Comparison of multi-gene integration strategies in CRISPR-based transformation of *Saccharomyces cerevisiae*.

UNIVERSITY *of the*
WESTERN CAPE

Odwa Jacob

A thesis submitted in partial fulfilment of the requirements for the degree of Magister Scientiae in the Department of Biotechnology, University of the Western Cape

Supervisor: Prof. R. Den Haan

December 2021

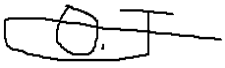
UNIVERSITY of the
WESTERN CAPE

Name: Odwa Jacob

Student number: 3339572

I declare that "**Comparison of multi-gene integration strategies in CRISPR-based transformation of *Saccharomyces cerevisiae***" is my own work, that it has not been submitted for any degree or examination in any other university, and that all sources I have used I have used have been indicated and acknowledge by complete references.

Signature:



Date: December 2021

CONTENTS

A. ACKNOWLEDGEMENTS

First and foremost, I would like to thank God for providing me with the strength to pursue my master's degree. A special thanks to my supervisor, **Prof. Riaan Den Haan**, for his unwavering support since the first day we began this project in the lab. I would also like to express my gratitude to him for his patience, support, and understanding during the challenging times in the lab.

I am also indebted to my incredible colleagues in the yeast lab, and I would like to express my heartfelt gratitude for your friendship and for providing a positive work environment.

I would like to express my gratitude to my grandmother, **Nomasomi Albertina Rasmeni**, and my mother, **Nomalungelo Khumalo**, for their unwavering efforts and sacrifices to see me prosper and achieve my goals. I owe them an inordinate amount of gratitude for this. I would not be where I am today without their love and support. I would also like to thank my older brother, **Aviwe Jacob**, and my uncles, **Thulani Rasmeni**, **Vuyisa Rasmeni**, **Sandiso Rasmeni**, and **Vusani Rasmeni**, for their words of encouragement and the role they played at home when I was growing up.

I will forever be indebted to my incredible friend **Anamhla Ndzulu** for all her support during this challenging time. Your constant encouragement and motivation aided me tremendously in completing this thesis. I would not have been able to do it without your additional support.

Finally, I would like to thank the **University of the Western Cape** for providing the necessary work environment, as well as the **National Research Foundation (NRF)** for financial assistance in completing this degree.

B. LIST OF ABBREVIATIONS

ARS – Autonomously Replicating Sequence

BGL – Beta (β) Glucosidase

Cas9 – CRISPR-associated nuclease 9

CBH – Cellobiohydrolase

CBP– Consolidated Bioprocessing

CEN – Centromere sequence

CMC – CarboxyMethyl Cellulose

CRISPR – Clusters of Regularly Interspaced Palindromic Repeats

DSBs – Double Stranded Breaks

EG – Endoglucanase

gRNA – guide RNA

HDR – Homologous Directed Repair

HR – Homologous Recombination

LB – Luria Bertani

NHEJ – Non-Homologous End Joining

PAM – Protospacer Adjacent Motif

pCas9 – Cas9 plasmid

qPCR – Quantitative Polymerase Chain Reaction

RNA – Ribonucleic acid

S.f.BGL1– *Saccharomycopsis fibuligera*. β-Glucosidase

T.e.cbh1– *Taloromyces emersonii*. Cellobiohydrolase I

T.r.eg2– *Trichoderma reesei*. Endoglucanase I

TALENs – Transcription Activator-Like Effector Nucleases

TB – Terrific Broth

URA3 – YEL021W gene

YPD – yeast extract, Peptone, D-glucose

ZFNs –Zinc finger nucleases

C. LIST OF FIGURES

D. LIST OF TABLES

E. ABSTRACT

Saccharomyces cerevisiae is an important host in industrial biotechnology. This yeast is the host of choice for the first and second-generation biofuels for ethanol production. Genome modification in *S. cerevisiae* has been extremely successful largely due to this yeast's highly efficient homology-directed DNA repair machinery. The advent of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) genome editing technology has made multi-gene editing in yeast more accessible. In this study, we aimed at targeting the Cas9 to multiple genomic positions for integrating multiple genes at different sites. We have developed two CRISPR-Cas9 systems, based on published one- and two-plasmid systems, for application in *S. cerevisiae* strains. In this study, these CRISPR-Cas9 systems were used to transform fungal heterologous genes into yeast using the electroporation transformation method. We first utilized the CRISPR systems for targeting the *T.r.eg2* gene to single locus chromosomal sites for single copy integration. Subsequently, we then targeted the same gene to repeated sequences in the genome, namely the delta sites, for multi-copy integration. The procedure was repeated with a different gene, *T.e.cbh1*, integrated into the same sites to ascertain reporter gene specific effects. High integration efficiency was achieved, since all the strains successfully integrated the genes. However, we discovered significant differences in enzyme activities between the two genes when targeted to different loci, as well as varying copy numbers as determined by qPCR. The *T.e.cbh1* gene was highly expressed by yeast transformants targeted at the repeated delta sequences used for multi-copy integration, reaching maximum levels of 248 mU/gDCW. The *T.r.eg2* gene was highly expressed in yeast transformants targeted to the single locus site on chromosome 12, reaching a maximum of 160U/gDCW, though it was shown that off-target integration likely occurred. We then used the information from these observations to construct a CBP yeast strain containing three cellulase genes: *T.r.eg2*, *T.e.cbh1*, and *S.f.BGL1*. Significant differences in enzyme activities were observed between the three genes, and it was shown that the *S.f.BGL1* gene was poorly expressed by the CBP yeast strain, whereas the *T.r.eg2* gene was highly expressed. Notably, due to the fact that marker containing plasmids could be cured from these strains, many additional genetic changes can still be made. Overall, our two CRISPR-Cas9 systems were efficient at engineering strains that produce recombinant proteins and can be used in future studies for a variety of applications, including metabolic engineering in *S. cerevisiae*.

CHAPTER 1

LITERATURE REVIEW

1.1. INTRODUCTION

Saccharomyces cerevisiae is a unicellular eukaryotic organism, commonly known as yeast. It is found on the cuticles of sugary fruits (grapes and plums) and is traditionally used in winemaking, baking, and brewing (Duina *et al.,* 2014; Borodina and Nielsen, 2014). This yeast has also been selected as a host for the development of first- and second-generation biofuel production, which utilizes either food biomass (corn starch, sugarcane) or lignocellulosic feedstocks (straw, corn stover, wood), respectively. *S. cerevisiae* is also used to make enzymes and pharmaceutical proteins (for example, insulin, hepatitis-, and human papillomavirus vaccines), and it is being explored to make advanced biofuels including farnesene and isobutanol, as well as fine chemical compounds such as resveratrol or nootkatone (Borodina and Nielsen, 2014). The production of fuels and chemicals in biorefineries necessitates the use of robust strains that are resistant to industrial stresses such as low pH, high ethanol concentrations, fluctuating temperatures, and the presence of different inhibitors (Demeke *et al.,* 2013).

Bioconversion of lignocellulosic biomass is gaining traction as a technique of generating biofuels (Kumar et *al.,* 2009). The following stages are involved in biomass processing to produce bioethanol: (i) chemical and physicochemical pre-treatment of biomass to make it susceptible to cellulolytic enzyme activity; (ii) enzymatic hydrolysis of pre-treated biomass components; and (iii) fermentation of the resulting hexose and pentose sugars (Hasunuma and Kondo 2012). The first conventional conversion technique was to separate hydrolysis and fermentation (SHF) (Figure 1.1). In SHF, each process is performed independently, allowing optimal conditions for the processes, however, it poses difficulties for industrial applications that include low conversion rates and the risk of contamination. As a result, combining the stages of lignocellulosic conversion was proposed (Oh and Jin, 2020). By combining enzymatic hydrolysis with microbial fermentation of hexoses, a simultaneous saccharification and fermentation (SSF) technique streamlines and shortens the process stages (Figure 1.1).

However, both SHF and SSF procedures need expensive exogenous enzymes, indicating that there is room for improvements of facility and operating costs.

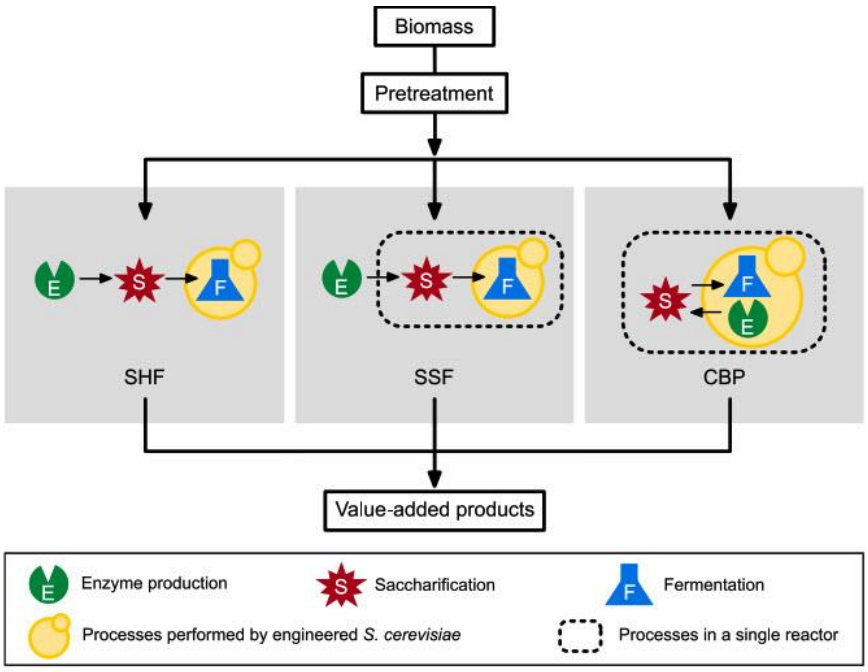


Figure 1.1. Biofuel production schemes (Oh and Jin, 2020). SHF, separate hydrolysis and fermentation; SSF, simultaneous saccharification and fermentation; CBP, consolidated bioprocess.

The most integrated bioconversion process is the consolidated bioprocess (CBP) (Figure 1.1). All processes, including enzyme synthesis, occur in a single reactor (Oh and Jin, 2020). CBP conversion of pre-treated lignocellulose using a single microorganism or microbial consortium capable of producing the required enzymes and fermenting the resulting sugars into value-added products may provide economic benefits by lowering enzyme costs. CBP requires a microbial workhorse with the necessary phenotypes for enzyme synthesis, saccharification, and productivity (Lynd *et al.,* 2005; Van Zyl *et al.,* 2007; Yamada *et al.,* 2013; Den Haan *et al.,* 2015).

The major challenges in developing *S. cerevisiae* for effective microbial conversion of lignocellulosic biomass include heterologous expression of cellulolytic enzymes, engineering co-fermentation of hexose and pentose sugars, and ensuring resilience to different stressors

(Den Haan *et al.,* 2015). Cellulose, hemicellulose, and lignin are the three major components of lignocellulosic biomass. Lignocellulose is composed mostly of cellulose, which are β-1,4 linked polymers of glucose, and glucose is the most abundant sugar in lignocellulose hydrolysates (Oh and Jin, 2020). The synthesis of glucose monomers from cellulose requires the synergistic and coordinated activity of three main types of cellulases namely: Endoglucanases (EGs), exoglucanases such as cellobiohydrolases (CBHs), and β-glucosidases (BGLs) (Yamada *et al.,* 2013). Hemicellulose is a heteropolymer made up of xylose, arabinose, galactose, and other sugars. The second most abundant sugar in lignocellulose hydrolysates is xylose, which is produced from hemicellulose hydrolysis (Oh and Jin, 2020).

Heterologous expression of genes encoding cellulases from fungi or bacteria has been shown in recombinant strains of *S. cerevisiae*, (Van Zyl *et al.,* 2007; Den Haan *et al.,* 2015). Van Rensburg *et al.,* (1998) co-expressed β-glucosidase from *Endomyces fibuliger*, EG from *Butyrivibrio fibrisolvens*, CBH from *Phanerochaete chrysosporium*, and cellodextrinase from *Ruminococcus flavefaciens*. The resulting strain generated physiologically active cellulases that hydrolysed cellulosic substrates. However, additional advancements and rapid utilization of lignocellulosic sugars by fermenting microorganisms are needed before adopting engineered yeasts for commercially viable bioconversion on an industrial scale.

To direct the metabolic flux towards the products of interest, strain improvement through metabolic engineering may require multiple rounds of genetic changes, such as (i) the introduction of heterologous genes or whole metabolic pathways, (ii) the complete or partial elimination of the activity of some endogenous proteins and/or (iii) the overproduction of certain endogenous proteins (Li and Borodina, 2015). It is thus clear that to enable CBP conversion of lignocellulosic sugars to products of interest, many genetic changes need to be engineered into this yeast. While several methods to enable genetic engineering in yeasts have been developed over the past four decades, multiple changes often require multiple rounds of engineering that become cumbersome or even impossible due to a lack of appropriate vectors or selection markers.

When compared to other eukaryotes, *S. cerevisiae* is the most easily controlled organism in a laboratory environment due to its ease of genetic manipulation, which is greatly aided by a preference for homologous recombination (HR) over non-homologous end joining (NHEJ) for

double stranded break (DSB) repair (Figure1.2a). HR and NHEJ are processes that cells use to repair double stranded DNA breaks that will be fatal to the cell if left unattended. To repair a DNA break without causing deletions or non-sense mutations, homologous recombination requires a DNA donor to be provided to the cell (Makarova *et al.,* 2015; Sander and Joung, 2014; Hsu *et al.,* 2014). Non-Homologous End-Joining (NHEJ) causes errors in the genome during the repair process that result in small insertions or deletions. Scientists have taken advantage of *S. cerevisiae*'s preference for HR over the years, as it allowed for site-specific installation of foreign genetic material into the yeast genome to produce desired recombinant yeast strains.

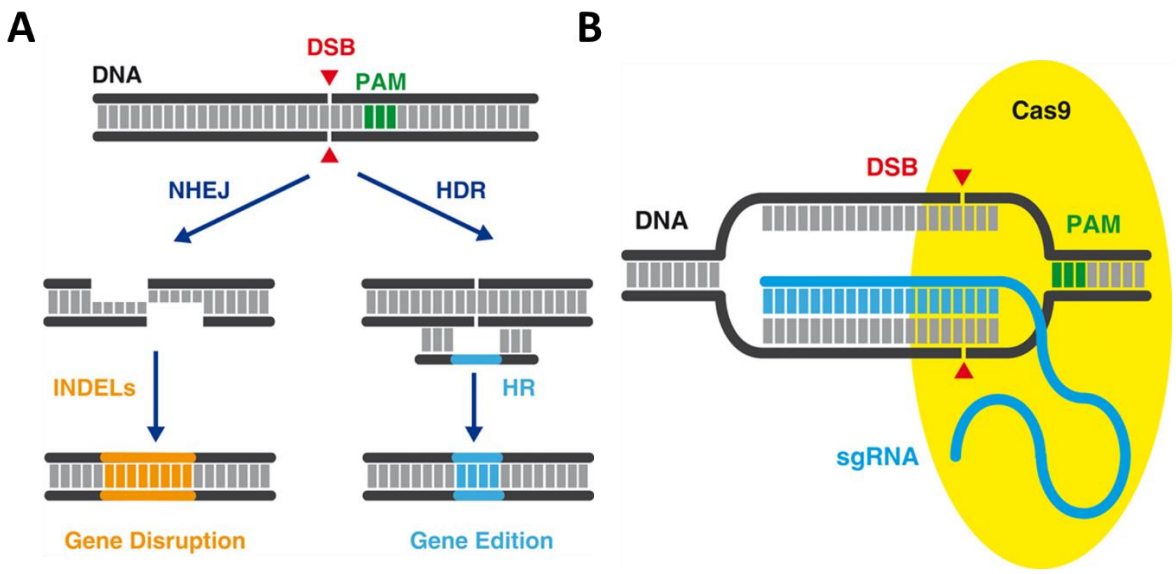


Figure 1.2: Schematic representation of the DSB repair mechanism in *S. cerevisiae* (Makarova *et al.,*2015). (a) Two repair pathways occurring after CRISPR-Cas9 action. The cell detects the breaks and attempts to repair them via either HDR or NHEJ. (b) CRISPR-Cas9 mechanism of action. When Cas9 and single strand gRNA interact, they create a Cas9-gRNA complex, which causes DSBs in the target sequence.

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat) is a type of adaptive immune system found in some prokaryotes (Mans *et al.,* 2015). When a prokaryote employs this mechanism, it preserves some of the virus's DNA in the CRISPR locus on its genome to aid in recognizing and defending against the same virus the next time it invades (Lander, 2016). Two research groups unravelled the system's mechanism in 2012, when they were able to

modify and reprogram the targeting function of the CRISPR-associated nuclease (Cas9) from *Streptococcus pyogenes* so that it could specifically induce double stranded breaks (DSBs) in a DNA target sequence *in vitro* (Gasiunas *et al.,* 2012). Cas9 needs (i) a guide RNA (gRNA) sequence to direct it to the target sequence and (ii) a protospacer adjacent motif (PAM) sequence linked to the 3' end of the target site to cut at the precise location (Figure 1.2b) (Makarova *et al.,* 2015; Sander and Joung, 2014). As mentioned, the cell must repair the Cas9 induced break by either homologous recombination (HR) or non-homologous end-joining (NHEJ) in order to survive.

The development of the CRISPR-Cas9 system resulted in the publication of numerous applications for *in vivo* editing in mammalian cells in early 2013 (Cong *et al.,* 2013). The first report to describe the exploitation of this CRISPR system and its applications in *S. cerevisiae* engineering was DiCarlo *et al.,* (2013). In the years afterwards, scientists have modified and refined the system such that it may be utilized for a variety of species.

Several methods for editing *S. cerevisiae* using CRISPR-Cas9 systems have been established. The bulk of them utilize different constructions to express the gRNA and Cas9 endonuclease (DiCarlo *et al.,* 2013; Stovicek *et al.,* 2015; Jakociunas *et al.,* 2015). The main aim of this review is to highlight the use of CRISPR/Cas9 in genetic manipulation of the industrially important yeast *S. cerevisiae*, as well as to detail some examples of the application of this system, ranging from simple gene knock-ins and knockouts to more complex processes such as synthetic heterologous pathway integration, protein overexpression, and stress tolerance. Additionally, novel techniques for circumventing some of the shortcomings of CRISPR-Cas9 will be discussed.

1.2. TRADITIONAL TRANSFORMATION METHODS

S. cerevisiae has in the past been engineered using several conventional techniques to generate the desired product (Duan *et al.,* 2017; Ryabova *et al.,* 2003). Most metabolic engineering approaches in *S. cerevisiae* have relied on plasmids, such as yeast episomal plasmids (YEps), or yeast integrative plasmids (YIps) (see Figure 1.3). Most *S. cerevisiae* episomal plasmids (YEps) contain a portion of the 2µ plasmid, that occurs naturally in some

strains, including the autonomously replicating sequence (ARS), five direct repeats and part of one of the inverted repeats (Misumi *et al.,* 2019). While they can yield high heterologous protein production, episomal 2µ plasmids require constant selective pressure to be maintained by the host strain, which is not economically feasible on an industrial scale.

On the other hand, yeast integrative plasmids (YIps) are introduced into the host cell genome (Manivasakam *et al.,* 1995). YIps use conventional homologous recombination, to allow site-specific integration of linear DNA segments flanked by homologous arms (Shi *et al.,* 2016; Romanos *et al.,* 1992). Once integrated, YIps are reproduced and transferred as part of a chromosome to successor cells. YIps are defined by their ability to target sequences that are homologous to chromosomal regions (Symington *et al.,* 2014; Kowalczykowski, 2015). Depending on the yeast strain and integration location, as little as 30 bp of homologous targeted sequence may be enough to generate proper transformants. However, targeted sequences are typically just a few hundred base pairs in length to ensure reliable and efficient integration. Integration happens through a single-crossover or a double-crossover recombination process, depending on the number and location of the targeted sequences (Rothstein, 1991) (see Figure 1.3b&c).

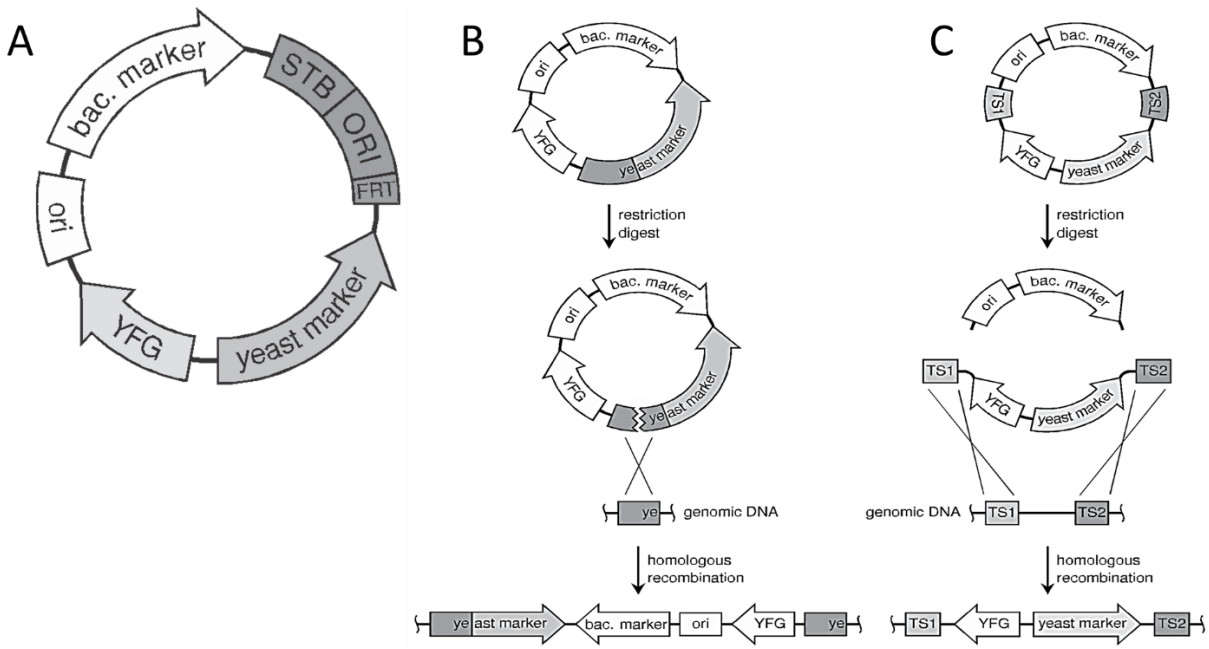


Figure 1.3. Yeast Plasmids (Gnügge and Rudolf, 2017). (a) Yeast episomal plasmid (YEp). For maintenance in yeast, 2µ plasmid-derived STB, FRT and ORI sequences are present. For maintenance in *E. coli*, both plasmid types contain a bacterial selection marker (bac. marker) and replication origin (ori). YFG, your favourite gene; the gene sequence for constructing recombinant strain. (b) Yeast integrative plasmids (YIps) that integrate via a single-crossover recombination mechanism carry a single continuous targeting sequence, which is often a part of the marker gene in case of YIps with auxotrophic markers. The YIp is linearized within the targeting sequence by restriction digest. After integration, the target site is duplicated. (c) YIps that integrate via a double-crossover mechanism contain two targeting sequences flanking the part of the vector that is to be integrated. Cutting outside the targeting sequences liberates the integrative part of the YIp. bac. marker, bacterial marker; ori, bacterial origin of replication; YFG, your favourite gene; TS, targeting sequence.

These conventional genome engineering techniques also depend largely on the use of a limited amount of selection markers for the confirmation and maintenance of integrated sequences (Duan *et al.,* 2017; Ryabova *et al.,* 2003). Additionally, conventional methods require several rounds of selection and screening to generate and identify positive clones, which takes time, leaves scars in the genome, and reduces genome stability (Da Silva and Srikrishnan, 2012; Karim *et al.,* 2013).

Targeted double-stranded breaks (DSBs) have been shown to substantially increase recombination efficiencies (Storici *et al.,* 2003; Wingler and Cornish, 2011). This discovery has prompted the adoption of different methods, such as *I-SceI* homing endonucleases, to induce

DSBs in the *S. cerevisiae* chromosome to promote homologous recombination (Wingler and Cornish, 2011). However, homing endonucleases are extremely specific and can only cut their cognate DNA target sites, limiting their flexibility (Figure 1.4) (Barzel *et al.,* 2011).

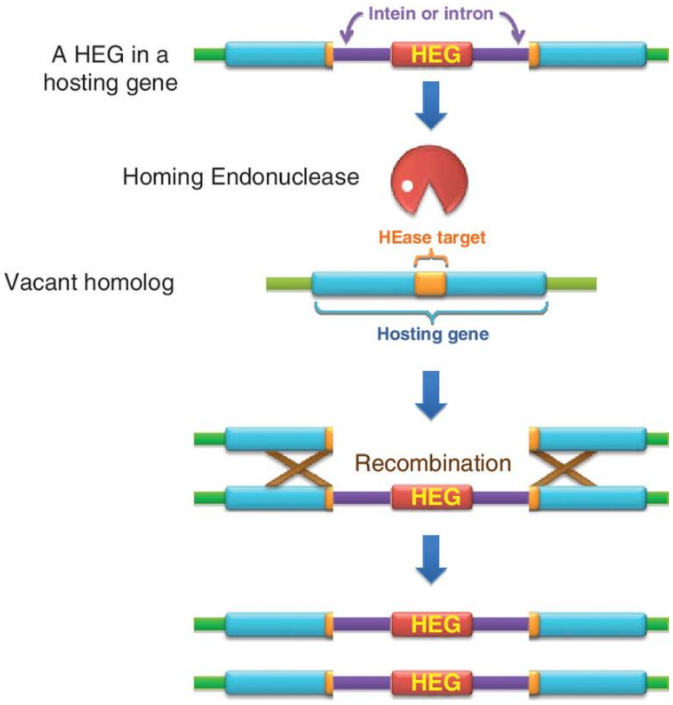


Figure 1.4: The Homing process (Barzel *et al.,* 2011). The homing endonuclease (HEase) enzyme is produced from the HEase gene (HEG) (red), which is in an intron or as an in-frame domain of an intein (purple) in a hosting gene (cyan). It cleaves the target site (orange) in the hosting gene's vacant homolog to induce homologous recombination, transforming the vacant homolog into a HEG-carrying one.

1.3. ADVANCED GENOMIC EDITING APPROACHES

Several methods of genetic editing have been explored in the past. These genetic techniques relied on the employment of endonucleases that were engineered to produce double stranded breaks (DSBs) and activate the cell repair machinery (Li *et al*., 2011). The techniques are described below, along with their benefits and drawbacks.

1.3.1. Zinc-finger nucleases & Transcription activator-like effector nucleases

Zinc finger nucleases (ZFNs) are non-specific nucleases that become specific by combining them with a DNA binding domain present in eukaryotes (Li *et al.,* 2011). These were among the first methods used for alteration of an organisms' genetic makeup. The type of DNA binding domain used is called a Zinc finger which normally comprises of ~ 30 amino acid modules that interact with nucleotide triplets. Tri-nucleotide specific zinc finger nucleases have been developed to identify all 64 potential triplet combinations, and by stringing together several zinc finger subunits, one may develop ZFNs that precisely target any DNA sequence (Li *et al.,* 2011). Each ZFN can distinguish 3-6 triplet patterns. They only operate as dimers, and thus it is necessary to utilize two pairs of ZFNs to target any locus: one that identifies the sequence upstream of the location to be changed, and the other that detects the sequence downstream of it (Urnov *et al.,* 2010).

TALENs (Transcription activator-like effector nucleases) are similar to ZFNs in that they utilize DNA binding motifs to guide the same non-specific nuclease to cleave the genome at a specific location, but each domain recognizes a single nucleotide rather than DNA triplets (Urnov *et al.,* 2010). The interactions of TALEN-derived DNA binding domains with their target nucleotides are less complicated than those of ZFNs with their target trinucleotides, therefore creating TALENs is usually easier than constructing ZFNs (Li *et al.,* 2011).

1.3.2. Advantages and Limitations

A key advantage of both techniques is that they are not limited to mutagenesis when applied to cells and tissues (Urnov *et al.,* 2010). Modifications to ZFNs and TALENs have been created to apply them in yeast, bacteria, mice, cattle, and even monarch butterflies (Li *et al.,* 2011). Desired ZFNs and TALENs may take time and resources to develop and are often difficult to construct (Urnov *et al.,* 2010). These two approaches are therefore proving impractical for metabolic engineering that often requires many changes.

1.4. THE CRISPR-Cas9 MECHANISM

In comparison to the ZFN and TALEN genetic techniques, CRISPR-associated nuclease 9 (CRISPR–Cas9) is a more recent genome engineering technology (Mali *et al.,* 2013). In the study of synthetic biology, this discovery contributed to a breakthrough because of its utilization of natural occurrences and RNA-guided nuclease activity. The bacterial CRISPR locus contains the following elements: protospacer DNA (which stores the viral DNA that invades the bacterial cell), a repetitive short palindromic segment of DNA (which separates the protospacer DNA sequences), and associated genes (which are transcribed to produce Cas9 nucleases, CRISPR RNA (crRNA), and trans-activating CRISPR-RNA) (tracrRNA)) (Stovicek *et al.,* 2015).

The *Streptococcus pyogenes* immune system type II RNA-guided endonuclease (Cas9) creates a complex with two short RNA molecules, allowing it to make precise cuts inside the viral DNA sequence (Horvath and Barrangou, 2010). A CRISPR array in the bacterial genome includes numerous distinct clusters of foreign DNAs inside its protospacer sequences DNA (see figure 1.5). CRISPR (cr) RNA is then transcribed from these clusters and is processed to make it stable before annealing to another RNA termed trans-activating (tracr) RNA. This RNA-complex then directs Cas9 to the target sequence to cause DSBs (Gasiunas *et al.,* 2012). For Cas9 to identify the exact cutting location in the genome sequence, it needs (i) a ~20-bp sequence inside the crRNA that base matches with the target genomic sequence and (ii) an NGG trinucleotide termed the protospacer adjacent motif (PAM) directly downstream of the target region on the genome (Gasiunas *et al.,* 2012). The two RNA molecules (crRNA and tracrRNA) are required for Cas9 action. However, they may be merged to form a single chimeric guide RNA termed guide RNA (gRNA), which is linked with the target sequence at its 5'-end. As a result, by altering the gRNA molecule's 5’-end, the system may be designed to target any desired sequence (Jinek *et al.,* 2012).

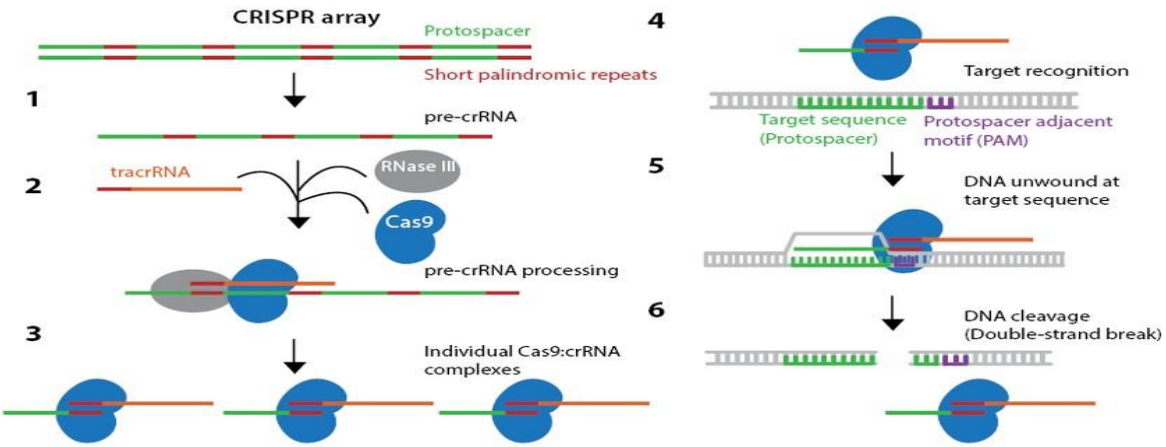


Figure 1.5: An overview of the endogenous Type II bacterial CRISPR system (Makarova *et al.,* 2015). A CRISPR array includes several distinct protospacer sequences that have similarity to foreign DNA inside the bacterial genome. Short palindromic repeat sequences divide protospacers. (1) The CRISPR array is transcribed to generate pre-CRISPR RNA (pre-crRNA). (2) A unique trans-activating crRNA (tracrRNA) with homology to the short palindromic repeat processes the pre-crRNA into separate crRNAs. The tracrRNA aids in the recruitment of the RNAse III and Cas9 enzymes, which work together to separate the various crRNAs. (3) Each distinct, unique crRNA forms a complex with the tracrRNA and Cas9 nuclease. (4) Each crRNA-tracrRNA-Cas9 complex searches for DNA sequences complementary to the crRNA. A possible target sequence in the Type II CRISPR systems is only acceptable if it includes a specific Protospacer Adjacent Motif (PAM) immediately after where the crRNA would attach. (5) Cas9 splits the double-stranded DNA target and cleaves both strands at the PAM once the complex binds. (6) Following the double-strand break, the crRNA-tracrRNA-Cas9 complex unbinds.

1.4.1. Advantages over existing approaches

The CRISPR-Cas9 technology surpasses the ZFN and TALEN mutagenesis techniques in several ways. The first is the simplicity of the target design. Because the target specificity is based on ribonucleotide complex building rather than protein/DNA recognition, gRNAs may be easily and inexpensively produced to target virtually any region in the genome (Zhang *et al.,* 2014). The second advantage is that the modifications may be delivered into the cell directly by introducing RNAs encoding the Cas9 protein and gRNA, or plasmids expressing them from appropriate expression cassettes into the cell (Stovicek *et al.,* 2015). Furthermore, the simultaneous introduction of multiple gRNAs and Cas9 nuclease can easily induce multiple DSBs, resulting in multiple instances of mutagenesis and chromosomal editing such as large deletions (Sakum *et al.,* 2014; Ota *et al.,* 2014), inversions (Li *et al.,* 2015; Ota *et al.,* 2014),

duplications (Li *et al.,* 2015), and translocations (Choi and Meyerson, 2014; Maddalo *et al.,* 2014).

1.5. CRISPR-CAS9 GENOME EDITING APPROACHES IN YEASTS

Since DiCarlo *et al.,* (2013) established CRISPR-Cas9-mediated genome editing in *S. cerevisiae* strains, a wide range of novel methods for gRNA and Cas9 expression have been developed. As mentioned previously, when the Cas9 protein and gRNA are expressed in yeast cells, Cas9 causes DSBs that must be mended by the cells to prevent cell death. This is achieved via non-homologous end joining (NHEJ) or homologous recombination (HR) (Liu *et al.,* 2015). In yeast, HR can repair DSBs with flexible donors carrying desired sequences, allowing for a variety of genetic manipulations such as gene deletion (e.g., entire coding sequence knockout) (Zhang *et al.,* 2019), gene mutation or disruption (DiCarlo *et al.,* 2013), and gene integration (Shi *et al.,* 2016; Roy *et al.,* 2018). Unlike other genome editing techniques, the CRISPR-Cas9 system eliminates the requirement for a selectable marker to be integrated with the genome edit. The two main advantages of CRISPR-Cas9 technology for yeast genome editing are thus relatively precise and flexible targeting and the removal of the requirement for positive selection.

1.5.1. *Developing CRISPR-Cas9 for use in Saccharomyces cerevisiae*

Previous research by DiCarlo *et al.,* (2013), Stovicek *et al.,* (2015), and Jakociunas *et al.,* (2015) demonstrated different methods of editing *S. cerevisiae* using CRISPR-Cas9. Most of them used different constructs to express the Cas9 nuclease and the gRNA. As a result, components were either expressed from episomal plasmids or the Cas9 expression cassette was stably integrated into the genome, with only the gRNA cassette expressed from a plasmid (Stovicek *et al.,* 2015). The following sections discuss ways in which the Cas9, gRNA and DNA repair templates have been delivered to *S. cerevisiae*.

1.5.1.1. Cas9 and gRNA delivery in yeast

The Cas9 encoding gene, which is commonly used in *S. cerevisiae* editing experiments, is derived from the bacteria *S. pyogenes* (Ryan *et al.,* 2014). For application in eukaryotes, the Cas9 is linked to a nucleolar localization sequence, which enables it to be translocated into the nucleus to access the chromosomes. Its DNA sequence is either native, yeast, or human codon-optimized (Bao *et al.,* 2015). The Cas9 gene is mostly expressed via constitutive promoters of various strengths and characteristics, using high-copy episomal vectors, low-copy centromeric vectors, or it is incorporated into the *S. cerevisiae* yeast genome (Zhang *et al.,* 2014). Cas9 toxicity has been shown in several studies, although it is readily avoided by using weaker or inducible promoters or using a 2-plasmid system where the Cas9 encoding gene is carried on a low-copy centromeric vector. The Cas9 expression system may be modified to meet the requirements of the study. For single modifications, a plasmid-based approach is preferable since the plasmid may be readily cured after the procedure. However, if numerous subsequent modifications are planned, genomic integration may provide more stable Cas9 expression.

To create effective CRISPR-Cas9 genome editing, it is essential to have an adequate understanding of the design and expression of the gRNA components (Stovicek *et al.,* 2015). In *S. cerevisiae*, a chimeric gRNA molecule is typically expressed from a high-copy vector, allowing it to be continuously strongly expressed. The ends of the gRNA molecule must be distinct to provide a functional and accurate Cas9-gRNA interaction. The production and functioning of the gRNA have been realized via the use of promoters requiring RNA polymerase III or II, depending on what the ends of the gRNA are flanked with (Gao and Zhao, 2014).

The alteration of the 20 bp target specific section of the gRNA that will transport the Cas9-gRNA complex to the target site and produce precise DSBs is required for the engineering of the target site of interest on the *S. cerevisiae* genome (DiCarlo *et al.,* 2013). Figure 1.6 depicts different methods for obtaining an expression vector containing the modified gRNA molecule with the 20 bp target specific sequence of interest (Stovicek *et al.,* 2017). Whole vector amplification is one technique that can be used to modify the target recognizing sequence of the gRNA vector (Tsai *et al.,* 2015). Here, complimentary 20 bp primers defining the target region are used to create the gRNA's new recognition sequence. Alternatively, the vector may

be circularized to allow one of three ligation techniques described in previous research (Tsai *et al.,* 2015). The first technique was a PCR method in which phosphorylated primers were employed to ligate the vector in yeast *in vivo* (Generoso *et al.,* 2016). The Gibson assembly technique, which utilizes two oligonucleotides that overlap at the target region, was the second method used. The third method was restriction-free cloning, which employs two complementary oligonucleotides , both of which contain the target sequence (Laughery *et al.,* 2015).

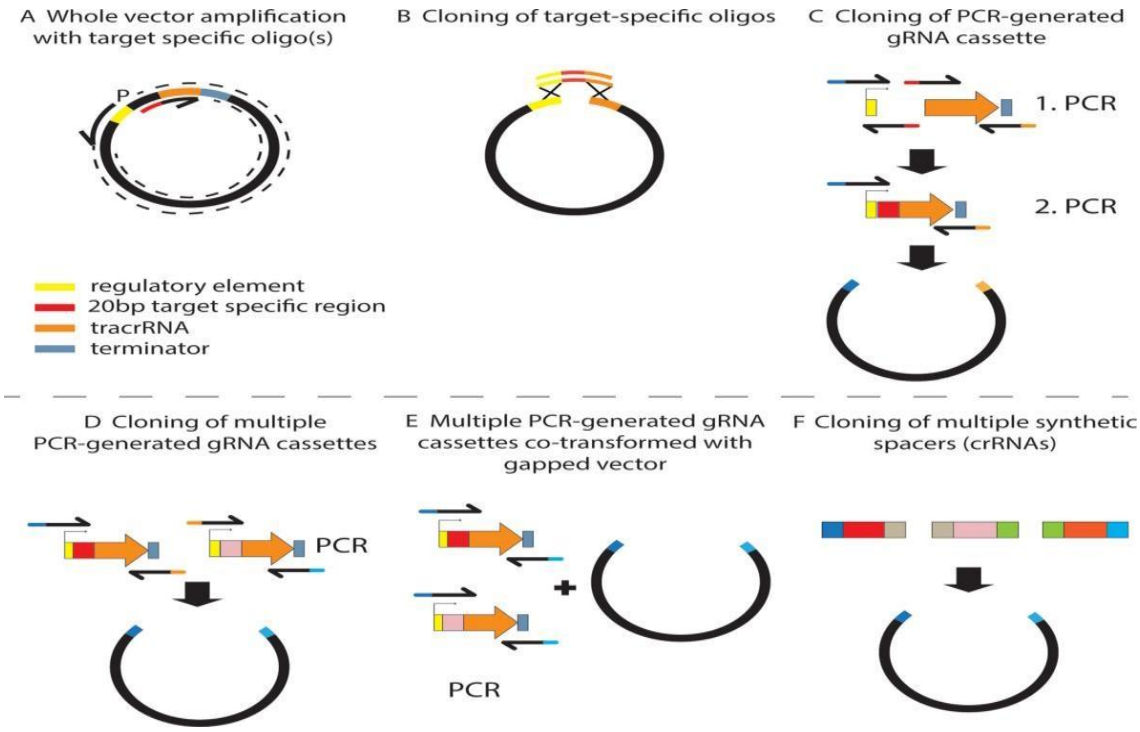


Figure 1.6: Different techniques for creating gRNA expression cassettes. (Stovicek *et al.* 2017). (A) Whole vector amplification with ligation of one phosphorylated oligonucleotide to circularize the vector (Tsai *et al.,* 2015). (B) Either restriction cloning, or Gibson assembly can be used to clone the oligonucleotides in the vector. (C) Cloning of two gRNA cassettes produced by PCR using restriction cloning or Gibson assembly. (D) Cloning of several gRNA cassettes derived from PCR using restriction cloning or Gibson assembly. (E) *In vivo* recombination of several PCR-generated gRNA cassettes with a gapped vector. (F) crRNA array cloning through Golden gate assembly of short synthetic segments with homologous overlaps (Bao *et al.,* 2015).

Generoso *et al.,* (2016) created several vectors for simultaneous expression of Cas9 and gRNA in *S. cerevisiae*, allowing for the simple and rapid removal of numerous genes at the same

time. This was accomplished by codon-optimizing the *S. pyogenes* Cas9 gene sequence and attaching a nuclear localization signal to it (Wiedemann and Boles, 2008). After that, the Cas9 was cloned into the backbone of a multi-copy vector (*pRS62K*) carrying the selectable marker gene *KanMX* enabling Geneticin G418 resistance (Farwick *et al.,* 2014). Preliminary results failed because the expression of *cas9* under the control of a strong promoter (truncated *HXT7*) impacted the rate at which *S. cerevisiae* grew. As a result, a different strategy was used, and three new promoters were evaluated (Generoso *et al.,* 2016). It was discovered that *cas9* expression under the control of a weak promoter (in this instance, *ROX3*) had minimal impact on the growth rate of *S. cerevisiae.* Following that, two CRISPR-Cas9 vectors with different gRNA content were synthesized. One vector had a single gRNA cassette, whereas the other contained two gRNA cassettes separated by a 2μ-sequence (see figure 1.7) (Generoso *et al.,* 2016). For gRNA transcriptional control, the promoter *SNR52* and terminator *SUP4* were both utilized, and the *natMX* gene resistance selectable marker (enabling CloNat resistance) was included in both CRISPR-Cas9 vectors (DiCarlo *et al.,* 2013).

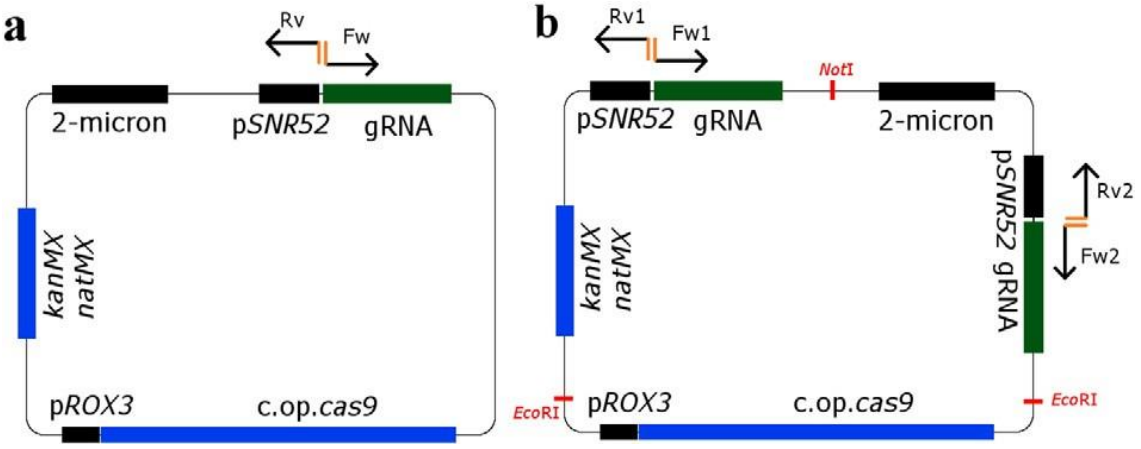


Figure 1.7: Schematic illustration of multi-copy vectors harboring codon optimized *cas9,* expressed under the control of a weak promoter (*pROX3*) and the *kanMX*/*natMX* resistance genes (Generoso *et al.,* 2016). (A) A plasmid that carries one gRNA. (B) The plasmid that carries two gRNAs. Fw1/2 and Rv1/2 are the primers required to insert the first and second gene's protospacers, respectively.

The effectiveness of the two CRISPR-Cas9 vectors was evaluated in *S. cerevisiae* by removing two gene loci (Generoso *et al.,* 2016). Protospacer sequences were chosen for detecting the

ILV1 gene and the *LEU4* gene. The protospacer sequences were cloned individually at the 5'end of the vector carrying one gRNA. However, both protospacers were cloned together in the vector carrying two gRNA for simultaneous deletion of the two genes. The vectors constructed in this study provided a 90% gene deletion efficiency, even with simultaneous recombination cloning of the plasmid and deletion of the target genes in industrial strains.

1.5.1.2. Multi-gene editing with gRNA expression

The CRISPR/Cas9 system is suitable for simultaneous multigene editing in *S. cerevisiae* because of the high HR rate (Jakociunas *et al.,* 2015; Ding *et al.,* 2020). The execution of multigene editing requires the expression of multiple gRNAs and repair templates being delivered inside the cell to fix those specific breaks defined by the introduced gRNAs. Multiplexing of gRNAs in *S. cerevisiae* has been successfully achieved using different expression vectors: (i) those with various selectable markers but carrying one gRNA, (Ryan *et al.,* 2014) (ii) a single expression plasmid containing cassettes of different gRNAs, (Mans *et al.,* 2015) (iii) vectors with an array of various interspaced gRNAs (Bao *et al.,* 2015) or (iv) a single gapped expression vector with multiple PCR-generated gRNAs for integration into the genome (see figure 1.6e) (Horwitz *et al.,* 2015). The efficiency of three different vectors with different selectable markers, each carrying two expression cassettes of gRNA in achieving gene deletions was tested after being transformed into the yeast cell. An efficiency percentage of 100% was achieved for two gene deletions, 70% was achieved for four gene deletions and 65% was achieved for six gene deletions using these expression vectors (Bao *et al.,* 2015).

Plasmids with multi-cassettes, each containing different gRNAs with unique promoters, were created by Jakociunas *et al.,* (2015). This method effectively altered five genes in a single step, resulting in a 41-fold increase in mevalonate synthesis. By combining *in vivo* assembly and targeted editing, Jakociunas *et al.,* (2015) expanded and upgraded this technique to "CasEMBLR" (Cas9 aided multi-loci DNA integration assembler); CasEMBLR enabled marker-free integration of 15 exogenous DNA components in one step. Similarly, Ronda *et al.,* (2015) used CrEdit to edit three genomic DNAs by creating three gRNAs, which allowed them to

accomplish simultaneous triple insertions of a non-native pathway for β-carotene synthesis in *S. cerevisiae* without selection, with up to 84 % targeting efficiency.

Using a single-transcript approach for producing gRNAs, homology-integrated CRISPR-Cas (HI-CRISPR) was developed with an efficiency ranging from 27 to 87 percent for disrupting three genes sequentially in the artificial hydrocortisone biosynthesis pathway (Bao *et al.,* 2015). The pre-crRNAs were produced by a single promoter and subsequently processed into various crRNAs by host RNase III and unknown nuclease(s). Recently, Ferreira *et al.,* (2018) co-expressed the bacterial endoribonuclease Csy4 along with a single transcript comprising several gRNAs coupled with Csy4-cleavable RNA, resulting in a 96% efficient quadruple deletion. Zhang *et al.,* (2019) recently created a gRNA–tRNA array for CRISPR-Cas9 (GTR-CRISPR) utilizing endogenous tRNA for gRNA processing. This technique allowed the interruption of eight genes with 87% efficiency in one step, making it the best demonstration of multigene editing reported to date. GTR-CRISPR was used to achieve a 30-fold increase in free fatty acid synthesis in only 10 days.

The yeast background strain used will be a significant factor in the success of genome editing. A single expression vector encoding hepatitis delta virus (HDV)-gRNA cassettes was utilized for gene deletions of haploid and diploid yeast strains in the research performed by Ryan *et al.,* (2014). The efficiency was 86% and 81% for three gene deletions in haploid strains, whereas in diploid strains it was 43% and 19%.

These studies demonstrated the ability to edit several genes at the same time, with varying degrees of efficiency. To ensure a high efficiency, it seems that the choice of gRNA sequences and the efficient expression of these sequences are essential, though the host strain used is also significant. As a result, it is anticipated that the analysis of gRNA design and efficient expression would significantly benefit multiple-gene modification of CRISPR-mediated techniques and applications. The use of CRISPR technologies for multi-loci editing has significantly shortened the operational timeframe of metabolic engineering workflows. For instance, the conventional approach takes around six weeks to alter three genomic loci (Horwitz *et al.,* 2015), while multiplexed CRISPR-Cas9 takes just one week with one transformation step. Furthermore, by omitting the cloning phase in *Escherichia coli*, the GTR-CRISPR produced six gene disruptions in three days (Zhang *et al.,* 2019).

1.5.1.3. DNA repair templates

Homologous repair in *S. cerevisiae* is the dominant mode for the repair of double stranded breaks when a DNA donor template is supplied to the cell in the processes (Mans *et al.,* 2015). Generoso *et al.,* (2016) discovered that short, single-stranded donor oligos with identical homology to the target region may be effectively utilized as the simplest repair template. DiCarlo *et al.,* (2013), on the other hand, demonstrated that double-stranded DNA oligos with similarity to the target region may also serve as the repair template in *S. cerevisiae*. Donor DNA lengths vary depending on the kind of modification required, ranging from a stop-codon insertion to a full heterologous biosynthetic pathway. It is critical to ensure that the DNA donor repair template used to repair Cas9 breaks must be devoid of PAM sites, to prevent Cas9 from cutting it (DiCarlo *et al.,* 2013).

The CRISPR system was also coupled with *in vivo* assembly of different DNA fragments, which eliminated the requirement for cloning procedures. For example, Tsai *et al.,* (2015) constructed *in vivo* a metabolic pathway comprising six genes with 300 bp homology arms comprised of four DNA fragments and utilized it as a DSB repair template. Apel *et al.,* 2017 utilized three DNA oligos containing a genetic pathway with 1 kb homologous arms, resulting in 40% integration efficiency in a specified region. The combination of donor *in vivo* assembly and CRISPR-Cas9 DSB repair is linked with low efficiencies, despite being flexible and quicker. Efficiencies are improved when DNA fragments are pre-assembled (Stovicek *et al.,* 2017). Plasmids with both gRNA and a DNA repair template can also be utilized (Garst *et al.,* 2017). Given that the 5' end region of gRNA or crRNA is cleaved by 5' exonuclease and other endogenous nucleases during maturation, the researchers hypothesized that this site might be used to store the HR disruption donor DNA, as shown in previous studies (Bao *et al.,* 2015). When the combined DNA components are short enough to be synthesized and incorporated into the CRISPR plasmid, this is advantageous.

In terms of the length of homology arms to attain successful recombination, the minimal homology required has been determined in studies that predate CRISPR-Cas9. Early research in genetic integration established a minimum homology of 30 bp at either end of a DNA segment for recombination (Hua *et al.,* 1997). Homology arms of greater length (200 to 1000 bp) resulted in higher recombination efficiencies, particularly when extensive heterologous pathways must be integrated.

1.5.1.4. Multicopy gene integration

Recombinant protein synthesis has often been aided by the introduction of multiple copies of a gene of interest into host cells (Malcı *et al.,* 2020). To produce industrial quantities of a recombinant protein in yeast, a large copy number of the genes encoding the desired protein should preferably be incorporated into the yeast. The advantages of utilizing genome integration rather than plasmids include increased stability of gene expression, reduced level of heterogeneity, and the elimination of the need for selective growth media after validation. To that aim, multiple-copy integration techniques, such as delta-integration and rDNA-integration, which are both simple and reliable have been utilized in *S. cerevisiae* (Malcı *et al.,* 2020).

The delta sequences on the yeast genome are the result of the yeast retrotransposon Ty, which leaves a sequence scar at the original site after transposition (Sakai *et al.,* 1990). The delta-integration technique, which allows for multi-copy integration into the repeated delta-sequences that exists in over 400 copies on the *S. cerevisiae* chromosome, has been utilized effectively for the creation of different recombinant proteins and metabolic engineering to date (Yamada *et al.,* 2010). Ribosomal DNA (rDNA) is a DNA sequence that encodes ribosomal RNA (Kobayashi and Sasaki, 2017). rDNA in yeasts is usually composed of many identical repeats grouped in a head-to-tail tandem array (Kobayashi and Sasaki, 2017; Cutler *et al.,* 2018). Each rDNA repeat is made up of two transcribed regions that code for 35S precursor rRNA and 5S rRNA, as well as two non-transcribed sections called NTS1 and NTS2. The amount and length of rDNA repeats varies across species. The rDNA locus in *S. cerevisiae* is made up of 150–200 tandem copies of a 9.1-kb rDNA repeat, providing a target site for multi-copy integration of heterologous genes.

Several researchers have conducted studies stating that multi-copy integration through delta-integration was effective for recombinant protein synthesis, and some of them have shown that recombinant proteins generated in this approach had higher expression levels (Shi *et al.,* 2016; Hanasaki and Masumoto, 2019; Huang and Geng, 2020). Conversely, it has also been reported that for delta integration, integration efficiencies and the number of integrated copies decreased rapidly with the size of the integrated donor DNA, limiting the use of this approach for integrating large biochemical pathways (Shi *et al.,* 2016).

An innovative method to multiple-copy integration of large pathways into the yeast genome was devised by Shi *et al.,* (2016). This platform's specifications were multi-copy, high efficiency, one-step and, preferably, marker-less integration. To accomplish this, they used the CRISPR-Cas9 system to precisely induce DSBs at delta regions in *S. cerevisiae* chromosomes to enhance homologous recombination efficiency (Figure 1.8). They used Di-CRISPR (delta integration CRISPR-Cas) to accomplish highly efficient and marker-less integration of extensive biochemical pathways, as well as a remarkable 18-copy genomic integration of a 24 kb DNA fragment containing a xylose utilization pathway and a butanediol synthesis pathway. The modified strain consumed more xylose and produced more butanediol than the control strain with single copy integration.

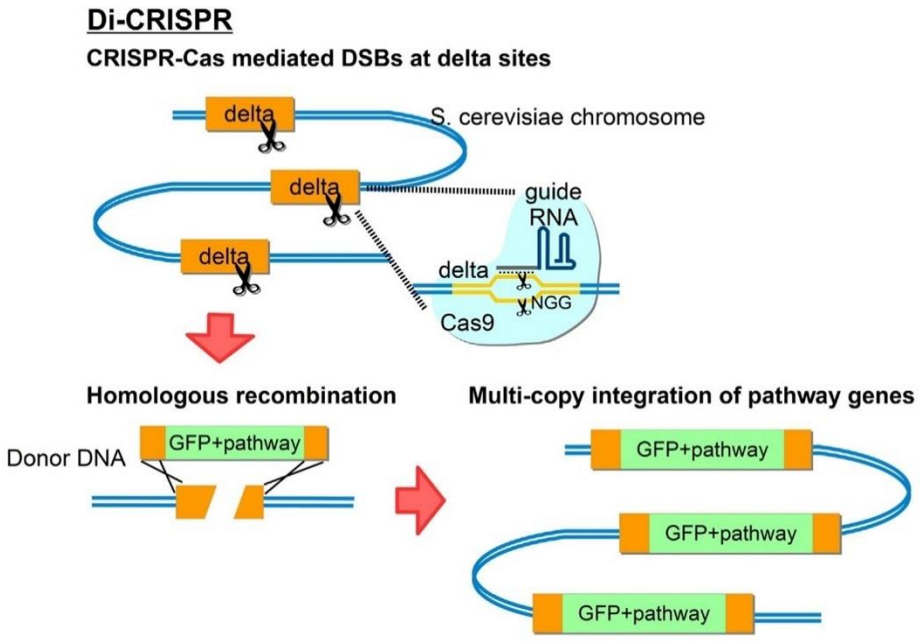


Figure 1.8: A simplified illustration of the Di-CRISPR system (Shi *et al.,*2016). CRISPR-Cas generated several DSBs at the delta sites to enable homologous recombination of metabolic processes at these locations. This system allowed the highly efficient single-step, marker-less, and multicopy chromosomal integration of entire metabolic pathways in *S. cerevisiae.*

Although the copy number of an integrative gene via delta-integration varies greatly based on its length, transformants with 1 to 80 copies of the gene of interest have been observed (Yamada *et al.,* 2010). Sasaki *et al.,* (2019) reported that the CRISPR-delta-integration technique they established was highly efficient for increasing integrated gene copy number

when compared to previous delta-integration methods, and it was also effective for increasing CMCase enzyme activity. As a result, the Di-CRISPR platform may potentially offer a better alternative to high copy plasmids for genome editing and metabolic engineering in *S. cerevisiae.*

1.5.1.5. Single chromosomal target sites

Single-chromosomal editing is becoming more significant in genome functional analysis. In the past, traditional techniques such as PCR-mediated chromosomal splitting (PCS) were employed to target specific chromosomes in the *S. cerevisiae* genome for editing (Sugiyama *et al.,* 2005). Sasano *et al.,* (2016) developed the CRISPR-PCS technique, a combination of CRISPR/Cas9 and the PCS systems, to perform simultaneous and multiple chromosomal splitting in *S. cerevisiae.* They began by attempting to create a split at position C16-P1 on Chromosome XVI (Sugiyama *et al.,* 2005). A gRNA sequence that targeted the desired split site was designed and the FY834-Cas9 strain was transformed simultaneously with dividing modules containing the *Candida glabrata LEU2* gene (*CgLEU2*) gene for position C16-P1 and a gRNA expression plasmid with a target sequence near C16-P1. A total of 680 Leu^+ transformants were produced after transformation. In comparison, traditional PCS, which used the identical transformation conditions but did not include a gRNA expression plasmid, produced just three Leu^+ transformants (Sugiyama *et al.,* 2005). They selected ten transformants from those produced using CRISPR-PCS and utilized pulsed field gel electrophoresis (PFGE) and subsequent Southern blot analysis to confirm that the desired locus was split. All transformants were split at the anticipated location with a high frequency of transformation. There was no indication of genomic rearrangements when CRISPR-PCS was used, since the PFGE band pattern remained consistent except for split products.

1.5.2. CRISPR–Cas9 genome editing in non-conventional yeasts.

Based on the success in *S. cerevisiae*, the CRISPR–Cas9 system has already been used in several non-conventional yeasts. Although there are additional optimizations to be made, this system has already displayed excellent genome editing potential in these yeasts. The CRISPR-Cas9 system's applications in several non-conventional yeasts are described further below.

1.5.2.1. Kluyveromyces lactis

K. lactis is used in industry to produce recombinant proteins, some metabolites, and fermented dairy products such as cheese (Spohner *et al.,* 2016). Horwitz *et al.,* (2015) discovered that CRISPR-Cas9 genome editing may be applied in an industrial strain of *K. lactis*. The expression vector's 2µ-element was replaced with the *K. lactis* pKD1 vector-stabilizing element. The authors inactivated the *YKU80* gene in *K. lactis*, thus decreasing NHEJ activity. While the method had a modest effectiveness (2.3%), three six-gene DNA segments were successfully integrated into three different chromosomal sites.

1.5.2.2. Komagataella phaffii (formerly Pichia pastoris)

Because of its superior protein folding and yielding abilities, *K. phaffii* (*P. pastoris*) is arguably the most important recombinant protein producing yeast (Weninger *et al.,* 2016). However, it has low homologous recombination, making genetic modification challenging. Weninger *et al.,* (2016) investigated several ways of expression of the Cas9 encoding gene and gRNA molecules in this yeast. The use of a low copy *ARS*-element vector with a bidirectional native *HXT1* promoter driving the expression of human codon-optimized Cas9 and an HH/HDV-ribozyme-flanked gRNA transcript resulted in up to 90% of single-gene knockout mutations. When two genes were targeted, knockout mutations in both ORFs were found with a frequency of 69 %. Even though a donor template with 1-kb homologous arms was available, only a very modest integration effectiveness (2%) was observed, indicating that NHEJ remained the dominating method of DSB repair in these strains of *Komagataella phaffii* (Weninger *et al.,* 2016).

1.5.2.3. *Schizosaccharomyces pombe*

The fission yeast *Sch. pombe* is an important model organism for studying eukaryotic cellular biology, particularly cell cycle regulation (Hoffman *et al.,* 2015). Jacobs *et al.,* (2014) utilized the *rrk1* promoter for gRNA molecule expression in this yeast because it provided a defined 5'-leader that is cleaved during maturation. As *rrk1* is a Pol II promoter, the 3'-end of the gRNA molecule was joined to the HH ribozyme, resulting in polyadenylation of mature RNAs. When a PCR-amplified mutant allele was employed as donor template, expression of gRNA and Cas9

independently on two low-copy vectors or simultaneously on one vector, resulted in a target modification efficiency of 85–98%. A comparable method allowed the creation of a single-gene deletion with 33% efficiency (Fernandez and Berro, 2016).

1.5.3. Drawbacks of the CRISPR-Cas 9 system

Despite its advantages in genome editing, CRISPR-Cas9 has been linked to several concerns such as the large protein size of the Cas9, restriction of target sequence due to the requirements of the PAM sequence, and unintended off-target mutagenesis (Nakade *et al.,* 2017; Ran *et al.,* 2015). As mentioned earlier, for CRISPR- Cas9 components to bind to target DNA, crRNA and the target strand of dsDNA must be base-paired, but Cas9 nuclease must also interact with a few bases of the target DNA sequence, known as a protospacer adjacent motif (PAM) (Mohanraju *et al.,* 2016). There are no evident restrictions in crRNA target sequence, however, there have been some reports of relationships between target base components and nuclease activity, which may limit the range of target DNA sequence (Doench *et al.,* 2014). More significantly, a specific PAM sequence, whose base specificity varies across the derived species, is necessary for Cas9 nuclease binding to the target sites (Gasiunas *et al.,* 2012). This limitation is not particularly restrictive, but in certain instances, this motif is difficult to find within the target genomic region, particularly when the environment is very AT-rich (Deltcheva *et al.,* 2011).

A further limitation of CRISPR-Cas9 is that there is a wide variance in efficiency when targeting various loci, perhaps due to the positional impact of the target area (Smith *et al.,* 2016). In addition, there seems to be a limit to the number of modifications that may be introduced simultaneously, as every new DSB created, reduces the total yield of surviving clones (Mans *et al.,* 2015). Moreover, CRISPR-Cas9 multiplexing leads to a substantial increase in workload to identify the required clones containing all the required genome edits.

Additionally, the engineering efficiency of different yeast strains vary (Ryan *et al.,* 2014). As previously stated, haploid strains outperformed diploid strains in terms of engineering efficiency. As a result, it may be difficult to engineer diploid or polyploidy strains, making multiplexing in industrial yeast strain backgrounds more complex and time intensive.

Researchers may choose from a variety of cloning systems to produce a target gRNA molecule and benefit from online resources that facilitate the specific cloning design (Laughery *et al.,* 2015) to simplify procedures (Ryan *et al.,* 2014). However, even in its most basic form, CRISPR-Cas9 engineering depends on the production of a gRNA vector, which may be time-consuming and expensive. Horwitz *et al.,* (2015) devised a gap repair method that avoids the cloning phase. However, it requires longer DSB repair templates, a high level of HR efficiency in the strain, and may result in non-equal molar expression of the gRNAs when multiplexing. It has also been observed that engineering using vectors based on *in vivo* assembly has a lower efficiency (Mans *et al.,* 2015).

1.5.4. Novel approaches and applications of the CRISPR-Cas9 system

The CRISPR-Cas9 system still has room for improvement in terms of target design flexibility, introduction of the CRISPR-Cas9 components into the cell, and DSB introduction specificity. Given these shortcomings, advances or alternative CRISPR-Cas9 systems are required, of which some are discussed below.

1.5.4.1. gRNA design and decreased the off-target effects.

The most difficult challenge for using the CRISPR– Cas9 system in genome editing is the off-target effect, which may result in undesired sequence cleavage (Chen, 2019). In an off-target effect, Cas9 cleaves a DNA site other than the intended target, resulting in a mutation. As a result, the possible off-target effect should be identified to improve cutting efficiency at the targeted locus.

A sequence must contain 20bp near the PAM site in the genome to be selected as the gRNA targeting sequence (Mohr *et al.,* 2016). The reason for precise gRNA selection is to reduce the likelihood of Cas9-mediated cleavage at undesired locations in the genome (off-target effects) while increasing cutting efficiency at the selected location (on-target activity) (Rainha *et al.,* 2020). The design and expression of gRNAs is a critical element that has a significant impact on gene editing efficiency (Thyme *et al.,* 2016). Ideally, multiple gRNAs should be evaluated for a new target; however, validating the target efficiency of each gRNA requires time. As a

result, the predictability must be improved further. Several web-based tools have been created to help and automate the creation of gRNA targets. CRISPy (Jako *et al.,* 2015), CRISPy-web (Blin *et al.,* 2016), CRISPR-ERA (Liu *et al.,* 2015), and CHOPCHOP v2 (Labun *et al.,* 2016) are a few examples. The primary goal of these tools is to provide guide sequences that minimize the likelihood of off-target effects by matching all potential targets within the specified parameters against the reference genome (Mohr *et al.,* 2016).

Techniques for identifying CRISPR-Cas9 off-target locations have been established and are detailed elsewhere (Martin *et al.,* 2016). Nevertheless, these types of studies are often used in genomes that are considerably more complicated, such as mammalian cells. Waldrip *et al.,* (2020) utilized cross-linking chromatin immunoprecipitation (ChIP) sequencing to identify off-target sites in *S. cerevisiae* and discovered that, as anticipated, Cas9 is extremely specific and has almost no off-target sites throughout the yeast genome.

1.5.4.2. CRISPR/Cpf1

Cpf1 is a recently described single-stranded RNA guided endonuclease with characteristics distinct from Cas9 (Table 1) (Makarova *et al.,* 2015). Both Cpf1 and Cas9 endonucleases are present in bacteria and play a role in immunological protection against invading pathogens. As previously stated, Cas9 is directed to the target sequence by two RNA transcripts, while Cpf1 is guided to the target sequence by one RNA transcript (Zetsche *et al.,* 2015). Both endonucleases induce double-stranded breaks in the DNA molecules when they reach the target sequence. In the case of Cpf1, however, the resultant breaks leave sticky ends, while Cas9 leaves blunt ends. CRISPR-Cpf1 has the potential to be the next genome editing method, addressing some of the limitations of CRISPR-Cas9 owing to its unique characteristics listed below.

Verwaal *et al.,* (2018) evaluated the functionality of three distinct Cpf1 proteins in *S. cerevisiae* for genome editing. The first protein (AsCpf1) came from *Acidaminococcus spp*., the second (LbCpf1) from a bacterium called *Lachnospiraceae*, and the third (FnCpf1) from *Francisella novicida*. The two plasmid-based genome editing method was then used in conjunction with these Cpf1-based systems along with linear donor DNA (Verwaal *et al.,* 2018). The editing efficiencies of FnCpf1 and LbCpf1 were comparable to the CRISPR-Cas9

system, while AsCpf1 editing efficiency was extremely low. Furthermore, when multiplex genome editing was used, the CRISPR-LbCpf1 system was found to be functional in yeast (Zetsche *et al.,* 2017). This research showed that CRISPR-Cpf1 increased the availability of tools for genome engineering in *S. cerevisiae*, possibly bridging problems identified in earlier studies that utilized CRISPR-Cas9.

Table 1.1: Major differences between Cas9 and Cpf1 endonucleases (Zetsche *et al.,* 2015)

Features	**CRISPR-CPf1**	**CRISPR-Cas9**
Structure	Made up of one RNA molecule (crRNA)	Made up of two RNA molecules ((crRNA and tracrRNA = gRNA)
Cutting operation	Produces sticky/ staggered ends	Produces only blunt ends
Cutting region	Cuts far from the PAM site	Cuts Near the PAM site
Target region	T-rich PAM sequence	G-rich PAM sequence

1.5.4.3. Novel CRISPR Toolkits for S. cerevisiae

Several new methods based on CRISPR-Cas9 methodology has recently been described, showing the versatility of this technology (Rainha *et al.,* 2020). One such technique, EvolvR, combines the target specificity of CRISPR-Cas9 technology with the error-prone capacity of a mutant DNA polymerase to allow for *in vivo* targeted nucleotide diversification in a single gene (Halperin *et al.,* 2018). The method utilizes a nicking Cas9 variant (nCas9) that cuts just one DNA strand which evades native homology repair. Instead, DNA polymerase uses the nick as a starting point to begin mutant base insertion. This technique was recently used in *S. cerevisiae* as yEvolvR (Tou *et al.,* 2020). The study revealed that yEvolvR could insert random mutations in both directions of the target sequence, and that it was also capable of targeting two different genomic loci at the same time. In basic eukaryotic research, such as determination of protein function or protein interactions, or in genetic processes research, where yeast is often employed as a model organism, this technique may be critical to the advancement of knowledge. Furthermore, it has the potential to be used in strain tolerance engineering for industrial applications.

To enhance the performance of industrial strains, imparting greater tolerance to an external stress, such as high temperature or oxidation through genetic engineering, is a highly valuable strategy. CRISPR has already been successfully implemented for the purpose of random mutagenesis and genome shuffling. It was shown by Mitsui *et al.,* (2019a) that CRISPR-Cas9

could be used to cleave the delta sequences of the chromosome to fragment it and that large-scale modifications, including as gene amplification, translocation, and deletion, may occur during the repair of these DNA fragments. Mitsui *et al.,* (2019a) triggered the repair of the fragmented DNA under high-temperature conditions. Upon completion of DNA repair, the modified yeast was able to thrive at 39^{o}C and demonstrated greater ethanol and acid tolerance than the original strain.

Targeted gene expression regulation is essential in both metabolic engineering and functional genomics (Stovicek *et al.,* 2017). In yeast, genetic expression is often regulated via the use of defined gene promoters of varying strengths; nevertheless, predicting the quantity of expression remains a challenging task. A mutant Cas9 enzyme, produced by Qi *et al.,* (2013), was found to have mutations in the nuclease active site that classify it as a "dead Cas9" (dCas9). Apart from lacking DNA cleavage activity, this dCas9 mutant is still capable of functioning as a simple specific DNA binding complex. The use of this version of Cas9 to target a coding region in the *Escherichia coli* genome resulted in transcriptional gene silencing. The dCas9 protein binds to the target sequence and inhibits the activity of RNA polymerase. CRISPR interference (CRISPRi) is the term used to describe this method.

The same research group demonstrated remarkable success with the system for gene suppression in *S. cerevisiae* (Gilbert *et al.,* 2013). In this instance, the dCas9 was directed to a specific promoter, resulting in effective gene silencing. Additionally, by attaching a transcriptional repressor domain to dCas9, the suppression effect was increased. Farzadfard *et al.,* (2013) conducted an alternative study in which they linked dCas9 to an activator domain and discovered that it either activated or repressed the gene depending on the targeting site. When the target was outside the TATA box, CRISPR activation (CRISPRa) was observed and when the target was near to the TATA box, gene suppression occurred (CRISPRi). Additional dCas9 proteins with activator domains have been generated, with greater degrees of regulation (Chavez *et al.,* 2015).

Zalatan *et al.,* (2015) developed a new approach for up/downregulating a target gene. Instead of including fusion domains into dCas9, the researchers incorporated effector protein recruitment RNA domains into the gRNA, thus turning it into a scaffold RNA. The scaffold RNA hairpins may attract a single RNA-binding protein, such as an activator or a repressor, which

may subsequently be used to regulate gene expression at a specific locus. Additionally, Jensen *et al.,* (2017) evaluated two other CRISPR/dCas9-mediated systems: one that employed inducible gRNA expression and CRISPR/dCas9 fusions with either the repressor or activator domain, and another that used constitutive scaffold RNA synthesis to recruit effector molecules. When employed at single or multiplex level, the two approaches achieved similar modifications in the activation or repression of the targeted promoters.

Deaner and Alper, (2017) also discovered that genetic expression may be regulated at different levels of intensity. The positioning of dCas9-based regulators with regards to the core of the promoter was shown to be correlated with the array of genetic expression levels. The application of grade regulation marks a notable advancement in the ability to regulate metabolic pathways in *S. cerevisiae*. CRISPRi/a has a wide range of potential applications, and it has already been utilized to enhance a yeast cell factory that produces β-amyrin (Yu *et al.,* 2018).

Apart from using a modified CRISPR protein, another technique for guiding effector molecules without generating a DSB is to change the length of the gRNA molecule. Truncated gRNAs, which are normally 14 nucleotides in length, have been demonstrated to be capable of directing Cas9 to the target sequence without yielding a double-strand break (DSB) (Kiani *et al.,* 2015; Dahlman *et al.,* 2015). This allows for the use of short gRNAs for transcriptional control while also enabling the use of long gRNAs for genome editing via the use of a single Cas9 protein. These characteristics enabled researchers to create versatile systems capable of CRISPRi, CRISPRa, or CRISPR editing in a single step. Lian *et al.,* (2017) developed the first trifunctional CRISPR system in *S. cerevisiae* for simultaneous gene inactivation, activation, and editing, termed CRISPR-AID. They achieved this using shorter gRNAs for CRISPRa and CRISPRi. However, the utilization of three different PAM-recognizing CRISPR proteins was essential to avoid gRNAs competing for the same Cas9 protein.

Dong *et al.,* (2020) developed a CRISPR system that also used a single Cas9 protein to execute three distinct functions. The method, which was designated CRISPR-ARE, made use of a Cas9 gene that had been fused to a VP64-p65-Rta (VPR) activation domain. The authors showed the application of CRISPR-ARE by improving the production of α-santalene. The method utilized shortened gRNAs to activate or repress genes as required, and full-length gRNAs to

edit a single gene by co-transforming the donor repair DNA. The editing efficiency was 100%, and the presence or absence of gene activation or repression was confirmed using a reporter protein. α-Santalene biosynthesis, was improved by 2.66-fold in the resultant strain.

1.6. ADVANCED CRISPR-Cas9 APPLICATIONS IN YEAST

1.6.1. Metabolic engineering

As described above, CRISPR-Cas9 has been successfully applied to yeasts for the silencing and activation of gene transcription (Stovicek et *al.,* 2017). For the development of yeast cell factories through metabolic engineering, controlled gene regulation is among the most essential aspects to consider. The process of creating a strain is often done via many repeated build-test-learn cycles, which the CRISPR method can facilitate in a flexible multiplex way. It enables simple multiplexing of genes without interfering with the intrinsic gene transcription mechanism (Stovicek *et al.,* 2017). Figure 1.9 shows four examples of how the CRISPR-Cas9 (multiplex technique) was effectively utilized for yeast cell factory creation.

Shi *et al.,* (2016) genetically modified *S. cerevisiae* using CRISPR-Cas9 with the aim of attaining the non-native product (R, R)-2, 3- butanediol (BDO) from a non-native substrate (xylose) in a single round of transformation (Figure 1.9a). A 24-kb integration construct encoding six gene expression cassettes was integrated (three for the xylose consumption pathway and three for the BDO biosynthesis pathway) to the delta sequences, allowing (R, R)-2, 3- butanediol production from xylose as carbon source (Shi *et al.,* 2016).

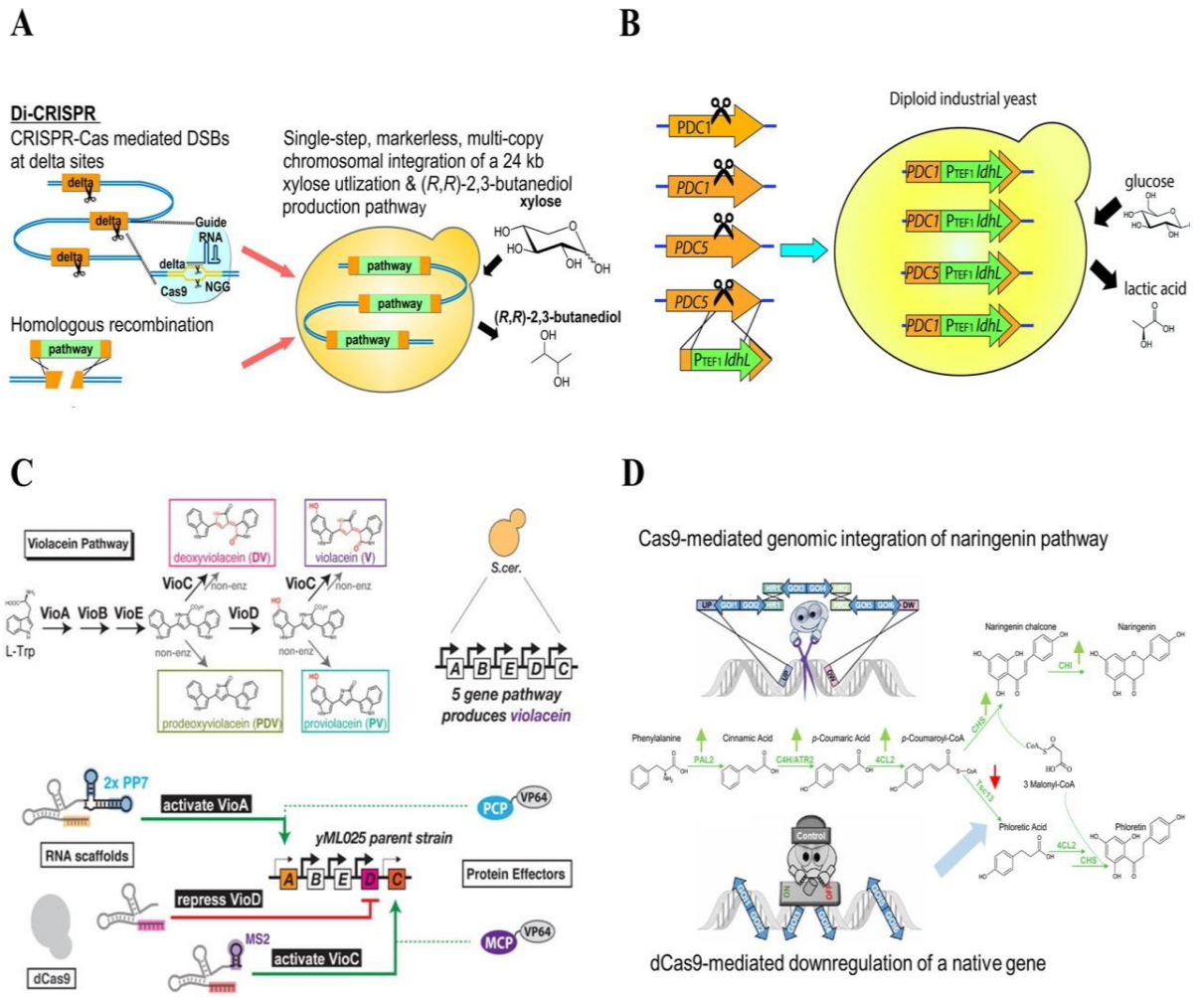


Figure 1.9. Application of CRISPR-Cas9 systems for engineering of yeast cell factories. (Stovicek *et al.,* 2017) (A) Production of (R, R)-2, 3-butanediol from xylose. Multicopy one-step integration of the xylose utilization and (R, R)-2, 3-butanediol pathways into delta sites in the genome (Shi et *al.,* 2016). (B) Production of lactic acid from glucose in an industrial yeast strain, one-step disruption of two genes in diploid strain and simultaneous integration of lactate dehydrogenase genes from *L. plantarum* (ldhL) (Stovicek *et al.,*2015). (C) Production of deoxyviolacein, violacein, prodeoxyviolacein and proviolacein from glucose (Zalatan *et al.,* 2015). (D) Production of naringenin from glucose. (Vanegas *et al.,* 2017).

The diploid industrial *S. cerevisiae* strain Ethanol Red, which is used in many first-generation ethanol plants, was engineered to produce lactic acid by replacing both alleles of the pyruvate decarboxylase genes *PDC1* and *PDC5* with the L-lactate dehydrogenase encoding gene (*ldhL*) from *Lactobacillus plantarum* (Figure 1.9b) (Stovicek *et al.,* 2015). The genetic modification

was accomplished in a single transformation step, generating a strain that produced 2.5 g/L lactic acid with a yield of 0.49 g of lactic acid per gram of (Stovicek *et al.,* 2015).

The synthesis of the bacterial pigment violacein in *S. cerevisiae* was optimised by transcriptional regulation using the CRISPR system (Figure 1.9c) (Zalatan *et al.,* 2015). CRISPR RNA scaffolds were utilized to attract transcriptional activators and repressors, either alone or in combination, to promoter sites, allowing for precise control of transcriptional activation and repression in the target genes. Altering the RNA scaffolds of the same strain may allow it to generate various ratios of the pathway products deoxyviolacein, violacein, pro-deoxyviolacein, and proviolacein, all without changing the strain's genetic makeup. Combining these RNA-encoded circuits with conditional expression of Cas9, resulted in the development of a mechanism for transitioning from the growth to the production phase (Zalatan *et al.,* 2015).

S. cerevisiae was genetically engineered to produce the flavonoid precursor naringenin (Figure 1.9d), demonstrating the effectiveness of a combined Cas9 genome editing and Cas9 transcriptional regulation strategy (Vanegas *et al.,* 2017). To begin, Cas9 was utilized to integrate a multi-gene pathway into an intragenic locus, leading to the synthesis of naringenin from phenylalanine. Following that, naringenin synthesis was boosted through dCas9-mediated silencing of *TSC13* to avoid the creation of the by-product phloretic acid (Vanegas *at al.,* 2017).

1.6.2. Overexpression of proteins

S. cerevisiae cell factories are used to commercialize a broad variety of recombinant proteins (Wang *et al.,* 2017). To accomplish this objective, it is critical to develop efficient and cost-effective techniques for recombinant protein production. The delta-integration method has been successfully applied to a number of metabolic engineering efforts and the production of a wide variety of recombinant proteins (Yamada *et al.,* 2010). Additionally, it has been shown that pre-breakdown of this insertion site by the CRISPR system improves the efficiency of gene integration through HR. The CRISPR-delta-integration method was examined by Sasaki *et al.,* (2019), who found that the copy number of delta-integrated *TrEG* (encoding *Trichoderma*

reesei endoglucanase II) was significantly enhanced. This method resulted in the integration of up to 40 copies of the 3.5 kb donor DNA, and the CMCase activity in the culture media of transformants after 168 hours was determined to be 559 U/L. When compared to previous delta-integration studies, the integrated gene copy number rose by 2.9-fold. Additionally, this research demonstrated a 17.3-fold increase in CMCase activity when compared to that produced by *S. cerevisiae* in a prior study using an episomal (YEp) vector (Qi *et al.,* 2013). They also demonstrated that this method could be easily applied to the expression of various recombinant protein types, such as pharmaceutical proteins. In this way, it is a technology that has great potential for future study and other uses.

1.7. Objectives of the Study

The conversion of lignocellulosic biomass to biofuels via microbial fermentation is a promising alternative for economically and sustainably replacing petroleum-based production. This prompted the use of *S. cerevisiae* in our laboratory to accomplish the goal of producing biofuels from lignocellulosic substrates in a CBP configuration. We chose this microbe due to its adaptability to genetic manipulation systems and robustness in the production of biofuels. (Borodina and Nielsen, 2014; Lynd *et al.,* 2005; van Zyl *et al.,* 2007; Yamada, Hasunuma and Kondo 2013; Den Haan *et al.,* 2015). However, significant challenges have been reported in the efficient microbiological conversion of lignocellulosic biomass by modified *S. cerevisiae*. These include heterologous cellulolytic enzyme production, co-fermentation of hexose and pentose sugars, and resistance to a variety of stresses. As a result, an effective engineering strategy is needed to produce CBP *S. cerevisiae* strains capable of using cellulose and cellobiose (a glucose dimer).

CRISPR-based transformation strategies are gaining traction in yeast metabolic engineering, and our lab has only recently established the technology. In earlier studies, students in our group have integrated genes into one of two loci, in laboratory and industrial yeast strains. To construct a CRISPR-based yeast CBP strain, we require information on the following: i) the effect of integrating at different sites on gene expression/protein production; ii) the effect of adding additional copies of a gene at a different site; and iii) the efficacy of multi-locus targeting to delta sequences via CRISPR. Thus, this study will offer insight on how we can

effectively use the CRISPR-Cas9 technology to express multiple genes with few rounds of transformation in yeasts intended for cellulose CBP.

The purpose of this study was to optimize the CRISPR tool for heterologous cellulase expression in *S. cerevisiae* laboratory and industrial strain transformants. To accomplish our aim, we pursued the following objectives:

- Designed PCR primers to construct CRISPR-Cas9 gRNA integration cassettes, targeting heterologous genes to repeated sequences in the genome, namely the delta sequences.
- Designed PCR primers to construct CRISPR-Cas9 gRNA integration cassettes, targeting a single genetic locus (on chromosome X, XI or XII) in the yeast genome.
- Designed PCR primers to create expression cassettes (promoter-gene-terminator) that allow targeted integration to the sites of DNA breaks provided by the Cas9 nuclease (homology repair partners for gene integration). The reporter genes used were *T. reesei eg2* (*T.r.eg2*) and *Talaromyces emersonii* (now named *Rasamsonia emersonii*) *cel7A* (*T.e.cbh1*).
- Transformed all the CRISPR elements to achieve heterologous gene integration to the target sites through homologous recombination.
- Validated transformants and perform comparative enzyme assays.
- Validate integrated gene copy numbers with qPCR.
- Build a putative CBP yeast strain for biofuel production with *T.r.eg2*, *S.f.BGL1* and *T.e.cbh1*, using the CRISPR tool.

CHAPTER 2

MATERIALS AND METHODS

2.1. PLASMIDS, MICROBIAL STRAINS, AND PRIMERS USED IN THE STUDY

Tables 2.1 and 2.2 summarize the origins and details of the plasmids and *S. cerevisiae* strains used in this study. All primers used in the study including their names, sequences, annealing temperatures and their applications are described in table 2.3.

Table 2.1: Description of Plasmids used in this study

PLASMIDS	DESCRIPTION	REFERENCE
pRDH180	Also abbreviated as *eg2* plasmid, carrying the *ENO1* promoter, terminator and *T.r.eg2* (Used to produce the PCR product carrying the *eg2* cassette)	Brevnova *et al.,* 2011
pMI529	Also abbreviated as *cbh1* plasmid, carrying the *ENO1* promoter, terminator, and *T.e.cbh1* (Used to produce the PCR product carrying the *cbh1* cassette)	Ilmén *et al.,* 2011
pMUSD1	Also abbreviated as *bgl1* plasmid, carrying the *ENO1* promoter, terminator, and *S.f.BGL1* (Used to produce the PCR product carrying the *bgl1* cassette)	Davison *et al.,* 2019
pCas9NAT	*CEN6/ARS4* plasmid, *TEF1* promoter, *CYC1* terminator, SV40 Nuclear Localization Sequence, human codon optimized *S.p.Cas9;* CloNAT resistance (Low copy plasmid carrying the *cas9* encoding gene in the 2-plasmid CRISPR system)	Addgene
pRS42-G_ ChX	Guide RNA expression plasmid for the 2-plasmid system targeting Chromosome X intergenic region; G418 resistance; contains the *SNR52* promoter and *SUP4* terminator for gRNA expression	This lab
pRS42-G-DELTA	Similar to pRS42H_ChX but targeting the yeast DELTA sequences	This lab
pRSCG_ChXI	pRS423-cas9-gRNA-G418 targeting Ch. XI intergenic region protospacer; G418 resistance; this plasmid also contains the *S.p.Cas9* under *TEF1* promoter and *CYC1* terminator – 1-plasmid system.	This lab
pRSCG_ChXII	Similar to pRSCG_ChXI but targets a Ch.XII intergenic region protospacer	This lab

Table 2.2: Description of yeast strains used in this study

STRAIN	Abbreviation	DESCRIPTION	REFERENCE
S. cerevisiae MH1000	MH1000	Industrial yeast strain, diploid, no auxotrophy	Davison *et al.,* 2016
S. cerevisiae M1744	M1744	Lab yeast strain, auxotrophic for uracil (Δ*ura3*)	Brevnova *et al.,* 2011
S. cerevisiae MH1000 + pCas9	MH1000-Cas9	MH1000 strain containing pCas9NAT	Thompson, 2017
S. cerevisiae M1744 + pCas9	M1744-Cas9	M1744 strain containing pCas9NAT	Thompson, 2017
S. cerevisiae Y294 + pRDH147::fur1	Y294-*eg2-Pos+*	*S. cerevisiae* Y294 with pRDH147, *FUR1* disrupted, *ENO1* promoter and terminator, *T. reesei eg2*, *S. cerevisiae fur::LEU2*	Brevnova *et al.,* 2011
S. cerevisiae Y294 + pMI529::fur1	Y294-*cbh1-Pos+*	*S. cerevisiae* Y294 with pMI529, *FUR1* disrupted, *ENO1* promoter and terminator, *T.e.cbh1*; *S. cer fur1::LEU2*	Ilmén *et al.,* 2011
S. cerevisiae Y294 + ySFI::fur1	Y294-*bgl-Pos+*	*S. cerevisiae* Y294 with ySFI, *FUR1* disrupted, *ENO1* promoter and terminator, *S.f.bgl1* (cel3A); *S. cer fur1::LEU2*	Den Haan *et al.,* 2007
S. cerevisiae M1744 + pCas9 **+** pRS42-G_ChX + *T.r.eg2*	M1744-CH10-EG2	*S. cerevisiae* M1744 with the *T.r.eg2* integrated at the chromosome X intergenic site using pCas9NAT and pRS42H_ChX)	This study
S. cerevisiae M1744 + pRSCG_ChXI + *T.r.eg2*	M1744-CH11-EG2	*S. cerevisiae* M1744 with the *T.r.eg2* integrated at the chromosome XI intergenic site using pRSCG_ChXI)	This study
S. cerevisiae M1744 + pRSCG_ChXII + *T.r.eg2*	M1744-CH12-EG2	*S. cerevisiae* M1744 with the *T.r.eg2* integrated at the chromosome XII intergenic site using pRSCG_ChXII)	This study
S. cerevisiae MH1000 + pCas9 + pRS42-G_ChX + *T.r.eg2*	MH1000-CH10-EG2	*S. cerevisiae* MH1000 with the *T.r.eg2* integrated at the chromosome X intergenic site using pCas9NAT and pRS42H_ChX)	This study
S. cerevisiae MH1000 + pRSCG_ChXI + *T.r.eg2*	MH1000-CH11-EG2	*S. cerevisiae* MH1000 with the *T.r.eg2* integrated at the chromosome XI intergenic site using pRSCG_ChXI)	This study
S. cerevisiae MH1000 + pRSCG_ChXII + *T.r.eg2*	MH1000-CH12-EG2	*S. cerevisiae* MH1000 with the *T.r.eg2* integrated at the chromosome XII intergenic site using pRSCG_ChXII)	This study

S. cerevisiae M1744 + pCas9 + pRS42-G_ ChX + *T.e.cbh1*	M1744-CBH-CH10	*S. cerevisiae* M1744 with the *T.e.cbh1* integrated at the chromosome X intergenic site using pCas9NAT and pRS42H_ChX)	This study
S. cerevisiae M1744 + pRSCG_ChXI + *T.e.cbh1*	M1744-CBH-CH11	*S. cerevisiae* M1744 with the *T.e.cbh1* integrated at the chromosome XI intergenic site using pRSCG_ChXI)	This study
S. cerevisiae M1744 + pRSCG_ChXII + *T.e.cbh1*	M1744-CBH-CH12	*S. cerevisiae* M1744 with the *T.e.cbh1* integrated at the chromosome XII intergenic site using pRSCG_ChXII)	This study
S. cerevisiae M1744 + pRS42-G-DELTA + *T.r.eg2*	M1744-Δ-EG2	*S. cerevisiae* M1744 with the *T.r.eg2* integrated at delta site in the genome using pCas9NAT and pRS42G-DELTA)	This study
S. cerevisiae M1744 + pRS42-G-DELTA + *T.e.cbh1*	M1744-Δ-CBH1	*S. cerevisiae* M1744 with the *T.e.cbh1* integrated at delta site in the genome using pCas9NAT and pRS42-G-DELTA)	This study
S. cerevisiae MH1000 + pRS42-G-DELTA + *T.e.cbh1*	MH1000-Δ-EG2	*S. cerevisiae* MH1000 with the *T.e.cbh1* integrated at delta site in the genome using pCas9NAT and pRS42-G-DELTA)	This study
S. cerevisiae MH1000 + pRS42-G-DELTA + *T.e.cbh1*	MH1000-Δ-CBH1	*S. cerevisiae* MH1000 with the *T.e.cbh1* integrated at delta site in the genome using pCas9NAT and pRS42-G-DELTA)	This study
S. cerevisiae MH1000 + (*T.r.eg2* + *T.e.cbh1* + *S.f.bgl1*)	MH1000-CBP	*S. cerevisiae* MH1000 transformed with three cellulase genes (*T.r.eg2*; *T.e.cbh1*; and *S.f.bgl1*)	This study

2.2. MICROBIAL STRAIN CULTIVATIONS

All plasmids (Table 2.1) were propagated using *Escherichia coli* DH5α (Thermofisher) and were cultured overnight at 37°C on LB agar (0.5% yeast extract, 1% tryptone, 1% sodium chloride, and 2% agar) containing 100 µg/ml ampicillin. The bacterial colonies from overnight incubation were inoculated on TB media (2.4% yeast extract, 1.2% tryptone, and 0.4% glycerol) containing 100 µg/ml ampicillin and incubated at 37°C on a rotary wheel overnight, prior to plasmid DNA extraction. The yeast strains listed in Table 2.2 were obtained from glycerol stocks and cultivated on YPD (1% yeast extract, 2% glucose, 2% peptone, and 2% agar

when required) supplemented with 100 µg/ml CloNAT (WernerScientific) and/or 200µg/mL Geneticin (G418) disulphate (Invitrogen) as needed, at 30°C for 2-3 days.

2.3. PLASMID PREPARATION AND PCR AMPLIFICATION OF THE REPAIR TEMPLATES

All plasmids were extracted using the ZymoPure™ Plasmid Maxiprep kit (Zymo Research) as directed by the manufacturer. For all PCR analyses performed, Taq DNA Polymerase Master Mix RED (Ampliqon) was used according to the manufacturer's instructions in an Applied Biosystems thermocycler. The plasmids pRDH180, pMI529, and pMUSD1 were used to amplify the homology repair templates for *Trichoderma reesei eg2* (*T.r.eg2*), *Talaromyces emersonii* (now called *Rasamsonia emersonii*) *cbh1* (*T.e.cbh1*), and *Saccharomycopsis fibuligera CEL3A* (*S.f.bgl1*), respectively. Each of these plasmids, which were used to amplify a specific repair template, was assigned specific primers for amplification of the target gene (see Table 2.3). The PCR reactions for the templates pRDH180 (Eg2 plasmid) and pMI529 (CBH1 plasmid) was conducted as follows: an initial denaturation step at 95°C for 5 min, followed by 31 cycles of denaturation at 95°C for 30s, annealing at 58°C for 30s, and extension at 72°C for 2 min 45 45s; a final extension step of 7 min at 72°C was allowed. For pMUSD1(BGL1 plasmid) the conditions were optimized and set as follows: an initial denaturation step at 94°C for 5 min, followed by 31 cycles of denaturation at 94°C for 30s, annealing at 55°C for 30s, and extension at 72°C for 3 min 10s; a final extension step of 7 min at 72°C was allowed. PCR products were resolved by gel electrophoresis to confirm the amplification. The PCR products were then purified using standard Phenol: Chloroform: Isoamyl alcohol PCI extraction, followed by quantification on a Nanodrop2000 Spectrophotometer (ThermoScientific) to determine the concentration of the purified PCR product to be used for transformation.

Table 2.3: Description of the primers that were used in the study.

Primer Name	Primer sequence (in 5'-3' direction)	T_A used	Application
CHX_ENO-L	GCAGTTATCTCTGTGTCCAGATCCCTTTGA AGTAAAGTTTATTCAATTTTCTTCTAGGCG GGTTATCTACTG	57°C	Amplification of the gene (*T.r.eg2*) flanked by *ENO1* promoter and terminator to provide
CHX_ENO-R	CTACAGTAATTGTGCGGTGCAGGGAGGC AATGTTTAGTGCATCTCCTCCGTCGAACA		

	ACGTTCTATTAGG		ChX homology 3'and 5' ends
ChrXI_ENO-L	TGTAAAACAGGTATTGGCTGCTTCATAGTA CACCCAATTGCTTCTAGGCGGGTTATCTACTG	55°C	Amplification of the gene(*T.e.cbh1*) flanked by
ChrXI_ENO-R	GCAACTCTGAAATGTCAAAACGGTCGTGTATA AATAAATGCCGTCGAACAACGTTCTATTAGG		*ENO1*promoter and terminator to provide CHXI homology 3'and 5' ends
ChrXII_ENO-L	GCGTCCTACAGCGTGATGAAAATTTCGCCTGC TGCAAGATCTTCTAGGCGGGTTATCTACTG	53°C	Amplification of the gene(*S.f.bgl1*) flanked by *ENO1* promoter
ChrXII_ENO-R	CTGTCAAACTTCTGAGTTGCCGCTGATGTGACA CTGTGACCCGTCGAACAACGTTCTATTAGG		and terminator to provide CHXII homology 3'and 5' ends
DELTA-ENO1-L	CTTAAGATGCTCTTCTTATTCTATTAAAAATAGA AAATGACTTCTAGGCGGGTTATCTACTG	53°C	Amplification of the gene (*T.r.eg2*) flanked by *ENO1* promoter and terminator with
DELTA-ENO1-R	GTTTGTTTGCGAAACCCTATGCTCTGTTGTTCGG ATTTGACGTCGAACAACGTTCTATTAGG		DELTA homology 3'and 5' ends
ENO1-L	GTAACATCTCTCTTGTAATCCCTTATTCCTTCTAGC	57°C	Confirmation of the *eg2* cassette in
ENO1-R	GCAACCCTATATAGAATCATAAAACATTCGTGA		transformants
EGR-Rev	ATCTGGATTAGTAACTTGAGACAAAGCAG	59°C	To confirm presence of *eg2*, use with ENO1-L
CBH1R-Rev	TGTTGAGAGAAGTCGTCGGTGTCAC	59°C	To confirm presence of *cbh1*, use with ENO1-L
BGLR-Rev	GGTTCATCATGTAAGAGTTTTCGC	63°C	To confirm presence of *bgl1*, use with ENO1-L
Ch.10Check-L	GCAGTTATCTCTGTGTCCAGATCC	57°C	To confirm if genes
Ch.11Check-L	GCCTTCGATTTGACACATCTCTAAGC	55°C	were integrated into
Ch.12Check-L	GCCATTGAGTCAAGTTAGGTCATCC	53°C	the correct CRISPR
DeltaCheck-L	CTGTTGGAATAAAAATCCACTATCGTC	53°C	targeted site, use with ENO1-R
ALG9L	TGCATTTGCTGTGATTGTCA	60°C	qPCR primers for amplification of α-1,2-mannosyltransferase gene (ALG9) in the yeast genome as internal reference gene.
ALG9R	GCCAGATTCCTCACTTGCAT		

Eg2_L	TCTGCTGCTGCTTTGTCTCAAG	60°C	qPCR primers for amplification of
Eg2_R	CTCAACCAAGTAGCCAATGGAG		*T.r.eg2*
CBH_L	TCTAACAACGCTAACACTGGCA	60°C	qPCR primers for amplification of
CBH_R	TAAGTACCACCACAGTCATCGC		*T.e.cbh1*

2.4. YEAST TRANSFORMATION

All yeast transformations were carried out using the electroporation methods described by Cho *et al.,* (1999), with minor modifications to increase yeast cell permeabilization and thus improve transformation efficiency (Moriguchi *et al.,* 2016). Briefly, harvested cells were washed with deionized distilled water, followed by resuspension in LiOAc/TE (0.1 M LiOAc, 10 mM TrisHCl pH 8.0, and 1 mM EDTA) solution. Resuspended cells were then incubated at 30°C for 45 minutes, prior to the addition of 20 µL 1M DTT and further incubation with gentle shaking for 15 minutes at the same temperature. The mixture was then centrifuged, and cells were washed with deionized distilled water, followed by resuspension in the electroporation buffer (1 M sorbitol, 20 mM HEPES). Competent cells were transformed with ~10 µg repair template DNA and 1 µg CRISPR plasmid DNA under standard conditions (1.4 kV, 200 ohms, 25 µF) using a micropulser (BioRad). Following electroporation, cells were resuspended in 1 ml YPD broth media supplemented with 1M sorbitol, followed by overnight incubation at 30°C on an orbital shaker at 180 rpm. The transformation mixture was plated on YPDS solid media supplemented with CloNAT (100 µg/ml) and Geneticin (G418) (200 µg/ml) or Geneticin only as required for 2-3 days at 30°C. For the 2-plasmid system, *cas9* carrying yeast strains were first created by transforming with the pCas9NAT (Table 2.2). These were subsequently transformed with the plasmid containing the gRNA cassette and the relevant repair template. For the 1-plasmid system, *cas9*-free MH1000 and M1744 strains were transformed using the plasmids bearing both the cas9 and gRNA cassettes as well as the relevant repair template (Table 2.1).

2.4.1. Single and Multi-Copy gene integration.

Yeast strains were transformed with homology repair templates (*eg2* cassette / *cbh1* cassette), the pCas9-NAT plasmid, and a CRISPR plasmid as required targeting a specific intergenic region on Chromosome 10 (pRS42-G_ChX), Chromosome XI (pRSCG_ChXI), or

Chromosome XII (pRSCG_ChXII) for single gene integration (Table 2.4). The CRISPR plasmid pRS42-G-DELTA (targeting delta repeated sequences) was used to achieve multi-copy integration. Webtool E-CRISP was used to identify the protospacer sites (Heigwer *et al.,* 2014).

After sub-cultivation on selective plates, positive transformants were isolated and inoculated on YPD liquid media supplemented with CloNAT (100 µg/ml) and Geneticin (G418) (200 µg/ml) for strains transformed via the two-plasmid system, and with only Geneticin (G418) (200 µg/ml) for strains transformed via the one-plasmid system for further screening.

Table 2.4: Gene integration target sites on different chromosomes

Chromosome sites	**gRNA targeting sequence (5`-3`)**
Chr. X	GTAGCTACAAGAACATATGG
Chr. XI	GCACCTCTAAAACTGCTCCG
Chr. XII	GTCACTGACAGCCACCGCAG
Delta	GGAATATTGGGTCAGATGAA

2.4.2. CBP yeast construction

The preceding procedure was repeated, but this time for the purpose of introducing three genes (*T.r.eg2*, *T.e.cbh1*, and *S.f.bgl1*) into a diploid yeast strain (MH1000-Cas9) in successive rounds of transformation. The *T.r.eg2* gene was initially integrated into MH1000-Cas9 using the above-mentioned procedure with pRS42-G-ChX. The confirmed transformants were streaked onto selective YPD media containing 100 µg/ml CloNAT to maintain the Cas9 plasmid in the yeast while eliminating the gRNA plasmid, followed by overnight incubation at 30^{o}C. The procedure of sub-cultivation was repeated for six days. After day 6, colonies from the original selective plates and the day 6 re-streaked plate were streaked on a new selective plate (containing G418) to confirm the loss of the G418 selective gRNA plasmid in the sub-cultured strain. Following curing of the gRNA plasmid and maintenance of the Cas9 plasmid, a second gene (*T.e.cbh1*) was transformed into this strain using the same method as previously described, but this time with a different gRNA (pRS42-G-DELTA). After confirming *T.e.cbh1* integration with PCR, the Cas9 and gRNA plasmids were cured by re-streaking the transformants on non-selective YPD media to eliminate both plasmids. Following plasmid curing, a third gene (*S.f.bgl1*) was transformed into the same strain using pRSCG_ChXI, to target gene integration to the chromosome 11 intergenic site.

2.6. CMC PLATE SCREENING

The *T.r.eg2* transformants were inoculated on YPD liquid media at 30^{o}C overnight. Cultures were then spotted on CMC solid media (containing: 1% carboxymethyl cellulose (CMC), 1% yeast extract, 2% glucose, 2% peptone, and 2% agar) and incubated at 30^{o}C for 24 hours. Following incubation, the plate was stained with 0.1% Congo Red for 30min, followed by washing with 1.2M $NaCl_2$.

2.7. PCR CONFIRMATION OF GENE INTEGRATION AND POSITIONING

Hoffman and Winston's (1987), total yeast DNA extraction method was used on randomly selected yeast transformants. Following that, PCR confirmation of the transformants was performed using the extracted DNA as a template. Taq DNA Polymerase Master Mix RED was used as directed by the manufacturer, using annealing temperatures indicated in table 2.3. Different primers were used to confirm the presence of each transformed gene as well as the position of each gene in the genome. Table 2.3 contains detailed descriptions of the primers used to confirm the presence and position of each gene in the relevant genomic locus.

2.8. ENZYME ASSAYS

Transformed strains were cultivated in triplicate in 10 ml YP media supplemented with 2% glucose and cultivated at 30˚C for 48-72 hours, on an orbital shaker at 180 rpm for all assays. The OD_{600} readings were taken at a 48- and 72-hour cultivation times for all the respective strains and these values were used to calculate the dry cell weight (DCW) for the strains (Meinander *et al*., 1996). The cells were removed by centrifugation and supernatant was used for determination of the endoglucanase (EG) and cellobiohydrolase (CBH) activity, while for β-glucosidase (BGL) the total culture was used for determination of the activity.

The dinitrosalicylic acid EG assay was performed, as previously described by La Grange *et al.*, (2001) using 1% CMC (carboxymethylcellulose, Sigma Aldrich) as substrate. A standard curve was set using D-glucose at concentrations of 0.5-10 g/L (see appendix, supplementary figure S1). For all enzyme assays one unit was defined as the amount of enzyme required to produce one µmol of reducing sugar or equivalent in one minute under the assay conditions. All spectrophotometric readings for the enzymatic assays were taken at 540nm on a

SPECTROstar (BMG LABTECH) plate reader and media blanks were included. Two reference strains, one with multiple copies of the *T.r.eg2* gene and the other with no cellulase encoding gene was included in all assays to serve as positive and negative controls, respectively. Assays were performed in biological and technical triplicates and values are given as averages of these repeats with standard deviation indicated.

The CBH activity of transformants was also evaluated at 48 hours and 72 hours and the activity was measured on the soluble fluorescent substrate 4-methylumbelliferyl-β-D-lactoside (MULac; Sigma) using the method previously described by Ilmén *et al.,* (2011) with a reaction time of 30 minutes at 37 °C and compared to a 4-methylumbelliferone (MU) standard curve set between 0.63 µM and 20 µM (supplementary figure S2). Liberation of MU was detected by fluorescence measurement (excitation wavelength = 355nm, emission wavelength = 460 nm) with a FLUOstar Omega Microplate Reader (BMG LABTECH). To evaluate the BGL activity of the recombinant strains, assays were carried out using *p*-nitrophenyl-β-D-glucopyranoside (pNPG; Sigma) as substrate, at 50°C with reaction times of 30 minutes as was previously described (Van Zyl, *et al.,* 2014). The absorbance of 100uL of the supernatant of the assay mixture was measured at 400nm using a FLUOstar Omega Microplate Reader (BMG LABTECH) and compared with a standard curve set between 0.075 and 1.25 mM *p*NP to the determine how much of the substrate was cleaved (supplementary figure S3). For all cellulase activities, one unit of enzyme was defined as the amount required to release one µmol of reducing sugar or equivalent from the substrate under assay conditions.

2.9. qPCR GENE COPY NUMBER ANALYSIS

The *T.r.eg2* and *T.e.cbh1* genes integrated during transformation were quantified using real-time quantitative PCR (qPCR) by Inqaba Biotechnical Industries (Pty) Ltd (Pretoria, South Africa). Primers specific for amplification of the target genes were used to determine the copy number of each cellulase expression cassette. The gene encoding α-1,2-mannosyltransferase (*ALG9*) was selected to normalize the copy number of our genes of interest, since it is present as a single copy in the haploid and as two copies in the diploid complement of the *S. cerevisiae* genome (Teste *et al.,* 2009). All DNA concentrations measurements were carried out using the ND-1000 Spectrophotometer NanoDrop (Thermo-Fischer Scientific). A standard curve was generated using a serial dilution of the pRDH180 (for *T.r.eg2* gene) and pMI529 for (*T.e.cbh1*

gene) from 1 ng to 0.1 fg and with the parental strain for *ALG9* from 10 ng to 0.1 pg (see appendix 2 for details). qPCR was then performed in 96 well plates with Luna Universal qPCR Master Mix (New England Biolabs, Ipswich, MA, USA) using a dye-based qPCR assay. Each reaction contained 1 μl of DNA template, 0.25 μm forward and reverse primers and 1X Luna Universal qPCR Master mix. The reactions were run on a CFX96 Real-Time PCR System (Bio-Rad) following a standard two-step PCR program as suggested by the Luna Universal qPCR Master Mix manual. Three technical replicates were run for each DNA sample. Amplification of different input templates were evaluated based on quantification cycle (C_q) value. Following that, the absolute copy number was calculated using a formula (Absolute copy number = DNA (g) / (g to bp const. x genome size). The average C_q values were plotted against the absolute copy number of standards and standard curves were generated by a linear regression of the plotted points. Absolute copy number for the strains was calculated based on the standard curves (appendix 2). The efficiency of the PCR was determined using the formula (E = 10-1/slope – 1) and the efficiency of all primer pairs used was over 90%. Standard melting curve analysis was performed to check the specificity of the qPCR products.

2.10. SDS- PAGE

Gel electrophoresis was used to investigate the secretion of heterologous CBH enzymes in the supernatant as previously described (Ilmén *et al*., 2011). The strains were cultured for 72 hours in double strength SC medium, after which the OD_{600} value of each culture was determined. These were used to standardize the protein concentrations in the supernatant against the OD_{600} values of the individual cultures. These steps were taken to ensure that differences in culture density did not affect the comparison of proteins from different strains on the gel. Following that, 20 μl extracellular protein fractions were analysed using 10% sodium dodecyl polyacrylamide gel electrophoresis (SDS-PAGE), according to the Laemmli method (1970). Silver staining was used to visualize the separated protein bands (Kroukamp *et al.,* 2013). Protein deglycosylation was performed on samples using the Endo H kit (New England Biolabs) as directed by the manufacturer.

CHAPTER 3

RESULTS AND DISCUSSION

3.1 CONSTRUCTION AND EVALUATION OF THE RECOMBINANT YEAST STRAINS MADE VIA CRISPR-CAS9

To investigate the efficiency of targeting integration into the *S. cerevisiae* genome, we chose to transform haploid and diploid yeast strains using 2 different CRISPR systems. The first approach was termed the two-plasmid system, in which Cas9 was expressed from a constitutive promoter on a low copy *ARS4*/*CEN4*-based vector, while the gRNA that targets Cas9 to the chosen integration site was expressed from an episomal 2µ-based vector (Figure 3.1a) (Zhang *et al.,* 2014). The one-plasmid system was used in the second approach, in which Cas9 and gRNA were both expressed from a multi-copy plasmid (Figure 3.1b) (Van Wyk *et al.,* 2020). The second approach was chosen because of its versatility and recyclability, as well as its ability to reduce the number of transformation rounds and selectable marker usage. The CRISPR systems were designed to introduce double-strand breaks (DSBs) into specific chromosome intergenic sites, allowing yeast to repair the breaks via homologous recombination. This was accomplished by using a donor DNA that included the reporter genes under the *ENO1* gene promoter and terminator, flanked by sequences homologous to either side of the targeted chromosome intergenic site (Figure 3.1c&d).

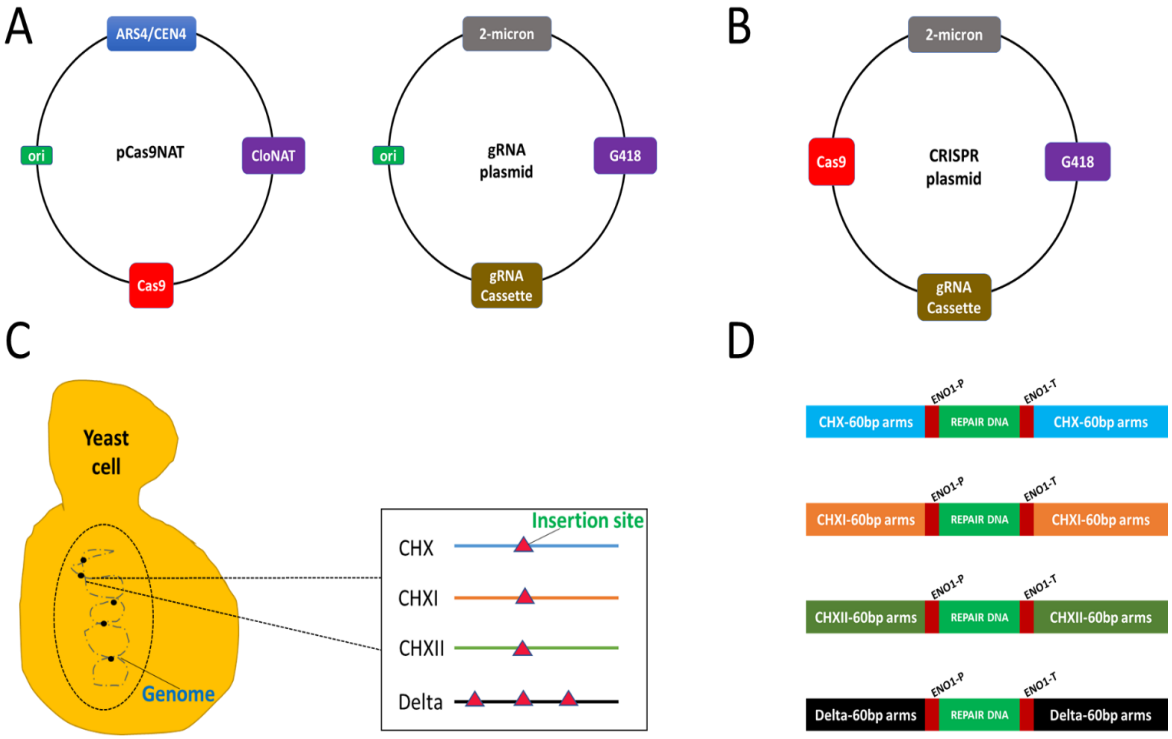


Figure 3.1: A schematic overview of the CRISPR-Cas9 system used to transform a diploid industrial and a haploid laboratory *S. cerevisiae* strain. (A) Schematic illustration of the two-plasmid CRISPR system. A low copy replicative (*CEN*/*ARS*-containing) plasmid contained the Cas9 encoding gene and the selection marker for CloNAT resistance, and a multi-copy plasmid (2µ) contained the gRNA expression cassette and the selection marker for G418 resistance. (B) Schematic illustration of the one-plasmid CRISPR system. A multi-copy (2µ) plasmid contained the Cas9 encoding gene, a gRNA cassette, and a selection marker for G418 resistance. (C) The CRISPR system targeted chromosomal intergenic sites for integration of different genomic repair expression cassettes for gene editing. "Delta" represents the repeated Ty delta elements dispersed in the yeast genome that allows multi-target integration. (D) The repair DNA cassettes contained reporter genes flanked by the *ENO1* promoter, *ENO1* terminator, and 60-bp homology arms, targeting integration to the various genomic sites.

The sequence of the donor DNA flanked by homology arms was evaluated using a sequence analysis tool to ensure that the recognition sequence and PAM site (protospacer-adjacent motif, i.e., the three nucleotides required for Cas9 recognition) was not present. This prevents Cas9 from cutting the donor DNA sequence after successful integration (Ronda *et al.,* 2015). In this study, the PAM site was located within a section of the genome that is deleted upon successful integration of the two interspaced homology arms into the targeted sites. Because the PAM sequence is deleted upon successful integration, we anticipated that this would have an additional beneficial effect on obtaining correct transformants, as Cas9 continues to cut in

cells where integration was not successful. Failing to repair the DSB is lethal for the cells (DiCarlo *et al.,* 2013). Successfully transformed yeast strains acquired specific antibiotic resistance and a heterologous gene integrated into their genome (Table 3.1). To ensure that the correct transformants were obtained, all transformants were re-streaked on YPD selective plates and random colonies were selected for further screening of the integrated reporter genes. Prior to PCR validation, the EG2 yeast transformants were first spotted on 1 % CMC agar plates to check for the production of the heterologous enzyme indicated by clearing halo formation (Figure 3.2a). The halo formation on CMC agar plate results from cleavage of CMC into fragments smaller than cellohexaose (an oligosaccharide, consisting of six glucose residues, formed by hydrolysis of cellulose) to which Congo Red does not bind (Sazci *et al.,* 1986).

Table 3.1: The following table summarizes the transformed yeast strains constructed in this study using the two- and one-plasmid systems.

Name of the yeast transformant	CRISPR plasmid system	Gene(s) integrated	Targeted site	Selection
Haploid strains:				
M1744-CH10-EG2	Two	*T.r.eg2*	10	CloNAT & G418
M1744-CH11-EG2	One	*T.r.eg2*	11	G418
M1744-CH12-EG2	One	*T.r.eg2*	12	G418
M1744-DELTA-EG2	Two	*T.r.eg2*	Delta	CloNAT & G418
M1744-CH10-CBH1	Two	*T.e.cbh1*	10	G418
M1744-CH11-CBH1	One	*T.e.cbh1*	11	G418
M1744-CH12-CBH1	One	*T.e.cbh1*	12	G418
M1744-DELTA-CBH1	Two	*T.e.cbh1*	Delta	CloNAT & G418
Diploid strains:				
MH1000-CH10-EG2	Two	*T.r.eg2*	10	CloNAT & G418
MH1000-CH11-EG2	One	*T.r.eg2*	11	G418
MH1000-CH12-EG2	One	*T.r.eg2*	12	G418
MH1000-DELTA-EG2	Two	*T.r.eg2*	Delta	CloNAT & G418
CBP yeast construction				
MH1000-EG2	Two	*T.r.eg2*	10	CloNAT & G418
MH1000-EG2+CBH1	Two	*T.r.eg2* *T.e.cbh1*	10 & Delta*	CloNAT & G418
MH1000-CBP	Two Two One	*T.r.eg2,* *T.e.cbh1* *S.f.bgl1*	10, Delta* & 11*	CloNAT & G418 CloNAT & G418 G418

* Prior to these construction steps, the plasmids containing CRISPR machinery were cured from the transformants

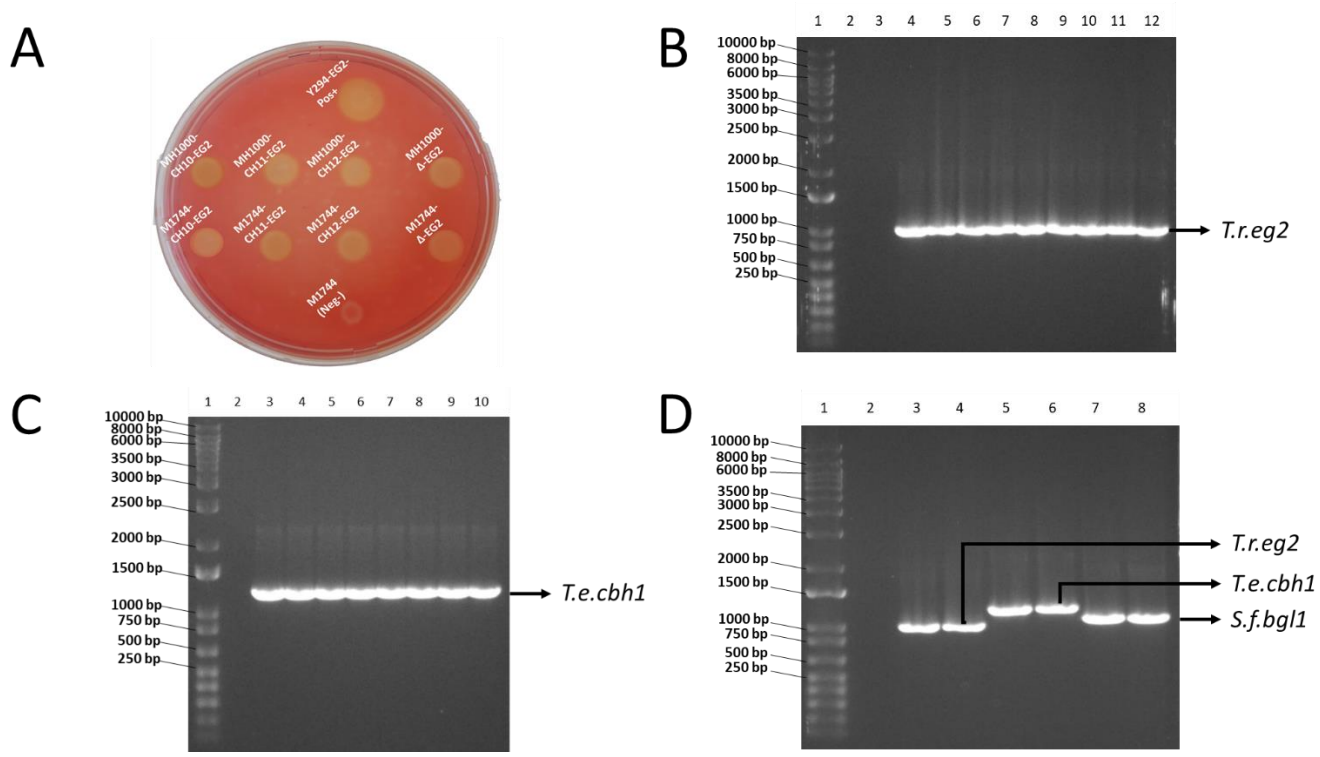


Figure 3.2: Confirmation of the cellulase genes integrated into haploid and diploid *S. cerevisiae* isolates. (A) Screening of *T.r.EG2* activity in EG2 yeast transformants on a 1% CMC agar plate. The plate was stained with 0.1% Congo Red, and the generated halos represented the EG2 active transformants. Y294 +pRDH147::fur1 and M1744 were used as positive and negative controls, respectively. This plate represents an example as several transformants for each integration locus was screened (B) Electrophoresis of *T.r.eg2* PCR products from CMC selected yeast transformants on a 1% agarose gel. Lane 1: 1kb Plus DNA Ladder (Invitrogen); Lane 2 and 3: M1744 and MH1000 (negative controls), respectively; Lane 4: Positive control (pRDH180); Lane 5 to 12: *T.r.eg2* yeast transformants. (C) Electrophoresis of *T.e.cbh1* PCR products from selected yeast transformants on a 1% agarose gel. Lane 1: 1kb Plus DNA Ladder (Invitrogen); Lane 2: M1744 (negative control); Lane 3: Positive control (pMI529); Lane 4 to 10: *T.e.cbh1* yeast transformants. (D) Electrophoresis of three distinct PCR products from the CBP MH1000 strain on a 1% agarose gel. Lane 1: 1kb Plus DNA Ladder (Invitrogen); Lane 2: MH1000 (negative control); Lane 3: pRDH180 (*T.r.eg2* positive control), Lane 4: *T.r.eg2* in CBP MH1000; Lane5: pMI529 (*T.e.cbh1* positive control); Lane 6: *T.e.cbh1* in CBP MH1000; Lane 7: pMUSD1 (*S.f.bgl1* positive control); Lane 8; *S.f.bgl1* in CBP MH1000

Following the successful screening of the *T.r.eg2* gene in EG2 yeast transformants (Figure 3.2b), we used PCR to confirm the presence of the *T.e.cbh1* gene in all CBH1 yeast transformants (Figure 3.2c). Finally, we PCR validated the presence of three genes (*T.r.eg2*,

T.e.cbh1 and *S.f.bgl1*) that were transformed in a single yeast strain (CBP MH1000) to ensure that the reporter genes were present and stable (Figure 3.2d). All the colonies that showed appropriate bands in the agarose gel electrophoresis were then picked and cultivated in flasks and their activities were subsequently evaluated through liquid assays.

3.2 ENDOGLUCANASE INTEGRATION AND ACTIVITY

3.2.1 Single locus gene integration

As a proof of concept for the CRISPR-Cas9 system's applicability, we targeted single locus integration of *T.r.eg2* to different chromosomal sites in diploid and haploid *S. cerevisiae* strains, to determine which sites are most suitable for integrating the endoglucanase gene. After the *T.r.eg2* gene was confirmed to be present in the transformed MH1000 and M1744 strains via PCR, strains were cultured in Erlenmeyer flasks for 72 hours to determine their endoglucanase activities, as shown in figure 3.3. The parental strain was included as negative control and resulted in no activity, as expected, since the *T.r.eg2* was not integrated. It was shown that both haploid and diploid *S. cerevisiae* strains had low activities when the gene was targeted to chromosomes 10 and 11, compared to chromosome 12. At 72 hours of cultivation, the highest activity for haploid M1744 strains was 160U/gDCW for the chromosome 12 targeted strain, and the lowest activity was 40U/gDCW for the chromosome 10 targeted strain. The highest activity observed in diploid MH1000 strains was 48U/gDCW for chromosome 12 targeted strain and the lowest activity observed was 18U/gDCW for the chromosome 11 targeted strain. Based on these findings, the haploid M1744 strain also had higher EG activities in all the targeted chromosome sites than the diploid MH1000 strain.

Due to the differences in activity observed, we were interested in determining the number of *T.r.eg2* copies integrated into the genome of each transformant in order to disambiguate locus and copy number effects on expression level. This was determined using qPCR (Table 3.2). We expected the *S. cerevisiae* M1744 transformants to have one copy of the *T.r.eg2* gene because they are haploid with only one set of chromosomes and thus one target site. Since MH1000 *S. cerevisiae* strains are diploid and have two sets of chromosomes, we expected

them to have at least two copies of the *T.r.eg2* gene. For targeting chromosomes 10 and 11, qPCR results matched our exceptions as we confirmed one copy of *T.r.eg2* in the M1744 yeast strains and two copies in the MH1000 yeast strains. Unfortunately, we were unable to determine the copy number of *T.r.eg2* genes targeted to chromosome 12 from the qPCR. Evaluating the position of integration in these strains using PCR (not shown) also indicated possible off-target integration.

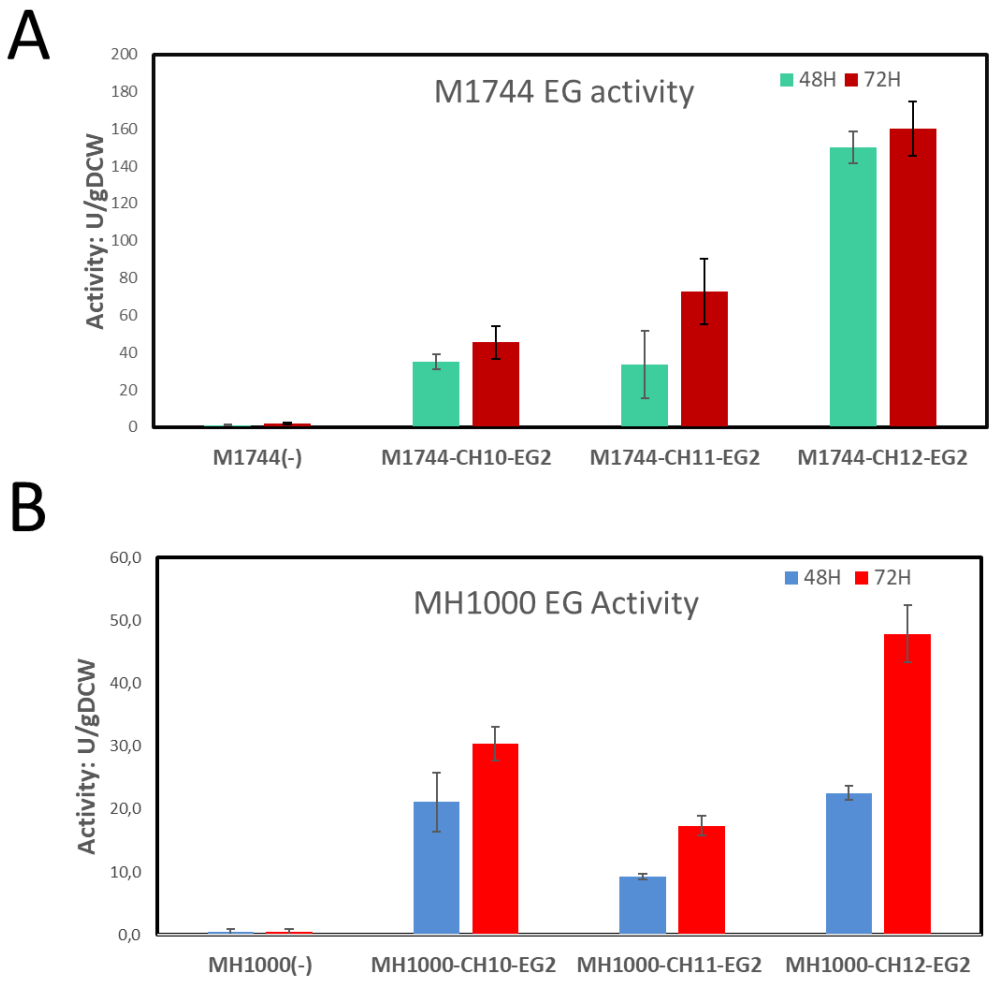


Figure 3.3: Enzyme activity profiles of recombinant yeast strains after 48 and 72h cultivation. (A) Activity of EG2 producing haploid strains on CMC. (B) Activity of EG2 producing diploid strains on CMC. Values obtained were normalized with the dry cell weight of each specific yeast strain. The M1744 and MH1000 parental strains were used as negative control reference strains. Values given are the mean values of enzyme assays conducted in triplicate. Error bars indicate standard deviation from the mean value for each strain.

Ronda *et al.,* (2015) previously reported that expression cassettes may integrate elsewhere in the genome via break-induced replication (BIR), resulting in strains where the gene of interest is difficult to localize. This is a known side-effect of some protospacer sequences in CRISPR applications. Despite using the E-CRISP predictive tool to avoid this problem, it is clear that it still occurred for the chromosome 12 targeting site, though not for targeting to the chromosome 10 or 11 sites.

Table 3.2: A summary of the copy numbers of the *T.r.eg2* cassettes integrated at different chromosomal sites.

Stains	EG2 copy number
M1744-CH10-EG2 (Haploid)	1
M1744-CH11-EG2	1
M1744-CH12-EG2	3
M1744-DELTA1-EG2	1
M1744-DELTA2-EG2	1
MH1000-CH10-EG2 (Diploid)	2
MH1000-CH11-EG2	2
MH1000-CH12-EG2	N.D.
MH1000-DELTA1-EG2	3
MH1000-DELTA2-EG2	1

CRISPR-Cas9 has also been used in different strain backgrounds for targeted single gene integration in *S. cerevisiae* via homology-directed repair of double-strand breaks (DSBs) with short oligonucleotides as repair donors (DiCarlo *et al.,* 2013; Bao *et al.,* 2015; Ryan *et al.,* 2014, Jakociunas *et al.,* 2015; Mans *et al.,* 2015; Stovicek *et al.,* 2015). Ryan *et al.,* (2014) successfully integrated a three-part DNA assembly into a single chromosomal locus, and Mans *et al.,* (2015) performed a complete deletion of the *ACS2* locus in conjunction with a six-part DNA assembly that resulted in the deletion of the *ACS1*. In comparison, our CRISPR systems achieved integration efficiencies of 55–65 % for transformants expressing *T.r.eg2* at four different sites in the yeast genome. We observed, in agreement with previous reports, that the DSB created by the guide RNA-targeted Cas9 endonuclease was critical for correct

integration at a significantly higher efficiency than endogenous homologous recombination alone.

3.2.2 Multi-locus gene integration

Following the single locus integration of *T.r.eg2*, we targeted the delta-sequences for multi-copy integration of the gene in both haploid and diploid *S. cerevisiae* strains. Our CRISPR-delta-integration method involved creating DSBs in the delta-sequences on the yeast chromosome by the CRISPR system, followed by integration of *T.r.eg2*, to hopefully increase the integrated gene copy number. After confirmation of transformation (similar to single gene integration), EG enzyme activities of the delta-integrated transformants were determined (Figure 3.4).

As expression levels can vary significantly between different chromosomal regions (Flagfeldt *et al.,* 2009), we tested 8 positive transformants for each strain background to account for possible clonal variation, due to the different positions in the genome where genes targeted to delta sequences may have integrated. Our CRISPR-delta-integration method was successful in integrating the *T.r.eg2* into *S. cerevisiae* haploid and diploid strains. The EG activities of the M1744 and MH1000 after 48 and 72 hours were determined, as shown in figure 3.4. At 72 hours of cultivation the best performing haploid M1744-Δ-EG strain had an EG activity of 65 U/gDCW, while the best diploid MH1000-Δ-EG strain had an activity of 110 U/gDCW. The lowest EG activity observed in the haploid M1744 strains was 15 U/gDCW, while the lowest EG activity observed in the diploid MH1000 strains was 36 U/gDCW. Based on these findings, the diploid MH1000 strain outperformed the haploid M1744 strain in terms of EG activity levels when the gene was targeted to the delta-sequences. It is also clear that significant clonal variation was evident in these transformants. Interestingly, single gene targeting to chromosomes 10 and 11 in the haploid M1744 background led to similar and even higher EG activities compared to the delta-targeted strains. However, in the diploid strain background, EG activities of 3- to 5-fold higher were observed for some of the delta-targeted transformants, compared to their chromosome 10 of chromosome 11 targeted counterparts. We used qPCR to determine the number of *T.r.eg2* copies integrated into the delta sites. Table 3.2 shows that the M1744 strains that integrated *T.r.eg2* at the delta-

sequences only contained 1 copy, while the highest copy number observed among MH1000 strains tested was 3 copies. As a diploid strain, MH1000 may be expected to offer more potential sites for *T.r.eg2* integration than the haploid M1744. The copy number data is thus in agreement with the observed higher levels of activity for the MH1000-DELTA1-EG2 strain compared to its single integration counterparts in MH1000 and the delta-targeted M1744 transformants. However, the different levels of activity of the single copy transformants, and the various delta-targeted transformants show that the locus of integration also has a significant effect on activity.

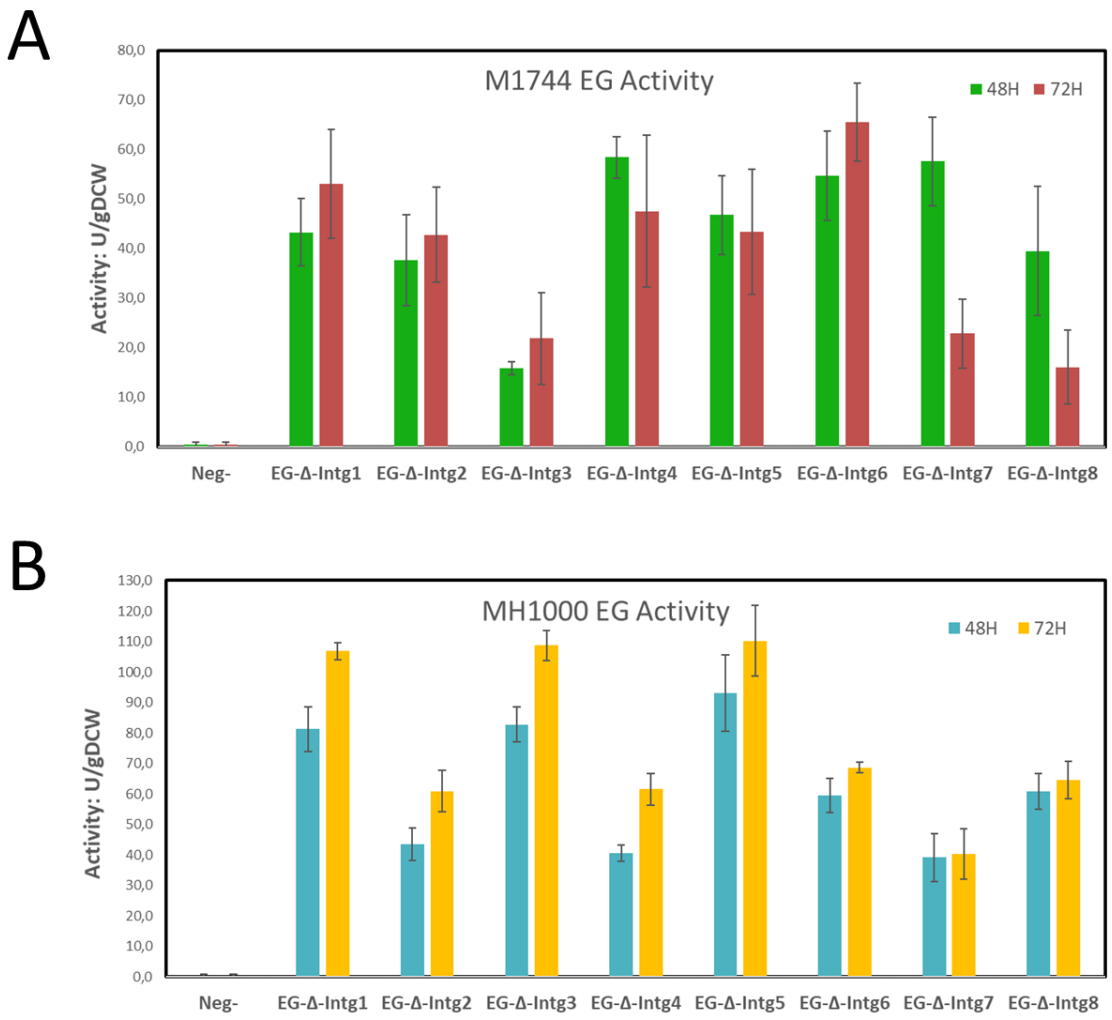


Figure 3.4 Enzyme activity profiles of recombinant yeast strains after 48 and 72h cultivation. (A) Activity of EG2 producing haploid strains and (B) diploid strains on CMC. Values obtained were normalized with the dry cell weight of each specific yeast strain. Values given are the mean values of enzyme assays conducted in triplicate. Error bars indicate standard deviation from the mean value.

Numerous researchers have demonstrated that multicopy integration via delta-integration is an effective method for producing recombinant proteins, demonstrating increased expression levels of recombinant proteins using this method (Mitsui *et al.,* 2019b). Although the number of integrative gene copies produced by delta-integration varies significantly according to its length, 1 to 80 copies have been reported. (Yamada *et al.,* 2010). Sasaki *et al.,* (2019) integrated 40 copies of a 3.5 kb donor DNA sequence into the delta-sequence using the CRISPR-Cas9 system of *S. thermophilus*. Shi et *al.,* (2016) also reported the breakdown of the delta-sequence by the CRISPR system and subsequent delta-integration. The highest copy numbers obtained with the 8, 16, and 24 kb donor DNAs were 11, 10, and 18, respectively. In our study, we were only able to obtain 1-3 copies with our ~2 kb donor DNA. These differences may be stem from differences in the strain backgrounds, differences on donor DNA length or the breakdown efficiency of the delta-sequences by CRISPR (Mitsui *et al.,* 2019b) and/or chromosomal rearrangements that are influenced by stressful environmental conditions (Fleiss *et al.,* 2019).

3.3 CELLOBIOHYDROLASE INTEGRATION AND ACTIVITY

To evaluate the effect of the reporter protein on activities observed for integration at various loci, we investigated whether integration of a different cellulase gene, *T.e.cbh1*, would result in different outcomes. We repeated the integration experiments and targeted the same chromosomal sites in the same strain backgrounds. Transformants were confirmed via PCR (Figure 3.2) and cultivated on YPD to determine CBH activity. According to the results shown in figure 3.5, the delta integrated *T.e.cbh1* strains had higher CBH1 activities than the single locus transformants where the gene was targeted to chromosome 10, -11, or -12. These findings revealed an inverse relationship between the *T.r.eg2* integration and the *T.e.cbh1* integration in the M1744 strain background, with sites leading to low EG2 activity having comparatively higher CBH1 activity and vice versa (Figures 3.3, 3.4a and 3.5). At 72 hours of cultivation, the highest CBH activity observed among the single locus targeted yeast strains was 110 mU/gDCW for the chromosome 12 targeted strain, while the lowest activity observed was 19 mU/gDCW for the chromosome 10 targeted strain. Our CRISPR system again yielded high integration efficiencies for targeting different sites of the yeast genome. In addition, all of the M1744 strains with *T.e.cbh1* targeted to the delta-sites for multi-copy integration had

higher activity than the single locus targeted M1744 CBH strains. The highest CBH activity observed at 72 hours among the delta-integrated M1744 strains was 248 mU/gDCW, while the lowest was 210 mU/gDCW.

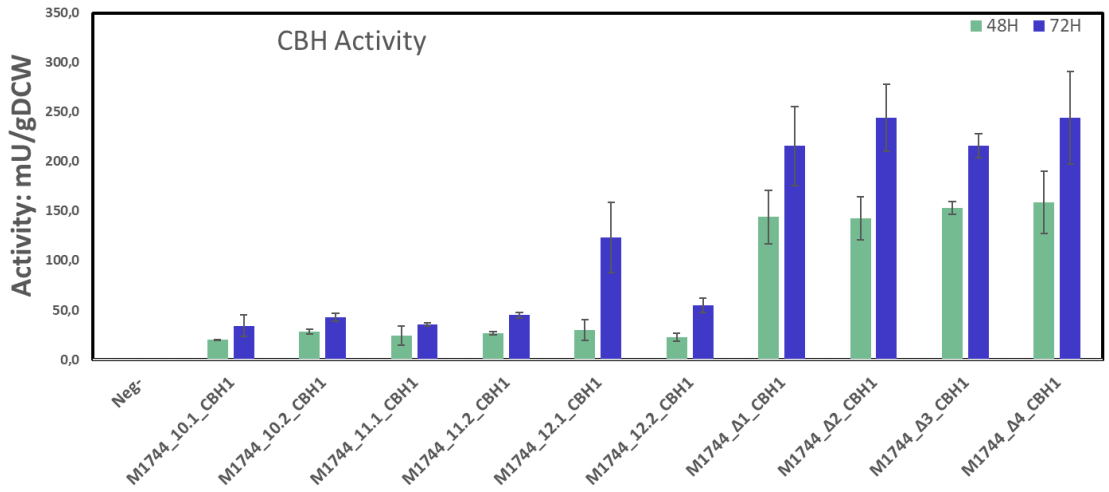


Figure 3.5: Enzyme activity profiles of CBH producing strains on MU-Lac after 48 and 72h cultivation. Values obtained were normalized with the dry cell weight of each specific yeast strain. Values given are the mean values of enzyme assays conducted in duplicates. Error bars indicate standard deviation from the mean value for each strain.

We again used qPCR to determine the number of *T.e.cbh1* copies integrated into these transformants (Table 3.3). Varying integrated copy numbers were detected in these strains. Delta sequence targeted transformants with higher activities (figure 3.5) were expected to have a higher number of gene copies compared to the other targets. Strains M1744-10-CBH1 and M1744-11-CBH, targeted for single locus integration of *T.e.cbh1* had one copy of the gene each, as expected, and displayed comparatively low CBH activity. However, the M1744-12-CBH1 strain, which was also designed for single locus *T.e.cbh1* integration, contained 4 copies of the gene. As with the integration of the *T.r.eg2*, the Chromosome 12 site was a challenge for obtaining the desired integration, and we suspected that this was due to off-target integration. Off-target integration can be confirmed with whole genome sequencing, however, that is beyond the scope of this study. Jensen *et al.,* 2014 developed an efficient set of vectors, which enabled simultaneous multiple integrations of genes into specific "safe sites of insertion". The insertion sites are located between essential elements, which limits the occurrence of chromosomal aberrations due to the lethal effect this would cause (Mikkelsen

et al., 2012). We concluded that the chromosome 12 targeting sequence we selected may not be unique in the genome, allowing unwanted off-target DSBs in areas that lacked the presence of essential elements required for stable gene integration. Interestingly, the M1744-12-CBH1 strain had double the amount of *T.e.cbh1* gene copies of the M1744 -DELTA1-CBH1 strains, which exhibited significantly higher CBH activity. This again illustrated that the locus of integration may be of greater importance than simple copy number with regards to successful heterologous protein production.

Table 3.3: A summary of the copy numbers of the *cbh1* cassettes integrated at different chromosomal site

Stains	***cbh1* copy number**
M1744-CBH-10	1
M1744-CBH-11	1
M1744-CBH-12	4
M1744-CBH-DELTA1	2
M1744-CBH-DELTA2	1
M1744-CBH-DELTA3	2
M1744-CBH-DELTA4	2

Following the successful expression of *T.e.cbh1* by our selected yeast transformants and the observation that different levels of activity was attained after integration at different loci, we turned our attention to the secretion of this CBH1. To investigate the secretion efficiency of CBH1 from *S. cerevisiae* M1744, the proteins from supernatants derived from selected transformants of *S. cerevisiae* were characterized by SDS-PAGE (Figure 3.6). SDS-PAGE analysis of untreated proteins from all *T.e.cbh1* transformants revealed very faint bands in the supernatants of transformants positive for secreted CBH1, with most of the protein visible as a smear in the size range of ~60 kDa to ~200 kDa, making it difficult to locate the single band that matched the size of the protein of interest. This is in accordance with previous observations of this protein that is hyper-glycosylated in heterologous expression in yeast (Den Haan *et al.,* 2013). As a result, we deglycosylated the CBH1 protein produced by our *S. cerevisiae* transformants to determine the extent of its N-linked glycosylation. The supernatant proteins were digested with endoglycosidase H (Endo H), which cleaved the

N-linked oligosaccharide chain of the glycoprotein. The molecular weight difference between protein samples, with and without Endo H treatment, reflects the extent of protein glycosylation mediated by the expression host. Figure 3.6 shows that the bands representing deglycosylated CBH1 from all CBH1 transformants were in the expected size range when compared to previously published results for the same protein (Den Haan *et al.,* 2013) and were distinguishable from those without Endo H treatment and were not present in the negative control.

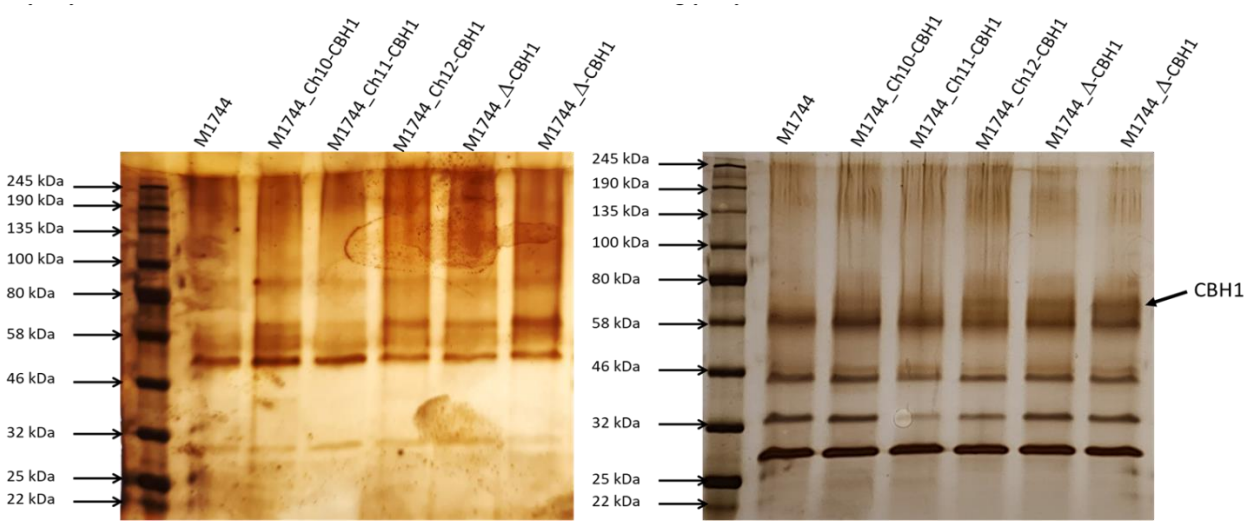


Figure 3.6: Silver stained 10% SDS-PAGE for analysis of M1744 CBH producing strains. The left gel shows untreated supernatant proteins, and the right gel shows proteins deglycosylated with Endo H. Lane 1: Molecular weight marker with sizes of the bands are indicated. Lane 2: Untransformed strain M1744 serves as a negative control. Lanes 2-5: CBH producing M1744 strains.

It has been reported that recombinant cellobiohydrolase enzymes in yeast exhibit variable levels of glycosylation (Boer *et al.,* 2000; Ilmen *et al.,* 2011; Den Haan *et al.,* 2007; Hong *et al.,* 2003; Godbole and Bowler, 1999). In some cases, hyper-glycosylation resulted in low activity of heterologous cellulases from yeast. In this study, we found smeared bands of CBH1 without using Endo H, implying that the recombinant protein was glycosylated to varying degrees. However, as we observed similar levels of glycosylation on the heterologous CBH1s from the various strains, we conclude that the different loci of expression did not lead to differences in the levels of glycosylation. Therefore, differences in the activities observed likely indicate differences in secreted protein level as opposed to differences in enzyme specific activities.

The two CRISPR systems used in this study were effective for producing recombinant proteins in both diploid and haploid strains of *S. cerevisiae*. However, we discovered that the two different genes (*T.e.cbh1* and *T.r.eg2*) we integrated in the yeast genome at different sites produced inverse results in the M1744 strain background, as the sites with high EG activity were found to have low CBH activity and vice versa. Furthermore, the delta targeted sites performed better than all other sites at expressing high levels of CBH. The CBH activity was expressed at a comparatively low level in yeast transformants when the CBH1 gene was integrated at single locus sites. However, in the EG2 transformants, the single locus targeted sites were more efficient at expressing EG than the CBH1 transformants that also had the gene integrated at the single locus sites. Wu *et al.,* (2017) discovered a correlation between the integration sites conferring the lowest and highest levels of expression. Low levels of expression are associated with the telomeres and centromeres, whereas high levels of expression have traditionally been associated with the ARS. In our study, intergenic chromosomal target sites were selected that were previously shown to support heterologous protein production at high levels (Mikkelsen *et al.,* 2012). However, it is clear that reporter protein specific factors exist in obtaining high levels of expression of a specific gene of interest.

3.4. CONSTRUCTING A YEAST STRAIN FOR CONSOLIDATED BIOPROCESSING OF CELLULOSE

Consolidated bioprocessing (CBP) is an improved process design to convert lignocellulosic feedstocks to biofuels in a single reactor using a single microbe. (Lynd *et al.,* 2005; Xu *et al.,* 2009). Heterologous expression of genes encoding cellulases from fungi has been reported in recombinant *S. cerevisiae* (Lynd *et al.,* 2005; van Zyl *et al.,* 2007; Yamada, Hasunuma and Kondo 2013; Den Haan *et al.,* 2015). Constructing a CBP yeast strain requires the insertion of multiple genes into the genome of the host organism. This is often hampered by the availability of markers for subsequent rounds of engineering. Furthermore, for some applications, it is desired to have no DNA from bacterial origin remain in the industrial transformants. As CRISPR-Cas9 allows the insertion of a defined sequence without the need for leaving the selection marker in the transformant, it is an ideal technique for industrial strain development.

The results shown in Figures 3.3 to 3.5 informed the selection of loci for the integration of the cellulase genes. It was previously reported that optimal synergy between cellulases rely on different ratios of the enzymes (Den Haan *et al.,* 2013; Van Dyk and Pletschke, 2012). We observed that integrating *T.r.eg2* and *T.e.cbh1* to different loci yielded different levels of activity. These levels can thus be used to optimise the ratio of the cellulases produced. We therefore attempted to insert multiple genes into the genome of the industrial strain isolate *S. cerevisiae* MH1000, after previously achieving highly efficient insertion of *T.r.eg2* or *T.e.cbh1* into different sites in the yeast genome using our CRISPR systems. We thus introduced non-native cellulase production in *S. cerevisiae* by combining the expression of three heterologous genes, *T.r.eg2* (EG), *T.e.cbh1* (CBH), and *S.f.bgl1* (BGL), all under the *ENO1* promoter to construct a CBP MH1000 *S. cerevisiae* using CRISPR-Cas9. The genes were integrated into the previously tested sites on chromosome 10 and chromosome 11 as well as the delta sequences. The *T.r.eg2* gene was first integrated into MH1000-Cas9 via a two-plasmid system at the chromosome 10 site (Figure 3.2a). The gRNA plasmid (targeting Chr. 10) was then cured from the transformants and a second gene, *T.e.cbh1*, was integrated at the delta sites also using the two-plasmid system. The Cas 9 and gRNA expression plasmids were then removed from the yeast strain, and a third gene, *S.f.bgl1*, was integrated at the chromosome 11 site via the one-plasmid system to create the CBP MH1000 yeast strain. PCR amplification was used to confirm the presence of the three types of integrated genes in the CBP MH1000 yeast strain (Figure 3.2d).

Our results showed that the CRISPR system was effective in producing the MH1000_CBP *S. cerevisiae* strain with all three genes present in its genome (Figures 3.2 and 3.7). However, β-glucosidase (BGL) activity was found to be low in our CBP yeast strain when compared to previous publications and the other genes tested (Lynd *et al.,* 2005; van Zyl *et al.,* 2007; Yamada, Hasunuma and Kondo 2013; Den Haan *et al.,* 2015). Since all three genes were expressed under the *ENO1* promoter and terminator, this may cause expression at lower levels. Numerous studies have demonstrated that modified organisms with multiple enzyme pathways often require precise control of the associated genes' expression levels (Borodina and Nielsen, 2014; Da Silva and Srikrishnan, 2012). As a result of this information, we concluded that expressing each gene with a unique promoter could circumvent the issue of low enzyme activity. The reduction in the expression of the *S.f.bgl1* gene could also be due to

concurrent expression of the other genes integrated in the yeast genome as was previously reported (Shi et *al.,* 2016). Furthermore, some studies have suggested that poor cellulase performance may be due to a lack of appropriate signal peptides, which are important for cellulase expression in yeast as they guide proteins to and through the endoplasmic reticulum and the secretory pathway (Nunnari and Walter, 1992; Sanders *et al.,* 1992). Nevertheless, we have shown that, in principle, an industrial *S. cerevisiae* strain could be successfully engineered with three cellulase encoding genes without the need to maintain any selectable marker in the final strain. This makes the strain amenable to subsequent engineering steps. Furthermore, the observed differences in cellulase activities after integration at the different sites could be used to optimise the ratio of cellulases produced.

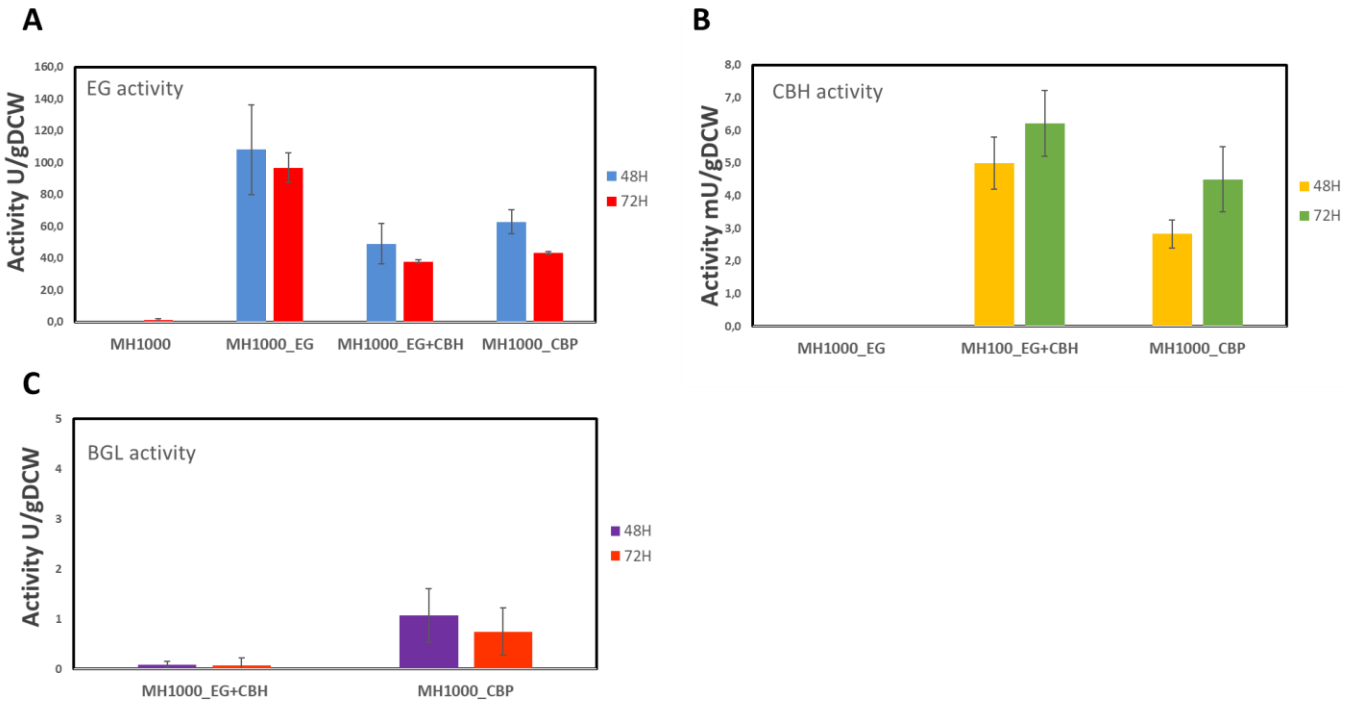


Figure 3.7: Enzyme activity profiles of recombinant yeast strains after 48 and 72h cultivation. (A) Activity of EG2 producing diploid strains on CMC. The untransformed MH1000 strain was used as negative control. (B) Enzyme activity profiles of CBH producing strains in MU-Lac. MH1000_EG was used as negative control reference. (C) Enzyme activity profiles of BGL producing strains in pNP-G. MH1000_EG+CBH strain was used as negative control. Values obtained were normalised using Dry Cell Weight (DCW) of each strain. All values represent mean values of assays conducted in triplicate with error bars indicating the standard deviation from the mean value for each strain.

Several studies have also shown that the CRISPR-Cas9 system in *S. cerevisiae* can be used to introduce multiple heterologous genes into its genome to produce various chemicals (Jakociunas *et al.,* 2015; Ronda *et al.,* 2015). In comparison to this study, sequential integration of three genes (*T.r.eg2*, *T.e.cbh1*, and *S.f.bgl1*) involved in the breakdown of cellulose for biofuel production demonstrated the CRISPR system's ability to insert genes of varying lengths, which is extremely advantageous for industrial metabolic engineering applications.

CHAPTER 4

SUMMARY AND CONCLUSION

The study aimed to use CRISPR-Cas9 to integrate multiple genes at different genomic locations. This was done to determine which sites were the most effective for integrating cellulase encoding genes and producing high heterologous enzyme activities. Our CRISPR-Cas9 systems enabled efficient gene integration in different chromosomal positions within the yeast genome. However, we observed significant differences in activity at different loci: (i) when the same gene (*T.r.eg2*) was integrated and (ii) when a different gene was integrated (*T.e.cbh1*). Differences in gene copy number were also observed between haploid and diploid yeast strains. We discovered that targeting genes to our selected Chr. 12 site yielded higher levels of activity in EG and CBH transformants than all other sites targeted for single copy integration. However, we were unable to confirm that the gene was integrated at the selected Chr. 12 position using PCR, most likely due to off-target integration, casting doubt over the usefulness of this particular locus for strain construction.

Delta sites were targeted for multi-copy integration of the same genes that were tested for single copy integration, and we discovered that delta integration was frequently found in 1 copy in haploid strains and 2 copies in diploid strains for both genes tested. Furthermore, *T.e.cbh1* transformants where the gene was targeted to delta sites, yielded higher activity than any *T.e.cbh1* transformants that were targeted for single copy integration. We also showed that while targeting *T.e.cbh1* to the delta sites yielded transformants with the highest CBH activity, *T.r.eg2* transformants had slightly lower EG activity when targeting the delta sites. Our results clearly show that there are gene specific, and locus specific factors involved in obtaining high levels of heterologous protein production and that these factors cannot necessarily be predicted prior to testing the various loci and heterologous genes of interest.

We then attempted to use the knowledge we gained to create a CBP industrial diploid yeast strain transformed with three genes in different rounds of transformation, to create a rudimentary cellulase system. We discovered that of the three genes integrated into the CBP yeast strain, *S.f.bgl1* yielded the lowest enzyme activity, while EG and CBH activities were also

negatively affected to some extent. Nevertheless, the principle that a rudimentary cellulase system could be engineered into an industrial yeast strain, without the need to maintain any markers in the final transformant strain, was shown. Based on these findings, we concluded that our CRISPR systems can be useful and easily applied in metabolic engineering of *S. cerevisiae* for biofuel production.

4.1. PERSPECTIVES

The development of multi-copy integration of heterologous genes in the yeast *S. cerevisiae* using the CRISPR-Cas9 genome editing tool has been revolutionary. It has facilitated the development and advancement of yeast cell factories capable of converting a wide range of substrates into a wide range of products ranging from fuels and chemicals to drugs (Demeke *et al.,* 2013). In this study, we directed the CRSISPR-Cas9 system to delta sites to achieve multi-copy integration of various reporter genes. Ribosomal DNA (rDNA), which is composed of 150–200 tandem copies of a 9.1-kb rDNA repeat, has also been used as a target site for multicopy gene integration (Wang *et al.,* 2018). In future, we could target these sites and compare them to the delta sites to determine if they are more effective at expressing the cellulase genes. Sasaki *et al.,* (2019) used CRISPR-delta-integration and multiple promoters shuffling methods, which involved conjugating a heterologous gene with different promoters and then delta-integrating it. We could also combine the multiple promoters shuffling method with our CRISPR systems in the future to improve the transcriptional levels and activity of the various genes that may be required to create a new metabolic pathway such as cellulose utilization.

Multiplex genome editing is another option to be considered as we performed the introduction of the three different genes into a single yeast strain over the course of several transformation rounds. Creating suitable gRNAs and an appropriate vector to express them is critical in achieving multiplex integration and minimizing transformation rounds. (DiCarlo *et al.,* 2013; Adiego-Pérez *et al.,* 2019). Despite the significant advancements in the development of numerous assisting tools for designing gRNAs for various Cas proteins, (Concordet and Haeussler, 2018; Labuhn *et al.,* 2018; Labun *et al.,* 2019; Liao *et al.,* 2019), the effectiveness of these gRNAs *in vivo* still have to be evaluated, which is even more critical for multiplex integration (Bourgeois *et al.,* 2018). Additionally, the target locus selection has a

substantial effect in the integration efficiency. In our study, we discovered that certain sites offered more activity for a certain gene than others. Baek *et al.,* (2021) discovered that specific target locations in gene-sparse regions were very ineffective due to restricted chromatin accessibility. Observations similar to these have been made elsewhere (Mans *et al.,* 2015; Bourgeois *et al.,* 2018; Verkuijl and Rots, 2019). The identification and characterization of effective gRNAs and target loci will significantly increase the performance and efficacy of multiplex integration (Apel *et al.,* 2017; Bourgeois *et al.,* 2018; Baek *et al.,* 2021).

Another factor that must be tuned to maximize multiplex integration is the homology directed repair (HDR), which functions as a DSB repair process in yeast. As previously reported, *S. cerevisiae* prefers HDR over non-homologous end joining (NHEJ) when mending double strand breaks (DSBs) (Sander and Joung, 2014). Producing Cas9 and gRNA in a cell may be harmful to yeast without donor DNA. However, several investigations of multiplex gene integration showed that certain yeast colonies may survive in the absence of donor DNA, suggesting that NHEJ is capable of repairing DSBs (Baek *et al.,* 2021). This may result in an increase in false positive colonies and a decrease in multiplex gene integration efficiency. Deletions of genes implicated in the NHEJ process, including *POL4*, *DNL4*, and *Ku70*, have been found to decrease false positive rates and improve HDR success (Lemos *et al.,* 2018; Yan and Finnigan, 2018). Because NHEJ-mediated chromosomal repairs generate a large number of background colonies, the removal of NHEJ genes may increase the capacity and efficiency of multiplex genome integration. Overcoming these constraints and creatively integrating numerous technologies will enable multiplex genome editing to increase in scope. As a result, the research and optimization of yeast's extensive specialized metabolic pathways will be accelerated.

5. REFERENCES

Adiego-Pérez, B., Randazzo, P., Daran, J. M., Verwaal, R., Roubos, J. A., Daran-Lapujade, P., & van der Oost, J. (2019). Multiplex genome editing of microorganisms using CRISPR-Cas. *FEMS microbiology letters, 366*(8), fnz086.

Apel, A. R., d'Espaux, L., Wehrs, M., Sachs, D., Li, R. A., Tong, G. J., Garber, M., Nnadi, O., Zhuang, W., Hillson, N. J., Keasling, J. D., & Mukhopadhyay, A. (2017). A Cas9-based toolkit to program gene expression in *Saccharomyces cerevisiae. Nucleic acids research, 45*(1), 496-508.

Baek, S., Utomo, J. C., Lee, J. Y., Dalal, K., Yoon, Y. J., and Ro, D.- K. (2021). The yeast platform engineered for synthetic gRNA-landing pads enables multiple gene integrations by a single gRNA/Cas9 system. *Metabolic engineering. 64*, 111–121.

Bao, Z., Xiao, H., Liang, J., Zhang, L., Xiong, X., Sun, N., Si, T., & Zhao, H. (2015). Homology-integrated CRISPR-Cas (HI-CRISPR) system for one-step multigene disruption in *Saccharomyces cerevisiae*. *ACS synthetic biology, 4*(5), 585–594.

Barzel, A., Obolski, U., Gogarten, J. P., Kupiec, M., & Hadany, L. (2011). Home and away-the evolutionary dynamics of homing endonucleases. *BMC evolutionary biology, 11*, 324.

Blin, K., Pedersen, L. E., Weber, T., & Lee, S. Y. (2016). CRISPy-web: An online resource to design sgRNAs for CRISPR applications. *Synthetic and systems biotechnology, 1*(2), 118–121.

Boer, H., Teeri, T. T., & Koivula, A. (2000). Characterization of *Trichoderma reesei* cellobiohydrolase Cel7A secreted from *Pichia pastoris* using two different promoters. *Biotechnology and bioengineering, 69*(5), 486–494.

Borodina, I., and Nielsen, J. (2014). Advances in metabolic engineering of yeast *Saccharomyces cerevisiae* for production of chemicals. *Biotechnology journal, 9*(5), 609–620.

Bourgeois, L., Pyne, M. E., and Martin, V. J. J. (2018). A highly characterized synthetic landing pad system for precise multicopy gene integration in yeast. *ACS Synthetic Biology. 7,* 2675–2685.

Brevnova E, Mcbride JE, Wiswall E, Wenger SK. (2011). Yeast expressing saccharolytic enzymes for consolidated bioprocessing using starch and cellulose' *Patent No*. W0/2011/153516

Chavez, A., Scheiman, J., Vora, S., Pruitt, B. W., Tuttle, M., P R Iyer, E., Lin, S., Kiani, S., Guzman, C. D., Wiegand, D. J., Ter-Ovanesyan, D., Braff, J. L., Davidsohn, N., Housden, B.

E., Perrimon, N., Weiss, R., Aach, J., Collins, J. J., & Church, G. M. (2015). Highly efficient Cas9-mediated transcriptional programming. *Nature methods, 12*(4), 326–328.

Chen S. J. (2019). Minimizing off-target effects in CRISPR-Cas9 genome editing. *Cell biology and toxicology, 35*(5), 399–401.

Cho, K.M., Yoo, Y.J., and Kang, K.S. (1999). Delta-integration of endo/exo-glucanase and beta- glucosidase genes into the yeast chromosomes for direct conversion of cellulose to ethanol. *Enzyme and Microbial Technology 25*, 23-30.

Choi, P. S., and Meyerson, M. (2014). Targeted genomic rearrangements using CRISPR/Cas technology. *Nature communications, 5*, 3728.

Concordet, J.-P., and Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Research. 46*, W242–W245.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science (New York, N.Y.), 339*(6121), 819–823.

Cutler, S., Lee, L. J., & Tsukiyama, T. (2018). Chromatin Remodeling Factors Isw2 and Ino80 Regulate Chromatin, Replication, and Copy Number of the *Saccharomyces cerevisiae* Ribosomal DNA Locus. *Genetics, 210*(4), 1543–1556.

Da Silva, N. A., & Srikrishnan, S. (2012). Introduction and expression of genes for metabolic engineering applications in *Saccharomyces cerevisiae*. *FEMS yeast research, 12*(2), 197–214.

Dahlman, J. E., Abudayyeh, O. O., Joung, J., Gootenberg, J. S., Zhang, F., & Konermann, S. (2015). Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. *Nature biotechnology, 33*(11), 1159–1161.

Davison, S. A., Den Haan, R., & Van Zyl, W. H. (2016). Heterologous expression of cellulase genes in natural *Saccharomyces cerevisiae* strains. *Applied microbiology and biotechnology, 100*(18), 8241–8254.

Davison, S. A., Keller, N. T., van Zyl, W. H., & Den Haan, R. (2019). Improved cellulase expression in diploid yeast strains enhanced consolidated bioprocessing of pretreated corn residues. *Enzyme and microbial technology, 131*, 109382.

Deaner, M., and Alper, H. S. (2017). Systematic testing of enzyme perturbation sensitivities via graded dCas9 modulation in *Saccharomyces cerevisiae*. *Metabolic engineering*, *40*, 14–22.

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., & Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature, 471*(7340), 602–607.

Demeke, M. M., Dietz, H., Li, Y., Foulquié-Moreno, M. R., Mutturi, S., Deprez, S., Den Abt, T., Bonini, B. M., Liden, G., Dumortier, F., Verplaetse, A., Boles, E., & Thevelein, J. M. (2013). Development of a D-xylose fermenting and inhibitor tolerant industrial *Saccharomyces cerevisiae* strain with high performance in lignocellulose hydrolysates using metabolic and evolutionary engineering. *Biotechnology for biofuels, 6*(1), 89.

Den Haan, R., Rose, S. H., Lynd, L. R., & van Zyl, W. H. (2007). Hydrolysis and fermentation of amorphous cellulose by recombinant *Saccharomyces cerevisiae*. *Metabolic engineering, 9*(1), 87–94.

Den Haan, R., van Rensburg, E., Rose, S. H., Görgens, J. F., & van Zyl, W. H. (2015). Progress and challenges in the engineering of non-cellulolytic microorganisms for consolidated bioprocessing. *Current opinion in biotechnology, 33*, 32–38.

Den Haan, R., van Zyl, J.M., Harms, T.M., and van Zyl, W.H. (2013). Modeling the minimum enzymatic requirements for optimal cellulose conversion. *Environmental Research Letters* 8.

DiCarlo, J. E., Norville, J. E., Mali, P., Rios, X., Aach, J., & Church, G. M. (2013). Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic acids research, 41*(7), 4336–4343.

Ding, W., Zhang, Y., & Shi, S. (2020). Development and Application of CRISPR/Cas in Microbial Biotechnology. *Frontiers in bioengineering and biotechnology, 8*, 711.

Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B. L., Xavier, R. J., & Root, D. E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature biotechnology, 32(12)*, 1262–1267.

Dong, C., Jiang, L., Xu, S., Huang, L., Cai, J., Lian, J., & Xu, Z. (2020). A Single Cas9-VPR Nuclease for Simultaneous Gene Activation, Repression, and Editing in *Saccharomyces cerevisiae*. *ACS synthetic biology, 9*(9), 2252–2257.

Duan, X., Gao, J., & Zhou, Y. (2017). Advances in engineering methylotrophic yeast for biosynthesis of valuable chemicals from methanol. *Chinese Chemical Letters, 29*, 681-686.

Duina, A. A., Miller, M. E., & Keeney, J. B. (2014). Budding yeast for budding geneticists: a primer on the *Saccharomyces cerevisiae* model system. *Genetics*, *197*(1), 33–48.

Farwick, A., Bruder, S., Schadeweg, V., Oreb, M., & Boles, E. (2014). Engineering of yeast hexose transporters to transport D-xylose without inhibition by D-glucose. *Proceedings of the National Academy of Sciences of the United States of America, 111*(14), 5159–5164.

Farzadfard, F., Perli, S. D., & Lu, T. K. (2013). Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. *ACS synthetic biology, 2*(10), 604–613.

Fernandez, R., and Berro, J. (2016). Use of a fluoride channel as a new selection marker for fission yeast plasmids and application to fast genome editing with CRISPR/Cas9. *Yeast (Chichester, England), 33*(10), 549–557.

Ferreira, R., Skrekas, C., Nielsen, J., & David, F. (2018). Multiplexed CRISPR/Cas9 Genome Editing and Gene Regulation Using Csy4 in *Saccharomyces cerevisiae*. *ACS synthetic biology, 7*(1), 10–15.

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

Fleiss, A., O'Donnell, S., Fournier, T., Lu, W., Agier, N., Delmas, S., Schacherer, J., & Fischer, G. (2019). Reshuffling yeast chromosomes with CRISPR/Cas9. *PLoS genetics, 15*(8), e1008332.

Gao, Y., and Zhao, Y. (2014). Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *Journal of integrative plant biology, 56*(4), 343–349.

Garst, A. D., Bassalo, M. C., Pines, G., Lynch, S. A., Halweg-Edwards, A. L., Liu, R., Liang, L., Wang, Z., Zeitoun, R., Alexander, W. G., & Gill, R. T. (2017). Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nature biotechnology, 35*(1), 48–55.

Gasiunas, G., Barrangou, R., Horvath, P., & Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America, 109*(39), E2579–E2586.

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

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A., Lim, W. A., Weissman, J. S., & Qi, L. S. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell, 154*(2), 442–451.

Gnügge, R., and Rudolf, F. (2017). *Saccharomyces cerevisiae* Shuttle vectors. *Yeast (Chichester, England)*, *34*(5), 205–221.

Godbole, S., & Bowler, B. E. (1999). Effect of pH on formation of a nativelike intermediate on the unfolding pathway of a Lys 73 --> His variant of yeast iso-1-cytochrome c. *Biochemistry, 38(1)*, 487–495.

Halperin, S. O., Tou, C. J., Wong, E. B., Modavi, C., Schaffer, D. V., & Dueber, J. E. (2018). CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature, 560*(7717), 248–252.

Hanasaki, M., and Masumoto, H. (2019). CRISPR/Transposon gene integration (CRITGI) can manage gene expression in a retrotransposon-dependent manner. *Scientific reports, 9*(1), 15300.

Hasunuma, T., & Kondo, A. (2012). Development of yeast cell factories for consolidated bioprocessing of lignocellulose to bioethanol through cell surface engineering. *Biotechnology advances, 30*(6), 1207–1218.

Heigwer, F., Kerr, G., & Boutros, M. (2014). E-CRISP: fast CRISPR target site identification. *Nature methods, 11*(2), 122–123.

Hoffman, C. S., and Winston, F. (1987). A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. *Gene, 57*(2-3), 267–272.

Hoffman, C. S., Wood, V., & Fantes, P. A. (2015). An Ancient Yeast for Young Geneticists: A Primer on the *Schizosaccharomyces pombe* Model System. *Genetics, 201*(2), 403–423.

Hong, Y., Ohishi, K., Kang, J. Y., Tanaka, S., Inoue, N., Nishimura, J., Maeda, Y., & Kinoshita, T. (2003). Human PIG-U and yeast Cdc91p are the fifth subunit of GPI transamidase that attaches GPI-anchors to proteins. *Molecular biology of the cell, 14*(5), 1780–1789.

Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science (New York, N.Y.), 327*(5962), 167–170.

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

Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell, 157*(6), 1262–1278.

Hua, S. B., Qiu, M., Chan, E., Zhu, L., & Luo, Y. (1997). Minimum length of sequence homology required for *in vivo* cloning by homologous recombination in yeast. *Plasmid, 38*(2), 91–96.

Huang, S., and Geng, A. (2020). High-copy genome integration of 2,3-butanediol biosynthesis pathway in *Saccharomyces cerevisiae* via in vivo DNA assembly and replicative CRISPR-Cas9 mediated delta integration. *Journal of biotechnology*. 310, 13–20.

Ilmén, M., Den Haan, R., Brevnova, E., McBride, J., Wiswall, E., Froehlich, A., Koivula, A., Voutilainen, S. P., Siika-Aho, M., la Grange, D. C., Thorngren, N., Ahlgren, S., Mellon, M., Deleault, K., Rajgarhia, V., van Zyl, W. H., & Penttilä, M. (2011). High level secretion of c*ellobiohydrolases* by *Saccharomyces cerevisiae*. *Biotechnology for biofuels, 4*, 30.

Jacobs, J. Z., Ciccaglione, K. M., Tournier, V., & Zaratiegui, M. (2014). Implementation of the CRISPR-Cas9 system in fission yeast. *Nature communications, 5*, 5344.

Jako, T., Bonde, I., Herrgård, M., Harrison, S.J., Kristensen, M., Pedersen, L.E., Jensen, M.K., Keasling, J.D., (2015). Multiplex metabolic pathway engineering using CRISPR/Cas9 in *Saccharomyces cerevisiae*. *Metabolic engineering. 28*, 213–222

Jakociunas, T., Bonde, I., Herrgård, M., Harrison, S.J., Kristensen, M., Pedersen, L.E., Jensen, M.K., Keasling, J.D., (2015). Multiplex metabolic pathway engineering using CRISPR/Cas9 in *Saccharomyces cerevisiae*. *Metabolic engineering.28*, 213–222.

Jensen, E. D., Ferreira, R., Jakociunas, T., Arsovska, D., Zhang, J., Ding, L., Smith, J. D., David, F., Nielsen, J., Jensen, M. K., & Keasling, J. D. (2017). Transcriptional reprogramming in yeast using dCas9 and combinatorial gRNA strategies. *Microbial cell factories, 16*(1), 46.

Jensen, N. B., Strucko, T., Kildegaard, K. R., David, F., Maury, J., Mortensen, U. H., Forster, J., Nielsen, J., & Borodina, I. (2014). EasyClone: method for iterative chromosomal integration of multiple genes in *Saccharomyces cerevisiae*. *FEMS yeast research, 14*(2), 238–248.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna J.A, Charpentier E. (2012). A programmable dual-RNA- guided DNA endonuclease in adaptive bacterial immunity. *Science. 337*(6096):816-21.

Karim, A.S., Curran, K.A., Alper, H.S., (2013). Characterization of plasmid burden and copy number in *Saccharomyces cerevisiae* for optimization of metabolic engineering applications. *FEMS yeast research. 13*, 107–116.

Kiani, S., Chavez, A., Tuttle, M., Hall, R. N., Chari, R., Ter-Ovanesyan, D., Qian, J., Pruitt, B. W., Beal, J., Vora, S., Buchthal, J., Kowal, E. J., Ebrahimkhani, M. R., Collins, J. J., Weiss, R., & Church, G. (2015). Cas9 gRNA engineering for genome editing, activation and repression. *Nature methods, 12*(11), 1051–1054.

Kobayashi, T., and Sasaki, M. (2017). Ribosomal DNA stability is supported by many 'buffer genes'-introduction to the Yeast rDNA Stability Database. *FEMS yeast research, 17*(1), 10.1093/femsyr/fox001.

Kowalczykowski S. C. (2015). An Overview of the Molecular Mechanisms of Recombinational DNA Repair. *Cold Spring Harbor perspectives in biology, 7*(11), a016410.

Kroukamp, H., Den Haan, R., Van Wyk, N. & Van Zyl, W. H. (2013). Overexpression of native PSE1 and SOD1 in *Saccharomyces cerevisiae* improved heterologous cellulase secretion. *Applied Energy, 102*, 150-156.

Kumar, S., Singh, S.P., Mishra, I.M., & Adhikari, D.K. (2009). Recent Advances in Production of Bioethanol from Lignocellulosic Biomass. *Chemical Engineering & Technology, 32*, 517-526.

La Grange, D. C., Pretorius, I. S., Claeyssens, M., & van Zyl, W. H. (2001). Degradation of xylan to D-xylose by recombinant *Saccharomyces cerevisiae* coexpressing the *Aspergillus niger* beta-xylosidase (xlnD) and the *Trichoderma reesei xylanase* II (xyn2) genes. *Applied and environmental microbiology, 67*(12), 5512–5519.

Labuhn, M., Adams, F. F., Ng, M., Knoess, S., Schambach, A., Charpentier, E. M., Schwarzer, A., Mateo, J. L., Klusmann, J. H., & Heckl, D. (2018). Refined sgRNA efficacy prediction improves large- and small-scale CRISPR-Cas9 applications. *Nucleic acids research*, 46(3), 1375–1385.

Labun, K., Montague, T. G., Gagnon, J. A., Thyme, S. B., & Valen, E. (2016). CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic acids research, 44*(W1), W272–W276.

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

Laemmli U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature, 227*(5259), 680–685.

Lander E. S. (2016). The Heroes of CRISPR. *Cell, 164*(1-2), 18–28.

Laughery, M. F., Hunter, T., Brown, A., Hoopes, J., Ostbye, T., Shumaker, T., & Wyrick, J. J. (2015). New vectors for simple and streamlined CRISPR-Cas9 genome editing in *Saccharomyces cerevisiae*. *Yeast (Chichester, England), 32*(12), 711–720.

Lemos, B. R., Kaplan, A. C., Bae, J. E., Ferrazzoli, A. E., Kuo, J., Anand, R. P., Waterman, D. P., & Haber, J. E. (2018). CRISPR/Cas9 cleavages in budding yeast reveal templated insertions and strand-specific insertion/deletion profiles. *Proceedings of the National Academy of Sciences of the United States of America, 115*(9), E2040–E2047.

Li, J., Shou, J., Guo, Y., Tang, Y., Wu, Y., Jia, Z., Zhai, Y., Chen, Z., Xu, Q., & Wu, Q. (2015). Efficient inversions and duplications of mammalian regulatory DNA elements and gene clusters by CRISPR/Cas9. *Journal of molecular cell biology, 7*(4), 284–298.

Li, M., and Borodina, I., (2015). Application of synthetic biology for production of chemicals in yeast *S. cerevisiae*. *FEMS Yeast Res. 15*, 1–12.

Li, T., Huang, S., Zhao, X., Wright, D. A., Carpenter, S., Spalding, M. H., Weeks, D. P., & Yang, B. (2011). Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic acids research, 39*(14), 6315–6325.

Lian, J., HamediRad, M., Hu, S., & Zhao, H. (2017). Combinatorial metabolic engineering using an orthogonal tri-functional CRISPR system. *Nature communications, 8*(1), 1688.

Liao, C., Ttofali, F., Slotkowski, R. A., Denny, S. R., Cecil, T. D., Leenay, R. T., Keung, A. J., & Beisel, C. L. (2019). Modular one-pot assembly of CRISPR arrays enables library generation and reveals factors influencing crRNA biogenesis. *Nature communications, 10*(1), 2948.

Liu, H., Wei, Z., Dominguez, A., Li, Y., Wang, X., & Qi, L. S. (2015). CRISPR-ERA: a comprehensive design tool for CRISPR-mediated gene editing, repression and activation. *Bioinformatics (Oxford, England), 31*(22), 3676–3678.

Lynd, L. R., van Zyl, W. H., McBride, J. E., & Laser, M. (2005). Consolidated bioprocessing of cellulosic biomass: an update. *Current opinion in biotechnology, 16*(5), 577–583.

Maddalo, D., Manchado, E., Concepcion, C. P., Bonetti, C., Vidigal, J. A., Han, Y. C., Ogrodowski, P., Crippa, A., Rekhtman, N., de Stanchina, E., Lowe, S. W., & Ventura, A. (2014). *In vivo* engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature, 516*(7531), 423–427.

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., Barrangou, R., Brouns, S. J., Charpentier, E., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J., Terns, R. M., Terns, M. P., White, M. F., Yakunin, A. F., Garrett, R. A., van der Oost, J., Backofen, R., Koonin, E. V. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nature reviews. Microbiology*, *13*(11), 722–736.

Malcı, K., Walls, L. E., & Rios-Solis, L. (2020). Multiplex Genome Engineering Methods for Yeast Cell Factory Development. *Frontiers in bioengineering and biotechnology, 8*, 589468.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., & Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science (New York, N.Y.)*, *339*(6121), 823–826.

Manivasakam, P., Weber, S. C., McElver, J., & Schiestl, R. H. (1995). Micro-homology mediated PCR targeting in *Saccharomyces cerevisiae*. *Nucleic acids research, 23*(14), 2799–2800.

Mans, R., van Rossum, H. M., Wijsman, M., Backx, A., Kuijpers, N. G., van den Broek, M., Daran-Lapujade, P., Pronk, J. T., van Maris, A. J., & Daran, J. M. (2015). CRISPR/Cas9: a

molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS yeast research, 15*(2),

Martin, F., Sánchez-Hernández, S., Gutiérrez-Guerrero, A., Pinedo-Gomez, J., & Benabdellah, K. (2016). Biased and Unbiased Methods for the Detection of Off-Target Cleavage by CRISPR/Cas9: An Overview. *International journal of molecular sciences, 17*(9), 1507.

Meinander, N., Zacchi, G., & Hahn-Hägerdal, B. (1996). A heterologous reductase affects the redox balance of recombinant *Saccharomyces cerevisiae*. *Microbiology (Reading, England), 142* (Pt 1), 165–172.

Mikkelsen, M. D., Buron, L. D., Salomonsen, B., Olsen, C. E., Hansen, B. G., Mortensen, U. H., & Halkier, B. A. (2012). Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. *Metabolic engineering, 14*(2), 104–111.

Misumi, Y., Nishioka, S., Fukuda, A., Uemura, T., Nakamura, M., Hoshida, H., & Akada, R. (2019). YHp as a highly stable, hyper-copy, hyper-expression plasmid constructed using a full 2-µm circle sequence in cir0 strains of *Saccharomyces cerevisiae*. *Yeast (Chichester, England), 36*(5), 249–257.

Mitsui, R., Yamada, R., & Ogino, H. (2019a). CRISPR system in the yeast *Saccharomyces cerevisiae* and its application in the bioproduction of useful chemicals. *World journal of microbiology & biotechnology, 35*(7), 111.

Mitsui, R., Yamada, R., & Ogino, H. (2019b). Improved Stress Tolerance of *Saccharomyces cerevisiae* by CRISPR-Cas-Mediated Genome Evolution. *Applied biochemistry and biotechnology, 189*(3), 810–821.

Mohanraju, P., Makarova, K. S., Zetsche, B., Zhang, F., Koonin, E. V., & van der Oost, J. (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science (New York, N.Y.), 353*(6299), aad5147.

Mohr, S. E., Hu, Y., Ewen-Campen, B., Housden, B. E., Viswanatha, R., & Perrimon, N. (2016). CRISPR guide RNA design for research applications. *The FEBS journal*, *283*(17), 3232–3238.

Moriguchi, K., Yamamoto, S., Ohmine, Y., & Suzuki, K. (2016). A Fast and Practical Yeast Transformation Method Mediated by *Escherichia coli* Based on a Trans-Kingdom Conjugal Transfer System: Just Mix Two Cultures and Wait One Hour. *PloS one, 11*(2), e0148989.

Nakade, S., Yamamoto, T., & Sakuma, T. 2017. Cas9, Cpf1 and C2c1/2/3-What's next?. *Bioengineered, 8*(3), 265–273.

Nunnari, J., & Walter, P. (1992). Protein targeting to and translocation across the membrane of the endoplasmic reticulum. *Current opinion in cell biology, 4*(4), 573–580.

Oh, E. J., and Jin, Y. S. (2020). Engineering of *Saccharomyces cerevisiae* for efficient fermentation of cellulose. *FEMS yeast research, 20*(1), foz089.

Ota, S., Hisano, Y., Ikawa, Y., & Kawahara, A. (2014). Multiple genome modifications by the CRISPR/Cas9 system in zebrafish. *Genes to cells : devoted to molecular & cellular mechanisms, 19*(7), 555–564.

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell, 152*(5), 1173–1183.

Rainha, J., Rodrigues, J. L., & Rodrigues, L. R. (2020). CRISPR-Cas9: A Powerful Tool to Efficiently Engineer *Saccharomyces cerevisiae*. *Life (Basel, Switzerland), 11*(1), 13.

Ran, F. A., Cong, L., Yan, W. X., Scott, D. A., Gootenberg, J. S., Kriz, A. J., Zetsche, B., Shalem, O., Wu, X., Makarova, K. S., Koonin, E. V., Sharp, P. A., & Zhang, F. (2015*). In vivo* genome editing using *Staphylococcus aureus* Cas9. *Nature, 520*(7546), 186–191.

Romanos, M.A., Scorer, C.A., Clare, J.J., (1992). Foreign gene expression in yeast: a review. *Yeast 8,* 423–488.

Ronda, C., Maury, J., Jakociunas, T., Jacobsen, S. A., Germann, S. M., Harrison, S. J., Borodina, I., Keasling, J. D., Jensen, M. K., & Nielsen, A. T. (2015). CrEdit: CRISPR mediated multi-loci gene integration in *Saccharomyces cerevisiae*. *Microbial cell factories, 14*, 97.

Rothstein R. (1991). Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. Methods in enzymology, 194, 281–301.

Roy, B., Zhao, J., Yang, C., Luo, W., Xiong, T., Li, Y., Fang, X., Gao, G., Singh, C. O., Madsen, L., Zhou, Y., & Kristiansen, K. (2018). CRISPR/Cascade 9-Mediated Genome Editing-Challenges and Opportunities. *Frontiers in genetics, 9*, 240.

Ryabova, O. B., Chmil, O. M., & Sibirny, A. A. (2003). Xylose and cellobiose fermentation to ethanol by the thermotolerant methylotrophic yeast *Hansenula polymorpha*. *FEMS yeast research*, 4(2), 157–164.

Ryan, O. W., Skerker, J. M., Maurer, M. J., Li, X., Tsai, J. C., Poddar, S., Lee, M. E., DeLoache, W., Dueber, J. E., Arkin, A. P., & Cate, J. H. (2014). Selection of chromosomal DNA libraries using a multiplex CRISPR system. *eLife*, *3*, e03703.

Sakai, A., Shimizu, Y., & Hishinuma, F. (1990). Integration of heterologous genes into the chromosome of *Saccharomyces cerevisiae* using a delta sequence of yeast retrotransposon Ty. *Applied microbiology and biotechnology, 33*(3), 302–306.

Sakuma, T., Nishikawa, A., Kume, S., Chayama, K., & Yamamoto, T. 2014. Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Scientific reports, 4*, 5400.

Sander JD, Joung JK. 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature biotechnology. 32*(3):279-84.

Sanders, S. L., Whitfield, K. M., Vogel, J. P., Rose, M. D., & Schekman, R. W. (1992). Sec61p and BiP directly facilitate polypeptide translocation into the ER. *Cell, 69*(2), 353–365.

Sasaki, Y., Mitsui, R., Yamada, R., & Ogino, H. (2019). Secretory overexpression of the endoglucanase by *Saccharomyces cerevisiae* via CRISPR-δ-integration and multiple promoter shuffling. *Enzyme and microbial technology, 121*, 17–22.

Sasano, Y., Nagasawa, K., Kaboli, S., Sugiyama, M., & Harashima, S. (2016). CRISPR-PCS: a powerful new approach to inducing multiple chromosome splitting in *Saccharomyces cerevisiae*. *Scientific reports, 6*, 30278.

Sazci, A., Erenler, K. and Radford, A., 1986. Detection of cellulolytic fungi by using Congo red as an indicator: a comparative study with the dinitrosalicyclic acid reagent method. *Journal of Applied Bacteriology, 61*(6), pp. 559-562.

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

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

Spohner, S. C., Schaum, V., Quitmann, H., & Czermak, P. (2016). Kluyveromyces lactis: An emerging tool in biotechnology. *Journal of biotechnology*, *222*, 104–116.

Storici, F., Durham, C. L., Gordenin, D. A., & Resnick, M. A. (2003). Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(25), 14994–14999.

Stovicek, V., Holkenbrink, C., & Borodina, I. (2017). CRISPR/Cas system for yeast genome engineering: advances and applications. *FEMS yeast research*, *17*(5), fox030.

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

Sugiyama, M., Ikushima, S., Nakazawa, T., Kaneko, Y., & Harashima, S. (2005). PCR-mediated repeated chromosome splitting in *Saccharomyces cerevisiae. BioTechniques*, 38(6), 909–914.

Symington L.S, Rothstein R, Lisby M. 2014. Mechanisms and regulation of mitotic recombination in *Saccharomyces cerevisiae*. *Genetics 198*: 795–835.

Teste, M. A., Duquenne, M., François, J. M., & Parrou, J. L. (2009). Validation of reference genes for quantitative expression analysis by real-time RT-PCR in *Saccharomyces cerevisiae*. *BMC molecular biology, 10*, 99.

Thompson, R. 2017. Development of CRISPR based transformation stratergies for transformaning industrial and laboratory yeast strains.

Thyme, S. B., Akhmetova, L., Montague, T. G., Valen, E., & Schier, A. F. (2016). Internal guide RNA interactions interfere with Cas9-mediated cleavage. *Nature communications*, 7, 11750.

Tou, C. J., Schaffer, D. V., Dueber, J. E. (2020). Targeted Diversification in the *S. cerevisiae* Genome with CRISPR-Guided DNA Polymerase I. *ACS synthetic biology, 9*(7), 1911–1916.

Tsai C-S, Kong II, Lesmana A. 2015. Rapid and marker-free refactoring of xylose-fermenting yeast strains with Cas9/CRISPR. *Biotechnol Bioeng 112*: 2406–11.

Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S., & Gregory, P. D. (2010). Genome editing with engineered zinc finger nucleases. *Nature reviews. Genetics*, *11*(9), 636–646.

Van Dyk, J. S., & Pletschke, B. I. (2012). A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes--factors affecting enzymes, conversion and synergy. *Biotechnology advances*, *30*(6), 1458–1480.

Van Rensburg, P., Van Zyl, W. H., & Pretorius, I. S. (1998). Engineering yeast for efficient cellulose degradation. *Yeast (Chichester, England), 14*(1), 67–76.

Van Wyk, N., Kroukamp, H., Espinosa, M. I., von Wallbrunn, C., Wendland, J., & Pretorius, I. S. (2020). Blending wine yeast phenotypes with the aid of CRISPR DNA editing technologies. *International journal of food microbiology, 324*, 108615.

Van Zyl, J. H., Den Haan, R., & Van Zyl, W. H. (2014). Over-expression of native *Saccharomyces cerevisiae* exocytic SNARE genes increased heterologous cellulase secretion. *Applied microbiology and biotechnology, 98*(12), 5567–5578.

Van Zyl, W. H., Lynd, L. R., Den Haan, R., & McBride, J. E. (2007). Consolidated bioprocessing for bioethanol production using *Saccharomyces cerevisiae*. *Advances in biochemical engineering/biotechnology, 108*, 205–235.

Vanegas, K. G., Lehka, B. J., & Mortensen, U. H. (2017). SWITCH: a dynamic CRISPR tool for genome engineering and metabolic pathway control for cell factory construction in Saccharomyces cerevisiae. *Microbial cell factories*, *16*(1), 25.

Verkuijl, S. A., & Rots, M. G. (2019). The influence of eukaryotic chromatin state on CRISPR-Cas9 editing efficiencies. *Current opinion in biotechnology*, *55*, 68–73.

Verwaal, R., Wang, J., Meijnen, J. P., Visser, H., Sandmann, G., van den Berg, J. A., & van Ooyen, A. J. (2007). High-level production of beta-carotene in Saccharomyces cerevisiae by successive transformation with carotenogenic genes from Xanthophyllomyces dendrorhous. *Applied and environmental microbiology*, *73*(13), 4342–4350.

Waldrip, Z. J., Jenjaroenpun, P., DeYoung, O., Nookaew, I., Taverna, S. D., Raney, K. D., & Tackett, A. J. (2020). Genome-wide Cas9 binding specificity in *Saccharomyces cerevisiae*. *PeerJ, 8*, e9442.

Wang, G., Huang, M., & Nielsen, J. (2017). Exploring the potential of *Saccharomyces cerevisiae* for biopharmaceutical protein production*. Current opinion in biotechnology, 48*, 77–84.

Wang, L., Deng, A., Zhang, Y., Liu, S., Liang, Y., Bai, H., Cui, D., Qiu, Q., Shang, X., Yang, Z., He, X., & Wen, T. (2018). Efficient CRISPR-Cas9 mediated multiplex genome editing in yeasts. *Biotechnology for biofuels, 11*, 277.

Weninger, A., Hatzl, A. M., Schmid, C., Vogl, T., & Glieder, A. (2016). Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast Pichia pastoris. *Journal of biotechnology*, *235*, 139–149.

Wiedemann, B., and Boles, E. (2008). Codon-optimized bacterial genes improve L-Arabinose fermentation in recombinant Saccharomyces cerevisiae. *Applied and environmental microbiology*, *74*(7), 2043–2050. https://doi.org/10.1128/AEM.02395-07

Wingler, L. M., and Cornish, V. W. 2011. Reiterative Recombination for the *in vivo* assembly of libraries of multigene pathways. *Proceedings of the National Academy of Sciences of the United States of America, 108*(37), 15135–15140.

Wu, X. L., Li, B. Z., Zhang, W. Z., Song, K., Qi, H., Dai, J. B., & Yuan, Y. J. (2017). Genome-wide landscape of position effects on heterogeneous gene expression in *Saccharomyces cerevisiae*. *Biotechnology for biofuels, 10*, 189.

Xu, Q., Singh, A., & Himmel, M. E. (2009). Perspectives and new directions for the production of bioethanol using consolidated bioprocessing of lignocellulose. *Current opinion in biotechnology, 20*(3), 364–371.

Yamada, R., Hasunuma, T., & Kondo, A. (2013). Endowing non-cellulolytic microorganisms with cellulolytic activity aiming for consolidated bioprocessing. *Biotechnology advances, 31*(6), 754–763.

Yamada, R., Nakatani, Y., Ogino, C., & Kondo, A. (2013). Efficient direct ethanol production from cellulose by cellulase- and cellodextrin transporter-co-expressing *Saccharomyces cerevisiae*. *AMB Express, 3*, 34.

Yamada, R., Tanaka, T., Ogino, C., & Kondo, A. 2010. Gene copy number and polyploidy on products formation in yeast. *Applied microbiology and biotechnology, 88*(4), 849–857.

Yan, Y., and Finnigan, G. C. (2018). Development of a multi-locus CRISPR gene drive system in budding yeast. *Scientific reports*, *8*(1), 17277.

Yu, Y., Chang, P., Yu, H., Ren, H., Hong, D., Li, Z., Wang, Y., Song, H., Huo, Y., & Li, C. (2018). Productive Amyrin Synthases for Efficient α-Amyrin Synthesis in Engineered *Saccharomyces cerevisiae*. *ACS synthetic biology, 7(*10), 2391–2402.

Zalatan, J. G., Lee, M. E., Almeida, R., Gilbert, L. A., Whitehead, E. H., La Russa, M., Tsai, J. C., Weissman, J. S., Dueber, J. E., Qi, L. S., & Lim, W. A. (2015). Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell*, *160*(1-2), 339–350.

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., van der Oost, J., Regev, A., Koonin, E. V., & Zhang, F. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, *163*(3), 759–771.

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E. M., Winblad, N., Choudhury, S. R., Abudayyeh, O. O., Gootenberg, J. S., Wu, W. Y., Scott, D. A., Severinov, K., van der Oost, J., & Zhang, F. (2017). Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nature biotechnology*, *35*(1), 31–34.

Zhang, G. C., Kong, I. I., Kim, H., Liu, J. J., Cate, J. H., & Jin, Y. S. (2014). Construction of a quadruple auxotrophic mutant of an industrial polyploid saccharomyces cerevisiae strain by using RNA-guided Cas9 nuclease. *Applied and environmental microbiology*, *80*(24), 7694–7701.

Zhang, Y., Wang, J., Wang, Z., Zhang, Y., Shi, S., Nielsen, J., & Liu, Z. (2019). A gRNA-tRNA array for CRISPR-Cas9 based rapid multiplexed genome editing in *Saccharomyces cerevisiae*. *Nature communications, 10*(1), 1053.

6. APPENDIX 1: SUPPLEMENTARY FIGURES

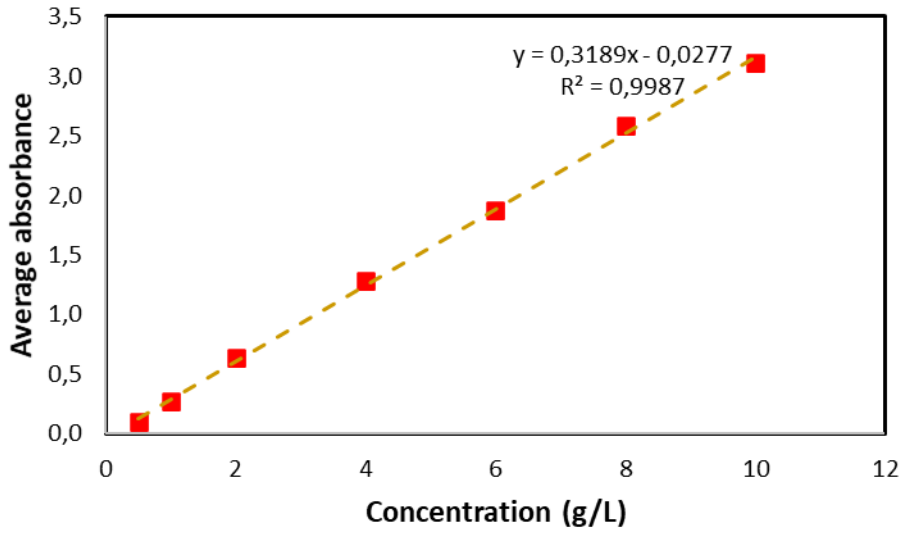


Figure S1: Calibration curve of glucose used to determine unknown concentrations of reducing sugars from which EG activity was determined

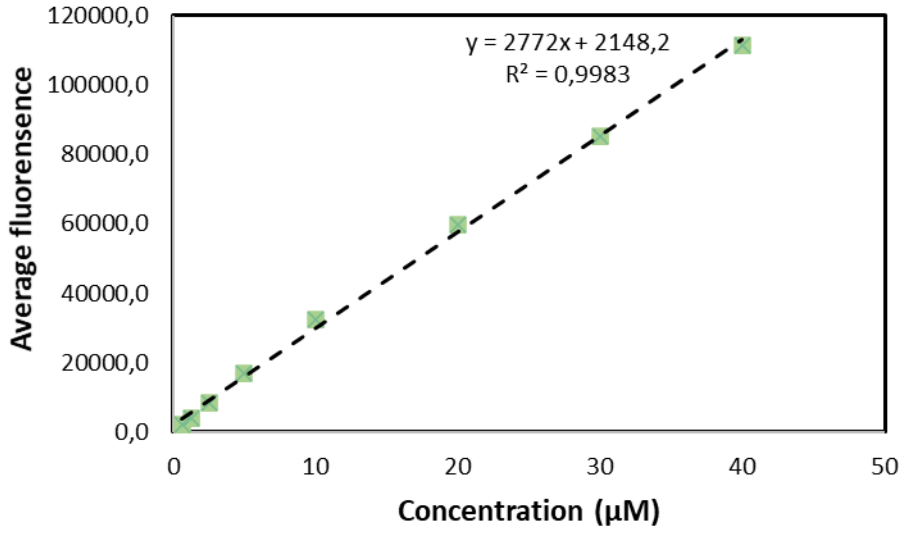


Figure S2: Calibration curve of 4-Methylumbelliferyl used to determine unknown concentrations of released methylumbelliferyl from which CBH activity was determined.

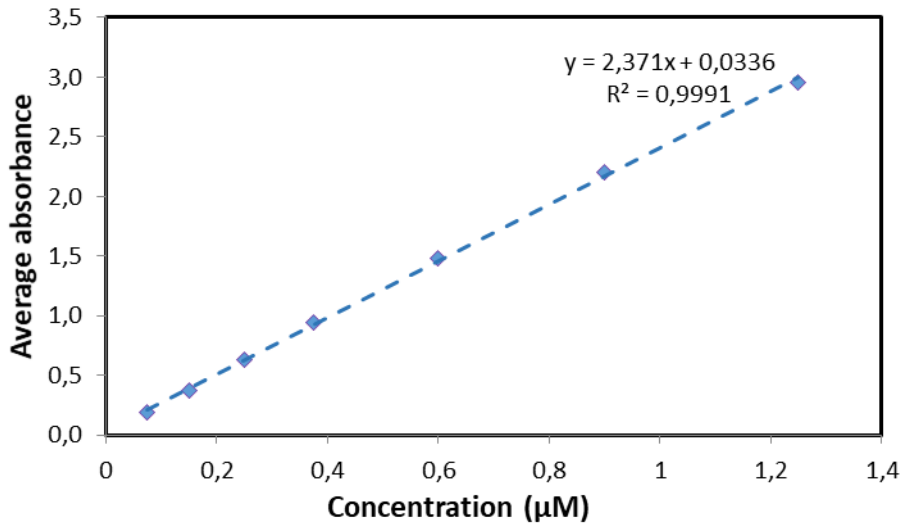


Figure S3: Calibration curve of 4-Nitrophenyl used to determine unknown concentrations of released 4-NP from which BGL activity was determined.

7. APPENDIX 2: ADDITIONAL qPCR DETAILS FOR CALCULATION OF GENE COPY NUMBER

Copy number calculation for *T.r.eg2*

Size used=	Plasmid	8691bp
	Insert	1260bp
	Total	**9951bp**

Calculation used

Genome copy # = DNA (g) / (g_to_bp const. x genome size)

With DNA (ng) = 1 ng

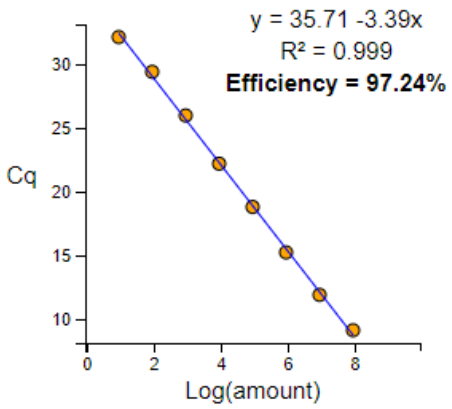
g to bp const = $1.096 \times 10^{\wedge -21}$ g

genome size = 9951bp

$9.169 \times 10^{\wedge 7}$ copies in 1ng 91690000

Copies	Concentration
9.169	0.01fg
91.69	1fg
916.9	0.01pg
9169	0.1pg
91690	1pg
916900	10pg
9169000	0.1ng
91690000	1ng

Standard curve with the above dilutions and all the samples were calculated based on the below curve.



Copy number calculation for *T.e.cbh1*

Size used=	Plasmid	7881bp
	Insert	1550bp
	Total	**9431bp**

Calculation used

Genome copy # = DNA (g) / (g_to_bp const. x genome size)

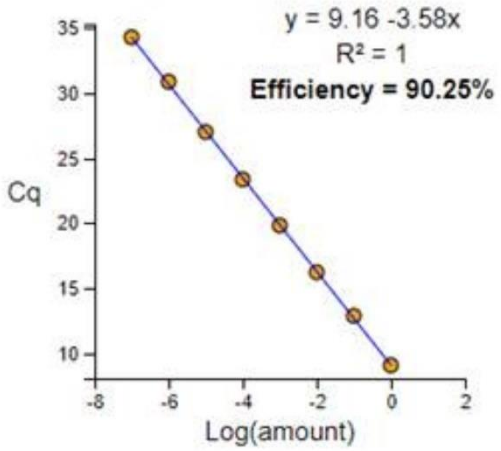
With DNA (ng) = 1 ng

g to bp const = 1.096 x $10^{\wedge -21}$ g

genome size = 9431bp

9.67 X $10^{\wedge 7}$ copies in 1ng

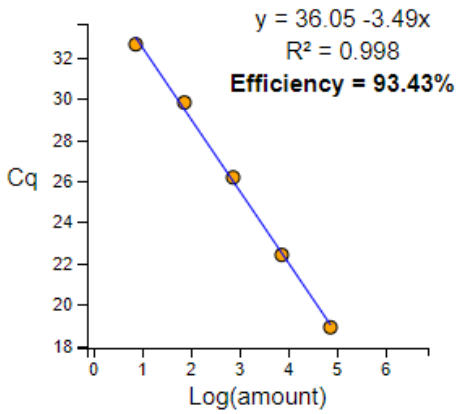
Standard curve with the 8 log dilutions (1ng to 0.1fg) and all the samples were calculated based on the below curve.



Internal control gene: ALG9

We used ~12Mb for the haploid genome.

Standard curve with the 5 log dilutions (1ng to 0.1pg) and all the samples were calculated based on the below curve.


y = 36.05 -3.49x
R² = 0.998
Efficiency = 93.43%
Cq
32
30
28
26
24
22
20
18
0
1
2
3
4
5
6
Log(amount)