBRsgRNA on the left and pWR63-Tef on the right.

The second alteration attempted was to test if the expression and effectiveness of CRISPR/Cas9 would change if Cas9 was stably integrated into the genome instead of being in a plasmid in the cells. To test if the integration of *Cas9* into the genome would increase its effectiveness, a plasmid was constructed by first PCR amplifying the P_{TEF1} from *S. stipitis* genomic DNA (primers Ptef1(Sbf)5 and Ptef1(Fse)3) and using SbfI and FseI to replace the *Gal1* promoter in pWR44, making pWR44-Tef. The second step was to remove the entire P_{TEF}-Cas9-terminator from pWR44-Tef and cloning it into pAllet-ShBle-coCBG99 using SbfI and SacII, making pWR44-Tef-coCBG99.

To test the integration strategy, pWR44-Tef-coCBG99 was transformed into *S. stipitis* and then plated on YPD+Zeocin. Resulting colonies were then patched onto a new YPD plate and incubated overnight. Samples from the resulting patches were then screened for successful integration. One of the colonies with the integrated plasmid was then cultured and transformed with pJBR-63-sgRNA and plated on YPD+hygromycin. Resulting colonies were then patched onto YPD-luciferin plates with colonies transformed using the initial CRISPR activity testing strategy. The plate was then observed for bioluminescence.

The experiment was run four times with the number of colonies resulting ranging from around 100-300. Despite the differences of the number of colonies at the end, the resulting patches showed the same results in the number of bioluminescent colonies compared to the non-bioluminescent colonies (Figure 19). There was also no noticeable difference between the integrated and the non-integrated Cas9. A test was also done using a plasmid carrying the sgRNA without Cas9 to verify the integrated Cas9 was being

expressed. The results of all transformations show no noticeable difference from the results of the experiment on CRISPR activity in *S. stipitis*.

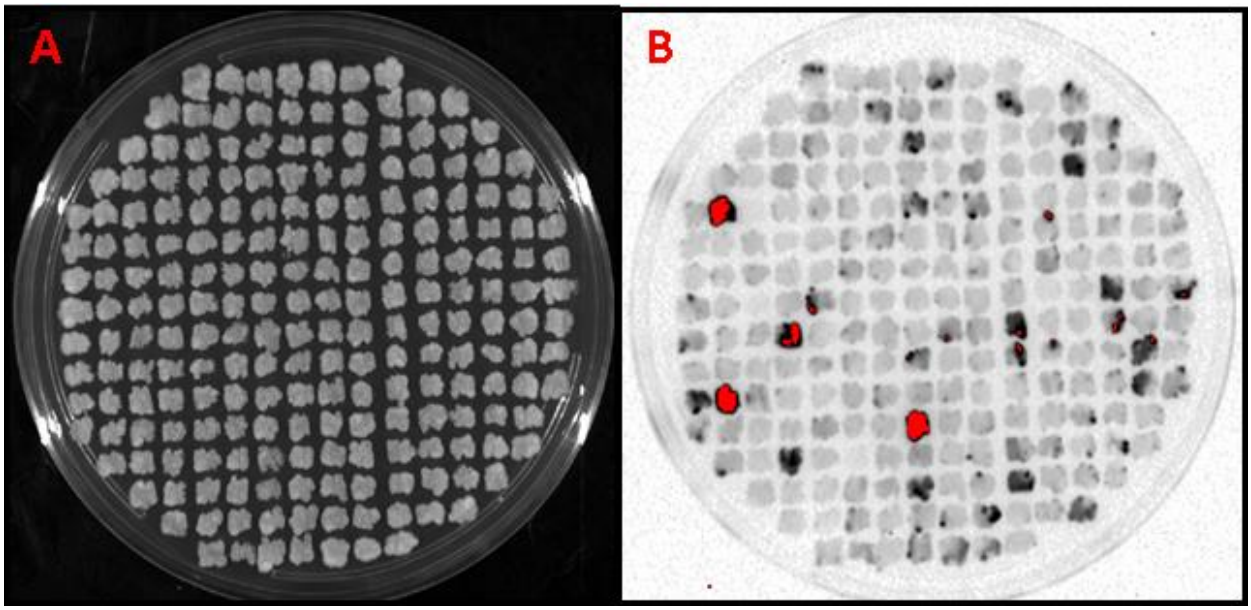

Figure 19: Transformation of WR44-Tef-coCBG99 in S. *stipitis*. **A)** Brightfield plate. **B)** Bioluminescent plate.

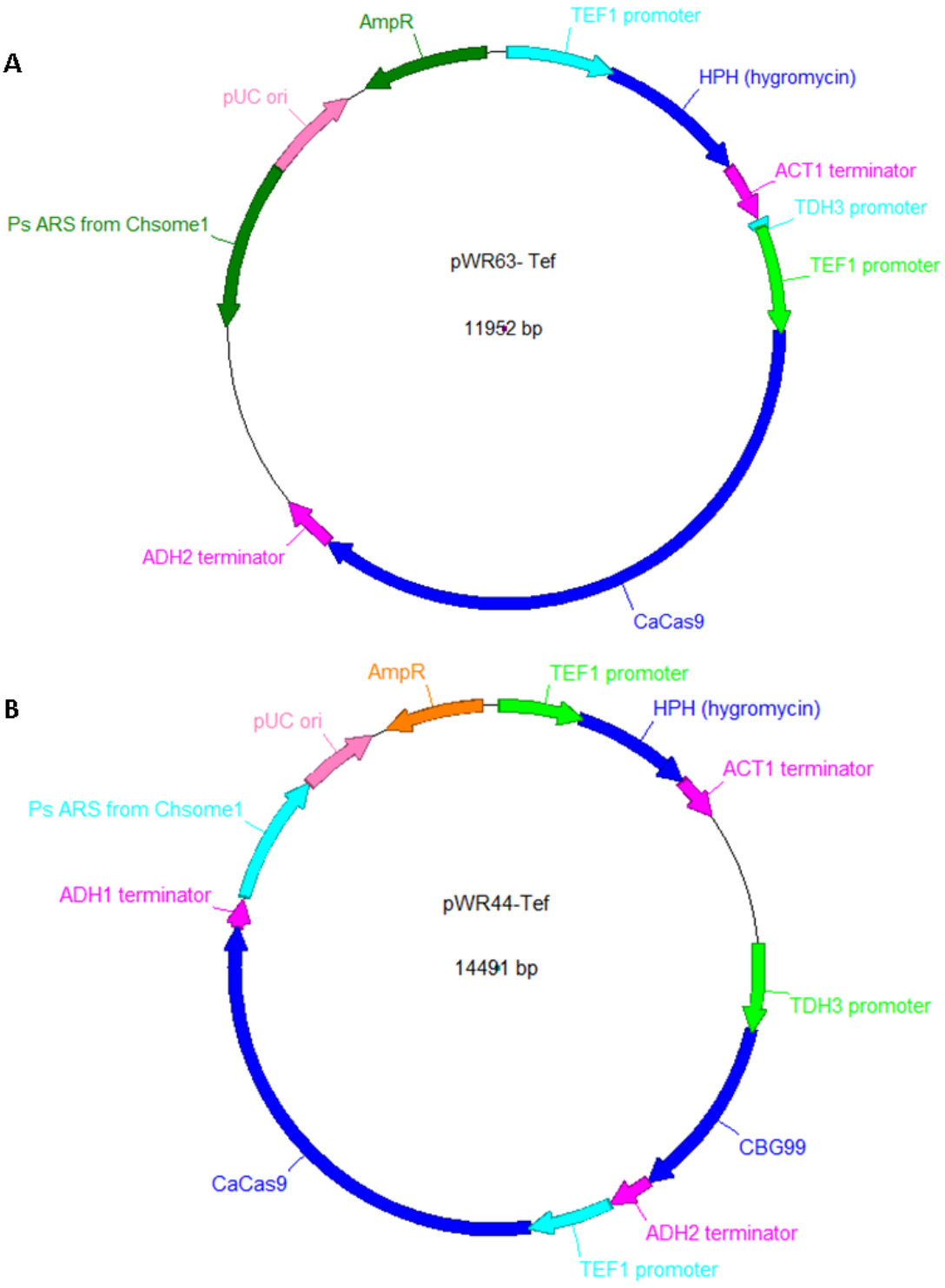

Figure 20: Plasmid layout designs. Both plasmids include a pUC origin of replication and ampicillin resistance. **A)** pWR63-Tef — Cas9 and Hygromycin resistance both with *TEF1* promoters, and JBRsgRNA element, with an ARS element from chromosome 1 for replication in *S. stipitis*. **B)** pWR44-Tef-coCBG99— *CBG99* interrupted by *ShBle* with Cas9, and a *URA3* integration element.

Table 9: Plasmids used for the improving efficiency strategies in this study. Plasmid nomenclature is indicated as "Backbone (promotor gene terminator) and/or (integration element)".

Plasmid	Description	Plasmid Origin
pWR44	pAllet (P_{TEF1}-Hph-T_{ACT}) (P_{TDH3}-co*CBG99*-T_{ADH1}) (P_{GAL1}-*Cas9*-T_{ADH1}) (*ARS2*)	
pWR44-Tef	pAllet (P_{TEF1}-Hph-T_{ACT}) (P_{TDH3}-co*CBG99*-T_{ADH1}) (P_{TEF1}-*Cas9*-T_{ADH1}) (*ARS2*)	
pWR63	pAllet (T_{ADH1}) (P_{TEF1}-co*HPH*-T_{ACT1}) (P_{TDH3}-*Cas9*-T_{ADH1}) (sgRNA element)	
pWR77	pAllet (P_{TEF1}-co*HPH*-T_{ACT1}) (P_{TEF1}-co*CBG99*-T_{ADH2}) (*ARS1*)	
pWR63-Tef	pAllet (T_{ADH1}) (P_{TEF1}-co*HPH*-T_{ACT1}) (P_{TEF1}-*Cas9*-T_{ADH1}) (sgRNA element)	
pAllet-ShBle-coCBG99	pAllet (P_{TDH3}-coCBG99- (P_{TEF1}-ShBle-T_{ACT})-T_{ADH2}) (URA3)	
pWR44-Tef-coCBG99	pAllet (P_{TDH3}-coCBG99- (P_{TEF1}-ShBle-T_{ACT})-T_{ADH2}) (P_{TEF1}-*Cas9*-T_{ADH1}) (URA3)	

Table 10: Primers used for the improving efficiency strategies in this study. Numerals at the end of the primer names indicate either 5´ or 3´ direction. Temp is the annealing temperature for PCR in degrees Celsius. Uppercase and underlined sequences indicate the binding sites. Lowercase letters indicate random added base pairs for length off of restriction sites for digestion, unless otherwise indicated in the descriptions.

Primer Name	Sequence	Temp (°C)	Description
Ptef1(Sbf)5	tgactaCCTGCAGGCCGTACACTTATTGTTAACTATGAA	61	SbfI sites
Ptef1(Fse)3	taccatGGCCGGCCTGTAGATAGACTTAGATTGTATGAAA	58	FseI sites

CHAPTER III
GENERAL METHODS

3.1 Introduction

The following chapter describes the methods for the general protocols used in this study.

3.2 Naming conventions

Plasmids, in this study, are typically named with a lower-case p, followed by the initials of either the one who constructed the plasmid or constructed the previous plasmids that were combined. One example is pWR83, plasmid Walter Reichard 83, so called because the base and pieces used for its construction mainly came from previous plasmids made by Walter Reichard. The numbering system is based on the order in which the plasmids were constructed. Additional information on the plasmids include promotor and terminator information indicated by a capital P or T followed by the gene it is normally associated with in subscript, such as: P_{TEF1}, which is the promotor from the *TEF1* gene. Synthesized genes that have been codon optimized are indicated with co before the name, such as co*CBG99*.

3.3 Culture media and methods

Luria-Bertani (LB) media was used for all bacteria cultures and plates in this study. Liquid LB was made by mixing 5 g of tryptone, 5 g NaCl, and 2.5 g Yeast extract in 500 ml of deionized H_2O. LB plates were made by adding 7.5 g of Agar to the liquid media

mix. All bacteria plates and cultures were incubated at 37°C. Yeast extract peptone dextrose (YPD) media was used for all yeast cultures and plates in this study. Liquid YPD was made by mixing 10 g of Peptone, 10 g of Glucose, and 5 g of yeast extract in 500 ml of deionized H_2O. YPD plates were made by adding 10 g agar to the liquid media. All yeast plates and cultures were incubated at 30°C.

After being mixed, all media was autoclaved. Liquid media was then left and stored at room temperature until use. Media containing agar for plates was moved into an incubator set at 55°C overnight, or left to cool to around 55°C, before any antibiotic and/or luciferin was added. Ampicillin was added at 1 µl/ml (concentration 50 mg/ml), zeocin was added at 1 µl/ml (concentration 100 mg/ml), hygromycin was added at 4 µl/ml (concentration 50 mg/ml), G418 was added at 4 µl/ml (concentration 50mg/ml), and luciferin (Beetle D-Luciferin potassium salt) was added at 1 µl/ml (concentration 100 mM). After the antibiotics and/or luciferin was added, the media was then poured into sterilized 100x15 mm petri plates in a UV sterilized hood, and left for roughly an hour before being stored at 4°C.

3.4 Plasmid Construction

The Robertson lab has several different plasmids built and optimized for *S. stipitis* from previous projects. Construction steps for this project used some of those previously made plasmids either for templates in PCR amplifications or restriction digests to produce pieces used in ligations.

The restriction digests were done for both the vector and insert in 20 µl reactions that used 2 µl of Buffer (specific NEB buffer depended on the enzymes that were to be used), 1000 ng (roughly) of plasmid DNA, 0.5 µl of each restriction enzyme (New England Biolabs, NEB), and molecular grade H_2O in the amount needed to bring the reaction to 20 µl. The reactions were incubated 2 hours at 37°C (or the optimal temperature for the enzyme used if it was different). After the first hour of incubation 0.5 µl of rSAP (NEB) was added to the vectors, and which were then incubated for the second hour. The resulting reactions were run on a gel using gel electrophoresis. The bands at the desired lengths for the vector and insert were then cut out from the gel and the DNA was purified out using GeneJet Gel Purification Kit column purification as described by the manufacturer (ThermoFisher). Following this, a 20 µl ligation reaction was made by adding the vector and insert in a 1:3 ration with 50 ng of vector. The reaction also included 2 µl T4 Ligase Buffer, 0.5 µl Ligase, and any molecular grade H_2O needed to bring the reaction up to 20 µl. The resulting reaction was left at 4°C overnight before it was transformed into *E. coli*. Success of the plasmid constructions were screened for by either using a restriction digest or by PCR followed by gel electrophoresis.

PCR amplifications were done by constructing primers with the desired restriction sites for integration into the vector plasmid and then running a PCR reaction. PCR reactions were made at 50 µl with 10 µl gotaq buffer, 8 µl $MgCl_2$, 1 µl dNTPs (concentration of 10 mM for each), 1 µl of primer mix (0.5 µl of each primer), 500 ng (roughly) of template plasmid, 0.25 µl goTaq DNA polymerase (Promega), and molecular grade H_2O needed to bring the reaction up to 50 µl. The reaction was then put in a thermocycler set for 94°C for 2 minutes, then 30 cycles of 94°C for 30 seconds,

followed by the lower primer temperature for 30 seconds before running at 72°C for 2-3 minutes (depending on the length of the sequence being amplified). After the 30 cycles, the reaction was left at 72°C for 3 minutes. When the PCR reaction was concluded, the amplified DNA segments were purified out using GeneJet column purification as described by the manufacturer (ThermoFisher) and then added to a restriction digest followed by ligation and transformation into *E. coli*.

3.5 Oligo assembly protocol

The oligo assembly protocol used in this study was adapted from a paper by Valmik Vyas (Vyas, 2015) and served to create dimerized oligos that could ligate into the SapI locations of pWR63 and provide the code for making user-defined sgRNAs. The oligo sequences were synthesized by Genscript.

The first step was the phosphorylating and annealing of the oligos. This was done by adding 100 μM of both oligos to a PCR tube with 5 μl of 10X T4 Ligase Buffer, 1 μl T4 Polynucleotide Kinase, and enough molecular grade H_2O to scale the reaction to 50 μl. The PCR tube was then put in a thermocycler set for 37°C for 30 minutes, followed by 95°C for 5 minutes, and then cooled to 16°C at the ramp rate of 0.1°C/second.

The oligos were then ready to be added to a ligation. The ligation was done by adding into a PCR tube 1 μl of 10X T4 Ligase Buffer, 0.5 μl Ligase, 0.5 μl annealed oligo mix, 20-40 ng vector, and molecular H_2O in the amount needed to make the reaction 10 μl total. The PCR tube was then put in a thermocycler set for 16°C for 30 minutes, followed

by 65°C for 10 minutes, and then cooled to 25°C. The resulting plasmid was then ready to be transformed into *E. coli*.

3.6 *E. coli* Transformations

E. coli cells (strain Top10) were taken from the freezer stock and thawed on ice 5-15 minutes. During that time the tube for the transformation was chilled on ice. 100 µl of *E. coli* cells and 2-5 µl of the plasmid being transformed (either already purified plasmid or ligation) were then placed in the chilled tube and incubated on ice for 30 minutes. The cells were then heat-shocked at 42°C for 45 seconds, before being put back on ice for 2 minutes to recover. The next step was to add 400 µl of LB to the cells and incubate them at 37°C continuously shaking for 1 hour. The cells were then plated on LB+Amp and incubated at 37°C overnight. The amount of cell culture plated varied between transformations, with the most common amount being between 20-50 µl. If the success of the transformation was expected to be low, then two plates were inoculated, one with 20 µl and the other with 200 µl.

3.7 *S. stipitis* Transformations

The *S. stipitis* transformation protocol used in this study was designed from several protocols and troubleshot by Walter Reichard (Reichard, 2017).

S. stipitis cells were grown to full cell density on the first day (the strain differed depending on which plasmids were being transformed). The cells grown to maximum density were inoculated into 20 ml of YPD in a 50 ml flask at a dilution level between 1:1500-1:2000. The cultures were then incubated overnight (approximately 16 hours) at

30°C to grow to an OD_{600} of 0.8-1.0. The cells were portioned into 9 ml volumes and pelleted at 3000 rpm for 5 minutes. The supernatant was then discarded and the cells were resuspended in 500 µl of 1X TE and 0.1 M lithium acetate to wash the cells. The cells were pelleted again and then resuspended into 100 µl of 1X TE and 0.1 M lithium acetate and incubated while constantly inverting for one hour at 30°C. During the hour incubation, herring sperm DNA was heated in a thermocycler at 97°C for 20 minutes and then moved directly to ice (amount of sperm DNA changed depending on how many transformations were being done). Any plasmids needing linearized for homologous recombination during the transformations were digested using EcoRV (a unique restriction site in all integration elements in the plasmids) either overnight previously or during the hour incubation. After the hour, 5-20 µg of DNA was added with 15 µl of single-stranded herring sperm DNA and 700 µl of a mixture of 40% poly ethylene glycol, 1X TE, and 0.1 M lithium acetate to the cells. The cells were then incubated while constantly inverting at 30°C for 30 minutes before being heat shocked at 42°C for 5 minutes. The cells were then pelleted and resuspended in 3 ml of YPD to outgrow for 4 hours. Following this, the culture was pelleted at 3000 rpm for 5 minutes and resuspended in 100 ml of YPD and plated onto YPD containing a selective antibiotic (Zeocin or Hygromycin B) to prevent cells that were not successfully transformed from growing into colonies.

3.8 Screening for successful integration of plasmids into *S. stipitis*

To test the success of plasmid being integrated into *S. stipitis* after transformations, the resulting colonies from the transformation were patched onto a new YPD plate and

incubated at 30°C overnight. Samples from the resulting patches were then cultured in liquid YPD 16-18 hours. A “Smash ‘N’ Grab” technique was then used to purify out all the DNA from each culture, and then a PCR screening was run to determine if the plasmid had been integrated.

The “Smash ‘N’ Grab” technique was done by transferring 1.4 ml of cell culture into a 1.5 ml Eppendorf tube and pelleting the cells. The YPD media was then removed and the cells were resuspended in 0.2 ml of lysis buffer (10 mM Tris pH 8, 1 mM EDTA, 100 mM NaCl, 1% SDS, and 2% Triton X-100) with 0.3 g of glass beads and 0.2 ml of 1:1 mix of phenol and chloroform. The tube was then vortexed at top speed for 2 minutes in 30 second increments with 30 seconds on ice in between. The next step was to then add 0.2 ml of TE (10 mM tris, 1 mM EDTA, ph 8.0) and vortex again for a few seconds. The tube was then centrifuged at top speed for 5 minutes, before the aqueous (upper) phase was transferred to a new 1.5 ml Eppendorf tube. Two volumes of 100% ethanol (usually 700 µl) at room temperature were then added and mixed thoroughly by inverting the tube. The tube was then immediately centrifuged at full speed for 15 minutes. The supernatant was then discarded using an aspirator. The pellet was then washed with 0.5 ml of cold 70% ethanol and centrifuged for 5 minutes. The supernatant was then removed and the tube was vacuum dried before the pellet of purified DNA was resuspended in 30 µl of TE.

3.9 Plasmid expression observation

After successful transformations, the resulting single colonies were patched onto YPD+Luciferin plates and incubated 24-48 hours at 30°C. The colonies were then visualized for bioluminescence using a ChemiDoc MP imaging system with Image Lab software. The Image Lab software was used to take brightfield images at an exposure time of 0.3 seconds, and bioluminescent plate images every 30 seconds over the course of 5 minutes (300 seconds) for a total of 10 images per plate. The images with the brightest overall colonies with the lowest amount of maximum brightness areas were chosen for the data images.

3.10 Washout non-integrated plasmids

When it was necessary, non-integrated plasmids were washed out of S. *stipitis* by plating the transformed yeast onto YPD without antibiotics and streaking for single colonies. A single colony was used to inoculate a 3 ml YPD culture (without antibiotics) and incubated overnight at 30°C while continuously shaking. The next step was to make a new 3 ml YPD culture (without antibiotics) which was inoculated with 1 µl of the previous culture and incubated overnight in a similar manner. This process of dilution and culture was repeated once more before the cells were checked for sensitivity to the selectable marker carried in the plasmid that was removed.

CHAPTER IV
CONCLUSION

4.1 Summary

Biochemical and genomic studies of the model yeast *Saccharomyces cerevisiae* have established much of what is understood of eukaryotic biology. Many features of yeasts make them ideal platforms for biotechnological processes. However, *S. cerevisiae* is different from many other yeasts. Because of this, many important biotechnological applications and the highly divergent physiological capabilities of other lesser-known yeast species have not been fully exploited. Among the benefits of other yeast species are that they can grow on methanol or n-alkanesas sole carbon. In addition, they can ferment unconventional carbon sources, as well as thrive under acidic conditions (Robert, 2016).

Due to the increasing demand for biofuels, chemicals, medical technologies, and other genetically engineered products, the vast potential for *S. stipitis* usage has been highlighted. Development and testing of genetic tools optimized for *S. stipitis* must be pursued, since the current tools utilized in other yeast species do not function with *S. stipitis*. The development and optimization of the types of genetic tools used in other organisms for specific use in *S. stipitis* will enhance further studies and uncover new usages of *S. stipitis* and, potentially, other yeast species in the "CUG clade." Because of the potential usages for *S. stipitis* in the industry for production of fuel and other chemicals, having genetic tools that can be easily and efficiently used in *S. stipitis* is a necessity. Genetic editing tools like Cre-recombinase and CRISPR/Cas9 have been used in other species and have increased the types of genetic manipulations possible in those species. Cre-recombinase and CRISPR/Cas9 are both tools currently being pursued for

genetic editing in *S. stipitis* by other labs. The Cre-recombinase strategy used in this study was based on similar work that has been done utilizing Cre-recombinase in *S. stipits* (Laplaza, 2006). The work was done utilizing a codon optimized Cre-recombinase and ShBle selectable marker to inhibit the xylose pathway in *S. stipitis.* The study showed that Cre-recombinase can work in *S. stipitis,* the goal of the research conducted for this study was not only to use Cre-recombinase for gene deletions, but also to recover the selectable marker after the deletion. Work with CRISPR/Cas9 has been done for gene deletions as well, though it has not been pursued specifically for the recovery of selectable markers (Cao, 2018). The CRISPR/Cas9 sgRNA strategy that was pursued in this study was adopted from a study done in *Candia albicans* utilizing CRISPR/Cas9 for genetic deletions (Henry, 2017). Both of these tools are becoming more commonly used in genetic research and having a better understanding of how they work and their usages in these different yeast species will be beneficial to utilizing them in the production of biofuels and other chemicals.

Both Cre-recombinase and CRISPR/Cas9 were tested in this study, although Cre-recombinase proved to have issues in its usage that will need to be resolved before it can easily be used. CRISPR/Cas9 proved to be capable of recovering selectable markers, which increases the possibilities of genetic alterations which rely on the usage of selectable markers. The study also demonstrates how CRISPR/Cas9 can potentially be used to repair any damaged genes and return them to functionality. This would allow for the recovery of previously deleted genes, if so desired. These tools and strategies are useful in furthering the understanding of the genetics of *S. stipitis* and improving its usage for research.

4.2 Future Direction

This project concludes that CRISPR/Cas9 is the easier and more versatile tool to use in *S. stipitis* when compared to Cre-recombinase. One of the first directions for further study should be to determine the current efficiency of the CRISPR/Cas9 strategy and develop ways to improve the efficiency. Future testing can be done by doing more than two consecutive selectable marker recoveries and recoveries on the same chromosome, with the purpose of determining if the duplicated regions (CAR) for homologous recombination leads to random recombination and deletion of genes if the same regions are used in multiple selectable marker recoveries. Once the CRISPR/Cas9 strategy and its limitations are better understood, one of the next steps would be to use CRISPR/Cas9 for sequential gene deletions of genes needed for glucose metabolism to create a glucose null *S. stipitis* strain, which could then be tested for sugar metabolism of lignocellulosic biomass in combination with *S. cerevisiae*. Other future directions can include the testing of these strategies in other yeast species belonging to the "CUG clade."

REFERENCES

1. Baldwin, Thomas O. “Firefly Luciferase: The Structure Is Known, but the Mystery Remains.” Structure, vol. 4, no. 3, Elsevier Inc, 1996, pp. 223–28, doi:10.1016/S0969-2126(96)00026-3.

2. Cao, Mingfeng, et al. “CRISPR-Mediated Genome Editing and Gene Repression in *Scheffersomyces stipitis*.” Biotech Journal, 2018. Doi: 10.1002/biot.201700598

3. Capecchi, Mario R. “Altering the Genome by Homologous Recombination.” Science (American Association for the Advancement of Science), vol. 244, no. 4910, The American Association for the Advancement of Science, 1989, pp. 1288–92, doi:10.1126/science.2660260.

4. Cheng, T. H., et al. “Controlling Gene Expression in Yeast by Inducible Site-Specific Recombination.” Nucleic Acids Research, vol. 28, no. 24, Oxford University Press, 2000, pp. E108–108, doi:10.1093/nar/28.24.e108.

5. Cunha, Joana T., et al. “HAA1 and PRS3 Overexpression Boosts Yeast Tolerance Towards Acetic Acid Improving Xylose or Glucose Consumption: Unravelling the Underlying Mechanisms.” Applied Microbiology and Biotechnology, vol. 102, no. 10, Springer, 2018, pp. 4589–600, doi:10.1007/s00253-018-8955-z.

6. Davis, Chris R., et al. “A Bioluminescent Reporter for the Halophilic Archaeon Haloferax Volcanii.” Extremophiles : Life Under Extreme Conditions, vol. 24, no. 5, 2020, pp. 773–85, doi:10.1007/s00792-020-01193-x.

7. Feng, Xueyang, and Huimin Zhao. “Investigating Glucose and Xylose Metabolism in Saccharomyces Cerevisiae and Scheffersomyces Stipitis via 13C Metabolic Flux Analysis.” AIChE Journal, vol. 59, no. 9, Wiley Subscription Services, Inc, 2013, pp. 3195–202, doi:10.1002/aic.14182.

8. Feng, Yong-Qing, et al. “Site-Specific Chromosomal Integration in Mammalian Cells: Highly Efficient CRE Recombinase-Mediated Cassette Exchange.” Journal of Molecular Biology, vol. 292, no. 4, Elsevier Ltd, 1999, pp. 779–85, doi:10.1006/jmbi.1999.3113.

9. Henry Ng, et al. “Dramatic Improvement of CRISPR/Cas9 Editing in Candida Albicans by Increased Single Guide RNA Expression.” MSphere, vol. 2, no. 2, Apr. 2017. EBSCOhost, doi:10.1128/mSphere.00385-16.

10. Ilic, Danica Djuric, et al. “Integration of Biofuel Production into District Heating – Part I: An Evaluation of Biofuel Production Costs Using Four Types of Biofuel Production Plants as Case Studies.” Journal of Cleaner Production, vol. 69, Elsevier Ltd, 2014, pp. 176–87, doi:10.1016/j.jclepro.2014.01.035.

11. Janke, Carsten et al. “A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes.” Yeast (Chichester, England) vol. 21,11 (2004): 947-62. doi:10.1002/yea.1142

12. Jimenez-Quero, Amparo, et al. “Fungal Fermentation of Lignocellulosic Biomass for Itaconic and Fumaric Acid Production.” Journal of Microbiology and Biotechnology, vol. 27, no. 1, 한국미생물·생명공학회, 2017, pp. 1–8, doi:10.4014/jmb.1607.07057.

13. Kim, Soo Rin, et al. "Simultaneous co-fermentation of mixed sugars: a promising strategy for producing cellulosic ethanol." Trends in biotechnology 30.5 (2012): 274-282.

14. Krishnamoorthy, Archana, and J. Brian Robertson. "Dual-Color Monitoring Overcomes the Limitations of Single Bioluminescent Reporters in Fast-Growing Microbes and Reveals Phase-Dependent Protein Productivity during the Metabolic Rhythms of Saccharomyces cerevisiae." Applied and environmental microbiology 81.18 (2015): 6484-6495.

15. Kurtzman, Cletus P., and Motofumi Suzuki. "Phylogenetic analysis of ascomycete yeasts that form coenzyme Q-9 and the proposal of the new genera Babjeviella, Meyerozyma, Millerozyma, Priceomyces, and Scheffersomyces." Mycoscience 51.1 (2010): 2-14.

16. Laplaza, Jose M., et al. “Sh Ble and Cre Adapted for Functional Genomics and Metabolic Engineering of Pichia Stipitis.” Enzyme and Microbial Technology, vol. 38, no. 6, Elsevier Inc, 2006, pp. 741–47, doi:10.1016/j.enzmictec.2005.07.024.

17. Li, Xin, et al. “Production of Fumaric Acid by Rhizopus Oryzae in Simultaneous Saccharification and Fermentation Using Xylo-Oligosaccharides Manufacturing Waste Residue.” Bioresources, vol. 11, no. 4, North Carolina State University, 2016, pp. 8831–43, doi:10.15376/biores.11.4.8831-8843.

18. Loureiro, Alexandre, and Gabriela Jorge da Silva. “CRISPR-Cas: Converting A Bacterial Defence Mechanism into A State-of-the-Art Genetic Manipulation Tool.” Antibiotics (Basel), vol. 8, no. 1, MDPI AG, 2019, p. 18–, doi:10.3390/antibiotics8010018.

19. Lu, Xiaoming, et al. “Biomass Logistics Analysis for Large Scale Biofuel Production: Case Study of Loblolly Pine and Switchgrass.” Bioresource Technology, vol. 183, Elsevier Ltd, 2015, pp. 1–9, doi:10.1016/j.biortech.2015.02.032.

20. Machado, Heather E., et al. “Pervasive Strong Selection at the Level of Codon Usage Bias in Drosophila Melanogaster .” Genetics, vol. 214, no. 2, Feb. 2020, pp. 511–528. EBSCOhost, doi:10.1534/genetics.119.302542.

21. Maitra, P. K., and Zita Lobo. “GENETICS OF YEAST GLUCOKINASE.” Genetics (Austin), vol. 105, no. 3, Genetics Soc America, 1983, pp. 501–15, doi:10.1093/genetics/105.3.501.

22. Massey, Steven E., et al. "Comparative evolutionary genomics unveils the molecular mechanism of reassignment of the CTG codon in Candida spp." Genome research 13.4 (2003): 544-557.

23. Mathew, Anil Kuruvilla, et al. “Continuous Bioethanol Production from Oilseed Rape Straw Hydrosylate Using Immobilised Saccharomyces Cerevisiae Cells.” Bioresource Technology, vol. 154, Elsevier Ltd, 2014, pp. 248–53, doi:10.1016/j.biortech.2013.12.044.

24. Mezzanotte, Laura, et al. "Sensitive dual color in vivo bioluminescence imaging using a new red codon optimized firefly luciferase and a green click beetle luciferase." PloS one 6.4 (2011): e19277.

25. Morgan, Jeffrey T., et al. “Excised Linear Introns Regulate Growth in Yeast.” Nature (London), vol. 565, no. 7741, Nature Publishing Group, 2019, pp. 606–11, doi:10.1038/s41586-018-0828-1.

26. Nigam, J. N. "Ethanol production from wheat straw hemicellulose hydrolysate by Pichia stipitis." Journal of biotechnology 87.1 (2001): 17-27.

27. Parenteau, Julie, et al. “Deletion of Many Yeast Introns Reveals a Minority of Genes That Require Splicing for Function.” Molecular Biology of the Cell, vol. 19, no. 5, The American Society for Cell Biology, 2008, pp. 1932–41, doi:10.1091/mbc.E07-12-1254.

28. Pathania, Shruti, et al. “Immobilization of Co-Culture of Saccharomyces Cerevisiae and Scheffersomyces Stipitis in Sodium Alginate for Bioethanol Production Using Hydrolysate of Apple Pomace Under Separate Hydrolysis and Fermentation.” Biocatalysis and Biotransformation, vol. 35, no. 6, Taylor & Francis, 2017, pp. 450–59, doi:10.1080/10242422.2017.1368497.

29. Priest, Shelby J., and Michael C. Lorenz. "Characterization of virulence-related phenotypes in Candida species of the CUG clade." Eukaryotic cell 14.9 (2015): 931-940.

30. Reichard, Walter. “Designing and Demonstrating the Use of a Bioluminescent Genetic Tool Set in Scheffersomyces Stipitis.” ProQuest Dissertations Publishing, 2017.

31. Robert, Riley, et al. “Comparative Genomics of Biotechnologically Important Yeasts.” Proceedings of the National Academy of Sciences of the United States of America, vol. 113, no. 35, Aug. 2016, pp. 9882–9887. EBSCOhost, search.ebscohost.com/login.aspx?direct=true&db=edsjsr&AN=edsjsr.26471720&site=eds-live&scope=site.

32. Robertson, J. Brian, et al. "Real-time luminescence monitoring of cell-cycle and respiratory oscillations in yeast." Proceedings of the National Academy of Sciences 105.46 (2008): 17988-17993.

33. Sharp, Paul M., et al. "Codon usage patterns in Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster and Homo sapiens; a review of the considerable within-species diversity." Nucleic acids research 16.17 (1988): 8207-8211.

34. Singh Digvijay, et al. “Real-Time Observation of DNA Target Interrogation and Product Release by the RNA-Guided Endonuclease CRISPR Cpf1 (Cas12a).” Proceedings of the National Academy of Sciences of the United States of America, vol. 115, no. 21, May 2018,pp.5444–5449.EBSCOhost,search.ebscohost.com/login.aspx?direct=true&db=edsjsr&AN=edsjsr.26509841&site=eds-live&scope=site.

35. Slininger, Patricia J., et al. "Repression of Xylose-specific enzymes by ethanol in Scheffersomyces (Pichia) stipitis and utility of repitching Xylose-grown populations to eliminate diauxic lag." Biotechnology and bioengineering 108.8 (2011): 1801-1815.

36. Suh, S.O., Marshall, C.J., McHugh, J.V. & Blackwell, M. Wood ingestion by passalid beetles in the presence of Xylose-fermenting gut yeasts. Mol. Ecol. 12, 3137–3145 (2003).

37. Tong, C. et al. “Production of P53 Gene Knockout Rats by Homologous Recombination in Embryonic Stem Cells.” Nature Reviews. Genetics, vol. 11, no. 9, Nature Publishing Group, 2010.

38. Van Duyne, Gregory D. “A STRUCTURAL VIEW OF Cre-loxP SITE-SPECIFIC RECOMBINATION.” Annual Review of Biophysics and Biomolecular Structure, vol. 30, no. 1, Annual Reviews, 2001, pp. 87–104, doi:10.1146/annurev.biophys.30.1.87.

39. Vijai, Singh, and Pawan K. Dhar. “Genome Engineering via CRISPR-Cas9 System.” Academic Press, 2020

40. Wei, Liang, et al. “Engineering Scheffersomyces Stipitis for Fumaric Acid Production from Xylose.” Bioresource Technology, vol. 187, Elsevier B.V, 2015, pp. 246–54, doi:10.1016/j.biortech.2015.03.122.

41. Xia, Wenjing, et al. “Population Genomics Reveals Structure at the Individual, Host-tree Scale and Persistence of Genotypic Variants of the Undomesticated Yeast Saccharomyces Paradoxus in a Natural Woodland.” Molecular Ecology, vol. 26, no. 4, Wiley Subscription Services, Inc, 2017, pp. 995–1007, doi:10.1111/mec.13954.