The development of a wide range CRISPR-Cas9 gene editing system

by

Eduvan Bisschoff

Submitted in fulfilment of the requirements in respect of the degree

Magister Scientiae

in the

Department of Microbial, Biochemical and Food Biotechnology; Faculty of Natural and Agricultural Sciences

at the

University of the Free State

P.O. Box 339

Bloemfontein

9301

South Africa

Supervisor: Prof. J. Albertyn

Co-supervisors: Prof. C. H. Pohl-Albertyn

January 2019

UNIVERSITY OF THE
FREE STATE
UNIVERSITEIT VAN DIE
VRYSTAAT
YUNIVESITHI YA
FREISTATA

UFS·UV
NATURAL AND
AGRICULTURAL SCIENCES
NATUUR– EN
LANDBOUWETENSKAPPE

MICROBIOLOGY
MIKROBIOLOGIE

Acknowledgements

I would like to thank and acknowledge the following:

- Prof. Jacobus Albertyn for all the guidance, support, patience, motivation and hard work throughout this study; as well as for being a role model and inspiration as a scientist.
- Prof. Carlien Pohl for all the motivation, valuable input and everyday support.
- Marianka de Jonge for all the immense support, patience, motivation and inspiration.
- Marnus Du Plooy, Ruan Fourie, JTR Brink, Bianca Pieterse and Johan Klinck for valuable input as well as discussions.
- The Pathogenic Yeast Research Group for all the support and help.
- Friends and family for all the love and support throughout the last few years.

Table of Contents

Declaration

I, Eduvan Bisschoff, declare that the Master's Degree research dissertation or publishable, interrelated articles, or coursework Master's Degree mini-dissertation that I herewith submit for the Master's Degree qualification at the University of the Free State is my independent work, and that I have not previously submitted it for a qualification at another institution of higher education.

I, Eduvan Bisschoff, hereby declare that I am aware that the copyright is vested in the University of the Free State.

I, Eduvan Bisschoff, hereby declare that all royalties as regards intellectual property that was developed during the course of and/or in connection with the study at the University of the Free State, will accrue to the University.

I, Eduvan Bisschoff, hereby declare that I am aware that the research may only be published with the dean's approval.

05/ 02/ 2019

Signature   **Date**

Dissertation Summary

CRISPR is a revolutionary method to effectively and efficiently alter the genomic make-up of an organism. Unlike any other genetic engineering tool or technique, CRISPR is remarkably cheaper, simpler and faster to perform. In biotechnology, the best eukaryotic organism for research is yeast, due to their fast growth rate and ease of manipulation compared to multicellular organisms. Hence the aim of the study was the development of a wide range CRISPR-Cas9 system for a wide variety of different yeasts for easy and fast gene editing. The system were validated in *Saccharomyces cerevisiae* and six other non-conventional yeast. System construction began with the incorporation (separately) of three different optimized *CAS9* (optimized for expression in *Pichia pastoris*, *Candida albicans* and *Homo sapiens*) genes into the wide range pKM180 vector. The three different *CAS9* construct were then tested for correct expression of the Cas9 protein and the effects thereof in all the yeasts. Through western blot analysis it was observed that all three of the different Cas9 proteins were expressed successfully in the different yeasts. However, all of the Cas9 proteins had a negative effect on the growth of the yeast. For the completion of the CRISPR-Cas9 system, a Ribozyme-gRNA-Ribozyme cassette was incorporated into the wide range *CAS9* vector, containing the *C. albicans* optimized *CAS9*. The system was then validated with successful disruption of the *ADE2* gene in all of the yeasts. This proved that the wide range CRISPR-Cas9 system was applicable in a wide variety of different yeasts, thus allowing for rapid, cost-effective genetic manipulation of biotechnologically relevant yeast strains.

Keywords: *Arxula adeninivorans* ; *Debaryomyces hansenii*; *Kluyveromyces lactis*; *Ogataea polymorpha*; *Komagataella phaffii*; *Saccharomyces cerevisiae*; *Yarrowia lipolytica*; *Pichia pastoris*; *Candida albicans* optimized *CAS9*; CRISPR-Cas9; Wide range CRISPR-Cas9 system.

Chapter 1

Literature review

CRISPR-CAS9 strategies used for efficient gene targeting as well as gene disruption in yeast.

1. Introduction

The direct manipulation of an organism's genome by the usage of certain biotechnology techniques is referred to as genetic manipulation. These techniques are used to alter the genetic makeup of cells by the transference or deletion of genes within or across species boundaries, which may result in improved or novel organisms (Lamont & Lacey 2013). New DNA may be introduced into the hosts' genome by making use of molecular cloning methods, which entails isolating and copying techniques to generate a DNA sequence (Berg & Mertz 2010). The DNA can also be synthesized for the purpose of inserting the construct into the host (Kaufman & Nixon 1996). The DNA can be incorporated either indirectly with the use of a vector system or directly through micro-injection and micro-encapsulation techniques (Lacal *et al.* 1999; Chen & Dubnau 2004; Singh *et al.* 2010; Zaretsky *et al.* 2011). Genes may be altered by either gene knockout (removal) or by gene knockdown (reduced expression) using a nuclease and may also be changed through the use of gene targeting (Westphal & Leder 1997; Summerton 2007). This technique (gene trageting) is different in that it uses homologous recombination to alter a gene, and can thus be used in the removal of a gene or exons, addition of a gene, or to introduce point mutations. An example is the Cre-Lox recombination. This system consists of a single enzyme, the Cre recombinase, which recombines a pair of short target sequences called the Lox sequences. Placing Lox sequences appropriately, allow genes to be activated, repressed, or exchanged for other genes (Sauer 1987; Egener *et al.* 2002).

Genetic engineering techniques have been applied in numerous fields including research, agriculture, industrial biotechnology and medicine. Genetic engineering can be used to produce desirable traits in organisms, which can be used in many different application areas to benefit humanity. For instance genetic engineering has been used in medical applications to mass-produce insulin, human growth hormones, follistim,

human albumin, monoclonal antibodies, antihemophilic factors, vaccines and many other drugs (Walsh 2006; Waegeman *et al.* 2013; Berlec & Štrukelj 2013). Gene therapy is also an application where in (only one) defective genes in humans can be replaced with effective ones (Knudson 1967; Krauss 1992).

Genetic engineering techniques can be used to manipulate most plants, animals and microorganisms. Bacteria, the first organisms to be genetically modified in 1973, can have plasmid DNA inserted containing new genes that code for pharmaceuticals (such as insulin and human growth hormone) or enzymes such as in laundry detergent, enzymes that process food and other substrates (Jones *et al.* 1968; Cohen & Chang 1973; Williams *et al.* 1982; Gray *et al.* 1985). Plants have been modified forinsect protection, herbicide resistance, virus resistance, enhanced nutrition, tolerance to environmental pressures and the production of edible vaccines. In addition modified crops have been commercialized (Dookun 2001; Li *et al.* 2001; ISAAA 2017). Experimental genetically modified (GM) cell lines and GM animals, such as mice or zebrafish, have been used for research, model animals and the production of agricultural or pharmaceutical products. The genetically modified animals include animals with genes deleted, increased susceptibility to disease, hormones for extra growth and the ability to express proteins in their milk (Dodd *et al.* 2000; Olswang *et al.* 2002; Yoder *et al.* 2002; Hanson & Hakimi 2008).

In biotechnology, one of the best eukaryotic model organism for research is yeast due to their fast growth rate and ease of manipulation compared to multicellular organisms (Madzak *et al.* 2004; Gerngross 2004; Nevoigt 2008; Stovicek *et al.* 2015; Skrzypek *et al.* 2017). *Saccharomyces cerevisiae* is characterized as the most well-known eukaryotic system due to the amount of data available regarding the yeasts' genetics, physiology and molecular biology (Nevoigt 2008). Due to this, *S. cerevisiae* is widely used as a model organism for research. The fast growth rate, ease of use, maintenance and manipulation of this yeast makes it ideal for the development of new genetically modified strains for research and industrial purposes. However, non-conventional yeasts, i.e. non *S. cerevisiae* species, have also attracted much attention and have been widely used as cell factories to produce a number of different recombinant proteins and biomolecules for various research and industrial purposes (Van der Walt *et al.* 1990; Hsieh & Da Silva 1998; van Ooyen *et al.* 2006; Gasser *et al.* 2013). However, compared to *S. cerevisiae*, engineering many of these yeast

species are challenging, because of the limited data available on the genetics, physiology and molecular biology and/or less defined molecular tools. Thus, improved knowledge on the biology and engineering tool development for many of the non-conventional yeasts could greatly improve industrial and research applications. This review will briefly present the key developments in genome engineering and major genetic engineering tools, which include Meganucleasese, zink finger nucleases (ZFN), transcription activator-like effector nucleases (TALEN). However, most of this review will focus on the CRISPR technology and its development for use in *S. cerevisiae* and non-conventional yeast.

2. Meganucleases

Meganucleases are found in a large number of organisms and are endodeoxyribonucleases characterized by a large recognition site (double-stranded DNA sequences of 12 to 40 base pairs) (Chevalier 2001; Chevalier *et al.* 2002). Due to the exceptional long length of the recognition sequence, this region generally occurs only once in any given genome (Chevalier 2001). This property therefore characterizes meganucleases to be the most specific, natural occuring restriction enzyme.

Meganucleases also known as homing endonucleases currently have six known structural families, which are classified on their structure (LAGLIDADG; GIY-YIG; His-Cys box; H-N-H; PD-(D/E)xK and Vsr-like) (Stoddard 2011). Among the homing endonucleases, the LAGLIDADG family has become a valuable tool for genome engineering and - studying (Chevalier *et al.* 2005). By using protein engineering, through mutagenesis, to change the recognition sequence, any target in any genome can be cleaved by the enzyme (Seligman 2002). This however can pose a drawback, meaning protein-engineering needs to be performed to allow recognition of the nuclease for a specific nucleotide sequence and therefore to obtain the desired target site. In addition, protein engineering can be difficult and time consuming (Seligman 2002; Sussman *et al.* 2004; Rosen *et al.* 2006). The best characterized meganucleases, which are mostly used for genome engineering, include *I-SceI*, *I-CreI* and *I-DmoI* (Chevalier *et al.* 2005).

3. Zinc finger nucleases

Zinc fingers are defined as zinc ion-regulated small protein motifs, which bind in a sequence-specific manner to DNA (Laity *et al.* 2001). Each one of the zinc fingers recognize a 3 bp DNA sequence (Klug & Rhodes 1987). Therefore, unlike meganucleases, multiple zinc fingers (about 6-8) could be combined to achieve a larger complex which contributes to a more specific DNA binding capability.

Zinc-finger nucleases are synthetic restriction endonucleases generated by combining a zinc finger domain (DNA-binding domain) to a restriction endonuclease domain (DNA-cleavage domain). The zinc finger DNA-binding domains commonly consist of three to six separate zinc finger repeats, which can recognize between 9 and 18 bp each (Gupta & Musunuru 2014). The non-specific cleavage domain from *Fok*I (type IIs restriction endonuclease) is usually used for the ZFNs' cleavage domain (Kim *et al.* 1996). Dimerization of the cleavage domain must occur to ensure the cleavage of the targeted DNA and thus a pair of ZFNs are needed for the targeted sites (Bitinaite *et al.* 1998). ZFNs fuse the cleavage domain to the C-terminus of each zinc finger domain. In order to allow for the dimerization of the two cleavage domains to perform DNA cleavage, the two individual ZFNs must bind to the complementary DNA strands. Zinc finger (DNA binding) domains can be modified to allow targeting of a specific desired DNA sequence (Gupta & Musunuru 2014). Thus, enabling the modified ZFNs to recognize and cleave any target sequence within a desired genome. This however is time consuming due to the protein engineering that needs to be performed to alter the binding site.

4. Transcription activator-like effector nucleases

The TALEs (Transcription Activator-Like Effector) are proteins, which are secreted by the bacteria *Xanthomonas* to help them infect plants (Boch & Bonas 2010). The bacteria inject the TALEs directly into plant cells by using a type III secretion system, followed by the specific binding of the TALEs to the plant genes, which regulates it to facilitate the bacterial colonization. Transcription activator-like effector nucleases (TALEN) are restriction endonucleases that can be modified to perform cleavage on a specific target DNA sequence (Boch 2011). They are generated by combining a transcription activator-like effector (TALE) DNA-binding domain to a DNA cleavage

domain (*Fok*I) (Boch & Bonas 2010). The TALE DNA binding domain commonly consists of 33 to 34 highly conserved repeated amino acid sequence (Boch *et al.* 2009; Moscou & Bogdanove 2009). TALEs can be modified, at the 12th and 13th amino acid to bind to nearly any target DNA sequence. Thus when combined with the restriction endonuclease domain, the DNA can be cleaved at a specific location (Boch 2011). TALEN also has the same drawbacks as ZFN and meganucleases, which includes protein engineering.

5. CRISPR-Cas9

Although the discovery of the three genetic engineering tools, meganucleases followed by ZFNs and TALENs, continuously increased the efficacy of genome editing, the targeting of different sites in the host genome required redesigning and rebuilding of new sets of proteins (Seligman 2002; Boch 2011; Gupta & Musunuru 2014). This drawback partially prevented ZFNs and TALENs to be broadly accepted by the scientific community (Sakuma & Woltjen 2014). In this respect, CRISPR has reformed the genetic engineering field due to its phenomenal editing efficiency compared to the existing tools (Figure 1) (Adli 2018). More importantly, CRISPR is also much simpler, more flexible, faster, and cheaper to use. The CRISPR gene-editing tool consists of a programmable endonuclease whose DNA-targeting specificity and cutting activity depends on a short guide RNA.
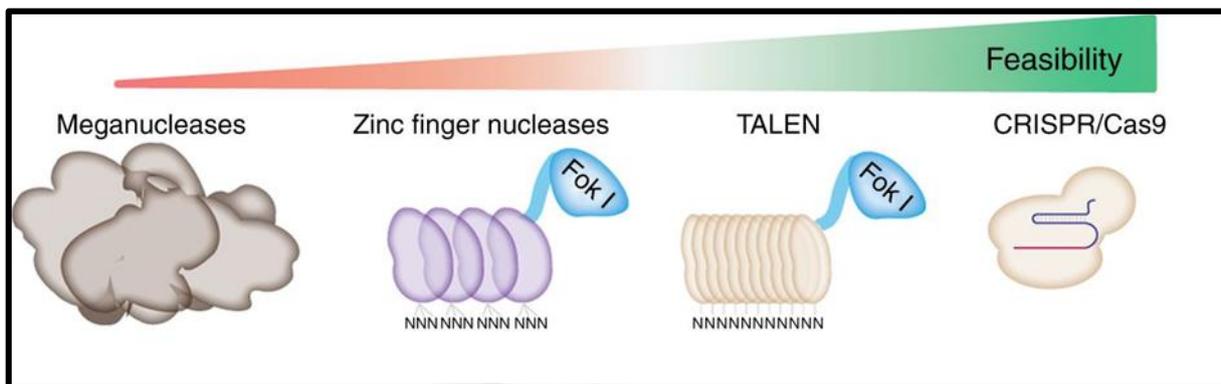
Figure1: Feasibility of the major genetic engineering tools (Adli 2018).

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) is a sequence in the bacterial genome which contains parts of foreign genetic material, such as those found within plasmids and bacteriophages, which have invaded a bacterium (Terns & Terns 2011; Bhaya *et al.* 2011; Wiedenheft *et al.* 2012). These foreign DNA parts (spacer) are then used by the host bacterium to target DNA from similar viruses during

following attacks by using Cas (CRISPR Associated Systems) proteins. This system is known as CRISPR-Cas and it plays a key role in the bacterial defence system. The application of the recently discovered CRISPR-Cas9 system is an innovative, genetic engineering tool that allows scientists to change, at will, any DNA sequence of any living organism in a specific manner (Jinek *et al.* 2012; Barrangou 2015). The Cas9 protein can be modified with a single guide RNA (sgRNA) to generate site-specific DNA breaks.

5.1. Mechanism

The CRISPR-Cas9 system helps to protect bacteria in three steps, namely adaptation, production of gRNA/Cas complex and finally targeting (Figure 2) (Terns & Terns 2011; Bhaya *et al.* 2011; Wiedenheft *et al.* 2012). Adaptation or spacer acquisition is the process where the Cas enzymes (Cas1 and Cas2) recognize the invading DNA and cleave it to produce a protospacer (Figure 2A). This protospacer is then ligated and extended into the CRISPR array adjacent to the leader sequence, to the direct repeat thus creating a new CRISPR part to serve as a tool for targeting similar foreign DNA in later infections.

The production of the gRNA/Cas complex step starts with the transcription of the CRISPR array to yield the pre-crRNA as indicated by Figure 2B (CRISPR RNA) (Marraffini & Sontheimer 2010; Dugar *et al.* 2013). This is followed by the binding of various trans-activating (tracr) RNAs to the pre-crRNA, resulting in double stranded RNAs. The Cas9 protein then associates with the tracrRNA-pre-crRNA strand, which is then recognised and cleaved, to generate small crRNAs, by RNaseIII. The crRNAs then undergo secondary trimming, at either the 5-prime or 3-prime ends to produce mature crRNAs (Marraffini & Sontheimer 2010; Dugar *et al.* 2013; Karvelis *et al.* 2013). Both the mature crRNAs and tracrRNA must associate with the Cas9 protein to form an active interference complex (Jinek *et al.* 2012). The crRNA-Cas9 complex then finds the invading DNA (Figure 2C), binds to it, recognises the PAM (Protospacer adjacent motif) sequence downstream of the protospacer, and cleave the invading DNA (Garneau *et al.* 2010; Jinek *et al.* 2012). The PAM site is an essential targeting component, which is not found in the bacterial genome (Mali *et al.* 2013a). This enables the bacteria to distinguish between his own DNA from foreign DNA, hence avoiding the possibility of self-targeting and degradation of the CRISPR locus by Cas

enzymes. Researchers saw the mechanism and modified it to serve as a genetic engineering tool (Jinek *et al.* 2012). However this prosess took approximately 30 years from the discovery of its first components (Ishino *et al.* 1987).
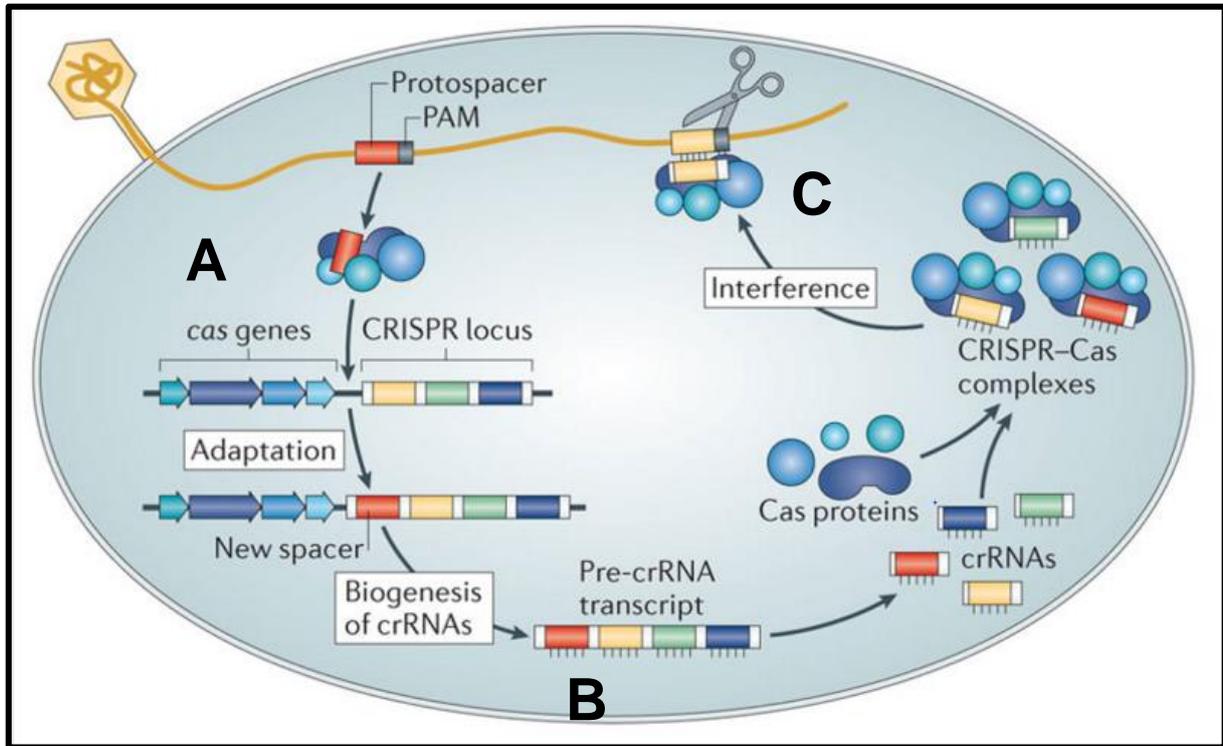
Figure 2: CRISPR-Cas9 adaptive immunity (Samson *et al.* 2013). The three crucial steps are (A) adaptation, (B) production of gRNA/Cas complex and then (C) targeting

5.2. History

Notably, CRISPR had been simply known as prokaryotic DNA repeat elements, from the genome of *Escherichia coli*, since it was first discovered in the late 1980s (Ishino *et al.* 1987). Later it was detected that CRISPRs are present in numerous bacteria and archaea (Mojica *et al.* 2000). Shortly after, the association of *CAS* genes with CRISPR were identified (Jansen *et al.* 2002). In 2005 an observation was made, which describes that many spacer sequences within CRISPR are derived from viral and/or plasmid origins (Mojica *et al.* 2005; Pourcel *et al.* 2005). These findings later led to the discovery that recognised CRISPR-Cas as a bacterial immune system (Barrangou *et al.* 2007). Later, in 2008, it was revealed that the Cas proteins are guided by small mature CRISPR RNAs (crRNAs), transcribed from spacer sequences, towards the

invading DNA to interfere with proliferation in *E. coli* (Brouns *et al.* 2008; Marraffini & Sontheimer 2008).

Around 2010, Garneau and colleagues discovered that Cas9 is the only enzyme within the gene cluster that can cleave DNA (Garneau *et al.* 2010). Next a key component in the biogenesis and processing of crRNA in a CRISPR system was revealed, describing a noncoding trans-activating crRNA (tracrRNA) hybridizing with crRNA to facilitate RNA-guided targeting of Cas9 (Deltcheva *et al.* 2011). Therefore, the Cas9 and RNase III enzymes are crucial for the processing of mature crRNA. These two above mentioned findings suggested that the Cas9, together with the mature crRNA and trancrRNA are crucial elements of the CRISPR-Cas9 system (Garneau *et al.* 2010; Deltcheva *et al.* 2011). The idea subsequently arose that this system, if engineered correctly, may be utilised for genome editing. In 2011, Sapranauskas and co-workers proved that the type II CRISPR system from *Streptococcus thermophiles* can be transferred to *Escherichia coli* and still perform its function (Sapranauskas *et al.* 2011). In 2012 the groups of Charpentier, Doudna, and Siksnys showed that the *CAS9* can be purified from *Streptococcus thermophilus* or *Streptococcus pyogenes* and can be guided by a site-specific crRNA fused to a tracrRNA to cleave a target DNA *in vitro* (Jinek *et al.* 2012; Gasiunas *et al.* 2012). In 2013 CRISPR was used to accomplish genome editing in human and mouse cells (Cong *et al.* 2013; Mali *et al.* 2013b). Since these initial studies, the CRISPR-Cas9 tool has been widely used for genome engineering in various model systems.

5.3. CRISPR –Cas9 strategies

The Cas9 enzyme consist of two nuclease domains, namely RuvC and HNH, which is based on homology of known nuclease structures (Haft *et al.* 2005; Makarova *et al.* 2006). The RuvC domain cleaves the non-target DNA strand, where the PAM site are located, and the HNH domain cleaves the target strand of DNA, strand to which the gRNA binds (Barrangou *et al.* 2007; Gasiunas *et al.* 2012).

Normally the *CAS9* and the gRNA are expressed in different expression cassettes. The DNA sequence of *CAS9* gene has been optimized by various authors to test for optimal expression and function. The *CAS9* gene have showed some success when it was either native (from *Streptococcus pyogenes*), *Homo sapiens* codon-optimized or yeast codon-optimized over a variety of different yeast, respectively (DiCarlo *et al.*

2013; Gao & Zhao 2014; Ryan *et al.* 2014; Zhang *et al.* 2014; Bao *et al.* 2015; Horwitz *et al.* 2015; Generoso *et al.* 2016). The *CAS9* expression cassette need to consist of a compatible promoter and terminator, to drive and terminate expression respectively, and a NLS (nuclear localisation signal), which can be linked to the 3-prime or both 5- and 3-prime side of the *CAS9,* to transport the Cas9 enzyme to the nucleus.

The gRNA is a short synthetic RNA that consist of scaffold part, which is necessary for the association with the Cas9 enzyme, and a user-specified targeting sequence (approximately 20 bp of length) which allows targeting of the desired DNA region in the host genome to be modified (Jinek *et al.* 2012). Therefore, by just changing the approximately 20 bp target sequence in the gRNA, one can simply change the genomic target. The gRNA expression cassette is usually driven by a non-mRNA polymerase promoter, for example, RNA polymerase III (Pol III) promoters (U3 and U6) (DiCarlo *et al.* 2013; Zhang *et al.* 2014; Jakočiūnas *et al.* 2015; Mans *et al.* 2015; Bao *et al.* 2015). The promoter is followed by a specific target sequence, the gRNA scaffold and lastly the terminator. However the Pol III promoters have limitations when it comes to the expression of gRNA in different hosts (Gao & Zhao 2014). They are very specific for each organism and not all these promoters have been characterized in many organisms, thus making it difficult to choose the correct Pol III promoters for CRISPR. Furthermore, the Pol III promoters limit the CRISPR target sequences to G (N20) GG and A (N20) GG. Therefore, another strategy was developed and used successfully for the correct expression of the gRNA. Gao and Zhao developed a strategy where they took advantage of the nuclease activity of ribozymes (Gao & Zhao 2014). They designed an artificial gene, RGR (Ribozyme-gRNA-Ribozyme), where the RGR gene undergoes self-catalysed cleavage to yield the gRNA without any modifications. The RGR gene was expressed under an alcohol dehydrogenase 1 (*ADH*1) promoter and was successful in guiding the Cas9 to the target site. In addition, this RGR strategy diminished the limit to what the target sequence must contain. The gRNA expression construct can thus contain any RNA polymerase II (Pol II) promoter followed by a Hammerhead (HH) self-splicing ribozyme, the target sequence, the gRNA scaffold, a Hepatitis delta virus (HDV) self-splicing ribozyme and the terminator. Hence, the self-splicing ribozymes were used in the construct to remove any post-transcriptional modifications on both ends. The HH promotes cleavage on its 3-prime side and the HDV promotes cleavage on its 5-prime side. On the 5-prime side of the

Hammerhead, six base pairs are added that are the reverse complement of the first six base pair of the CRISPR site. These two regions will subsequently combine allowing the spicing of the gRNA, resulting in a RNA strand that does not contain any post transcriptional modifications.

5.4. Double stranded break (DSB) repair

After cleavage of the DNA by the Cas9, the host cell will attempt to repair the DSB (Figure 3). Due to the fact that the DSBs are harmful to the host cell, the cell contains mainly two classes of DNA repair mechanisms (Pâques & Haber 1999; Sung & Klein 2006; Bétermier *et al.* 2014). These include non-homologous end joining (NHEJ) and homology directed repair (HDR) (Figure 3). NHEJ can be used to knockout genes when the aim is to introduce a mutation (Pitcher *et al.* 2007; Bétermier *et al.* 2014). These mutations can be of variable lengths due to the insertion and/or deletion of bases (indels). This repair method is by far the most common when it comes to DSB (when no repair template is available) and it also has an advantage in nuclease-induced breaks. The effectiveness of mutations are more likely to happen seeing that perfectly re-joined breaks will most likely be cleaved again, until they obtain an indel (DiCarlo *et al.* 2013; Bétermier *et al.* 2014; Mans *et al.* 2015). Insertion of an indel results in the change of the sequence, which in turn can no longer be cleaved. However, this repair method decreases the survival of the cells.
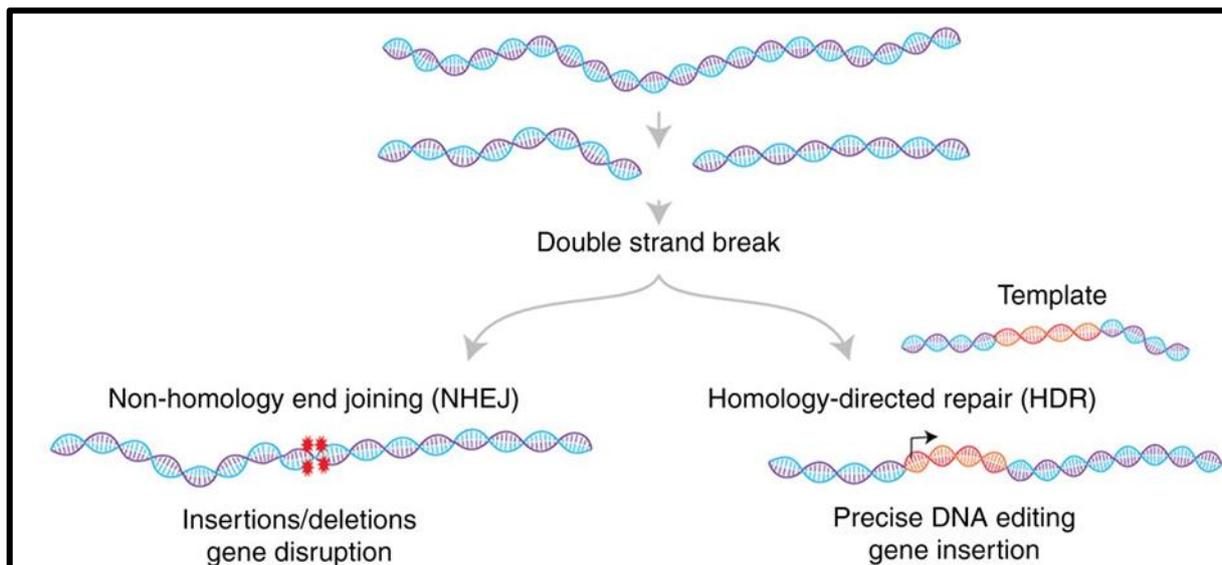
Figure 3: Double strand break repair via HDR and NHEJ (Adli 2018).

HDR relies on homologous recombination when a homologous donor DNA template is provided (Pâques & Haber 1999; Sung & Klein 2006). The donor template can be of various lengths (with homology to the target sequence) and can be modified to contain any sequence to disrupt or repair the gene of interest (DiCarlo *et al.* 2013; Blazeck *et al.* 2014; Horwitz *et al.* 2015; Richardson *et al.* 2016; Schwartz *et al.* 2016). Therefore, this can lead to the introduction of precise alterations to the genome, which are specified by the template. However, the donor template should not contain the target sequence or the PAM site, to prevent cleavage from the Cas9 (DiCarlo *et al.* 2013). In addition, the template can be introduced as a separate entity or transported as part of an expression vector (DiCarlo *et al.* 2013; Bao *et al.* 2015; Horwitz *et al.* 2015; Garst *et al.* 2016).

5.5. CRISPR interference (CRISPRi)

Another technique, CRISPR interference (CRISPRi), was reported by Qi and co-workers 2013 (Qi *et al.* 2013). They generated an inactive version of the Cas9 enzyme, which does not contain any nuclease activity. They mutated both the nuclease sites (D10A and H840A), and observed that this inactive Cas9 enzyme can still bind to a target sequence in the genome. Thus, the idea arose to use this inactive Cas9 (dCas9), when expressed with a gRNA, to generate a DNA complex to interfere with transcription, whether it be elongation, binding of the RNA polymerase or the transcription factor. CRISPRi can be modified to be used for both activation and repression of multiple targets simultaneously, and its effects are reversible (Qi *et al.* 2013; Farzadfard *et al.* 2013; Zalatan *et al.* 2015; Chavez *et al.* 2015). Thus, this system provides an ideal approach to disturb gene expression for study and have also been adapted for the use in yeast. In this sense, CRISPRi shares traits with the small interfering RNA (siRNA) tool for gene silencing in eukaryotes (Unniyampurath *et al.* 2016). However, instead of preventing the transcription factors and enzymes from binding, the siRNA binds to and cleaves the mature mRNA within the cytosol. Nevertheless both tools serves as valid candidates for gene silencing.

6. CRISPR-Cas9 strategies in various yeast

Due to aforementioned reasons, various strategies have been tested in various yeast to develop a functional CRISPR-Cas9 gene editing system. Here we will discuss

strategies used in *Saccharomyces cerevisiae, Kluyveromyces lactis, Yarrowia lipolytica, Komagataella phaffii, Kluyveromyces marxianus, Ogataea polymorpha* and *Ogataea parapolymorpha*.

6.1. *Saccharomyces cerevisiae*

6.1.1. Cas9

In *S. cerevisiae*, researchers preferentially use the *CAS9* gene from *S. pyogenes* (DiCarlo *et al.* 2013; Gao & Zhao 2014; Ryan *et al.* 2014; Zhang *et al.* 2014; Jakočiūnas *et al.* 2015; Mans *et al.* 2015; Bao *et al.* 2015; Horwitz *et al.* 2015; Generoso *et al.* 2016). Only Xu and co-workers described the use of *Streptococcus thermophilus* CRISPR3 loci-encoded *CAS9* (recognizing a different PAM site ,NGGNG), albeit with much lower engineering efficiency (Xu *et al.* 2015). Most commonly, the expression of *CAS9* was placed under the control of different strength constitutive promoters from either self-replicating low-copy centromeric -, high-copy 2μ - or integration vectors (DiCarlo *et al.* 2013; Gao & Zhao 2014; Ryan *et al.* 2014; Zhang *et al.* 2014; Jakočiūnas *et al.* 2015; Mans *et al.* 2015; Bao *et al.* 2015; Horwitz *et al.* 2015; Generoso *et al.* 2016). However, expression by a high-copy vector, using strong constitutive promoter, resulted in a poor growth of some yeast strains (Ryan *et al.* 2014; Generoso *et al.* 2016). However, this problem was not observed in other studies when the same strategy for *CAS9* expression was used (Gao & Zhao 2014; Bao *et al.* 2015). Nonetheless the toxicity problem the Cas9 protein poses could be avoided by expressing with weaker promoters (Ryan *et al.* 2014; Generoso *et al.* 2016). Apart from the toxicity, the way of expression and optimization of *CAS9* does not seem to be crucial for CRISPR-Cas9 engineering in *S. cerevisiae*.

6.1.2. gRNA

Design, expression, and delivery of the gRNA components are crucial parameters for successful CRISPR-Cas9 engineering (Jinek *et al.* 2012). For this yeast, to ensure abundant expression of the chimeric gRNA molecule, the gRNA construct was most commonly expressed using a high-copy vector (DiCarlo et al. 2013; Gao and Zhao 2014; Ryan et al. 2014; Zhang et al. 2014; Jakočinas et al. 2015; Mans et al. 2015; Bao et al. 2015; Horwitz et al. 2015; Generoso et al. 2016). A requirement for a functional Cas9-gRNA complex is that both ends of the gRNA molecule must be highly

specific (Jinek *et al.* 2012). Three most common strategies has been reported to successfully transcribe the gRNA molecule. Firstly a Pol III promoter were provided with a transcript containing a leader sequence, which are cleaved during gRNA maturation, were used (DiCarlo *et al.* 2013; Farzadfard *et al.* 2013). This strategy of expressing the gRNA cassette using a Pol III promoter was demonstrated when expression were driven using a *SNR52* promoter and terminated with a *SUP4* terminator. This resulted in a RNA molecule containing no post transcriptional modifications, similar to that found in prokaryotes (Wang & Wang 2008). This was used for the successful targeting of a single gene in a haploid or diploid laboratory strains (engineering efficiencies up to 100%), various industrial strains (engineering efficiencies from 65-78%) and polyploid strains (between 15-60% engineering efficiencies) (DiCarlo *et al.* 2013; Zhang *et al.* 2014; Jakočiūnas *et al.* 2015; Mans *et al.* 2015; Horwitz *et al.* 2015; Laughery *et al.* 2015; Generoso *et al.* 2016). It is noteworthy that engineering efficiencies discussed are defined as the number of clones with the desired edit compared to the number of transformants obtained. The second strategy includes the use of a Pol III promoter, which contains *cis*-regulatory elements within the mature RNA molecule (tRNA) combined with a ribozyme (Ryan *et al.* 2014). This results in the cleavage of the transcript on its 5-prime end. This was performed with the expression of a gRNA molecule fused to a *Hepatitis delta* virus (HDV) self-splicing ribozyme, which is driven by a tRNA promoter and *SNR52* terminator. This construct resulted in almost 100% gene knock-out efficiency in a diploid strain (laboratory), and for the polyploid strain (industrial), more than 90% efficiency was achieved. For the third mentioned strategy, a Pol II promoter followed by the flanking of the gRNA by two self-splicing ribozymes was used (Gao & Zhao 2014). This was demonstrated in a laboratory strain, with efficient gene disruption, with a construct containing a gRNA molecule flanked with a Hammerhead (HH) on its 5-prime side and HDV ribozymes on the 3-prime end. In addition, an *ADH1* promoter was used to regulate expression. Besides the chimeric gRNA approach, separate expression of a targeting crRNA array driven by a Pol III promoter, processed by native RNA processing enzymes, and tracrRNA transcribed from another Pol III promoter has been reported (Bao *et al.* 2015). This method resulted in engineering efficiencies from 76%-100% in a laboratory strain.

6.1.3. Multiplexing

The efficient HDR mechanism in *S. cerevisiae* allows for multiple CRISPR-Cas9 targets simultaneously (Ryan *et al.* 2014; Jakočiūnas *et al.* 2015; Mans *et al.* 2015; Bao *et al.* 2015; Horwitz *et al.* 2015; Lee *et al.* 2015). However, for each genome target, an individual gRNA and a donor template are required for successful editing. There were several strategies demonstrated to express multiple gRNAs. First Mans and co-workers constructed three different vectors, each containing a different selection marker and up to two different gRNA expression cassettes (Mans *et al.* 2015). This was co-transformed with donor DNA, which resulted in 100%, 70% and 65% gene knock out efficiencies of two, four or six genes targeted, respectively. The second strategy demonstrated was a single expression construct, containing several gRNA cassettes (Ryan *et al.* 2014; Jakočiūnas *et al.* 2015). A single expression vector containing 5 separate gRNAs resulted in efficiencies ranging between 50-100% (Jakočiūnas *et al.* 2015). Ryan and co-workers reported efficiencies of 86% and 81% in haploid and 43% and 19% in diploid strains when two or three genes were targeted, respectively, with a HDV-gRNA expression cassette (Ryan *et al.* 2014). Note that the expression vector was co-transformed with donor DNA, which had 50 bp homologous overlaps corresponding to the target locus. Bao and colleagues also demonstrated a multiplexing method where an array of different interspaced crRNAs were expressed (Bao *et al.* 2015). They targeted three different genes, which resulted in engineering efficiencies ranging from 27-100%. Lastly Horwitz and co-workers used a method where they transform different linear gRNA expression cassettes together with a single gapped expression vector (Horwitz *et al.* 2015). They used the gap repair method to transform three different gRNA cassettes together with the single open vector and donor DNA, with 500 bp homologous ends, which resulted in three-gene deletion mutants (64% efficiency).

6.1.4. CRISPRi

In the context of metabolic engineering and functional genomics, targeted regulation of gene expression is important. Since the initial development of the CRISPR method advances have been made to adapt this system for activation and repression of gene transcription in *S. cerevisiae.* Gilbert et al. 2013 showed that repression can be further enhanced by combining a repressor domain to the dCas9 (Gilbert *et al.* 2013). They

tested the method when they targeted a *TEF1* promoter driving *GFP* expression and observed an 18-fold reduction in GFP fluorescence. When they fused the dCas9 to the mammalian transcriptional repressor domain Mxi1, a 53-fold reduction in fluorescence was observed. Another group instead fused an activator domain (VP64) to the dCas9 (Farzadfard *et al.* 2013). This allowed for both repression and activation of the gene targeted, depending on the targeting site in the promoter region. When this complex targeted a region upstream the TATA box of the minimal *CYC1m* promoter, the promoter was activated to achieve an activation level of max 2.5 fold. For a higher activation level, a synthetic promoter was created by arraying a number of operators upstream of the *CYC1m* promoter. The activation level increased proportionally to the number of operators. When 12 operators were used, the activation reached a level of 70 fold. On the other hand, when the target sites changed to adjacent to the TATA box or the transcriptional start site, the expression of the *CYC1m* promoter was repressed. Chavez and co-workers fused a tripartite activator consisting of VP64, p65, and Rta (VRP) with the dCas9 (Chavez *et al.* 2015). This fusion was tested on the *HED1* and *GAL7* promoters and resulted in a 38- and 78 fold increase in activation, respectively. Fusion of dCas9 with VP64 only gave 9- and 14 fold activation of the same promoters. Zalatan and co-workers approached to targeted up- and down-regulation, using the dCas9, differently (Zalatan *et al.* 2015). They combined the gRNA with effector protein recruitment domains and expressed the dCas9 and regulation proteins, which are fused to RNA-binding domains. They called this gRNA complex, containing protein recruitment properties, the scaffold RNA (scRNA). This complex resulted in a 20-50 fold increase, when scRNA binding VP64 activation domain was used. This is much higher compared to the achieved level when the dCas9- VP64 fusion complex was used. They also showed that when several hairpins are combined in a single scRNA, it could amplify activation or combine activation and repression of different sites.

6.2. *Kluyveromyces lactis*

Kluyveromyces lactis is well known for its ability to produce β-galactosidase and has also been used as an expression host for the production of the milk clotting enzyme bovine chymosin (van den Berg *et al.* 1990). This yeast is also used to commercially produce the native enzyme lactase and some metabolites (van Ooyen *et al.* 2006).

Horwitz and co-workers validated CRISPR-Cas9 editing in an industrial strain of *K. lactis*. The 2μ element present in an *S. cerevisiae* expression vector was exchanged for the *K. lactis* specific pKD1 vector-stabilizing element (Horwitz *et al.* 2015). They deleted the *KU80* gene in the yeast to reduce the NHEJ activity. Numerous studies have shown that when the *KU70* and *KU80* (critical genes for NHEJ) genes are deleted, a decrease in NHEJ and an increase in HDR are observed (Boulton 1996; Daley *et al.* 2005; Verbeke *et al.* 2013; Kretzschmar *et al.* 2013; Juergens *et al.* 2018a). The method allowed for integration of three six-gene-DNA parts, with low efficiency, into three separate chromosomal loci (Horwitz *et al.* 2015).

6.3. Yarrowia lipolitica

Yarrowia lipolytica is the most studied oleaginous yeast and has attracted the attention of industry and researchers due to its extraordinary biotechnological potential and its application in the biotechnology industry with the production of several types of metabolites, such as mannitol, γ-decalactone citric acid, intracellular lipids, and lipase (Gonçalves *et al.* 2014).

Several recent studies have revealed the great potential the yeast has in CRISPR-Cas9 technology (Schwartz *et al.* 2016, 2017a; Gao *et al.* 2016). Schwartz and colleagues constructed a centromeric vector containing a hybrid *SCR1´*-tRNA promoter for gRNA expression and a *Yarrowia* codon-optimized *CAS9* (Schwartz *et al.* 2016). This vector successfully deleted the *KU80* gene with high efficiency. HDR-mediated deletions, with high efficiency, were also obtained when donor DNA homologous (1000 bp homologous overlapping ends) to the gene of interest were transformed with the vector into a *KU70* mutant strain, lacking the ability to perform NHEJ (Schwartz *et al.* 2016). Multiplex gene deletion in *Y. lipolytica* was also demonstrated (Gao *et al.* 2016). A vector was designed to carry a *Yarrowia* codon-optimized *CAS9* gene, driven by a *TEF1* promoter, together with gRNAs flanked with the HH and HDV ribozymes, which was also driven by a *TEF1* promoter. However, in the absence of donor DNA, NHEJ-mediated gene mutations occurred with decreasing efficiencies for the increasing number of targeted genes. When a donor template was included on the *CAS9*/gRNA vector, HDR-mediated gene disruption was shown to be successful, with higher rates in the *ku80* mutants.

CRISPR-Cas9 technology also resulted in the development of a toolkit (Schwartz *et al.* 2017b). This toolkit allows for integration of donor cassettes, which are delivered into the host yeast with the usage of a separate replicative vector that requires separate selection during the transformation. In a strain that contains the NHEJ repair mechanism, 17 locations were tested. Five, three, and nine sites showed integration efficiencies of 48-69%, 6% and 0%, respectively. Sequential marker-less integration of a metabolic pathway into the described loci was shown.

Another CRISPR tool (CRISPRi) was created by Schwartz and colleagues to repress several genes, which is involved in NHEJ, to increase the rate of HDR (Schwartz *et al.* 2017a). They showed that when using their CRISPRi tool to repress expression of the *KU70* and *KU80* genes, separately (56% and 73%) and together (approx. 90%), noteworthy increases in HDR are obtained when a donor fragment containing 1 kb homology was used (Schwartz *et al.* 2017a). They achieved these levels of HDR when they repressed the genes with a dCas9 fused to an Mxi1 repressor. In addition, the HDR rates obtained were comparable to the rates of a *Y. lipolitica ku70* mutant and their system proved to be successful in the repression of multiple, up to eight, genes.

6.4. *Komagataella phaffii* (formerly *Pichia pastoris*)

K. phaffii belong to a group of methylotrophic yeasts and is extensively used in protein production, due to its excellent folding and secretion capability, by means of recombinant DNA techniques. Although this yeast has a major limitation namely poor HDR, which makes it very difficult to engineer it is still extensively used in genetic and biochemical research as well as in the biotechnical industry (Gasser *et al.* 2013).

Weninger and co-workers tested a wide variety of differently expressed optimized *CAS9* genes and gRNA (Weninger *et al.* 2016). They tested over 90 constructs containing different, codon optimized *CAS9* genes, various gRNA sequences and various RNA polymerase promoters. They established that when changing a single feature, for example the codon optimization of the *CAS9* gene, CRISPR-Cas9 functionality could be completely abolished. They concluded this after only 6% of the constructs tested successful and resulted in single gene non-sense mutations.

When using a vector containing a low copy ARS element together with a native bi-directional *HXT1* promoter, which drives the expression of the *H. sapiens* codon-

optimized *CAS9* and gRNA flanked by the HH- and HDV-ribozymes, the transcript mostly (90%) resulted in single gene non-sense mutations (Weninger *et al.* 2016). When two genes were targeted, they observed non-sense mutations in both of the open reading frames (ORF) at a high frequency. However, when they included donor DNA, very low integration efficiency occurred. Thus, suggesting that NHEJ remained the dominant way of DSB repair.

6.5. Wide range CRISPR-Cas9 for *Kluyveromyces* and *Ogataea*.

Juergens and co-workers developed a wide range CRISPR-Cas9 system to edit the genome of two *Kluyveromyces* and *Ogataea* spieces (*Kluyveromyces lactis, Kluyveromyces marxianus, Ogataea polymorpha* and *Ogataea parapolymorpha*) (Juergens *et al.* 2018a). They constructed a plasmid vector containing two constitutive expression cassettes for the *CAS9* gene and the gRNAs, flanked by HH and HDV ribozymes, and a pangenomic origin of replication (Ori). The system was validated with the disruption of the *ADE2* gene in each of the yeast species. In both *Kluyveromyces* species, very high (≥96%) targeting efficiencies were obtained, however only about 23% (*K. marxianus*) and 31% (*K. lactis*) of the colonies contained the repair fragment. In the two *Ogataea* species, after a prolonged incubation period, targeting efficiencies of 9% (*O. polymorpha*) and approx. 63% (*O. parapolymorpha*) were observed mediated by NHEJ. When an *O. parapolymorpha KU80* mutant was transformed with a 960 bp donor fragment to disrupt the *ADE2* gene, <1% targeting efficiency was observed.

7. Conclusion

Several genetic engineering tools have been developed over the last few decades. Some of which include meganucleases, ZFN and TALEN. However, with the continuously improvement of the tools to obtain better engineering efficiencies, they nevertheless were difficult to use due to the protein engineering required to target different sites. This limitation was the main reason why these tools were never broadly accepted by the scientific community. In this respect, in 2013 immerged the CRISPR-Cas9 genetic engineering tool, which can simply target different sites in the genome by just changing the gRNA target sequence.

Chapter 1

The CRISPR-Cas9 tool consist of two main components, the Cas9 enzyme (binds to the target DNA and cleaves it) and a gRNA (binds to the Cas9 enzyme and guides it to the target site). In yeast, the expression of the *CAS9* has been indicated to not be a crucial element for the functionality of the system. However, some *CAS9* genes that have been optimized displayed enhanced editing efficiencies in certain yeast. For the gRNA, expression is very important to obtain the correct functionality, due to the fact that any Pol II promoters used will result in post-translational modifications (Gao & Zhao 2014). Thus, for correct expression a non-mRNA producing promoter needs to be used for expression. However, a recently discovered strategy displayed that any promoter can be used, by just making use of the self-catalysing characteristic of ribozymes. This will remove any post-transcriptional modifications, such as a poly(A) tail and a 7-methylguanylate cap, resulting in a functional gRNA.

Most of the yeasts tested have limitations when it comes to the deletion or repairment of genes. Thus, understanding different strategies performed of the different yeast and also the resulted efficiencies of the systems, may aid in the development of new systems to improve targeting as well as engineering efficiencies. Although the vast majority of CRISPR-Cas9 systems available for yeast, none can efficiently function in a wide variety of different yeast. Hence, using these strategies, which include the optimization of the *CAS9* and the gRNA expressing using the ribozymes, a wide range CRISPR-Cas9 system can be develop. This will provide a cost-effective system which can be used to test various aspects of different yeast through gene editing.

In this project the aim will be to construct such a system, which will be tested in a wide variety of different yeast. Three different CRISPR-Cas9 systems will be constructed by using the wide range pKM180 vector (Smit *et al.* 2012) for the backbone and a different optimized *CAS9* gene (optimized for *Pichia pastoris*, *Candida albicans* or *Homo sapiens*). A gRNA will also be incorporated in the vector. The RGR-gRNA (gRNA flanked with ribozymes) strategy will be used for the expression of the gRNA. The ribozyme strategy is important to use due to the wide range property of the system. Hence, the usage of a single promoter for the use in a variety of different yeast.

8. Purpose of study

Aim:

Chapter 1

To construct a single CRISPR-Cas9 system for easy and efficient gene editing in a wide variety of different yeast species and to explore the systems' potential for gene disruption in numerous yeast including *Saccharomyces cerevisiae* and six other non-conventional yeasts.

Objective 1:

To perform an extensive literature study on the different CRISPR-Cas9 strategies used for efficient gene targeting as well as gene disruption in non-pathogenic yeast.

Objective 2:

To construct a set of wide range vectors containing three different optimized *CAS9* genes and to test these for correct expression of the Cas9 protein and the effects thereof on the selected yeasts.

Objective 3:

To incorporate a Ribozyme-gRNA-Ribozyme -gRNA construct into the *CAS9* vectors to yield the complete CRISPR-Cas9 systems. This will provide the vector with gene targeting characteristics, which will allow for gene targeting in the selected yeasts.

Objective 4:

To validate the CRISPR-Cas9 system by disrupting the *ADE2* gene in *Saccharomyces cerevisiae* and six other non-conventional yeast.

9. References

Adli, M., (2018) 'The CRISPR tool kit for genome editing and beyond'. *Nat Commun* **9**, 1911. doi: 10.1038/s41467-018-04252-2

Bao, Z.,Xiao, H.,Liang, J.,*et al.*, (2015) 'Homology-Integrated CRISPR–Cas (HI-CRISPR) System for One-Step Multigene Disruption in *Saccharomyces cerevisiae*'. *ACS Synth Biol* **4**, 585–594. doi: 10.1021/sb500255k

Barrangou, R., (2015) 'The roles of CRISPR–Cas systems in adaptive immunity and beyond'. *Curr Opin Immunol* **32**, 36–41. doi: 10.1016/j.coi.2014.12.008

Chapter 1

Barrangou, R.,Fremaux, C.,Deveau, H.,*et al.*, (2007) 'CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes'. *Science (80-)* **315**, 1709–1712. doi: 10.1126/science.1138140

Berg, P.,& Mertz, J.E., (2010) 'Personal Reflections on the Origins and Emergence of Recombinant DNA Technology'. *Genetics* **184**, 9–17. doi: 10.1534/genetics.109.112144

Berlec, A.,& Štrukelj, B., (2013) 'Current state and recent advances in biopharmaceutical production in *Escherichia coli*, yeasts and mammalian cells'. *J Ind Microbiol Biotechnol* **40**, 257–274. doi: 10.1007/s10295-013-1235-0

Bétermier, M.,Bertrand, P.,& Lopez, B.S., (2014) 'Is Non-Homologous End-Joining Really an Inherently Error-Prone Process?'. *PLoS Genet* **10**, e1004086. doi: 10.1371/journal.pgen.1004086

Bhaya, D.,Davison, M.,& Barrangou, R., (2011) 'CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation'. *Annu Rev Genet* **45**, 273–297. doi: 10.1146/annurev-genet-110410-132430

Bitinaite, J.,Wah, D.A.,Aggarwal, A.K.,& Schildkraut, I., (1998) '*Fok*I dimerization is required for DNA cleavage'. *Proc Natl Acad Sci* **95**, 10570–10575. doi: 10.1073/pnas.95.18.10570

Blazeck, J.,Hill, A.,Liu, L.,*et al.*, (2014) 'Harnessing *Yarrowia lipolytica* lipogenesis to create a platform for lipid and biofuel production'. *Nat Commun* **5**, 3131. doi: 10.1038/ncomms4131

Boch, J., (2011) 'TALEs of genome targeting'. *Nat Biotechnol* **29**, 135–136. doi: 10.1038/nbt.1767

Boch, J.,& Bonas, U., (2010) '*Xanthomonas* AvrBs3 Family-Type III Effectors: Discovery and Function'. *Annu Rev Phytopathol* **48**, 419–436. doi: 10.1146/annurev-phyto-080508-081936

Boch, J.,Scholze, H.,Schornack, S.,*et al.*, (2009) 'Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors'. *Science (80-)* **326**, 1509–1512. doi: 10.1126/science.1178811

Boulton, S., (1996) 'Identification of a *Saccharomyces cerevisiae Ku80* homologue: roles in DNA double strand break rejoining and in telomeric maintenance'. *Nucleic*

Acids Res **24**, 4639–4648. doi: 10.1093/nar/24.23.4639

Brouns, S.J.J.,Jore, M.M.,Lundgren, M.,*et al.,* (2008) 'Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes'. *Science (80-)* **321**, 960–964. doi: 10.1126/science.1159689

Chavez, A.,Scheiman, J.,Vora, S.,*et al.,* (2015) 'Highly efficient Cas9-mediated transcriptional programming'. *Nat Methods* **12**, 326–328. doi: 10.1038/nmeth.3312

Chen, I.,& Dubnau, D., (2004) 'DNA uptake during bacterial transformation'. *Nat Rev Microbiol* **2**, 241–249. doi: 10.1038/nrmicro844

Chevalier, B.,Monnat, R.J.,& Stoddard, B.L., (2005) 'The LAGLIDADG Homing Endonuclease Family'. *Homing Endonucleases and Inteins* **16**, 33–47. doi: 10.1007/3-540-29474-0_3

Chevalier, B.S.,Kortemme, T.,Chadsey, M.S.,*et al.,* (2002) 'Design, Activity, and Structure of a Highly Specific Artificial Endonuclease'. *Mol Cell* **10**, 895–905. doi: 10.1016/S1097-2765(02)00690-1

Chevalier, B.S.,& Stoddard, B.L., (2001) 'Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility.'. *Nucleic Acids Res* **29**, 3757–74. doi: 10.1093/nar/29.18.3757

Cohen, S.N.,& Chang, A.C.Y., (1973) 'Recircularization and Autonomous Replication of a Sheared R-Factor DNA Segment in *Escherichia coli* Transformants'. *Proc Natl Acad Sci* **70**, 1293–1297. doi: 10.1073/pnas.70.5.1293

Cong, L.,Ran, F.A.,Cox, D.,*et al.,* (2013) 'Multiplex genome engineering using CRISPR/Cas systems'. *Science (80-)* **339**, 819–823. doi: 10.1126/science.1231143

Daley, J.M.,Palmbos, P.L.,Wu, D.,& Wilson, T.E., (2005) 'Nonhomologous End Joining in Yeast'. *Annu Rev Genet* **39**, 431–451. doi: 10.1146/annurev.genet.39.073003.113340

Deltcheva, E.,Chylinski, K.,Sharma, C.M.,*et al.,* (2011) 'CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III'. *Nature* **471**, 602–607. doi: 10.1038/nature09886

Chapter 1

DiCarlo, J.E.,Norville, J.E.,Mali, P.,*et al.*, (2013) 'Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems'. *Nucleic Acids Res* **41**, 4336–4343. doi: 10.1093/nar/gkt135

Dodd, A.,Curtis, P.M.,Williams, L.C.,& Love, D.R., (2000) 'Zebrafish: bridging the gap between development and disease.'. *Hum Mol Genet* **9**, 2443–2449. doi: 10.1093/hmg/9.16.2443

Dookun, A., (2001) 'Agricultural biotechnology in developing countries'. *Biotechnol Annu Rev* **7**, 261–285. doi: 10.1016/S1387-2656(01)07040-5

Dugar, G.,Herbig, A.,Förstner, K.U.,*et al.*, (2013) 'High-Resolution Transcriptome Maps Reveal Strain-Specific Regulatory Features of Multiple *Campylobacter jejuni* Isolates'. *PLoS Genet* **9**, e1003495. doi: 10.1371/journal.pgen.1003495

Egener, T.,Granado, J.,Guitton, M.-C.,*et al.*, (2002) 'High frequency of phenotypic deviations in *Physcomitrella* patens plants transformed with a gene-disruption library.'. *BMC Plant Biol* **2**, 1–6. doi: 10.1186/1471-2229-2-6

Farzadfard, F.,Perli, S.D.,& Lu, T.K., (2013) 'Tunable and Multifunctional Eukaryotic Transcription Factors Based on CRISPR/Cas'. *ACS Synth Biol* **2**, 604–613. doi: 10.1021/sb400081r

Gao, S.,Tong, Y.,Wen, Z.,*et al.*, (2016) 'Multiplex gene editing of the *Yarrowia lipolytica* genome using the CRISPR-Cas9 system'. *J Ind Microbiol Biotechnol* **43**, 1085–1093. doi: 10.1007/s10295-016-1789-8

Gao, Y.,& Zhao, Y., (2014) 'Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing'. *J Integr Plant Biol* **56**, 343–349. doi: 10.1111/jipb.12152

Garneau, J.E.,Dupuis, M.-È.,Villion, M.,*et al.*, (2010) 'The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA'. *Nature* **468**, 67–71. doi: 10.1038/nature09523

Garst, A.D.,Bassalo, M.C.,Pines, G.,*et al.*, (2016) 'Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering'. *Nat Biotechnol* **35**, 48–55. doi: 10.1038/nbt.3718

Gasiunas, G.,Barrangou, R.,Horvath, P.,& Siksnys, V., (2012) 'Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity

in bacteria'. *Proc Natl Acad Sci* **109**, E2579–E2586. doi: 10.1073/pnas.1208507109

Gasser, B.,Prielhofer, R.,Marx, H.,*et al.*, (2013) 'Pichia pastoris : protein production host and model organism for biomedical research'. *Future Microbiol* **8**, 191–208. doi: 10.2217/fmb.12.133

Generoso, W.C.,Gottardi, M.,Oreb, M.,& Boles, E., (2016) 'Simplified CRISPR-Cas genome editing for *Saccharomyces cerevisiae*'. *J Microbiol Methods* **127**, 203–205. doi: 10.1016/j.mimet.2016.06.020

Gerngross, T.U., (2004) 'Advances in the production of human therapeutic proteins in yeasts and filamentous fungi.'. *Nat Biotechnol* **22**, 1409–1414. doi: 10.1038/nbt1204-1589e

Gilbert, L.A.,Larson, M.H.,Morsut, L.,*et al.*, (2013) 'CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes'. *Cell* **154**, 442–451. doi: 10.1016/j.cell.2013.06.044

Gonçalves, F.A.G.,Colen, G.,& Takahashi, J.A., (2014) '*Yarrowia lipolytica* and its multiple applications in the biotechnological industry'. *Sci World J* **2014**, 1–14. doi: 10.1155/2014/476207

Gray, G.L.,Baldridge, J.S.,McKeown, K.S.,*et al.*, (1985) 'Periplasmic production of correctly processed human growth hormone in *Escherichia coli*: natural and bacterial signal sequences are interchangeable'. *Gene* **39**, 247–254. doi: 10.1016/0378-1119(85)90319-1

Gupta, R.M.,& Musunuru, K., (2014) 'Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9'. *J Clin Invest* **124**, 4154–4161. doi: 10.1172/JCI72992

Haft, D.H.,Selengut, J.,Mongodin, E.F.,& Nelson, K.E., (2005) 'A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes.'. *PLoS Comput Biol* **1**, e60. doi: 10.1371/journal.pcbi.0010060

Hanson, R.W.,& Hakimi, P., (2008) 'Born to run; the story of the PEPCK-Cmus mouse.'. *Biochimie* **90**, 838–42. doi: 10.1016/j.biochi.2008.03.009

Horwitz, A.A.,Walter, J.M.,Schubert, M.G.,*et al.*, (2015) 'Efficient Multiplexed

Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas'. *Cell Syst* **1**, 88–96. doi: 10.1016/j.cels.2015.02.001

Hsieh, H.-P.,& Da Silva, N.A., (1998) 'Partial-pKD1 plasmids provide enhanced structural stability for heterologous protein production in *Kluyveromyces lactis*'. *Appl Microbiol Biotechnol* **49**, 411–416. doi: 10.1007/s002530051191

ISAAA., (2017) 'GM traits list'. http://www.isaaa.org/gmapprovaldatabase/gmtraitslist/default.asp. Accessed, May 10, (2017)

Ishino, Y.,Shinagawa, H.,Makino, K.,*et al.*, (1987) 'Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product.'. *J Bacteriol* **169**, 5429–5433. doi: 10.1128/jb.169.12.5429-5433.1987

Jakočiūnas, T.,Bonde, I.,Herrgård, M.,*et al.*, (2015) 'Multiplex metabolic pathway engineering using CRISPR/Cas9 in *Saccharomyces cerevisiae*'. *Metab Eng* **28**, 213–222. doi: 10.1016/j.ymben.2015.01.008

Jansen, R.,Embden, J.D.A. van.,Gaastra, W.,& Schouls, L.M., (2002) 'Identification of genes that are associated with DNA repeats in prokaryotes'. *Mol Microbiol* **43**, 1565–1575. doi: 10.1046/j.1365-2958.2002.02839.x

Jinek, M.,Chylinski, K.,Fonfara, I.,*et al.*, (2012) 'A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity'. *Science (80-)* **337**, 816–821. doi: 10.1126/science.1225829

Jones, J.,Collier, E.,& Siebert, J., (1968) 'Enzyme-Containing Detergent Compositions'. https://docs.google.com/viewer?url=patentimages.storage.googleapis.com/pdfs/US3635828.pdf%0Ahttps://docs.google.com/viewer?url=patentimages.storage.googleapis.com/pdfs/US3790482.pdf. Accessed, May 10, (2017)

Juergens, H.,Varela, J.A.,Gorter de Vries, A.R.,*et al.*, (2018) 'Genome editing in *Kluyveromyces* and *Ogataea* yeasts using a broad-host-range Cas9/gRNA co-expression plasmid'. *FEMS Yeast Res* **18**, . doi: 10.1093/femsyr/foy012

Karvelis, T.,Gasiunas, G.,Miksys, A.,*et al.*, (2013) 'crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus*'. *RNA Biol* **10**, 841–

851. doi: 10.4161/rna.24203

Kaufman, R.I.,& Nixon, B.T., (1996) 'Use of PCR to isolate genes encoding sigma54-dependent activators from diverse bacteria.'. *J Bacteriol* **178**, 3967–3970. doi: 10.1128/jb.178.13.3967-3970.1996

Kim, Y.G.,Cha, J.,& Chandrasegaran, S., (1996) 'Hybrid restriction enzymes: zinc finger fusions to *Fok I* cleavage domain.'. *Proc Natl Acad Sci U S A* **93**, 1156–60. doi: 10.1128/AEM.03246-13

Klug, A.,& Rhodes, D., (1987) '"Zinc fingers": a novel protein motif for nucleic acid recognition'. *Trends Biochem Sci* **12**, 464–469. doi: 10.1016/0968-0004(87)90231-3

Knudson, A.G., (1967) 'Genetics and the lipidoses'. *J Am Oil Chem Soc* **44**, 623–627. doi: 10.1007/BF02680030

Krauss, J., (1992) 'Hematopoietic stem cell gene replacement therapy'. *Biochim Biophys Acta - Rev Cancer* **1114**, 193–207. doi: 10.1016/0304-419X(92)90015-Q

Kretzschmar, A.,Otto, C.,Holz, M.,*et al.*, (2013) 'Increased homologous integration frequency in *Yarrowia lipolytica* strains defective in non-homologous end-joining'. *Curr Genet* **59**, 63–72. doi: 10.1007/s00294-013-0389-7

Lacal, J.C.,Rosario, P.,& Feramisco, J., (1999) 'Microinjection'. https://books.google.co.za/books?id=2QVmrx_CpVUC&pg=PA9&redir_esc=y#v=onepage&q&f=false. Accessed, May 08, (2017)

Laity, J.H.,Lee, B.M.,& Wright, P.E., (2001) 'Zinc finger proteins: new insights into structural and functional diversity'. *Curr Opin Struct Biol* **11**, 39–46. doi: 10.1016/S0959-440X(00)00167-6

Lamont, J.,& Lacey, J., (2013) "Genetically Modified Organisms." *Blackwell Publishing Ltd, Oxford, UK*

Laughery, M.F.,Hunter, T.,Brown, A.,*et al.*, (2015) 'New vectors for simple and streamlined CRISPR-Cas9 genome editing in *Saccharomyces cerevisiae*'. *Yeast* **32**, 711–720. doi: 10.1002/yea.3098

Lee, M.E.,DeLoache, W.C.,Cervantes, B.,& Dueber, J.E., (2015) 'A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly'. *ACS Synth Biol* **4**,

975–986. doi: 10.1021/sb500366v

Li, Y.,Wu, Y.H.,McAvoy, R.,& Duan, H., (2001) 'Transgenics in crops'. *Biotechnol Annu Rev* **7**, 239–260. doi: 10.1016/S1387-2656(01)07039-9

Madzak, C.,Gaillardin, C.,& Beckerich, J.-M., (2004) 'Heterologous protein expression and secretion in the non-conventional yeast *Yarrowia lipolytica* : a review'. *J Biotechnol* **109**, 63–81. doi: 10.1016/j.jbiotec.2003.10.027

Makarova, K.S.,Grishin, N. V.,Shabalina, S.A.,*et al.*, (2006) 'A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action'. *Biol Direct* **1**, 1–7. doi: 10.1186/1745-6150-1-7

Mali, P.,Esvelt, K.M.,& Church, G.M., (2013a) 'Cas9 as a versatile tool for engineering biology'. *Nat Methods* **10**, 957–963. doi: 10.1038/nmeth.2649

Mali, P.,Yang, L.,Esvelt, K.M.,*et al.*, (2013b) 'RNA-Guided Human Genome Engineering via Cas9'. *Science (80-)* **339**, 823–826. doi: 10.1126/science.1232033

Mans, R.,van Rossum, H.M.,Wijsman, M.,*et al.*, (2015) 'CRISPR/Cas9: A molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*'. *FEMS Yeast Res* **15**, 1–15. doi: 10.1093/femsyr/fov004

Marraffini, L.A.,& Sontheimer, E.J., (2010) 'CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea'. *Nat Rev Genet* **11**, 181–190. doi: 10.1038/nrg2749

Marraffini, L.A.,& Sontheimer, E.J., (2008) 'CRISPR Interference Limits Horizontal Gene Transfer in *Staphylococci* by Targeting DNA'. *Science (80-)* **322**, 1843–1845. doi: 10.1126/science.1165771

Mojica, F.J.M.,Díez-Villaseñor, C.,García-Martínez, J.,& Soria, E., (2005) 'Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements'. *J Mol Evol* **60**, 174–182. doi: 10.1007/s00239-004-0046-3

Mojica, F.J.M.,Diez-Villasenor, C.,Soria, E.,& Juez, G., (2000) 'Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and

mitochondria'. *Mol Microbiol* **36**, 244–246. doi: 10.1046/j.1365-2958.2000.01838.x

Moscou, M.J.,& Bogdanove, A.J., (2009) 'A Simple Cipher Governs DNA Recognition by TAL Effectors'. *Science (80-)* **326**, 1501–1501. doi: 10.1126/science.1178817

Nevoigt, E., (2008) 'Progress in Metabolic Engineering of *Saccharomyces cerevisiae*'. *Microbiol Mol Biol Rev* **72**, 379–412. doi: 10.1128/MMBR.00025-07

Olswang, Y.,Cohen, H.,Papo, O.,*et al.*, (2002) 'A mutation in the peroxisome proliferator-activated receptor gamma-binding site in the gene for the cytosolic form of phosphoenolpyruvate carboxykinase reduces adipose tissue size and fat content in mice.'. *Proc Natl Acad Sci U S A* **99**, 625–30. doi: 10.1073/pnas.022616299

Pâques, F.,& Haber, J.E., (1999) 'Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*.'. *Microbiol Mol Biol Rev* **63**, 349–404

Pitcher, R.S.,Brissett, N.C.,& Doherty, A.J., (2007) 'Nonhomologous End-Joining in Bacteria: A Microbial Perspective'. *Annu Rev Microbiol* **61**, 259–282. doi: 10.1146/annurev.micro.61.080706.093354

Pourcel, C.,Salvignol, G.,& Vergnaud, G., (2005) 'CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies'. *Microbiology* **151**, 653–663. doi: 10.1099/mic.0.27437-0

Qi, L.S.,Larson, M.H.,Gilbert, L.A.,*et al.*, (2013) 'Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression'. *Cell* **152**, 1173–1183. doi: 10.1016/j.cell.2013.02.022

Richardson, C.D.,Ray, G.J.,DeWitt, M.A.,*et al.*, (2016) 'Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA'. *Nat Biotechnol* **34**, 339–344. doi: 10.1038/nbt.3481

Rosen, L.E.,Morrison, H.A.,Masri, S.,*et al.*, (2006) 'Homing endonuclease *I-CreI* derivatives with novel DNA target specificities'. *Nucleic Acids Res* **34**, 4791–4800. doi: 10.1093/nar/gkl645

Ryan, O.W.,Skerker, J.M.,Maurer, M.J.,*et al.*, (2014) 'Selection of chromosomal DNA

libraries using a multiplex CRISPR system.'. *Elife* **3**, 1–15. doi: 10.7554/eLife.03703

Sakuma, T.,& Woltjen, K., (2014) 'Nuclease-mediated genome editing: At the front-line of functional genomics technology'. *Dev Growth Differ* **56**, 2–13. doi: 10.1111/dgd.12111

Samson, J.E.,Magadán, A.H.,Sabri, M.,& Moineau, S., (2013) 'Revenge of the phages: Defeating bacterial defences'. *Nat Rev Microbiol* **11**, 675–687. doi: 10.1038/nrmicro3096

Sapranauskas, R.,Gasiunas, G.,Fremaux, C.,*et al.*, (2011) 'The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*'. *Nucleic Acids Res* **39**, 9275–9282. doi: 10.1093/nar/gkr606

Sauer, B., (1987) 'Functional expression of the cre-lox site-specific recombination system in the yeast *Saccharomyces cerevisiae*.'. *Mol Cell Biol* **7**, 2087–2096. doi: 10.1128/MCB.7.6.2087.Updated

Schwartz, C.,Frogue, K.,Ramesh, A.,*et al.*, (2017a) 'CRISPRi repression of nonhomologous end-joining for enhanced genome engineering via homologous recombination in *Yarrowia lipolytica*'. *Biotechnol Bioeng* **114**, 2896–2906. doi: 10.1002/bit.26404

Schwartz, C.,Shabbir-Hussain, M.,Frogue, K.,*et al.*, (2017b) 'Standardized Markerless Gene Integration for Pathway Engineering in *Yarrowia lipolytica*'. *ACS Synth Biol* **6**, 402–409. doi: 10.1021/acssynbio.6b00285

Schwartz, C.M.,Hussain, M.S.,Blenner, M.,& Wheeldon, I., (2016) 'Synthetic RNA Polymerase III Promoters Facilitate High-Efficiency CRISPR–Cas9-Mediated Genome Editing in *Yarrowia lipolytica*'. *ACS Synth Biol* **5**, 356–359. doi: 10.1021/acssynbio.5b00162

Seligman, L.M., (2002) 'Mutations altering the cleavage specificity of a homing endonuclease'. *Nucleic Acids Res* **30**, 3870–3879. doi: 10.1093/nar/gkf495

Singh, M.N.,Hemant, K.S.Y.,Ram, M.,& Shivakumar, H.G., (2010) 'Microencapsulation: A promising technique for controlled drug delivery'. *Res Pharm Sci* **5**, 65–77

Skrzypek, M.S.,Binkley, J.,Binkley, G.,*et al.*, (2017) 'The *Candida* Genome Database

(CGD): Incorporation of Assembly 22, systematic identifiers and visualization of high throughput sequencing data'. *Nucleic Acids Res* **45**, D592–D596. doi: 10.1093/nar/gkw924

Smit, M.S.,Theron, C.W.,Albertyn, J.,& Labuschagné, M., (2012) 'Novel expression construct'. 04652

Stoddard, B.L., (2011) 'Homing endonucleases: From microbial genetic invaders to reagents for targeted DNA modification'. *Structure* **19**, 7–15. doi: 10.1016/j.str.2010.12.003

Stovicek, V.,Borja, G.M.,Forster, J.,& Borodina, I., (2015) 'EasyClone 2.0: expanded toolkit of integrative vectors for stable gene expression in industrial *Saccharomyces cerevisiae* strains'. *J Ind Microbiol Biotechnol* **42**, 1519–1531. doi: 10.1007/s10295-015-1684-8

Summerton, J.E., (2007) 'Morpholino, siRNA, and S-DNA compared: impact of structure and mechanism of action on off-target effects and sequence specificity.'. *Curr Top Med Chem* **7**, 651–660. doi: 10.2174/156802607780487740

Sung, P.,& Klein, H., (2006) 'Mechanism of homologous recombination: mediators and helicases take on regulatory functions'. *Nat Rev Mol Cell Biol* **7**, 739–750. doi: 10.1038/nrm2008

Sussman, D.,Chadsey, M.,Fauce, S.,*et al.*, (2004) 'Isolation and Characterization of New Homing Endonuclease Specificities at Individual Target Site Positions'. *J Mol Biol* **342**, 31–41. doi: 10.1016/j.jmb.2004.07.031

Terns, M.P.,& Terns, R.M., (2011) 'CRISPR-based adaptive immune systems'. *Curr Opin Microbiol* **14**, 321–327. doi: 10.1016/j.mib.2011.03.005

Unniyampurath, U.,Pilankatta, R.,& Krishnan, M.N., (2016) 'RNA interference in the age of CRISPR: Will CRISPR interfere with RNAI?'. Int. J. Mol. Sci. 17:291

van den Berg, J.A.,van der Laken, K.J.,van Ooyen, A.J.,*et al.*, (1990) '*Kluyveromyces* as a host for heterologous gene expression: expression and secretion of prochymosin.'. *Biotechnology (N Y)* **8**, 135–139. doi: 10.1038/nbt0290-135

Van der Walt, J.P.,Smith, M.T.,& Yamada, Y., (1990) 'Arxula gen. nov. (Candidaceae), a new anamorphic, arthroconidial yeast genus'. *Antonie van Leeuwenhoek, Int J Gen Mol Microbiol* **57**, 59–61. doi: 10.1007/BF00400338

Chapter 1

van Ooyen, A.J.J.,Dekker, P.,Huang, M.,*et al.*, (2006) 'Heterologous protein production in the yeast *Kluyveromyces lactis*'. *FEMS Yeast Res* **6**, 381–392. doi: 10.1111/j.1567-1364.2006.00049.x

Verbeke, J.,Beopoulos, A.,& Nicaud, J.-M., (2013) 'Efficient homologous recombination with short length flanking fragments in *Ku70* deficient *Yarrowia lipolytica* strains'. *Biotechnol Lett* **35**, 571–576. doi: 10.1007/s10529-012-1107-0

Waegeman, H.,De Lausnay, S.,Beauprez, J.,*et al.*, (2013) 'Increasing recombinant protein production in Escherichia coli K12 through metabolic engineering'. *N Biotechnol* **30**, 255–261. doi: 10.1016/j.nbt.2011.11.008

Walsh, G., (2006) 'Biopharmaceutical benchmarks 2006'. *Nat Biotechnol* **24**, 769–776. doi: 10.1038/nbt0706-769

Wang, Q.,& Wang, L., (2008) 'New Methods Enabling Efficient Incorporation of Unnatural Amino Acids in Yeast'. *J Am Chem Soc* **130**, 6066–6067. doi: 10.1021/ja800894n

Weninger, A.,Hatzl, A.-M.,Schmid, C.,*et al.*, (2016) 'Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast *Pichia pastoris*'. *J Biotechnol* **235**, 139–149. doi: 10.1016/j.jbiotec.2016.03.027

Westphal, C.H.,& Leder, P., (1997) 'Transposon-generated "knock-out" and "knock-in" gene-targeting constructs for use in mice'. *Curr Biol* **7**, 530–533. doi: 10.1016/S0960-9822(06)00224-7

Wiedenheft, B.,Sternberg, S.H.,& Doudna, J.A., (2012) 'RNA-guided genetic silencing systems in bacteria and archaea'. *Nature* **482**, 331–338. doi: 10.1038/nature10886

Williams, D.C.,Van Frank, R.M.,Muth, W.L.,& Burnett, J.P., (1982) 'Cytoplasmic inclusion bodies in *Escherichia coli* producing biosynthetic human insulin proteins'. *Science (80-)* **215**, 687–689. doi: 10.1126/science.7036343

Xu, K.,Ren, C.,Liu, Z.,*et al.*, (2015) 'Efficient genome engineering in eukaryotes using Cas9 from *Streptococcus thermophilus*'. *Cell Mol Life Sci* **72**, 383–399. doi: 10.1007/s00018-014-1679-z

Yoder, J.A.,Nielsen, M.E.,Amemiya, C.T.,& Litman, G.W., (2002) 'Zebrafish as an

immunological model system'. *Microbes Infect* **4**, 1469–1478. doi: 10.1016/S1286-4579(02)00029-1

Zalatan, J.G.,Lee, M.E.,Almeida, R.,*et al.*, (2015) 'Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds'. *Cell* **160**, 339–350. doi: 10.1016/j.cell.2014.11.052

Zaretsky, D. V.,Zaretskaia, M. V.,Rusyniak, D.E.,& DiMicco, J.A., (2011) 'Stress-free microinjections in conscious rats'. *J Neurosci Methods* **199**, 199–207. doi: 10.1016/j.jneumeth.2011.05.004

Zhang, G.-C.,Kong, I.I.,Kim, H.,*et al.*, (2014) 'Construction of a Quadruple Auxotrophic Mutant of an Industrial Polyploid *Saccharomyces cerevisiae* Strain by Using RNA-Guided Cas9 Nuclease'. *Appl Environ Microbiol* **80**, 7694–7701. doi: 10.1128/AEM.02310-14

Chapter 2

Expression of three different codon optimized *Cas9* Genes using the wide range pKM180 vector in seven different yeast species
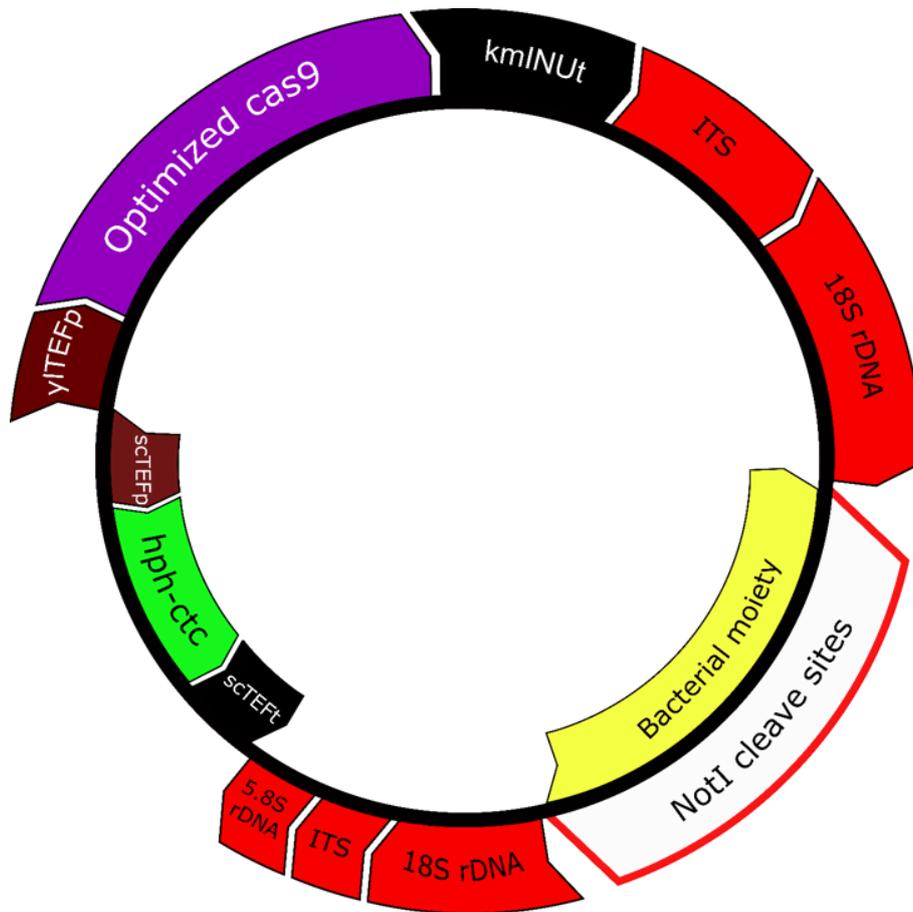
Chapter 2

1. Abstract

This study describes new adaptations to CRISPR-Cas9 systems to develop a new CRISPR-Cas9 system for easy and fast gene editing in a wide variety of different yeast species and to explore the systems' potential for numerous gene edits, simultaneously. Here we present three different pKM180- Cas9 expression constructs, each with a different codon optimized (*Pichia pastoris*, *Candida albicans* or *Homo sapiens*), constitutively driven, *CAS9*. Each of the *CAS9* genes are fused with a nuclear localization sequence to ensure transportation of the expressed protein into the nucleus. These expression constructs make use of the rRNA locus from *Kluyveromyces marxianus* for integration into the yeasts and selection of positive transformants is facilitated by the dominant selection of Hygromycin via the constitutively driven Hygromycin resistant (*HPH*) gene. The three different Cas9 constructs were transformed into seven different biotechnologically relevant yeasts using two different chemical transformation protocols. Integration of the expression vectors into the yeast genome were confirmed through PCR and gel electrophoresis. Protein expression of the Cas9 were confirmed via western blotting using a Cas9-antibody. This system would allow for rapid, cost-effective genetic manipulation of biotechnologically relevant yeast.

2. Introduction

In biotechnology the best eukaryotic organism for research is yeast, due to their fast growth rate and ease of manipulation compared to multicellular organisms (Madzak *et al.* 2004; Gerngross 2004; Nevoigt 2008; Stovicek *et al.* 2015; Skrzypek *et al.* 2017). Prokaryotic organisms are very easy to handle, but they have limitations when it comes to the production of eukaryotic proteins due to post-translational modifications. However, unlike bacteria, not much CRISPR-Cas9 research have been done on yeast. CRISPR-Cas9 is the recently discovered, innovative, genetic engineering tool that would allow scientists to change, at will, any DNA sequence of, presumably, any living organism in a specific manner (Jinek *et al.* 2012). The Cas9 can be programmed with a single guide RNA (sgRNA) to generate site-specific DNA breaks. Unlike any other genetic engineering tools or techniques, CRISPR is remarkably simpler, faster and cheaper to perform (Jinek *et al.* 2012; Adli 2018).

Chapter 2

Cas9 (CRISPR associated systems) is an RNA-guided DNA endonuclease enzyme which is associated with the adaptive immunity, CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats), of bacteria (Wang *et al.* 2011; Terns & Terns 2011; Bhaya *et al.* 2011; Jinek *et al.* 2012). The bacteria use Cas9 in the adaptation, CRISPR processing and interference steps in the adaptive immunity. The Cas9 enzymes binds to a guide RNA to form the interference complex, which will then check for sites complementary to the 20 bp spacer region of the guide RNA. If the foreign DNA is complementary to the guide RNA, the Cas9 binds to it, recognises the PAM site (NGG) and then cleaves the invading DNA.

Apart from the purpose Cas9 serves in bacterial immunity, the Cas9 enzyme has been modified to be utilized as a genome engineering tool, which can induce site-directed double strand DNA breaks when combined with a guide RNA (Jinek *et al.* 2012). While native Cas9 (*Streptococcus pyogenes*) requires a guide RNA composed of two different RNAs [CRISPR RNA (crRNA), and the trans-activating RNA (tracrRNA)] that interact and undergo processing to yield the mature guide RNA, Cas9 targeting has been simplified through the combination and modification of the crRNA and tracrRNA to create a chimeric single guide RNA. In addition, Cas9 has gained much popularity due to its property to cleave any sequence as long as it is complementary to the bounded guide RNA. Hence making CRISPR-Cas9 engineering very easy compared to other genetic engineering tools. The Cas9 nuclease binds to the DNA followed by the cleavage of the complementary DNA-strand (target strand) by the HNH domain of the Cas9 enzyme (Barrangou *et al.* 2007; Gasiunas *et al.* 2012). The other DNA strand (containing the PAM sequence) is cleaved by the RuvC-like domain, thus resulting in a DSB. This cleavage action is performed three nucleotides upstream of the PAM site.

Recently, numerous different approaches for CRISPR-Cas9 engineering have been described to alter genes in various yeasts (DiCarlo *et al.* 2013; Jacobs *et al.* 2014; Horwitz *et al.* 2015; Schwartz *et al.* 2016; Weninger *et al.* 2016; Juergens *et al.* 2018b). However, for the application in yeast, CRISPR-Cas9 engineering is not as straightforward as in bacteria. Unlike bacteria, yeast have a nuclease and other membrane enclosed organelles. Therefore, the *CAS9* needs to be transported from the cytoplasm to the nuclease to perform its function. This localization of the Cas9 protein can be done by linking the *CAS9* gene to a nuclear localization sequence

(NLS), which allows the transportation of the mature Cas9 protein to the nucleus of the host cell. This is needed for the Cas9 to get accessibility to the genomic material. The *CAS9* gene can not only be native but can, and in many cases have to be optimized according to the yeast codon bias to obtain the correct expression and optimal function (DiCarlo *et al.* 2013; Gao & Zhao 2014; Ryan *et al.* 2014; Zhang *et al.* 2014; Jakočiūnas *et al.* 2015; Mans *et al.* 2015; Bao *et al.* 2015; Horwitz *et al.* 2015; Generoso *et al.* 2016). In addition, some yeasts also showed enhanced function when the *CAS9* gene were *H. sapiens* optimized.

Usually, the expression of *CAS9* is under the control of constitutive promoters of different strength originating from either self-replicating low-copy centromeric -, high-copy 2μ - or integration vectors (DiCarlo *et al.* 2013; Gao & Zhao 2014; Ryan *et al.* 2014; Zhang *et al.* 2014; Jakočiūnas *et al.* 2015; Mans *et al.* 2015; Bao *et al.* 2015; Horwitz *et al.* 2015; Generoso *et al.* 2016). However, expression in a high-copy vector (if available), using a strong constitutive promoter may result in poor growth due to the toxic effect linked to the Cas9 protein in yeast (Ryan *et al.* 2014; Generoso *et al.* 2016). In some studies it was however observed that this toxicity does not pose a big problem and can be overcome with the usage of weaker promoters (Gao & Zhao 2014; Bao *et al.* 2015).

Therefore, the aim of this part of the study was the construction of a wide range vector containing either a *Pichia pastoris*, *Candida albicans* or *Homo sapiens* optimized *CAS9* gene. These constructs were tested for correct expression of the Cas9 and the effects thereof on *Saccharomyces cerevisiae* and six other non-conventional yeast. Please note that although the name of *Pichia pastoris* was changed to *Komagataella phaffi* for the purpose of this study the codon optimised *CAS9* gene is referred to as a *Pichia pastoris* optimised CAS9 (PCAS9) seeing that the sequence of the optimised gene was referred in the same way by Weninger *et al.* (2016) from whom this sequence was obtained.

3. Materials and methods

3.1. Rubidium chloride competent cells

Escherichia coli cells were inoculated in 5 ml Luria-Bertani (LB) (Table 1) media and grown over night (O.N.) at 37°C to yield the pre-inoculum (Turvey 1997).

Table 1. Media used for all bacterial transformations

Media or buffer	Composition
LB (Luria-Bertani)	1% typtone, 1% NaCl, 0.5% Yeast extract, pH 7 (adjusted with 5N NaOH), 1.5% agar (only for plates)

One millilitre of the pre-inoculum was transferred into 100 ml of Psi broth (Table 2) and grown at 37°C to an OD_{600} of 0.6. The cells were transferred to two 50 ml centrifuge tubes, chilled on ice for 15 minutes and centrifuged at 4 000 x g for 5 minutes at 4°C. After centrifugation the pellet was resuspended in 40 ml of ice cold transformation buffer 1 (TFB1) and incubated on ice for 15 minutes. The suspension was pelleted at 4 000 x g for 5 minutes at 4°C. The cells were resuspended in 4ml transformation buffer 2 (TFB2), incubated on ice for 60 minutes, aliquoted (50 µl per tube), snapped freezed and stored at -80°C.

Table 2. Media and buffers used for the making of Rubidium chloride competent cells

Media or Buffer	Composition
PSI broth	2% tryptone, 0.5% yeast extract, 0.5% $MgSO_4$, pH 7.6
TFB1	1.22% $RbCl^2$, 1% $MgCl_2$, 0.3% KOAc, 0.15% $CaCl_2$, 15% Glycerol, ph 5.8
TFB2	0.21% MOPS, 0.12% $RbCl_2$, 2.1% $CaCl^2$, 15% Glycerol, pH 6.5

3.2. Bacterial Transformation

All the bacterial transformations were performed as follows: 5 µl of the DNA suspension were added to the chemically-competent cells and incubated on ice for 1 hour, the cells were heat shocked at 42˚C for 30 seconds, incubated on ice for 2 minutes and incubated for 1 hour (with shaking) at 37˚C after 1 ml of LB media was added (Chung *et al.* 1989). The cells were then spun down, resuspended in 300 µl of LB media and plated out on LB plates containing [30 µg/ml] kanamycin, using the spread plate technique and incubated overnight at 37 ˚C. The colonies were picked up, inoculated in 5 ml LB media containing [50 µg/ml] kanamycin and incubated at 37˚C, with shaking, overnight. The plasmids were extracted and screening followed using restriction analysis.

3.3. Restriction digestion

Restriction enzymes where either obtained from New England Biolabs Inc or Thermo Fisher Scientific Inc (New England Biolabs Inc. 2014; Thermo Fisher Scientific Inc 2015a). For all restriction digestion reaction the mixtures, as depicted in Table 3, were prepared, depending on the enzyme manufacturer.

Table 3. The reaction composition for all restriction digestion reactions from both Thermo science and NEB

Component	Reaction mixture	
	Thermo science	NEB
DNA	500- 1000 ng	500- 1000 ng
Enzyme	5% (10 U – 20 U)	2% (10 U- 40 U)
Enzyme specific buffer	10%	10%
dH$_2$O	Up to 100%	Up to 100%

The mixture were incubated for at least 60 minutes at 37˚C, followed by analysis using agarose gel electrophoresis. Prior to the loading of the samples onto the gel, the samples were mixed with 6 x DNA loading dye in a ratio of 5:1.

3.4. Agarose gel electrophoresis

For DNA analysis using agarose gel electrophoresis a 0.8% agarose gel was prepared with 1 x TAE buffer (Tris base, acetic acid and EDTA) and stained with SYBR Safe DNA Stain at a dilution of 1:10,000 (Thermo Fisher Scientific). DNA samples were separated for 30 minutes at 90 volts and visualisation of gels were performed in a Bio-Rad Gel Doc™ XR+. For the molecular weight marker the O'GeneRuler DNA 10 kb ladder mix was used for each of the gels in this study.

3.5. Gel extraction

The gel piece containing the fragment of interest was excised from the gel and collected in a tube (Thermo Fisher Scientific Inc 2015b). The gel was then incubated after the addition of Binding Buffer, in a ratio of 1:1 (volume: weight), at 60˚C until the gel was completely dissolved. This was followed by the addition of 1 volume 100% isopropanol. Eight hundred microliters of the mixture was then transferred to a GeneJET purification column and centrifuged for 1 min at 14 000 x *g*. the column was washed with 700 µl of wash buffer and transferred to a new 1.5 ml microcentrifuge tube. The DNA was eluted after the wash step with the addition of 50 µl elution buffer (10 mM Tris-HCl, pH 8.5).

3.6. Ligation and cloning

For the purpose of cloning and ligation the DNA vectors and fragments were prepared with the desired restriction enzymes (Devi *et al.* 2009). Ligation was then performed with the preparation of the reaction mixture (Table 4) and incubated O.N at 4˚C.

Table 4. The T4 ligation sticky end protocol (**Thermo Fisher Scientific)**.

Component	Reaction mixture
Linear vector DNA	20-100 ng

Insert DNA	5:1 molar ratio over vector
10x T4 DNA Ligase buffer	10%
Thermo Scientific T4 DNA Ligase	5% (1 Weiss U)
Water, nuclease-free	To 100%

Following ligation, 5 µl reaction product was transformed into competent *E. coli* XL10-Gold (Stratagene) cells. The transformed cells were then plated out on LB plates (Table 5) containing [30 µg/ml] kanamycin and screened for positive transformants.

3.7. Plasmid extraction

The plasmid extraction method used is based on the alkaline lysis principle (Table 5) (Birnboim & Doly 1979; Ish-Horowicz & Burke 1981)

Table 5. The buffers used for the plasmid extraction for the purpose of screening or cloning

Buffer	Composition
Resuspension Buffer	50mM glucose, 25mM Tris-HCl (pH 8), EDTA (pH 8)
Lysis Buffer	0.2M NaOH, 1% SDS
Neutralization Buffer	60% 5 M potassium acetate, 11.5% acetic acid.
TE buffer	10mM Tris-HCl (pH 8), 1mM EDTA (pH 8).

Cells were harvested by centrifugation at top speed (12 000 x g) for 30 seconds, after which it was resuspended in a 100 µl of ice-cold resuspension buffer. A lysis buffer (200 µl) were added, inverted to mix, left for 5 minutes and then neutralized with the addition of 150 µl ice-cold neutralization buffer. The suspension was centrifuged at top speed (12000 x g) for 10 minutes at 4˚C and the supernatant was transferred to a new 1.5ml tube. The DNA was then precipitated with the addition of 1 ml, 100% ethanol and incubated for 2 minutes at room temperature. The DNA suspension was then centrifuged at 12 000 x g for 5 minutes at 4˚C, followed by the removal of the supernatant, washing with 1ml of 70% ethanol and resuspension in TE buffer containing [20 µg/ml] RNase A. The DNA concentration was then determined using a NanoDrop™ One/OneC Microvolume UV Spectrophotometer.

3.8. Construction of the pKM180 containing the codon optimized *CAS9* genes

For the construction of the expression system, three different codon optimised *CAS9* genes (as reported by other authors) were obtained from Genscript (GenScript 2002). The *Candida albicans* optimized *CAS9* (CCAS9) (Vyas *et al.* 2015) were obtained in plasmid pESC-URA. The *Pichia pastoris* (Weninger *et al.* 2016) and *Homo sapiens* (Human) optimized *CAS9* (HCAS9) (Gao & Zhao 2014) were obtained in a pAO815 vector.

For the purpose of successful ligation and cloning the different optimised *CAS9* genes as well as the pKM180 were cleaved with different restriction enzymes. For the PCAS9*, Nco*I and *Hind*III were used to remove the *CAS9* from the pAO815. For CCAS9 and HCAS9, *Nco*I + *Bam*HI were used to liberate it from pESC-URA and pAO815 vectors, respectively. After the restriction digestion, the fragments were purified through gel extraction. The purified fragments were then ligated into pKM180 that was digested with the same restriction enzymes, transformed into competent *E. coli* cells and screened for successful transformants containing the desired plasmids. For screening the plasmids were digested with restriction enzymes from thermo scientific. The pKM::PCAS9 were digested with *Afl*II, pKM::CCAS9 with *Acl*I and pKM::HCAS9 with *Nru*I and *Sal*I.

Chapter 2

3.9. Yeast transformation

In a previous study (E. Bisschoff; B.Sc. Honours research project performed in 2016) the optimal transformation protocol was determined for each yeast specie (Table 6). The different yeasts were therefore transformed using one of two different alkali-ion based protocols namely the bicine or a one-step (lithium acetate) transformation protocols (Klebe *et al.* 1983; Chen *et al.* 1992) depending on which of these two methods showed the highest transformation efficiency for each yeast specie. Optimal hygromycin concentration was also determined for each yeast respective to the optimized *CAS9* (Table 6).

Table 6. The different yeast used in this study and also the hygromycin concentration used for each yeast respective to the vector used for transformation

Yeast	Transformation Hygromycin [mg/L]			
	pKM180*	PCAS9	CCAS9	HCAS9
Saccharomyces cerevisiae CENPK42	600	600	600	150
Kluyveromyces lactis UOFS-Y1167	600	600	600	150
Yarrowia lipolytica Po1f	600	600	600	600
Debaryomyces hansenii UOFS-Y0610	600	300	150	150
Arxula adeninivorans UOFS-Y1220	600	600	300	300
Arxula adeninivorans UOFS-Y1219	600	300	300	300
Arxula adeninivorans Ls3	600	600	600	150
Ogataea polymorpha UOFS Y1507	600	300	300	150
Komagataella phaffii UOFS-Y1552	600	600	600	600
Komagataella phaffii SMD 1168	600	600	600	150

Komagataella phaffii Km.71	600	600	600	150
Komagataella phaffii GS115	600	600	600	150

*Empty plasmid acting positive control.

Before transformation, the three different *CAS9* plasmids were linearized by *Not*I (NEB) digestion to remove the bacterial moiety and therefore exposing the rDNA regions to allow homologous recombination to take place. For a positive control the normal pKM180 were transformed into all the yeast. For both transformation methods used, transformed yeast were incubated for up to 7 days at 30˚C.

For transformation using the one-step transformation method (Chen *et al.* 1992), cells were grown overnight in 5 ml YPD media (Table 7). The cell suspension (0.8 ml) was centrifuged at 12 000 x *g* for 4 seconds, the supernatant were aspirated and 0.5 µg plasmid and 50 µg carrier DNA (salmon sperm DNA) were added. The mixture were resuspended in 100 µl One-step buffer (Table 7) and left to incubate for 30 minutes at 42˚C.

Table 7. The media and buffers used for the One-step transformation protocol

Media or buffer	Composition
YPD (Yeast Extract–Peptone–Dextrose)	1% Yeast, 2% peptone, 2% glucose, 1.5% Agar (only for plates)
One-step buffer	2 M lithium acetate, 50 % PEG, 1 M DTT

After incubation the suspension were centrifuged at 12 000 x *g* for 8 seconds, the supernatant was aspirated and the cells were resuspended in 1 ml of YPD media and incubated at 30˚C, with shaking, for 3 hours. After incubation, the cell suspension were centrifuged at 5 000 x *g* for 2 minutes and resuspended in 300 µl of YPD. The cell suspension were then plated, spread-plate technique, on YPD plates containing hygromycin.

Chapter 2

The bicine transformation method (Klebe *et al.* 1983) benefits in that one can prepare competent yeast cells and these can be stored at -80˚C for later use. This is the only yeast protocol that allows the preparation of competent cells that can be stored for later use. The preparation of competent cells was done by the harvesting of the cells, grown in YPD media (Table 8), at an OD_{600} of 0.6- 0.8. Cells were centrifuged at 1 000 x *g* for 5 minutes to pellet the cells. The supernatant was then aspirated and the cells were resuspended in 50 ml of Buffer A (Table 8). The cells were then again pelleted at 1 000 x *g* for 5 minutes, resuspended in 2 ml of Buffer A, aliquoted in 200 µl per microcentrifuge tube, placed at -20˚C for 4 hours and then stored at -80˚C .

Table 8. The media and buffers used for the bicine transformation protocol

Media or buffer	Composition
YPD (Yeast Extract–Peptone–Dextrose)	1% Yeast, 2% peptone, 2% glucose, 1.5% Agar (only for plates)
Buffer A	1 M Sorbitol, 10 mM Bicine (pH 8.35), 3% ethylene glycol, 5% DMSO
Buffer B	40% PEG, 200 mM Bicine (pH 8)
Buffer C	0.15 M NaCl, 10 mM Bicine (pH 8.35)

For transformation the competent cells were incubated with 0.43 µg plasmid and 50 µg carrier DNA (salmon sperm DNA) for 15 seconds at 37 ˚C. The mixture was then suspended in 1.4 ml of Buffer B (Table 9) and incubated at 30˚C for 60 minutes. After incubation the cells were centrifuged at 3 000 x *g* for 5 seconds to form a pellet. The supernatant was aspirated and the pellet was suspended in 1 ml of Buffer C. The suspension were then centrifuged at 3 000 x *g* for 5 seconds, the supernatant were aspirated and the pellet was resuspended in 1 ml YPD and incubated at 30˚C, with shaking, for 3 hours. After incubation, the cells were spun down at 3 000 x g for 2 min and resuspended in 300 µl of YPD. The cell suspension was then plated out, spread-plate technique, on YPD plates containing hygromycin.

3.10. Plasmid integration confirmation

After transformation, plasmid integration was confirmed through amplification of a part of the *CAS9* gene. For the pKM::PCAS9 and pKM::CCAS9 transformed yeast, primers cacas9-1 forward and reverse were used for amplification and for the pKM::HCAS9, primers hscas9-1 forward and reverse was used for amplification (Table 9).

Table 9. Primers used with their corresponding properties

Primer name	Length (bp)	GC (%)	Tm (˚C)	Sequence
cacas9-1 forward	21	42.9	57.2	TTGCCAGAAAAAAGGATTGGG
cacas9-1 reverse	23	43.5	58.2	GATTGATGAATTAAGGTGGCGTC
hscas9-1 forward	20	50	57.3	TGGAGGAGTCCTTTTTGGTG
hscas9-1 reverse	20	45	57.7	AAAATTCCTCCTGGCTTGCT

The KAPA Taq PCR kit was used for the amplification with the reaction mixture prepared according to Table 10 and the reaction conditions as depicted in Table 11 (KAPA Biosystems Inc. 2014).

Table 10. The reaction mixture for the KAPA Taq PCR kit

Component	Reaction mixture
Template DNA	Up to 250 ng
10mM dNTP Mix	2%
10uM Forward primer	4%
10uM Reverse primer	4%

10X KAPA Taq Buffer	10%
KAPA Taq DNA Polymerase (5 U / ul)	2%
Water, nuclease-free	To 100%

Table 11. KAPA Taq PCR reaction conditions

Step	Temperature	Time	Cycles
Initial denaturation	95ºC	3 min	1
Denaturation	98ºC	20 sec	25
Annealing	55ºC	15 sec	
Elongation	72ºC	1 min/kb	
Final Elongation	72ºC	1 min/kb	1

3.11. SDS PAGE analysis

Transformed yeast cells were grown overnight in 5 ml YPD and supplemented with the appropriate concentration of hygromycin (Table 6). Cells (800 µl) were collected in a tube at 13 000 x g for 30 seconds. Cell breakage was then performed with the addition of approx. 200 µl glass beads and 300 µl lysis buffer (Tris 0.1% SDS). The mixture was shaken (in a microtube homogenizer) for four pulses with each pulse being 1 minutes long with incubation on ice between each pulse.

After cell breakage, 50 µl of the cell suspension were mixed with 50 µl of Laemmli sample buffer (5% β-mercaptoethanol, 95% Laemmli buffer), boiled at 96˚C for 10 minutes and centrifuged at 3 000 x g for 5 minutes. Twenty microliters of the sample were loaded on a SDS-PAGE gel, which was prepared according to the 10% TGX™

FastCast^{TM} Acrylamide Kit ,BIO-RAD protocol (Bio-Rad Laboratories 2019a). Protein separation was performed for approximately 40 minutes at 200 V.

3.12. Western blot analysis

Following SDS-PAGE separation of proteins, these were analysed through western blot analyses using a Cas9 monoclonal HRP-conjugated antibody (Abcam).

All the buffers used for the preparation and performance of the western blot analysis are shown in Table 12.

Table 12. The buffers used for western blot analysis

Buffer	Composition
TGS	Tris glycine SDS
PBS (Phosphate Buffered Saline)	0.8% NaCl, 0.02% KCl, 0.144% Na$_2$HPO$_4$, 0.024% KH$_2$PO$_4$ (pH 7.4)
Towbin (Transfer) buffer	20% methanol, 10% 10 x TGS.
Wash buffer	0.05% Tween 20, 99.95% PBS
Blocking buffer	0.05% Tween 20, 5% milk powder, 200 ml PBS
Antibody solution buffer	blocking buffer, Cas9 monoclonal HRP-conjugated (1: 5 000 of antibody: blocking buffer)
Antibody substrate solution (Clarity Western ECl Substrate)	Luminol Enhancer Solution, Peroxide Solution (1:1 ratio)

The SDS page gel was equilibrated in Towbin transfer buffer (Bio-Rad Laboratories 2010). While the gel equilibration took place, 6 pieces of thick (0.8 mm) filter paper were soaked in Towbin buffer. For the PVDF membrane, the membrane was briefly transferred to 100% methanol for 30 seconds, then rinsed in water for 2 minutes

followed by the incubation in Towbin buffer for 10 minutes. The sandwich was staked on the anode in the order of: three filter papers, the membrane, the SDS PAGE gel and again three filter papers. The cathode were placed on top and the transferring of proteins took place at 25 V and 1 A for 30 minutes.

After the run, the membrane was briefly wetted in methanol and incubated in wash buffer for 2 minutes (Bio-Rad Laboratories 2019b). The membrane was then transferred to blocking buffer, protein side up, and incubated for 1 hour with continuous agitation. After the 1 hour incubation, the membrane was transferred to blocking buffer, containing the HRP-conjugated Cas9 antibody. The membrane was incubated overnight in the antibody solution followed by washing six times, with each wash consisting of a 5 minutes incubation step, in wash buffer with continuous agitation. The antibody substrate solution was prepared and added to the membrane and incubated at room temperature for 5 minutes. The solution was discarded, and the membrane was imaged and analysed using a Gel Doc™ XR+ (Bio-Rad).

4. Results and discussion

4.1. Construction of pKM180::CAS9

The wide range expression system pKM180 (Smit *et al.* 2012), was used as backbone for the development of a single CRISPR-Cas9 system (Figure 1A). The pKM180 uses a region of the rRNA locus from *Kluyveromyces marxianus* for integration into the yeast's genome. Selection of positive transformants is facilitated by the dominant selection of Hygromycin by the presence of the *HPH* gene, which is also modified to function in CTG-clade yeasts (e.g. *C. albicans*), which is driven and terminated by the *Saccharomyces cerevisiae TEF1* promoter and -terminator. The plasmid has *NotI* restriction sites to linearize the plasmid. For the cloning in bacteria, the pKM180 contains the normal bacterial moiety with the kanamycin resistant gene as a selection marker. The yeast part of the vector consists of a promoter region derived from the *Yarrowia lipolytica* transcription elongation factor (*TEF1*) gene that drives expression of a gene inserted in the pET28 derived multiple cloning site (MCS). Expression is terminated by the *K. marxianus* inulase terminator (*Km*INUt) sequence.
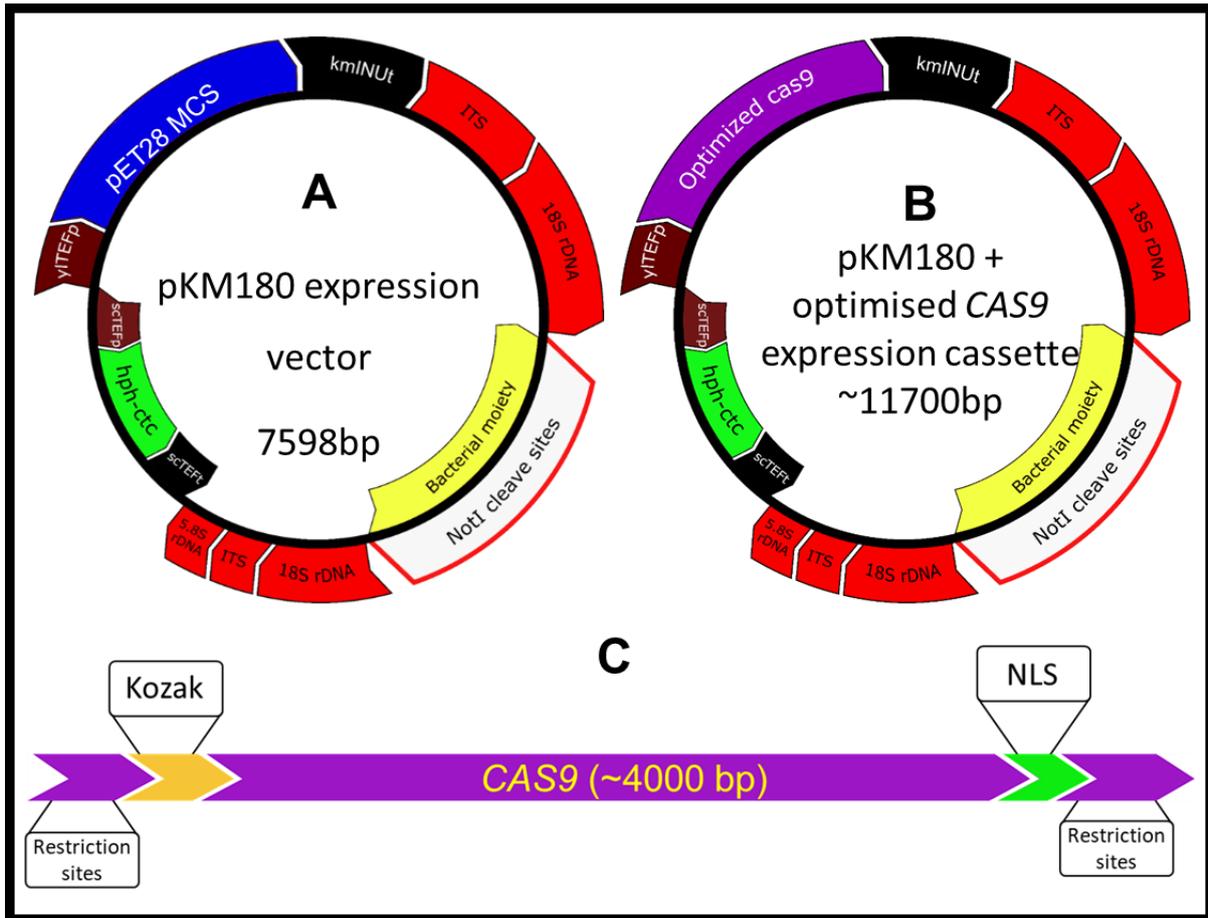
Figure 1: (A) the pKM 180 vector (7 598bp) that was used as the backbone for the construction of the (B) pKM180 + optimized *CAS9* vector (~11 700bp). (C) The general structure of the *CAS9* gene obtained consisting of two zones on both ends comprising several restriction sites, a Kozak sequence, the gene and a Nuclear localisation sequence (NLS).

For the construction of the expression cassettes the three different codons optimised *CAS9* genes were obtained from Genscript (GenScript 2002; Gao & Zhao 2014; Vyas *et al.* 2015; Weninger *et al.* 2016). The basis of each of the *CAS9* genes consist out of a region on both the 5- and 3-prime ends which contains various restriction sites for cloning purposes (Figure 1C). Following the restriction site region, on the 5-prime side of the gene, is a Kozak sequence to assist in translation initiation, this is followed by the *CAS9* gene and several SV40 nuclear localisation sequences (NLS), (1x NLS, 2x NLS and 3x NLS for the *P. pastoris*, *C. albicans* and *H. sapiens* optimised genes respectively). The purpose of the NLS is to target Cas9 localisation to the nucleus. In the case of the C*CAS9*, the gene is optimized to function in both the CTG and non-

CTG clade. In this case all the CTG codons were replaced with CTC codons. The reason for this is because some yeast including a number of *Candida* species have adopted a specific codon usage bias, where the CTG codon encodes serine instead of leucine (Translation table 12: Alternative Yeast Nuclear Code). This can be problematic due to the different properties the amino acids contain. Serine is commonly characterised as a polar amino acid, whereas leucine is characterized as a non-polar (hydrophobic) amino acid. In the formation of proteins, the hydrophobic amino acids tend to form bonds in the inside of the protein, where the environment is favourable for them, and the polar amino acids tends to be on the outside of the protein. This change can thus lead to the incorrect folding of the protein and thus altering its function. This modified version of the *CAS9* gene therefore allows selection in yeasts using the standard translation table as well as in yeast that use the alternative translation code. After the three different optimized Cas9 amino acid sequences were aligned to the native Cas9, using Geneious R10, (Figure 2) it was observed that there are several amino acids from the *P. pastoris* optimized Cas9 [which was obtained from Weninger and co-workers (2016)] that did not correspond to the native Cas9. This can be problematic when protein folding occurs within the cell and could influence the structure and function of the Cas9.
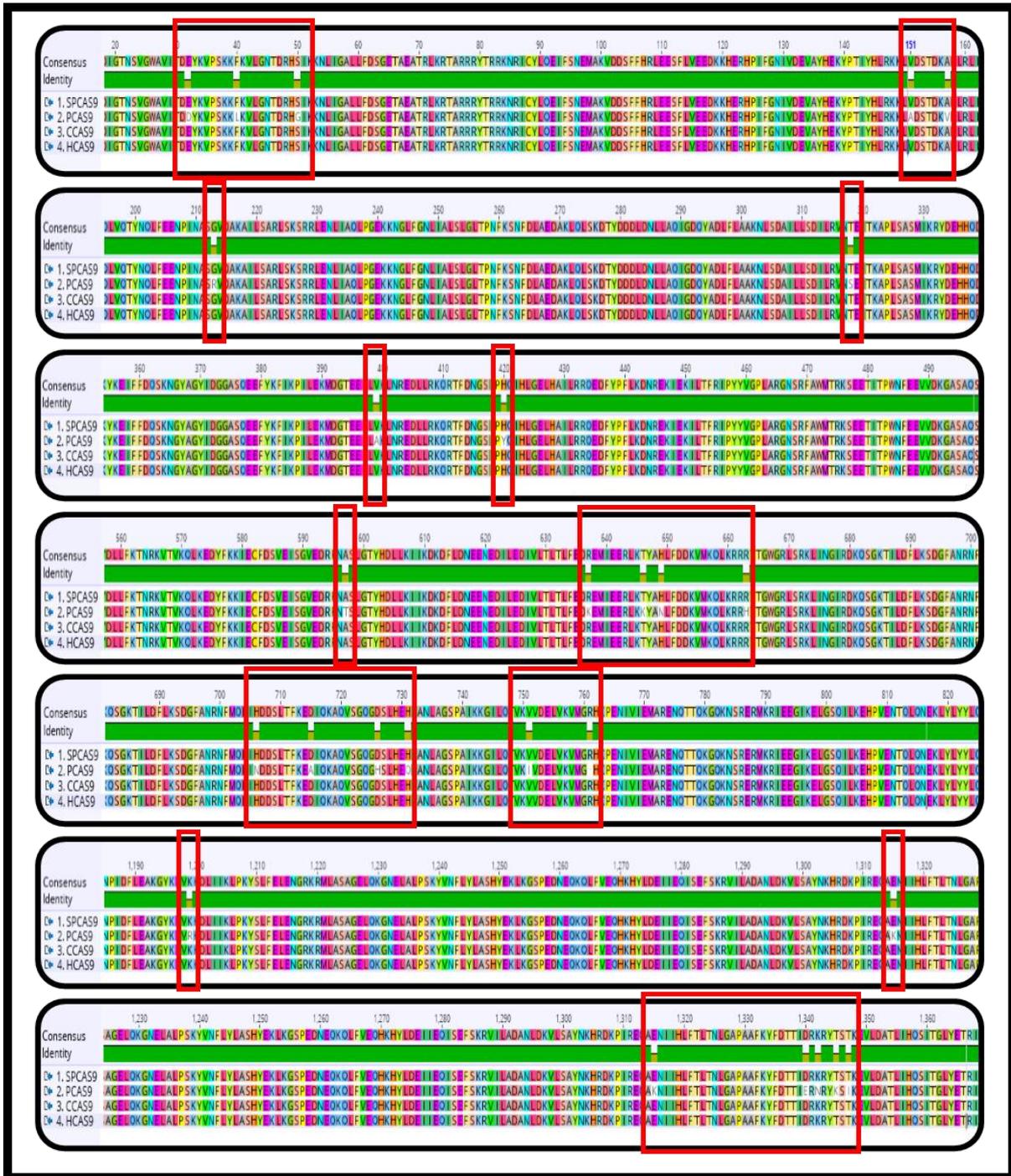
Figure 2: The alignment of the three different optimized *CAS9* translation (PCAS9 = *P. pastoris* CCAS9 = *C, albicans* and HCAS9 = *H. sapiens*) sequence to the native *CAS9* (SPCAS9= *Streptococcus pyogenes*) translation sequence. As can be seen from the alignment there are several amino acids (red boxes) which are different from the *P. pastoris* optimized Cas9 compared to the three other Cas9 proteins*.*

The 4 414 bp PCAS9, obtained in a pAO815 vector, was removed by restriction digestion by *Nco*I and *Hin*dIII (Figure 3B). The CCAS9 were obtained from the pESC-URA vector when cleaved by *Nco*I + *Bam*HI (Figure 4B), which yielded a fragment size of 4 213 bp. The HCAS9 (4 189bp) were removed from pAO815 with *Nco*I + *Bam*HI (Figure 5B). For the purpose of successful ligation and cloning, the pKM180 were cleaved with similar restriction enzymes respective to each optimized *CAS9* to yield fragment sizes of 7 506bp, 7 531bp and 7 531 bp for the PCAS9, CCAS9 and HCAS9, respectively (Figure 3A; 4A and 5A). DNA fragments were gel purified, pKM180 and the various *CAS9* genes ligated and the mixture transformed (as mentioned in chapter 2 section 3.8).

The colonies obtained from the transformation were screened with restriction enzymes which specifically cuts in the *CAS9* gene. pKM::PCAS9 were digested with *Afl*II, this yielded fragment sizes of 6 695bp, 3 779bp and 1 183bp for successful transformants and 7 598bp for negative transformants. As seen in Figure 3, three clones [(Figure 3C) sample 2, 3, 4] were obtained where all of them indicated the correct restriction profile.
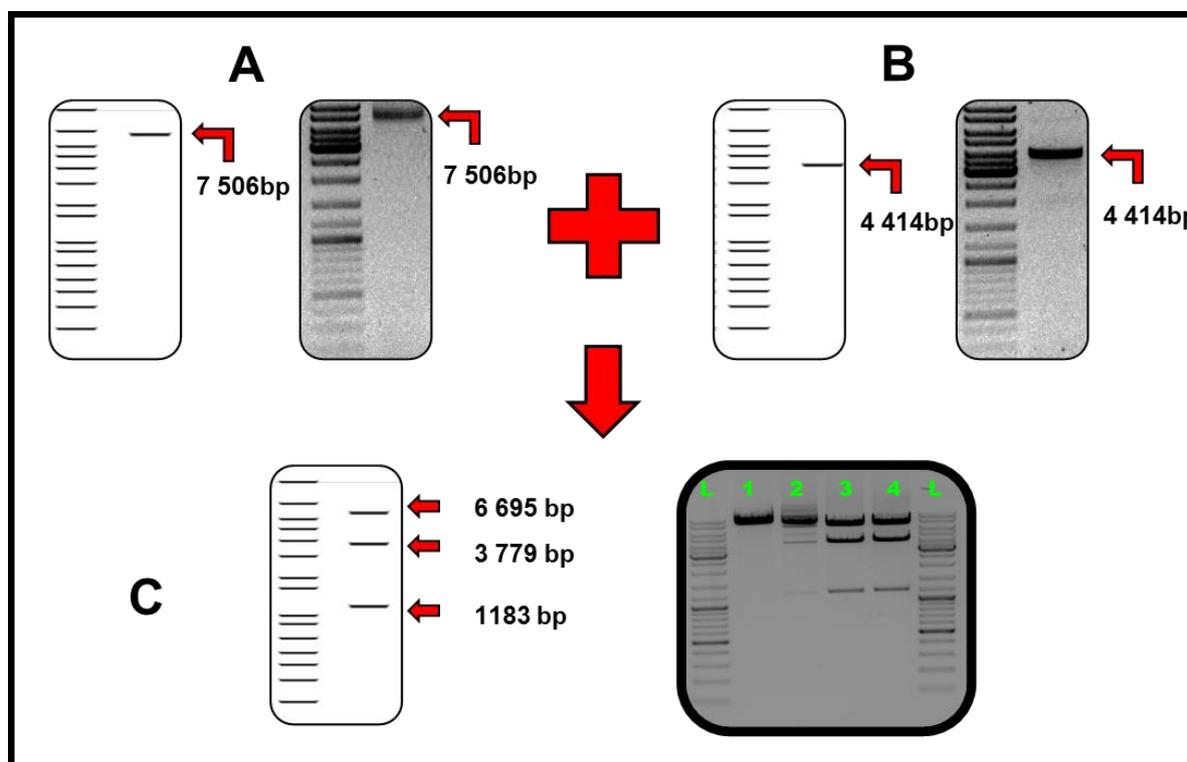

Figure 3: A) The pKM180 linearized with *Nco*I and *Hin*dIII (7 506bp). B) PCAS9, which was cleaved from the pAO815 vector with *Nco*I and *Hin*dIII (4 414bp). C) After ligation the colonies were screened with *Afl*II. For successful ligations fragment sizes of 6 695

bp, 3 779 bp and 1 183 bp were obtained. L represents the DNA ladder; 1 is the negative control (pKM180 without insert) and 2 to 4 pKM::PCAS9.

For the pKM::CCAS9 vector the transformants were screened with *Hin*dIII, which resulted in 8 530bp, 2 463bp and 755bp bands if successful and 7 598bp if unsuccessful. As seen in Figure 4C, several colonies were obtained after transformation, after which 14 were randomly screened. Samples 5, 6, 9 and 12 to 15 tested successful for ligation to obtain the pKM::CCAS9 vector
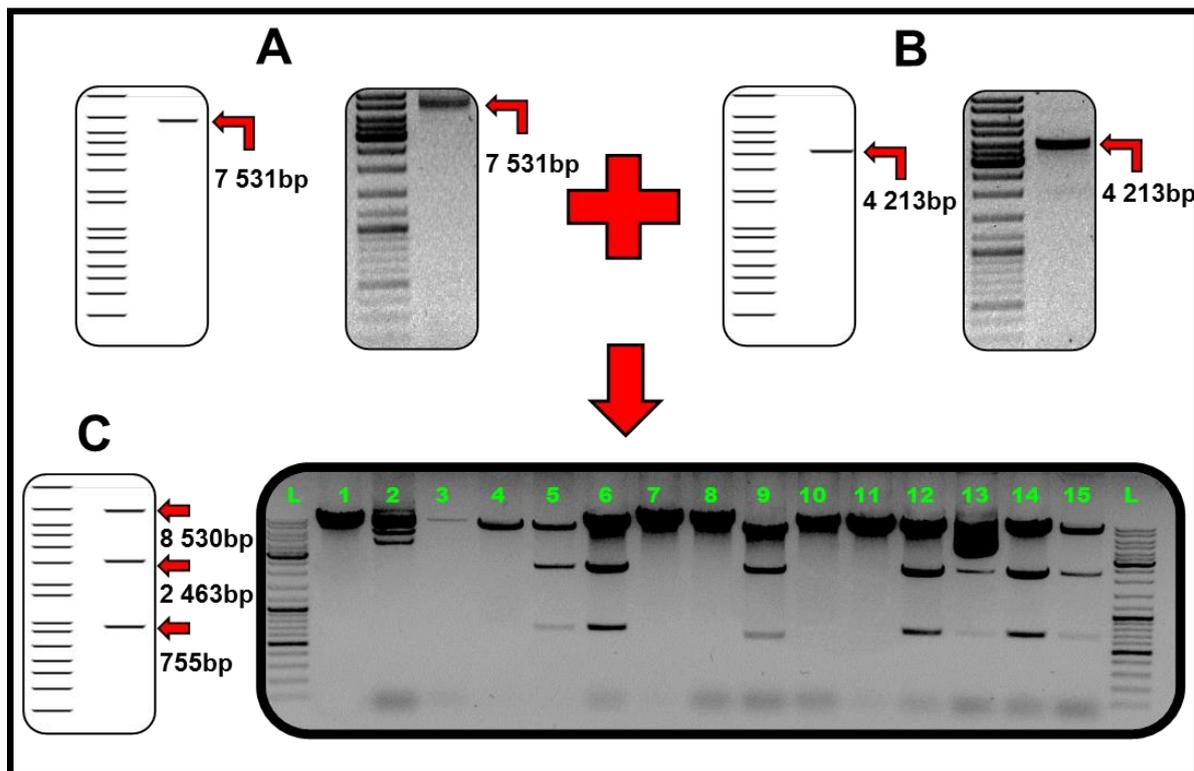

Figure 4: A) The pKM180 vector linearized with *Nco*I + *Bam*HI (7 531bp). B) CCAS9, which was cleaved from the pESC-URA vector with *Nco*I + *Bam*HI (4 213bp). C) After ligation the colonies were screened with *Hin*dIII. For successful ligations fragment sizes of 8 530bp, 2 463bp and 755bp have been obtained. L represents the DNA ladder; 1 is the negative control and 2 to 15 is the screened colonies.

Lastly, HCAS9 was digested with *Nru*I and *Sal*I to yield 10 766bp and 950bp when successful and 7 598bp if unsuccessful. As seen in Figure 5C, several colonies were obtained after transformation, after which 14 were randomly screened. Samples 5 to 7, 9,12, 13 and 15 tested successful for ligation to obtain the HCAS9 vector
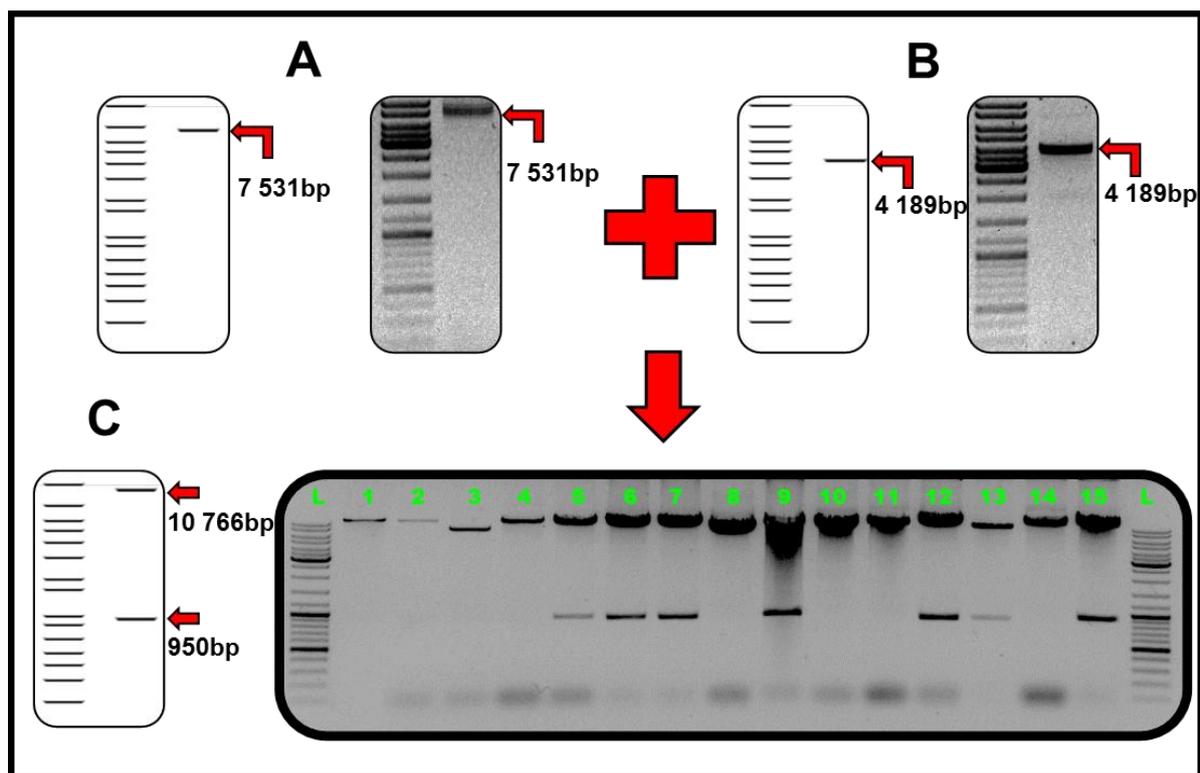
Figure 5: A) The pKM180 vector linearized with *Nco*I + *Bam*HI (7 531bp). B) HCAS9, which was cleaved from the pAO815 vector with *Nco*I + *Bam*HI (4 189bp). C) After ligation the colonies were screened with *Nru*I and *Sal*I. For successful ligations fragment sizes of 10 766bp and 950bp have been obtained. L represents the DNA ladder; 1 is the negative control and 2 to 15 is the screened colonies.

4.2. Yeast Transformation

After construction of the pKM180::CAS9 (Figure 1B), yielding the pKM::PCAS9 (11 645 bp), pKM::CCAS9 (11 736 bp) and pKM::HCAS9 (11 712bp), these (as well as the empty pKM180 plasmid as positive control) was transformed into 12 different yeasts (Table 7). Before transformation each of the plasmids were linearized by *Not*I digestion to remove the bacterial moiety and therefore exposing the rDNA regions to allow homologous recombination to take place. For *Saccharomyces cerevisiae* CENPK42, *Kluyveromyces lactis* UOFS-Y1167, *Yarrowia lipolytica* Po1f, *Debaryomyces hansenii* UOFS-Y0610, *Arxula adeninivorans* UOFS-Y1219, *Arxula adeninivorans* Ls3 and *Ogataea polymorpha* UOFS Y1507 the one-step transformation method was used for maximal transformation efficiency. For *Arxula adeninivorans* UOFS-Y1220, *Komagataella phaffi* UOFS-Y1552, *Komagataella phaffi*

Chapter 2

SMD 1168, *Komagataella phaffi* Km.71 and *Komagataella phaffi* GS115 transformation, the bicine method was used to achieve maximal transformation efficiency (Klebe *et al.*). Table 7 indicates potential toxicity of the CAS9 vectors through the increased susceptibility to hygromycin. Some of the yeast strains become more sensitive to Hygromycin following transformation of the pKM180::*CAS9* gene. This was most evident when the HCAS9 was transformed. For these strains, the Hygromycin was reduced to allow growth of transformed colonies (Table 7). The reason for this increased sensitivity is however not known.

After transformation, integration of the expression vectors into the transformed yeast genome were confirmed through PCR analysis (Figure 6). For the pKM::PCAS9 and pKM::CCAS9 primers cacas9-1 forward and reverse were used to amplify a 694bp part of the *CAS9*. For the pKM::HCAS9 primers hscas9-1 forward and reverse was used to amplify an 816 bp part of the *CAS9*.
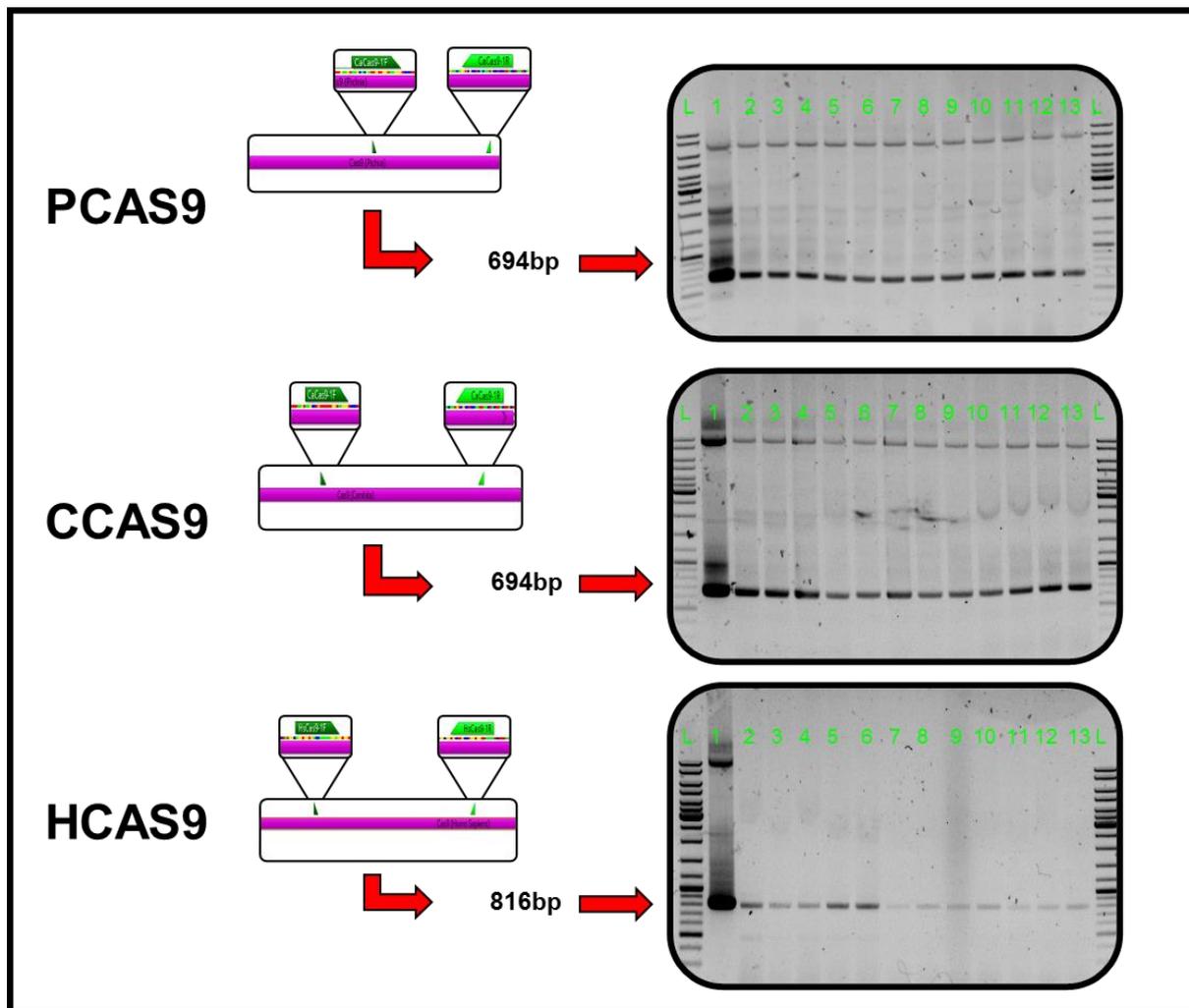
Figure 6: Vector integration into the host genome were successful in all the transformed yeast as displayed on each gel respective to the *CAS9* vector used. L represents the ladder, 1 represents the positive control (plasmid DNA) and 2 to 13 are the yeast samples, on all three of the gels. The primers binding site on each of the three different optimized *CAS9* genes are also depicted. The large amplicon in each lane may possibly represent impurities.

4.3. Western blot analysis

Western blot analysis was performed to confirm successful expression of the Cas9 protein. In the initial protocol, a centrifugation step (14 000 x *g* for 5 minute) was included to remove the cellular debris after the cell breakage with glass beads. It was determined that the glass beads used did not disrupt the nucleus of cells, and thus the nucleus was removed along with the cellular debris during the initial centrifugation step. This resulted in the removal of the Cas9 protein from the suspension and therefore no bands were obtained on the western blot membranes. Note that these membranes were not included due to the negative results. Therefore, the centrifugation step was removed and protein preparation for SDS PAGE was performed with the cellular debris.

Western blot analysis with the HRP conjugated Cas9 antibody confirmed that all three of the different optimized Cas9 proteins were expressed at their respective molecular weights when compared to the protein standard (Figure 7). However, as can be seen, in Figure 7C, first lane in panel 1 to 5; 9 and 12 (see Table 14 for the yeast referencing number) the pKM::PCAS9 had more than one band. However, the reason for this phenomenon is not known and difficult to determine. A possible reason may be that the yeasts started translation at a different ATG codon downstream, which resulted in different size proteins. This may also contribute to the reason for the zero targeting efficiency obtained by Weninger and co-workers, when they used the PCAS9 (Weninger *et al.* 2016).
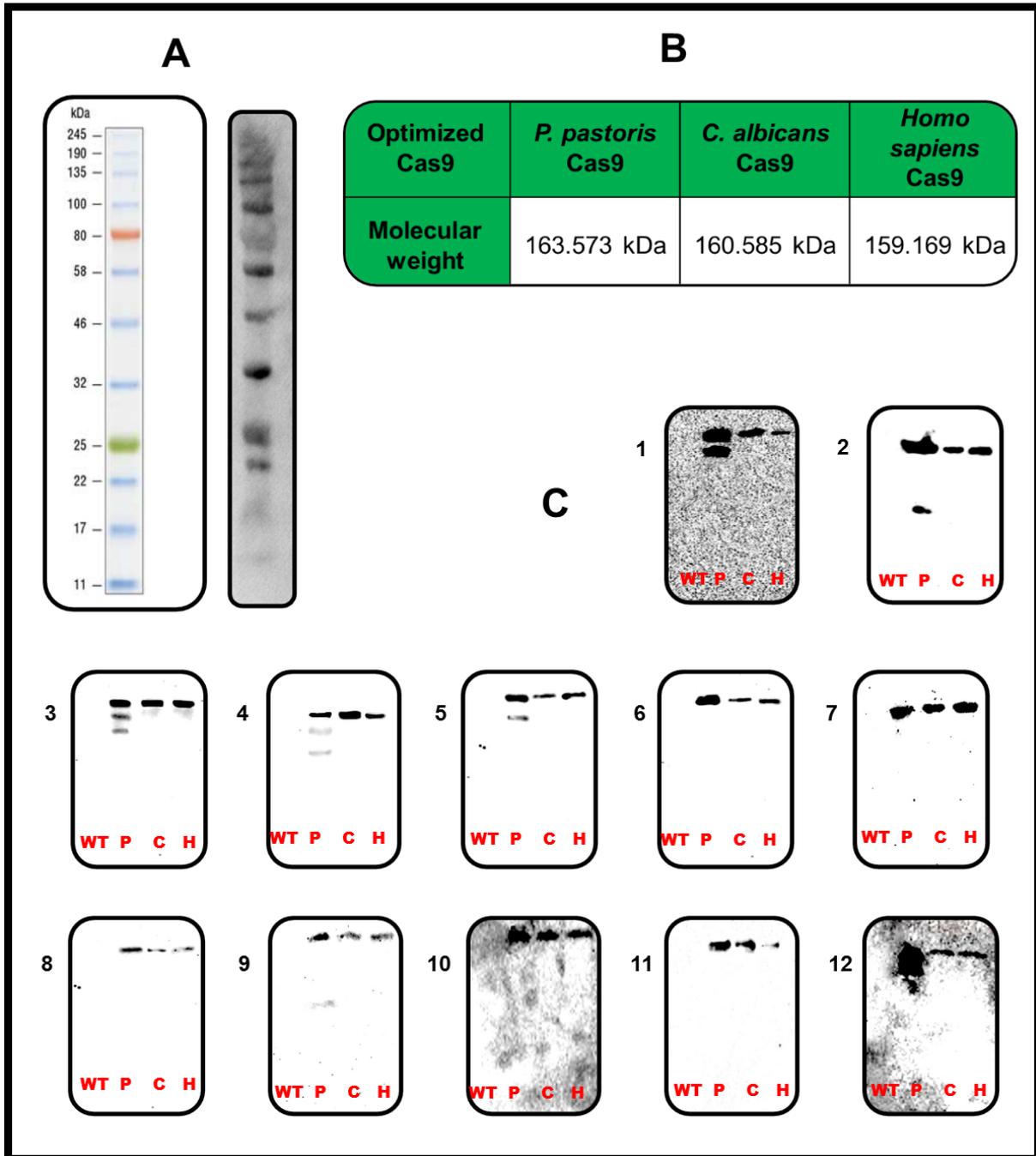
Figure 7: A) Protein ladder (colour prestained protein standard, broad range (11-245 kDa)) used for the determination of the protein band sizes. B) The calculated molecular weight of the three different optimized Cas9 proteins. (C) The different optimized Cas9 proteins visualised through western blot analysis with the HRP-conjugated Cas9 antibody, in each yeast (see Table 13 for the yeast reference number). WT represents wild type, P represents PCAS9, C represents CCAS9 and H represents HCAS9.

Table 13 The different yeast used in this study as well as their reference number.

Reference number	Yeast
1	*Saccharomyces cerevisiae* CENPK42
2	*Kluyveromyces lactis* UOFS-Y1167
3	*Yarrowia lipolytica* Po1f
4	*Debaryomyces hansenii* UOFS-Y0610
5	*Arxula adeninivorans* UOFS-Y1220
6	*Arxula adeninivorans* UOFS-Y1219
7	*Arxula adeninivorans* Ls3
8	*Ogataea polymorpha* UOFS Y1507
9	*Komagataella phaffii* UOFS-Y1552
10	*Komagataella phaffii* SMD 1168
11	*Komagataella phaffii* Km.71
12	*Komagataella phaffii* GS115

5. Conclusion

A number of yeast CRISPR-Cas9 systems have been developed and applied successfully (DiCarlo *et al.* 2013; Jacobs *et al.* 2014; Horwitz *et al.* 2015; Schwartz *et al.* 2016; Weninger *et al.* 2016; Juergens *et al.* 2018b). These systems are designed to satisfy the special demands of a specific yeast and they follow similar construction principles. The aim of this study was to construct a wide range yeast CRISPR-Cas9

system for the application in a wide variety of different yeasts and to explore the system's potential in regards to gene editing.

For the first part of the study three different pKM 180::Cas9 expression constructs were developed, each with a different codon optimized (*P. pastoris*, *C. albicans* and *H. sapiens*), constitutively expressed *CAS9*. Each of the *CAS9* genes were fused with a nuclear localization sequence (1x NLS, 2x NLS and 3x NLS for *P. pastoris*, *C. albicans* and *H. sapiens*, respectively), to ensure transportation of the expressed protein into the nucleus. These expression constructs make use of the rRNA locus from *Kluyveromyces marxianus* for integration into the yeasts and selection of positive transformants is facilitated by the dominant selection of Hygromycin. The three different *CAS9* constructs were transformed into seven different biotechnologically relevant yeasts (12 strains in total) using two different chemical transformation protocols. Some yeast showed reduced growth likely due to a toxic effect of the Cas9 protein, thus a lower hygromycin concentration was used to accommodate for the reduction in growth. Integration of the expression vectors into the yeast genome were confirmed through PCR and gel electrophoresis. This was followed by confirming successful expression of the Cas9 via western blot analysis using a Cas9 specific monoclonal HRP-conjugated antibody. In addition, the PCAS9 showed extra bands on some of the western blot membranes, which are possibly due to either incorrect translation or folding or possibly proteolysis of the PCas9 protein.

6. References

Adli, M., (2018) 'The CRISPR tool kit for genome editing and beyond'. *Nat Commun* **9**, 1911. doi: 10.1038/s41467-018-04252-2

Bao, Z.,Xiao, H.,Liang, J.,*et al.*, (2015) 'Homology-Integrated CRISPR–Cas (HI-CRISPR) System for One-Step Multigene Disruption in *Saccharomyces cerevisiae*'. *ACS Synth Biol* **4**, 585–594. doi: 10.1021/sb500255k

Barrangou, R.,Fremaux, C.,Deveau, H.,*et al.*, (2007) 'CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes'. *Science (80-)* **315**, 1709–1712. doi: 10.1126/science.1138140

Chapter 2

Bhaya, D.,Davison, M.,& Barrangou, R., (2011) 'CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation'. *Annu Rev Genet* **45**, 273–297. doi: 10.1146/annurev-genet-110410-132430

Bio-Rad Laboratories., (2019a) 'TGX ™ and TGX Stain-Free ™ FastCast ™ Acrylamide Kit and Starter Kit Instruction Manual'. http://www.bio-rad.com/. Accessed, August 23, (2018)

Bio-Rad Laboratories., (2010) 'Trans-Blot ® Turbo ™ Blotting System'. http://www.bio-rad.com/. Accessed, August 23, (2018)

Bio-Rad Laboratories., (2019b) 'Western ECL Substrates'. http://www.bio-rad.com/. Accessed, August 23, (2018)

Birnboim, H.C.,& Doly, J., (1979) 'Nucleic Acids Res. 1979 Birnboim'. *Nucleic Acid Res* **7**, 1–11. doi: 10.1093/nar/7.6.1513

Chen, D.C.,Yang, B.C.,& Kuo, T.T., (1992) 'One-step transformation of yeast in stationary phase'. *Curr Genet* **21**, 83–84. doi: 10.1007/BF00318659

Chung, C.T.,Niemela, S.L.,& Miller, R.H., (1989) 'One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution.'. *Proc Natl Acad Sci* **86**, 2172–2175. doi: 10.1073/pnas.86.7.2172

Devi, S.,Kedlaya, R.,Maddodi, N.,*et al.*, (2009) 'Calcium homeostasis in human melanocytes: role of transient receptor potential melastatin 1 (TRPM1) and its regulation by ultraviolet light'. *Am J Physiol Physiol* **297**, C679–C687. doi: 10.1152/ajpcell.00092.2009

DiCarlo, J.E.,Norville, J.E.,Mali, P.,*et al.*, (2013) 'Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems'. *Nucleic Acids Res* **41**, 4336–4343. doi: 10.1093/nar/gkt135

Gao, Y.,& Zhao, Y., (2014) 'Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing'. *J Integr Plant Biol* **56**, 343–349. doi: 10.1111/jipb.12152

Gasiunas, G.,Barrangou, R.,Horvath, P.,& Siksnys, V., (2012) 'Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria'. *Proc Natl Acad Sci* **109**, E2579–E2586. doi:

10.1073/pnas.1208507109

Generoso, W.C.,Gottardi, M.,Oreb, M.,& Boles, E., (2016) 'Simplified CRISPR-Cas genome editing for *Saccharomyces cerevisiae*'. *J Microbiol Methods* **127**, 203–205. doi: 10.1016/j.mimet.2016.06.020

GenScript., (2002) 'GenScript'. https://www.genscript.com/. Accessed, January 30, (2019)

Gerngross, T.U., (2004) 'Advances in the production of human therapeutic proteins in yeasts and filamentous fungi.'. *Nat Biotechnol* **22**, 1409–1414. doi: 10.1038/nbt1204-1589e

Horwitz, A.A.,Walter, J.M.,Schubert, M.G.,*et al.*, (2015) 'Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas'. *Cell Syst* **1**, 88–96. doi: 10.1016/j.cels.2015.02.001

Ish-Horowicz, D.,& Burke, J.F., (1981) 'Rapid and efficient cosmid cloning'. *Nucleic Acids Res* **9**, 2989–2898. doi: 10.1093/nar/9.13.2989

Jacobs, J.Z.,Ciccaglione, K.M.,Tournier, V.,& Zaratiegui, M., (2014) 'Implementation of the CRISPR-Cas9 system in fission yeast'. *Nat Commun* **5**, 5344. doi: 10.1038/ncomms6344

Jakočiūnas, T.,Bonde, I.,Herrgård, M.,*et al.*, (2015) 'Multiplex metabolic pathway engineering using CRISPR/Cas9 in *Saccharomyces cerevisiae*'. *Metab Eng* **28**, 213–222. doi: 10.1016/j.ymben.2015.01.008

Jinek, M.,Chylinski, K.,Fonfara, I.,*et al.*, (2012) 'A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity'. *Science (80-)* **337**, 816–821. doi: 10.1126/science.1225829

Juergens, H.,Varela, J.A.,Gorter de Vries, A.R.,*et al.*, (2018) 'Genome editing in Kluyveromyces and Ogataea yeasts using a broad-host-range Cas9/gRNA co-expression plasmid'. *FEMS Yeast Res* **18**, . doi: 10.1093/femsyr/foy012

KAPA Biosystems Inc., (2014) 'KAPA HiFi HotStart ReadyMix PCR Kit Technical Data Sheet'. https://www.kapabiosystems.com/product-applications/products/pcr-2/kapa-hifi-pcr-kits/. Accessed, January 28, (2019)

Chapter 2

Klebe, R.J.,Harriss, J. V.,Sharp, Z.D.,& Douglas, M.G., (1983) 'A general method for polyethylene-glycol-induced genetic transformation of bacteria and yeast.'. *Gene* **25**, 333–341

Madzak, C.,Gaillardin, C.,& Beckerich, J.-M., (2004) 'Heterologous protein expression and secretion in the non-conventional yeast *Yarrowia lipolytica* : a review'. *J Biotechnol* **109**, 63–81. doi: 10.1016/j.jbiotec.2003.10.027

Mans, R.,van Rossum, H.M.,Wijsman, M.,*et al*., (2015) 'CRISPR/Cas9: A molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*'. *FEMS Yeast Res* **15**, 1–15. doi: 10.1093/femsyr/fov004

Nevoigt, E., (2008) 'Progress in Metabolic Engineering of *Saccharomyces cerevisiae*'. *Microbiol Mol Biol Rev* **72**, 379–412. doi: 10.1128/MMBR.00025-07

New England Biolabs Inc., (2014) 'NEBcloner'. http://nebcloner.neb.com/#!/redigest. Accessed, January 07, (2019)

Ryan, O.W.,Skerker, J.M.,Maurer, M.J.,*et al*., (2014) 'Selection of chromosomal DNA libraries using a multiplex CRISPR system.'. *Elife* **3**, 1–15. doi: 10.7554/eLife.03703

Schwartz, C.M.,Hussain, M.S.,Blenner, M.,& Wheeldon, I., (2016) 'Synthetic RNA Polymerase III Promoters Facilitate High-Efficiency CRISPR–Cas9-Mediated Genome Editing in *Yarrowia lipolytica*'. *ACS Synth Biol* **5**, 356–359. doi: 10.1021/acssynbio.5b00162

Skrzypek, M.S.,Binkley, J.,Binkley, G.,*et al*., (2017) 'The *Candida* Genome Database (CGD): Incorporation of Assembly 22, systematic identifiers and visualization of high throughput sequencing data'. *Nucleic Acids Res* **45**, D592–D596. doi: 10.1093/nar/gkw924

Smit, M.S.,Theron, C.W.,Albertyn, J.,& Labuschagné, M., (2012) 'Novel expression construct'. 04652

Stovicek, V.,Borja, G.M.,Forster, J.,& Borodina, I., (2015) 'EasyClone 2.0: expanded toolkit of integrative vectors for stable gene expression in industrial *Saccharomyces cerevisiae* strains'. *J Ind Microbiol Biotechnol* **42**, 1519–1531.

doi: 10.1007/s10295-015-1684-8

Terns, M.P.,& Terns, R.M., (2011) 'CRISPR-based adaptive immune systems'. *Curr Opin Microbiol* **14**, 321–327. doi: 10.1016/j.mib.2011.03.005

Thermo Fisher Scientific Inc., (2015a) 'Restriction Enzymes–Thermo Scientific - ZA'. https://www.thermofisher.com/za/en/home/brands/thermo-scientific/molecular-biology/thermo-scientific-restriction-modifying-enzymes/restriction-enzymes-thermo-scientific.html. Accessed, January 07, (2019)

Thermo Fisher Scientific Inc., (2015b) 'Thermo Scientific GeneJET Gel Extraction Kit'. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0012661_GeneJET_Gel_Extraction_UG.pdf. Accessed, January 07, (2019)

Turvey, S., (1997) 'Competent cells : Rubidium Chloride'. **1**, 4000

Vyas, V.K.,Barrasa, M.I.,& Fink, G.R., (2015) 'A *Candida albicans* CRISPR system permits genetic engineering of essential genes and gene families'. *Sci Adv* **1**, e1500248. doi: 10.1126/sciadv.1500248

Wang, R.,Preamplume, G.,Terns, M.P.,*et al.*, (2011) 'Interaction of the Cas6 Riboendonuclease with CRISPR RNAs: Recognition and Cleavage'. *Structure* **19**, 257–264. doi: 10.1016/j.str.2010.11.014

Weninger, A.,Hatzl, A.-M.,Schmid, C.,*et al.*, (2016) 'Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast *Pichia pastoris*'. *J Biotechnol* **235**, 139–149. doi: 10.1016/j.jbiotec.2016.03.027

Zhang, G.-C.,Kong, I.I.,Kim, H.,*et al.*, (2014) 'Construction of a Quadruple Auxotrophic Mutant of an Industrial Polyploid *Saccharomyces cerevisiae* Strain by Using RNA-Guided Cas9 Nuclease'. *Appl Environ Microbiol* **80**, 7694–7701. doi: 10.1128/AEM.02310-14

Chapter 3

Development of the wide range CRISPR-Cas9 system by incorporating a RGR-gRNA construct and validation through *ADE2* disruption
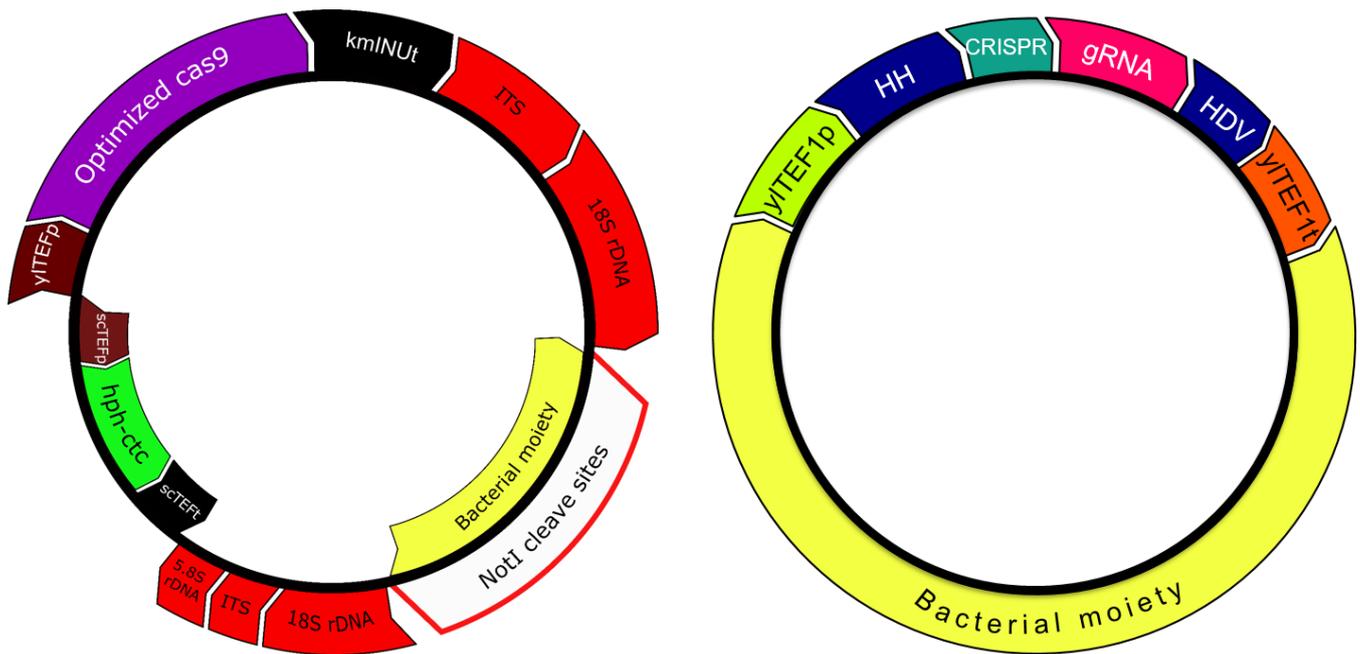
Chapter 3

1. Abstract

The wide range pKM180 consisting of a *C. albicans* optimized *CAS9* (CCAS9) gene and a gRNA construct flanked by ribozymes was developed. Both the *CAS9* and the gRNA expression were constitutively driven by a *Yarrowia lipolytica TEF1* promoter. The integration of the vector is facilitated by the rRNA locus from *Kluyveromyces marxianus* and selection of positive transformants is through hygromycin resistance. The vector was transformed into the different yeasts and the efficacy of the CRISPR-Cas9 gene editing system was determined by targeting the *ADE2* gene which if disrupted results in a red phenotype of colonies. When the vector was transformed without donor DNA, most of the yeast displayed moderate to high disruption efficiency through non-homologous end joining (NHEJ). When donor DNA was included in the transformation, most of the yeast displayed gene disruption through homologous recombination (HR), however at low to moderate efficiency. This proved that the wide range CRISPR-Cas9 system was applicable in a wide variety of different yeast thus allowing for rapid, cost-effective genetic manipulation of biotechnologically relevant yeast.

2. Introduction

With the application of CRISPR-Cas9 in yeast, there are two main strategies used for gRNA expression (Gao & Zhao 2014). The first strategy is the use of an RNA polymerase III (Pol III) promoter for example the U3 or U6 promoters. The second strategy makes use of the self-cleavage ability of ribozymes. Using the latter strategy allows the expression of the gRNA construct using any promoter. In this system the gRNA construct is flanked by self-catalysed ribozymes, which promotes self-cleavage and remove any modifications (poly(A) tail or 7-methylguanylate cap) that have been added during transcription. Here the application of this strategy is described in the development of a wide range CRISPR-Cas9 system where the use of a single promoter for the expression of the gRNA in all of the yeast is essential.

The CRISPR-Cas9 engineering tool has been successfully tested in a number of organisms (Jinek *et al.* 2012; Hwang *et al.* 2013; DiCarlo *et al.* 2013; Wang *et al.* 2013; Gratz *et al.* 2013; Friedland *et al.* 2013; Jiang *et al.* 2013; Vyas *et al.* 2015; Kistler *et al.* 2015; Yan *et al.* 2017). The system creates double stranded breaks through RNA-

guided DNA cleavage, resulting in a very easy method to achieve targeted gene disruption and insertion (Jinek *et al.* 2012; Horwitz *et al.* 2015). This system commonly consists of the Cas9 enzyme and a guide RNA (gRNA). The Cas9 enzyme performs the cleavage action of the double stranded DNA and consists of two nuclease domains, namely RuvC and HNH (Haft *et al.* 2005; Makarova *et al.* 2006). The RuvC domain cleaves the DNA strand that contains the PAM sequence and the HNH domain cleaves the target strand. The gRNA is a short synthetic RNA which originated when the site-specific crRNA was fused to a tracrRNA (scaffold part) (Jinek *et al.* 2012). The scaffold part is necessary for the association with the Cas9 enzyme and the user-specified crRNA part (approximately 20 bp of length), which states the target DNA in the host genome to be modified. Therefore, by just changing the target sequence in the gRNA, one can simply target any gene for modification.

The gRNA binds to the Cas9 nuclease and guides it to the specific target (Jinek *et al.* 2012). The Cas9 then binds to, and cleaves the target DNA to generate a double stranded break (DSB). In yeast, this DSB can lead to gene alteration that can inactivate the gene through non-homologous end joining (NHEJ) with the creation of indels (insertion or deletion of nucleotides) (Pitcher *et al.* 2007; Bétermier *et al.* 2014). These indels may create a frame shift in bases, which can create different amino acid sequence and thus will diminish proper protein folding and activity. This DSB can also introduce an opportunity for homologous recombination (HR) to insert a donor fragment (if provided) specified for the user's purpose (Pâques & Haber 1999; Sung & Klein 2006). However, there is one major challenge in using CRISPR for targeted gene editing, which is the production of the gRNAs (Gao & Zhao 2014).

The first 20- nucleotide sequence of the gRNA is used to guide targeted DNA cleavage (Jinek *et al.* 2012). Any additional bases or modifications at the ends of gRNA will abolish the gRNA's ability to guide DNA cleavage by Cas9 (Haurwitz *et al.* 2010; Jinek *et al.* 2012). Hence, the expression of the gRNA needs to be closely monitored for successful gene targeting. The most widely used and well-characterized promoters are RNA polymerase II promotors (Pol II) (Gao & Zhao 2014). However, these promoters cannot be used directly to produce gRNA for CRISPR, due to the fact that mRNAs that are transcribed through Pol II promoters undergo extensive processing and modification at both ends. Therefore, the gRNA expression cassette is usually driven by a non-mRNA polymerase promoter for example (Pol III) promoters (U3 and

U6), the promoter is followed by a specific target sequence, the gRNA scaffold and then the terminator. However, the Pol III promoters have limitations when it comes to the gRNA expression in different host. They are very specific for each organism and not all these promoters have been characterized in every organism, thus, making it difficult to choose the correct RNA polymerase promoters for CRISPR. Furthermore, the RNA polymerase promoters limit the CRISPR target sequences to G (N20) GG and A (N20) GG, respectively. Therefore, another strategy has been developed and used successfully for the correct expression of the gRNA. Gao and Zhao developed a strategy where they took advantage of the nuclease activity of ribozymes. They designed an artificial gene, RGR (Ribozyme-gRNA-Ribozyme), where the RGR gene undergoes self-catalysed cleavage to yield the gRNA without any modifications. This strategy leads to the successful expression of the gRNA using any promoter, thus resulting in successful guidance of the Cas9 enzyme to the target gene. In addition, this RGR strategy diminished the limitation to what the target sequence (in the gRNA) must contain. The gRNA expression construct can thus contain any RNA polymerase II promoter followed by a Hammerhead (HH) self-splicing ribozyme, the target sequence, the gRNA scaffold, a Hepatitis delta virus (HDV) self-splicing ribozyme and the terminator.

The RGR strategy needs to be applied to design an integrative wide range CRISPR-Cas9 system for the application in a wide variety of different yeast (Gao & Zhao 2014).

In the previous chapter, three different optimized *CAS9* vectors (pKM::PCAS9, pKM::CCAS9 and pKM::HCAS9) were constructed using the wide range pKM180 vector and three different optimized *CAS9* [*P. pastoris* (PCAS9)*, C. albicans* (CCAS9) and *H. sapiens* (HCAS9)]. The vectors were transformed into the yeast and expression of each of the different optimized *CAS9* were tested to observe if the Cas9 protein is being expressed. For further development of the wide range CRISPR-Cas9 system, a gRNA needs to be included together with the *CAS9*. This will provide the vector with gene targeting characteristics, which will allow for gene targeting in the different yeast tested. Therefore, the aim of this part of the study was to construct and validate a wide range CRISPR-Cas9 vector. This vector contained the optimized *CAS9* as well as an RGR ADE2 CRISPR target site and was used to test for successful and efficient gene disruption in *Saccharomyces cerevisiae* and six other non-conventional yeast.

3. Material and methods

3.1. *In silico* work

All *in silico* design and construction of plasmids were performed using Geneious R10 (Geneious 2017).

3.2. *E. coli* Rubidium chloride competent cells

Refer to chapter 2 section 3.1

3.3. Bacterial Transformation

Refer to chapter 2 section 3.2

3.4. Restriction digestion

Refer to chapter 2 section 3.3

3.5. Polymerase Chain Reaction

All PCR reactions performed were either perform using the KAPA Taq PCR kit or the KOD Hot Start PCR kit (Novagen 2011; KAPA Biosystems Inc. 2014). The PCR reaction mixtures were prepared according to either Table 1 or 3 and with reaction conditions as stated in either Table 2 or 4, respectively.

Table 1. The reaction mixture for the KAPA Taq PCR kit (KAPA Biosystems Inc. 2014)

Component	Reaction mixture
Template DNA	Up to 250 ng
10mM dNTP Mix	2%
10µM Forward primer	4%
10µM Reverse primer	4%

10X KAPA Taq Buffer	10%
KAPA Taq DNA Polymerase (5 U / µl)	2%
Water, nuclease-free	To 100%

Table 2. KAPA Taq PCR reaction conditions (KAPA Biosystems Inc. 2014)

Step	Temperature	Time	Cycles
Initial denaturation	95ºC	3 min	1
Denaturation	98ºC	20 sec	25
Annealing	Lowest Primer TmºC for 10 sec	15 sec	
Elongation	72ºC	1 min/kb	
Final Elongation	72ºC	1 min/kb	1

Table 3. The reaction mixture for the KOD Hot Start PCR kit (Novagen 2011)

Component	Reaction mixture
Template DNA	Up to 250 ng
2mM dNTP Mix	10%
25mM MgSO4	6%
10µM Forward primer	3%
10µM Reverse primer	3%

10X KOD Hot Start Buffer	10%
KOD Hot Start Polymerase (1 U / µl)	2%
Water, nuclease-free	To 100%

Table 4. KOD Hot Start PCR reaction conditions (Novagen 2011)

Step	Target size	Cycles
Pol. activation	95ºC	1
Denaturation	95ºC for 20 sec	30
Annealing	Lowest Primer TmºC for 10 sec	
Extension	70ºC 25 sec/kb	

All the primers used in this study are presented in Table 5. The first two letters represents the yeast that the primers were designed for, Aa= *Arxula adeninivorans* ; Dh= *Debaryomyces hansenii*; Kl= *Kluyveromyces lactis*; Op= *Ogataea polymorpha*; Pp= *Komagataella phaffii*; Sc= *Saccharomyces cerevisiae*; Yl= *Yarrowia lipolytica*.

The sequences that are in bold represent the *ADE2* CRISPR site. For the identification of the CRISPR target sites the CRISPR tool in Geneious R10 (www.geneious.com) was used. The *ADE2* ORF of each of the selected yeast species was analysed for a GN(20)NGG motif. Off target sites were eliminated by including the full genome sequence (if available) of the relevant yeast. Only CRISPR sites with an on-target activity of higher than 70% were selected.

Table 5. The primers used in this study with their corresponding properties

Name	Sequence	TmºC	GC content	Length
Aa donor 1F	GACTCAAAAACGGTGGGAAT	56.2	45	**20**

Aa donor 1R	GACAATGGAAATTCACATGCTCGATCTCAAT	64.4	38.7	31
Aa donor 2F	GCATGTGAATTTCCATTGTCCAGATGCCTC	66.6	46.7	30
Aa donor 2R	GTTTTTATATTCTTCATATCCAACCTTC	55.6	28.6	28
Aa gRNA ADE2 1F	GTTAAGCATTTCCTTCTGAGAAAGTTCTGAT GAGTCCGTGAGGACGAAACGAGTAAGCT	N/A*	44.1	59
Aa gRNA ADE2 1R	CTATTTCTAGCTCTAAAAC**TCGCTCTCTACTA CAAAGTT**GACGAGCTTACTCGTTTCGT	N/A	40.7	59
Dh donor 1F	GACGGAAAAACTATAGGTATATTAGG	55.1	34.6	26
Dh donor 1R	TTTGCACAATAATATCACATTTTCTGGCTAAT TCC	63.6	31.4	35
Dh donor 2F	ATGTGATATTATTGTGCAAATGCCAAGGGG	64.6	40	30
Dh donor 2R	TTGTAGTTTTCGTAGCCCAC	55.7	45	20
Dh gRNA ADE2 1F	GTTAAGCATTTCCTTCTGAGTCTTGACTGAT GAGTCCGTGAGGACGAAACGAGTAAGCT	N/A	45.8	59
Dh gRNA ADE2 1R	CTATTTCTAGCTCTAAAAC**TCGTATGCTAAA GTTCTTGA**GACGAGCTTACTCGTTTCGT	N/A	39	59
Kl donor 1F	GATCAAAGAACTGTCGGTATTTTA	55.2	33.3	24
Kl donor 1R	GAATGTAGAGTTTCAACAGTTAAGACATCAC AAA	62.8	32.4	34
Kl donor 2F	ACTGTTGAAACTCTACATTCCATTGTACAAAT GC	64.7	35.3	34
Kl donor 2R	ATTCTTGATATCCGATAGTTTCTAATT	54.9	35.9	27
Kl gRNA ADE2 1F	GTTAAGCATTTCCTTCTGAGATGGGCCTGAT GAGTCCGTGAGGACGAAACGAGTAAGCT	N/A	49.2	59
Kl gRNA ADE2 1R	CTATTTCTAGCTCTAAAAC**TTGTATGCCGAG AAATGGGC**GACGAGCTTACTCGTTTCGT	N/A	44.1	59
Op donor 1F	AAGGTCGTTGGAATTTTGGG	56.5	45	20
Op donor 1R	ATCGAATACAAACATCAACGTGCTCAATCT	63.9	36.7	30

Op donor 2F	CGTTGATGTTTGTATTCGATAGTTCAGATGCC	64.7	40.6	32
Op donor 2R	TTTATTAAGGTATTCTTCATAGCCAA	54.8	26.9	26
Op gRNA ADE2 1F	GTTAAGCATTTCCTTCTGAGTAGATGCTGATGAGTCCGTGAGGACGAAACGAGTAAGCT	N/A	45.8	59
Op gRNA ADE2 1R	CTATTTCTAGCTCTAAAAC**TCGTCTAGAGTACATAGATG**GACGAGCTTACTCGTTTCGT	N/A	40.7	59
Pp donor 1F	GATTCTCAGGTAATAGGTATTCTAGG	56	38.5	26
Pp donor 1R	GTTGTTGATATCAATCTCTACAGTGAGAACATC	62.3	36.4	33
Pp donor 2F	TAGAGATTGATATCAACAACGCCACCAATG	63.8	40	30
Pp donor 2R	AATTCAAAGACGATTCTTCAAATAGG	56.6	30.8	26
Pp gRNA ADE2 1F	GTTAAGCATTTCCTTCTGAGCCGTACCTGATGAGTCCGTGAGGACGAAACGAGTAAGCT	N/A	49.2	59
Pp gRNA ADE2 1R	CTATTTCTAGCTCTAAAAC**ATCGGGTGGTCTTTCCGTAC**GACGAGCTTACTCGTTTCGT	N/A	45.8	59
Sc donor 1F	GATTCTAGAACAGTTGGTATATTAGG	55	34.6	26
Sc donor 1R	GACAAGAACCGTTAGCACATCACATTTTTCAG	65.3	40.6	32
Sc donor 2F	ATGTGCTAACGGTTCTTGTCTAGATGGAGTAG	65.7	43.8	32
Sc donor 2R	GTTTTCTAGATAAGCTTCGTAACC	55.7	37.5	24
Sc gRNA ADE2 1F	GTTAAGCATTTCCTTCTGAGACAATTCTGATGAGTCCGTGAGGACGAAACGAGTAAGCT	N/A	44.1	59
Sc gRNA ADE2 1R	CTATTTCTAGCTCTAAAAC**TGTGGATAGTCTCTACAATT**GACGAGCTTACTCGTTTCGT	N/A	39	59
yl donor 1F	GACTCGAAAACAATTGGTATTCT	55.3	34.8	23
yl donor 1R	GTTGTTGACAAATTTCAACGGTTATGACGTC	64.2	38.7	31
yl donor 2F	CGTTGAAATTTGTCAACAACTCCACCAATG	64.3	40	30

yl donor 2R	TTGAGGTACTGCTCGTATCC	56.5	50	20
Yl gRNA ADE2 1F	GTTAAGCATTTCCTTCTGAGCCATGCCTGAT GAGTCCGTGAGGACGAAACGAGTAAGCT	N/A	49.2	59
Yl gRNA ADE2 1R	CTATTTCTAGCTCTAAAAC**TCGCTCGCATTC GTCCATGC**GACGAGCTTACTCGTTTCGT	N/A	47.5	59
pKM 180 (linearize)- F	AATGACGTCAGAATTCTCGAGATCCC	62.6	46.2	26
pKM 180 (linearize)- R	ATTATCAGTCGTACTGTTATTAAGTGCTGT	61	33.3	30
pKM 180 (gRNA)- F	AGTACGACTGATAATCCAGAGACCGGGTTG GCG	N/A	54.5	33
pKM 180 (gRNA)- R	AATTCTGACGTCATTGAATTCTGACGTCATTA TTATCAGTGTG	N/A	34.9	43

*Length of primer do not permit accurate Tm determination

3.6. Agarose gel electrophoresis

Refer to chapter 2 section 3.4

3.7. DNA extractions

For DNA gel extraction see chapter 2 section 3.5 and for DNA plasmid extraction chapter 2 section 3.7

3.8. Construction of the CRISPR-Cas9 vector

The *TEF1* gene was amplified from the *Yarrowia lipolytica* Po1f genome with primers ylTEF1-2F and - 2R using the Kapa PCR kit reaction protocol (KAPA Biosystems Inc. 2014). The PCR fragment was gel purified and inserted into the pMiniT 2.0 vector via the NEB ligation protocol to yield the pMini + TEF vector (Thermo Fisher Scientific Inc 2015b; New England BioLabs Inc. 2017). The ligated product (5 µl) were then transformed into NEB 10-Beta cells (Chung *et al.* 1989), plated on LB+ [100 µg/ml]

ampicillin and grew over night. The transformed colonies were then screened with *BgI*I for confirmation of successful ligation.

The pRS316-RGR-GFP vector containing the Ribozyme-gRNA-Ribozyme (RGR-gRNA) was obtained from Addgene (Gao & Zhao 2014; Addgene 2015). The RGR-gRNA construct was isolated by amplifying the fragment with the primer pair hammer 1F:ylTEF1 overlap and hammer 1R:ylTEF1 overlap (KAPA Biosystems Inc. 2014). The pMini + TEF1 vector was also linearized through amplification to remove the ORF of the *TEF1* gene with primers ylTEF1-3F and - 3R. Both the gRNA and vector fragments were gel purified and ligated using NEBuilder to yield the pMini+TEF1-RGR-gRNA vector (Thermo Fisher Scientific Inc 2015b; New England Biolabs 2017). NEB product (5 µl) was then transformed and plated out on LB+ [100 µg/ml] ampicillin plates and allowed to grow O.N. at 37˚C (Chung *et al.* 1989). The colonies obtained from the transformation were screened with *BgI*I, for confirmation of successful ligation (Thermo Fisher Scientific Inc 2015a).

For the pKM::PCAS9+CRISPR vector, the TEF1-RGR-gRNA construct was cleaved from the pMini+TEF1-RGR-gRNA vector with *Xho*I and *Sal*I and each of the CAS9 vectors were cleaved with *Xho*I (Thermo Fisher Scientific Inc 2015a). The fragments were gel extracted, ligated, transformed and screened for positive transformants with *BgI*II (Chung *et al.* 1989; Thermo Fisher Scientific Inc 2015a, b; New England BioLabs Inc. 2017). Alternatively, for the pKM::HCAS9 and the pKM::CCAS9, the vectors were linearized with primers (PKM180 (linearize)- F and -R) (Novagen 2011). The gRNA construct was also amplified out of the pMini+TEF1-RGR-gRNA vector with primers (pKM 180 (gRNA)-F and -R) (KAPA Biosystems Inc. 2014). The fragments (vectors as well as the gRNA construct) were ligated using NEBuilder, 5 µl of the product was then transformed into NEB 10-Beta competent cells and screening for successful transformants were done with the digestion using *BgI*II (Chung *et al.* 1989; Thermo Fisher Scientific Inc 2015a; New England Biolabs 2017).

The CCAS9 was isolated from the pKM::CCAS9 with *BgI*I and *Avr*II (New England Biolabs Inc. 2014; Thermo Fisher Scientific Inc 2015a). The PCAS9 were digested out of the pKM::PCAS9+CRISPR with *Acc*65I and *Hin*dIII. The fragments were then gel extracted and connected using NEBuilder. Five microliters of the product were transformed into NEB 10-Beta cells and the transformed colonies were then screened

with *Pst*I to find successful colonies (Chung *et al.* 1989; Thermo Fisher Scientific Inc 2015a, b; New England Biolabs 2017).

After the pKM::CCAS9+CRISPR was obtained, the *ADE2* target gene of each yeast were inserted into the pKM::CCAS9+CRISPR vector (Chung *et al.* 1989; KAPA Biosystems Inc. 2014; Thermo Fisher Scientific Inc 2015b; New England Biolabs 2017). For this purpose two 50 bp oligonucleotides were designed to insert the CRISPR site for each of the seven yeast species. The oligonucleotides were added together and elongated in a PCR reaction to obtain a 100 bp double stranded fragment. The pKM::CCAS9+CRISPR vector were linearized with primers ylTEF1-3R and AHO1099. The fragments were then purified through gel extraction, ligated using NEBuilder and transformed (5 µl of the NEBuilder product). This was done for each yeasts' *ADE2* target site, respectively, to yield pKM::CCAS9+ADE2.

3.9. Construction of the donor DNA and validation of the CRISPR-Cas9 vector.

The donor DNA was obtained by amplifying two separate regions of the *ADE2* gene from each yeast. These two regions encompass approximately 200 bp on the 5'-side and the 3'-side of the *ADE2* ORF. For fragment 1: a PCR was performed using each yeast genomic DNA with the respective donor 1f and donor 1r primers listed in Table 8. For fragment 2, a PCR was performed using genomic DNA from each yeast with the respective donor 2f and donor 1r primers listed in Table 5. In every case the reverse primer of fragment 1 had an overlap with the forward primer of fragment 2. The final donor fragment was then obtained when a PCR was performed using fragment 1 and 2 with the respective donor 1f and donor 2r primers from Table 5.

After obtaining the plasmids and donor DNA, transformation were performed on each yeast, with and without donor DNA (Klebe *et al.* 1983; Chen *et al.* 1992). After transformation single colonies were picked (from all the transformed yeast) and streaked out on YNB dropout plates (Table 6), lacking adenine, to confirm *ADE2* disruption mutants. Colony PCR were then performed on the successful mutants from the donor DNA transformed yeast to confirm successful integration of the donor fragment.

Table 6: The media used to test for *ADE2* mutants

Media or buffer	Composition
Yeast nitrogen base (YNB) dropout plates, lacking adenine	0.67% YNB; 2% glucose; 0,005% Histidine; 0,005% Uracil; 0,005% Tryptophan; 0,04% Leucine; 0,06% CSM;1.5% Agar

3.10. Ligation and cloning reactions

3.10.1. NEBuilder ligation and cloning (New England Biolabs 2017)

DNA vectors and fragments were prepared with the desired restriction enzymes or it was obtained through PCR amplification with the desired primers. Ligation reaction was then performed with the preparation of the reaction mixture as depicted in Table 7 and incubated for 60 minutes at 50°C (New England Biolabs 2017). This was followed by the transformation of 5 µl of the reaction product into competent *E. coli* NEB 10 beta cells. The transformed cells were plated on LB (Table 1 from chapter 2) plates containing [100 µg/ml] ampicillin and screened for positive transformants.

Table 7: NEBuilder reaction mixture (New England Biolabs 2017)

Component	Reaction mixture
Amount of fragments (vector: insert)	1: 2-5 pmols
NEBuilder HiFi DNA Assembly Master Mix	10 µl
Deionized H_2O	Up to 20 µl

3.10.2. T4 ligation and cloning (Devi *et al.* 2009)

Refer to chapter 2 section 3.6

Chapter 3

3.10.3. pMiniT 2.0 Ligation and cloning (New England BioLabs Inc. 2017)

For the purpose of successful cloning and ligation, the insert fragments were amplified through PCR using the respective primers. The ligation reaction was then performed with the preparation of the reaction mixture according to Table 8 and incubated at 25˚C for 15 minutes. The mixture was then incubated on ice for 2 minutes, transformed into NEB 10-Beta competent cells and plated on LB plates containing [100 µg/ml] ampicillin.

Table 8: pMiniT 2.0 Ligation protocol (New England BioLabs Inc. 2017)

Component	Reaction mixture
Linearized pMiniT 2.0 Vector (25 ng/µl)	1 µl (25 ng)
Insert	3:1 molar ratio over vector
H_2O	to 5 µl
Cloning Mix 1	4 µl
Cloning Mix 2	1 µl

3.11. Yeast transformation

Ready yeasts were transformed using two different alkali-ion based protocols to maximize transformation efficiency, transformation using bicine and a one-step (lithium acetate) transformation protocol (Klebe *et al.* 1983; Chen *et al.* 1992). Different hygromycin concentrations were also used for each yeast (Table 9), to maximize transformation efficiency. Before transformation of the pKM::CCAS9+CRISPR plasmids were performed, each of the plasmid were linearized by *Not*I (NEB) digestion to remove the bacterial moiety and therefore exposing the rDNA regions to allow homologous recombination to take place (New England Biolabs Inc. 2014). For both transformation methods the yeasts were incubated for up to 7 days at 30 ˚C. For transformation protocols see chapter 2 section 3.9

Table 9: The different yeasts used in this study and the hygromycin concentration used for each yeast respective to the pKM::CCAS9+CRISPR vector used for transformation.

Ref. no.	Yeast	Hygromycin [mg.l^{-1}]
		pKM::CCAS9+CRISPR
1	*Saccharomyces cerevisiae* CENPK42	150
2	*Kluyveromyces lactis* UOFS-Y1167	300
3	*Yarrowia lipolytica* Po1f	600
4	*Debaryomyces hansenii* UOFS-Y0610	150
5	*Arxula adeninivorans* UOFS-Y1220	150
6	*Arxula adeninivorans* UOFS-Y1219	150
7	*Arxula adeninivorans* Ls3	150
8	*Ogataea polymorpha* UOFS Y1507	150
9	*Komagataella phaffii* UOFS-Y1552	300
10	*Komagataella phaffii* SMD 1168	300
11	*Komagataella phaffii* Km.71	300
12	*Komagataella phaffii* GS115	300

4. Results and discussion

4.1. Constructing the CRISPR-Cas9 vectors.

In the first part of this study (Chapter 2) it was observed from the western blot analysis that the strong constitutive *TEF1* promoter from *Yarrowia lipolytica* was sufficient for *CAS9* expression ion all selected yeasts. It was therefore decided to use the same promoter for the expression of the gRNA cassette. For this purpose the *TEF1* gene was amplified from the *Yarrowia lipolytica* genome with primers ylTEF1-2F and - 2R to yield a 2 325 bp fragment (Figure 1). The fragment was gel purified and inserted

into the pMiniT 2.0 cloning vector (New England Biolabs) via ligation to yield the pMini+TEF vector (4 871 bp). The transformed colonies were screened with *Bgl*I to yield fragment sizes of 3 642 bp and 1 235 bp if successful for ligation (Figure 1C).
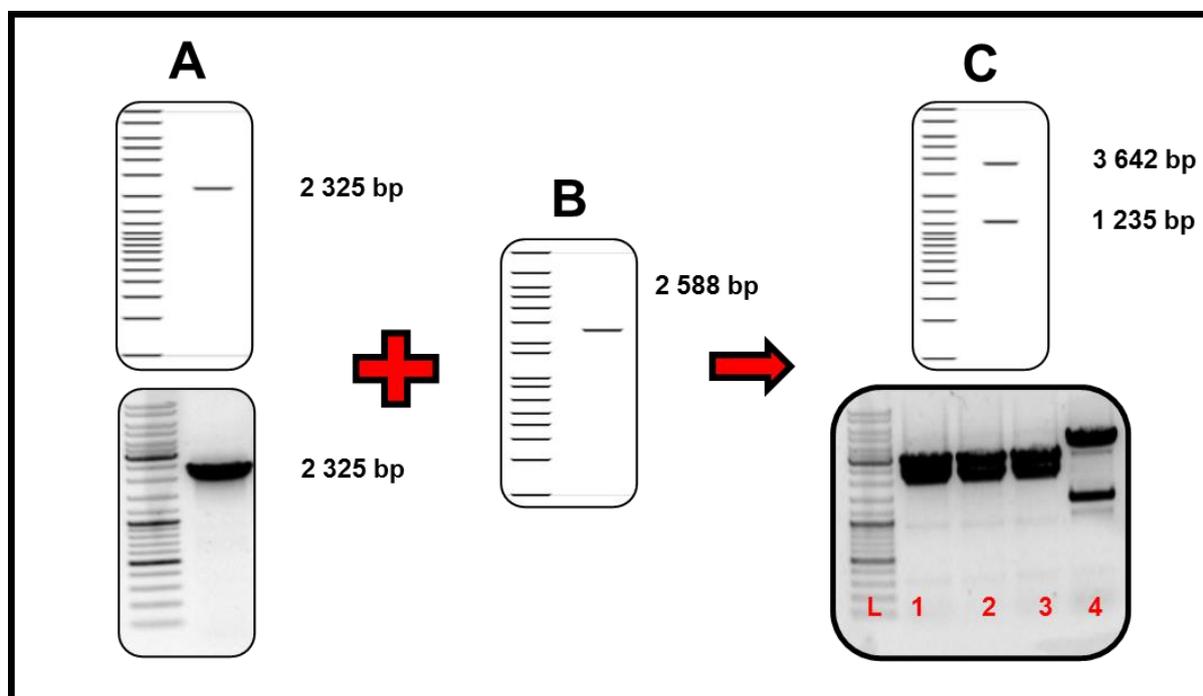

Figure 1: A) *Yarowia lipolytica TEF1* gene (2 325 bp), which was amplified from the yeast genome with primers ylTEF1-2F and - 2R. The *TEF1* gene, was ligated into the B) pMiniT 2.0 vector (2 588 bp). C) The transformed colonies were then screened with *Bgl*I. Successful ligation fragment sizes of 3 642 bp and 1 235 bp have been obtained. L represents the DNA ladder and 1 to 4 are the screened colonies with lane four indicating the correct fragments.

Plasmid pRS316-RGR-GFP (Gao & Zhao 2014; Addgene 2015) was obtained from Addgene (www.addgene.com). This plasmid contain the hammerhead (HH) ribozyme, gRNA scaffold and a HDV ribozyme as well as a CRISPR site specific for GFP (Figure 2A). The HH ribozyme promotes cleavage on its 3-prime side and the HDV ribozyme promotes cleavage on its 5-prime side. For the HH ribozyme to perform its function, the first 6 nucleotides must be reverse complementary to the first 6 bp of the CRISPR target sequence region (Figure 2B). Thus, for each of the different target sites it is necessary to modify the first 6 bp of the HH sequence, for the sole reason to maintain

ribozyme functionality. The reason for the cleavage is that the gRNA is an essential part of the CRISPR-Cas9 system, it must be highly specific without any post transcriptional modifications (i.e. a 5'-Cap and/or 3'-Poly A tail), otherwise the gRNA with the CRISPR site would not be able to bind to the Cas9 protein and therefore the interference complex won't form, thus abolishing its function. The post transcriptional modifications can be prevented by using a non mRNA producing polymerase promoter, which will ensure that no post transcriptional modifications take place, however, the use of these promoters have limitations. They are very specific for each organism and also only a limited number of all these promoters has been characterized in the selected yeasts. Importantly, these drawbacks prevent the use in a wide range system, with a universal promoter. For this reason, instead of using the more simpler approach by only using a Pol III promoter the more complex approach was used, where a self-splicing Ribozymes were used which promotes self-cleavage to remove any post transcriptional modifications on both ends.
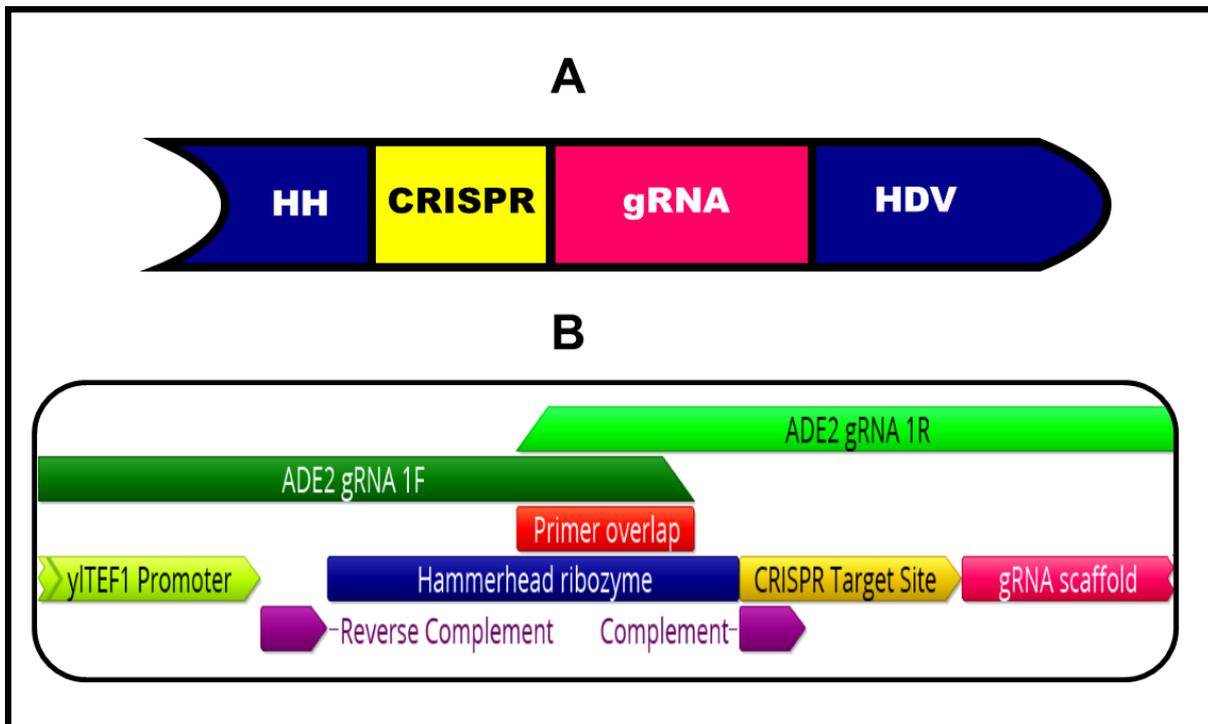

Figure 2: A) The generic RGR-gRNA, containing the CRISPR target site and the gRNA scaffold, which is flanked by self-splicing ribozymes (HH and HDV). B) The *ADE2* CRISPR sites for each yeast were created using two overlapping

oligonucleotide (gRNA 1F and 1R). The forward oligonucleotide contains the first 6 bp (that needs to be the reverse complementary of the first 6 bp of the *ADE2* CRISPR target site for each yeast). The reverse oligonucleotide contained the specific *ADE2* CRISPR site.

The RGR gRNA was removed from the pRS316-RGR-GFP plasmid by amplifying the fragment with primer pair hammer 1F:ylTEF1 overlap and hammer 1R:ylTEF1 overlap (Figure 3). This yielded a fragment size of 212 bp (RGR-gRNA). The primers were designed to create a 15 bp overlap that corresponds with the *TEF1* promotor's 3-prime end and the *TEF1* terminators 5-prime end. This overlap is necessary to perform the NEBuilder ligation reaction. The pMini+TEF vector was linearized through PCR amplification to remove the ORF of the *TEF1* gene but leaving the promoter and terminator regions intact. The vector was amplified with primers ylTEF1-3F and -3R to yield a 3 353 bp fragment. Both the gRNA and vector fragments were gel purified and ligated with NEBuilder to yield the pMini+TEF1-RGR-gRNA vector. The colonies obtained from the transformation were screened with *BgI*I, where 12 out of the 15 colonies were positive when restriction analysis displayed band sizes of 2 296 bp and 1253 bp. Some of the lanes do display three bands, this is a result of incomplete digestion, meaning the top band represent undigested plasmid (3 543 bp).
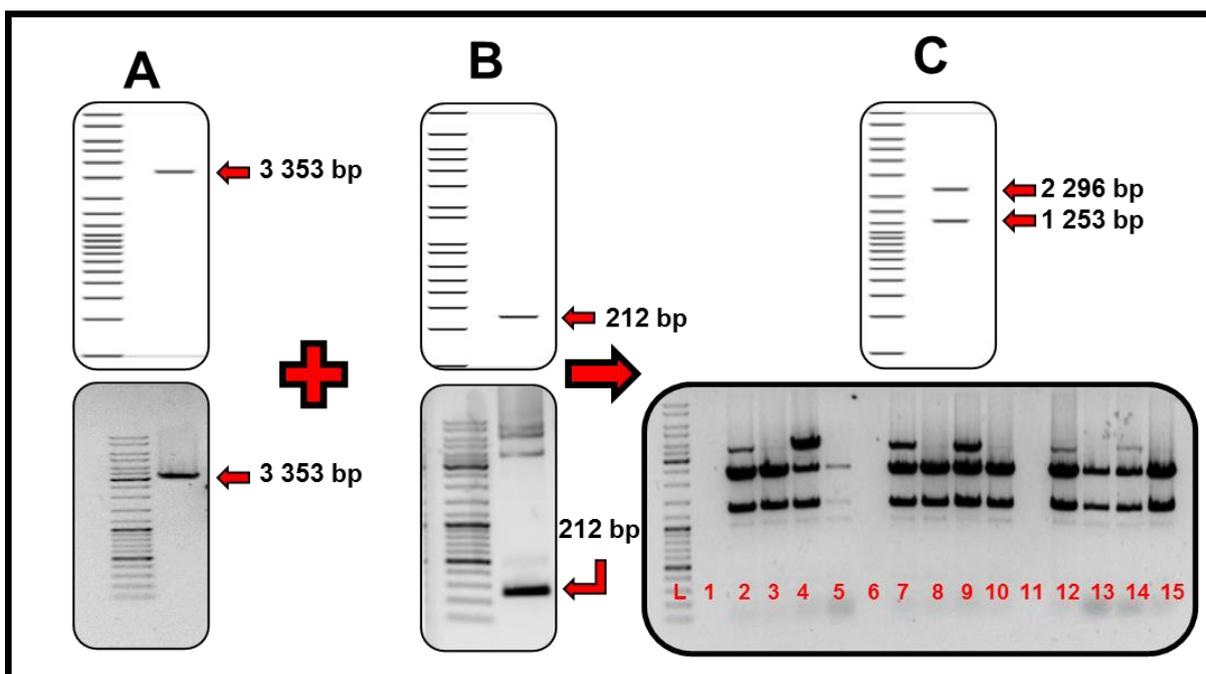
Figure 3: A) pMini + TEF1, where the *TEF1* ORF frame was removed to obtain a 3 353 bp fragment. B) The RGR gRNA was amplified from plasmid pRS316-RGR-GFP to obtain a 212 bp fragment. C) The two fragments were ligated and screening for positive transformants were done with the *Bgl*I enzyme to obtain sizes of 2 296 bp and 1 253 bp if successful. L represents the ladder and 1 to 15 the screened samples.

The next step comprised the construction of the complete pKM::CAS9+CRISPR vector containing either the PCAS9, CCAS9 or the HCAS9 gene (designated pKM::PCAS9+CRISPR, pKM::CCAS9+CRISPR or pKM::HCAS9+CRISPR, respectively). This was done by inserting the TEF1-RGR-gRNA into each of the pKM180::*CAS9* vectors resulting in either pKM::PCAS9+CRISPR, pKM::CCAS9+CRISPR or pKM::HCAS9+CRISPR. Initially the *CAS9* vectors was linearize by digesting it with *Xho*I, resulting in a ~11 700 bp linear fragment. The TEF1-RGR-gRNA were obtained by digesting the pMini + TEF1-RGR-gRNA vector with *Xho*I and *Sal*I to yield a 1 036 bp fragment. The different restriction enzymes used to obtain the TEF1-RGR-gRNA fragment would not pose any problem in the ligation reaction, because they share compatible ends to one another. The TEF1-RGR-gRNA fragment were then ligated into each of the different optimized linear *CAS9* fragments to yield the CRISPR-Cas9 vector (~12 700 bp). However when the transformed colonies were screened with *Bgl*II, only the pKM::PCAS9+CRISPR were obtained (Figure 4). For successful ligation there needed to be three bands, due to the TEF1-RGR-gRNA containing one of the *Bgl*II cleaving sites. For pKM::PCAS9+CRISPR the band sizes needed to be 5 950 bp, 4 219 bp and 2 508 bp for successful ligation and this profile was obtained for one of the colonies tested (Figure 4C, Lane 1).
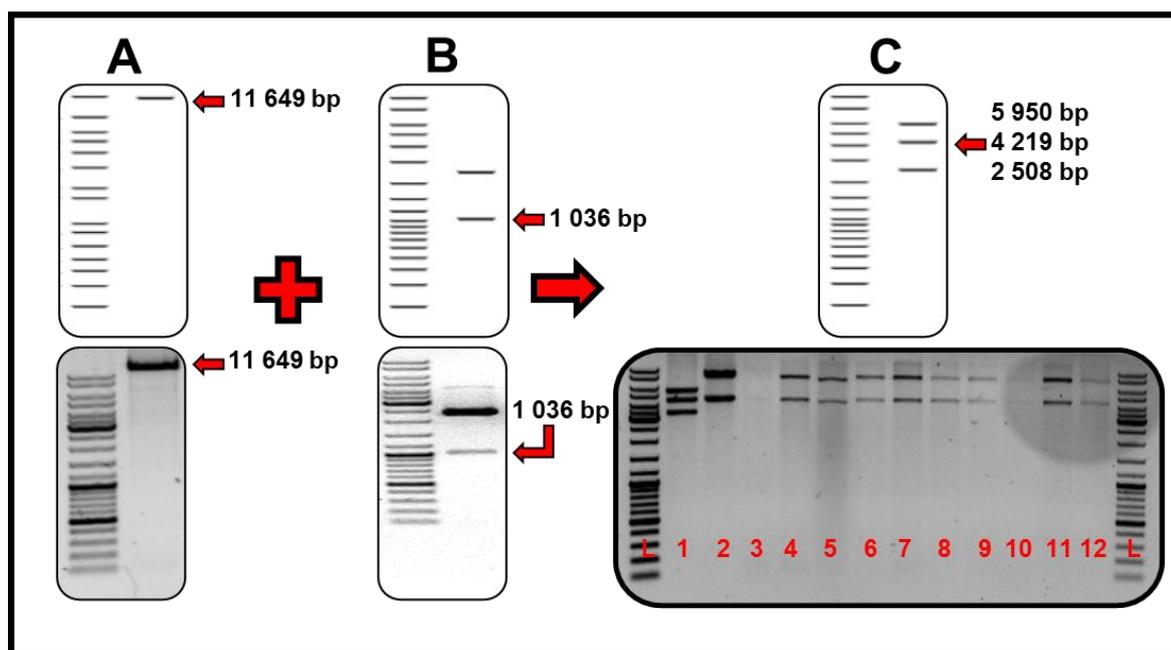
Figure 4: A) pKM::PCAS9 vector linearized with *Xho*I (11 649 bp). B) TEF1-RGR-gRNA fragment removed from pMini with *Xho*I and *Sal*I (1 036 bp). C) The fragments were ligated and screened for positive transfomants with *Bgl*II, which should yield fragment sizes of 5 950 bp, 4 219 bp and 2 508 bp if successful. Only one clone, Lane 1, was obtained with the correct restriction profile. L represents the DNA ladder and 1 to 12 the screened samples.

To obtain the pKM::CCAS9+CRISPR or pKM::HCAS9+CRISPR an alternative approach was followed (Figure 5). Linearization of the pKM::HCAS9 (11 712 bp) and the pKM::CCAS9 (11 736), linearizing of the vectors were performed with primers (pKM180 (linearize) - F and -R) and amplifying the TEF1-RGR-gRNA construct out of the pMini with primers that has 15nt extensions (pKM 180 (gRNA)-F and -R) to yield a 1 062 bp fragment. This gave the TEF1-RGR-gRNA construct overlapping end which corresponded with the ends of the linearized pKM::CAS9 vector. This is a requirement so that NEBuilder can be performed. NEBuilder reactions were then performed on the fragments, however no positive colonies were obtained after screening with *Bgl*II. The reason for the unsuccessful ligation is likely due to the NEBuilder reaction that is performed at 50˚C and a crucial requirement is that the overlapping regions must have a Tm of at least 48˚C. In this case the Tm of the overlapping regions of the gRNA were 40˚C, therefore the overlaps of the gRNA did possible not hybridize to the ends of the vectors.
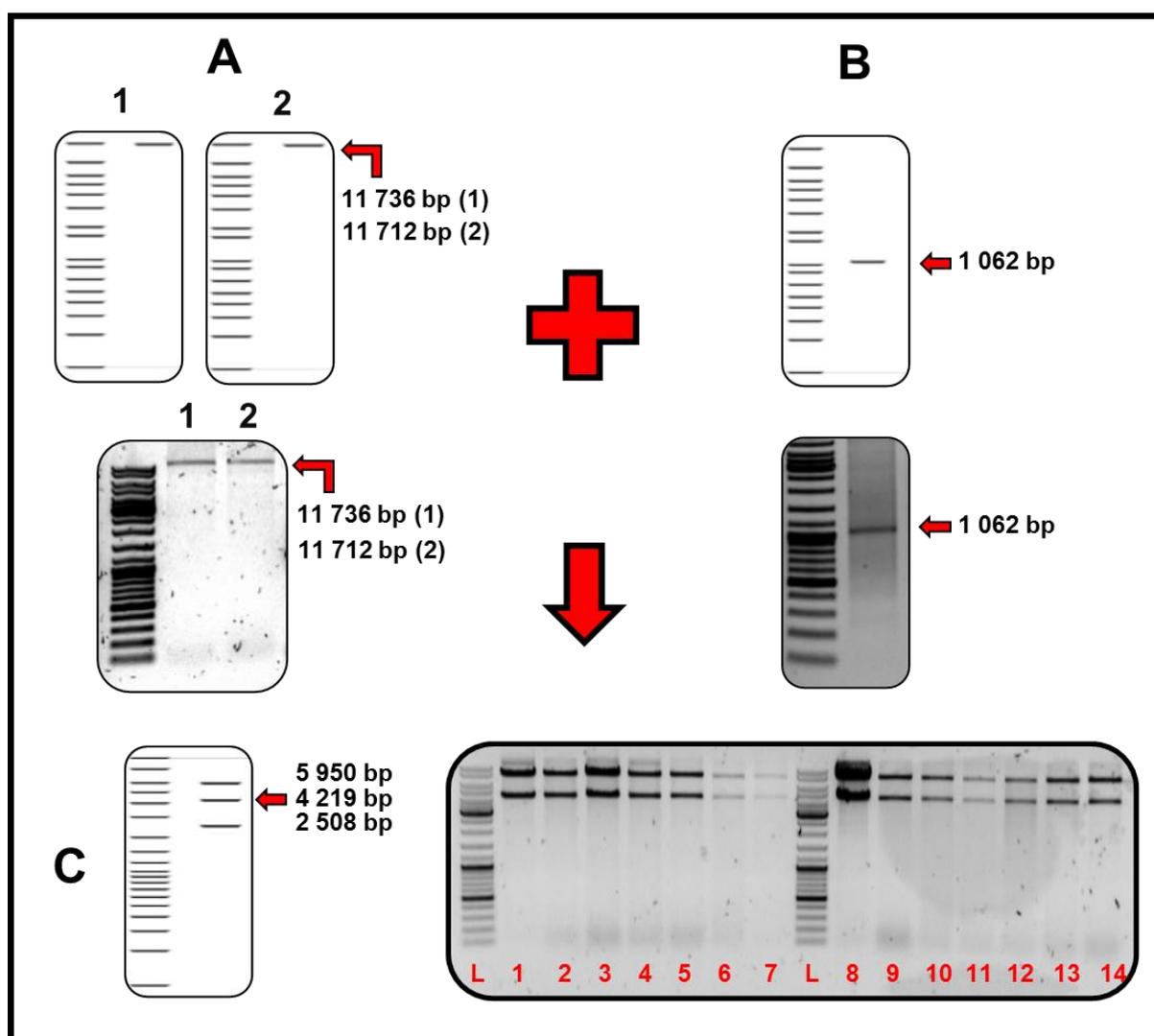
Figure 5: A) pKM::CCAS9 (1 = 11 736 bp) and pKM::HCAS9 (2 = 11 712 bp) vectors which were linearized through amplifying with primers pKM180 (linearize) - F and –R. B) The TEF1-RGR-gRNA (1 062 bp) was amplified with primers pKM 180 (gRNA)-F and –R form the pMini. C) The fragments were ligated and the screened for positive transformants with *Bgl*II, that should yield product sizes of 5 950 bp, 4 219 bp and 2 508 bp if successful. L represents the ladder, 1 to 7 the pKM::CCAS9+CRISPR screening and 8 to 14 the pKM::HCAS9+CRISPR screening.

At this stage of the study it was decided to only use the pKM::CCAS9 for completion of the wide range CRISPR-Cas9 system. The main reason was due to the fact it was observed that the pKM::HCAS9 had a negative effect on the yeasts in term of both growth as well as hygromycin resistance in comparison to the other two Cas9 proteins.

In addition it was also observed that the codon optimization as reported by (Weninger *et al.* 2016) of the PCAS was not optimal. This was evident when the amino acid sequences of the three Cas9 proteins were aligned with the native *CAS9* (Chapter 2; Figure 2). This can be crucial seeing that the different amino acids can abolish or change the function of the Cas9 protein.

At this stage only the pKM::PCAS9+CRISPR plasmid were successfully constructed. For completion of the pKM::CCAS9+CRISPR the PCAS9 was replaced (swapped) with the CCAS9 gene (Figure 6).

The C*CAS9* was isolated from the pKM::CCAS9 with *Bgl*II and *Avr*II, which resulted in a 4 203 bp fragment containing the *CAS9* gene. The PCAS9 were removed by restriction digestion from the pKM::PCAS9+CRISPR with *Acc*65I and *Hin*dIII to yield an 8 160 bp fragment. This resulted in an overlap between the *CAS9* fragment (CCAS9) and the CRISPR-Cas9 backbone containing the gRNA construct. The fragments were then gel extracted and fused using NEBuilder. In addition, this time the Tm of the overlapping regions were well above 50˚C. The transformed colonies were then screened with *Pst*I to yield fragment sizes of 5 136 bp, 4 226 bp, 3 016 bp and 406 bp if successful (Figure 6C, Lane 1 to 6 and 8).
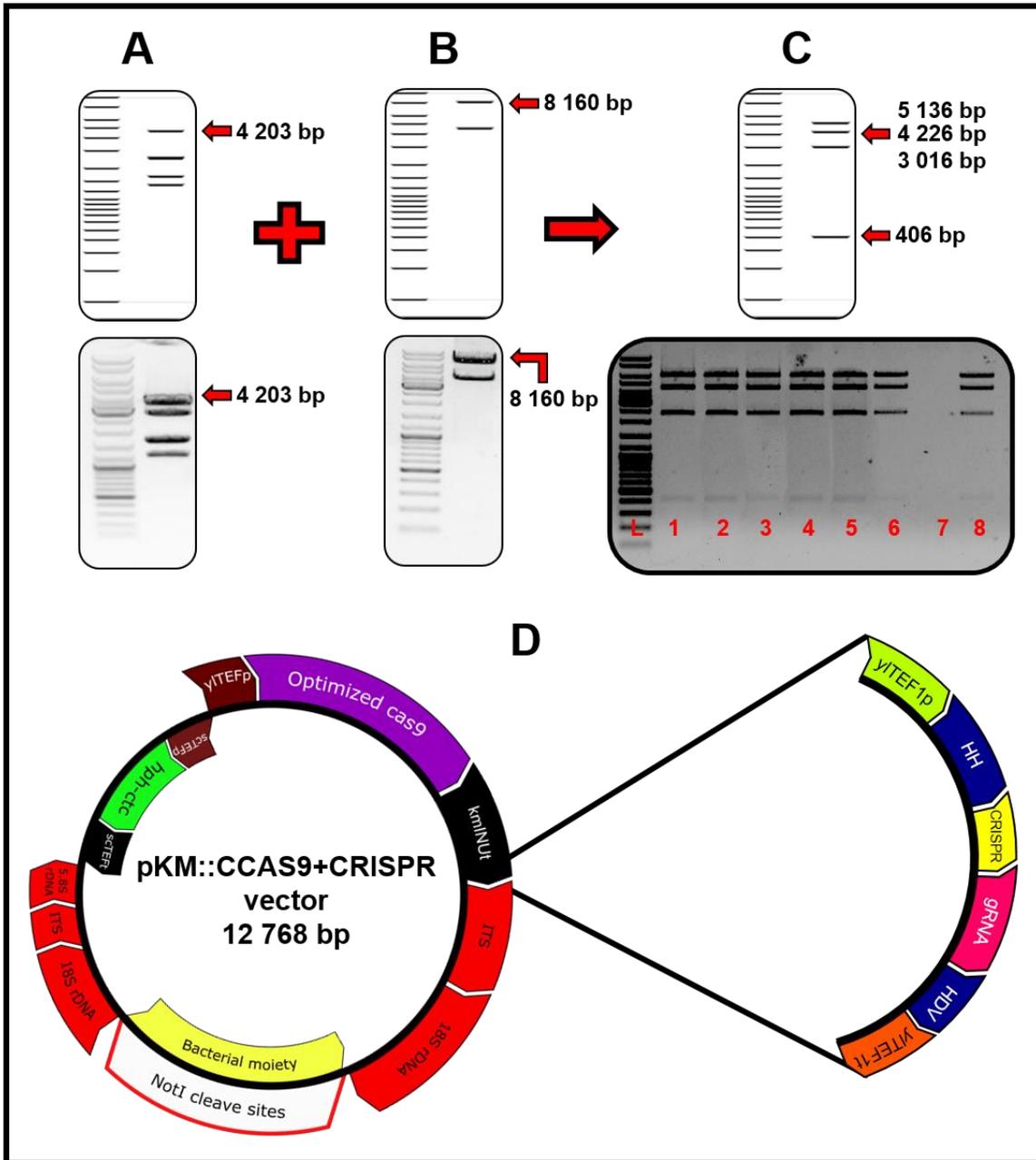
Figure 6: A) Isolation of the CCAS9 (4 203 bp). B) PCAS9 were removed from the pKM::PCAS9+CRISPR by restriction enzyme digestion to yield an 8 160 bp (backbone) fragment. C) The fragments were ligated and screened for positive clones with *Pst*I. Expected bans sizes for positive clones were 5 136 bp, 4 226 bp, 3 016 bp and 406 bp. L represents the ladder and 1 to 8 the pKM::CCAS9+CRISPR that were screened. All clones screened (apart for lane 7) indicated correct band sizes. D) Plasmid map depicting the complete wide range pKM::CCAS9+CRISPR plasmid .

Chapter 3

A plasmid map of the complete pKM::CCAS9+CRISPR expression vector is depicted in Figure 6D. As described in Chapter 2 (section 3.8) the pKM180 contains a region of the rRNA locus from *Kluyveromyces marxianus* for integration into the yeast's host genome. Selection is facilitated through Hygromycin by the presence of the *HPH* gene which is regulated and terminated by the *Saccharomyces cerevisiae TEF1* promoter and -terminator. The cloning region contains the *Yarrowia lipolytica* transcription elongation factor (*TEF1*) gene. This promoter will drive expression of any gene that is inserted in the pET28 derived multiple cloning site (MCS). For this study the *CAS9* gene was inserted into this multiple cloning site (Chapter 2). Expression is terminated by the *K. marxianus* inulase terminator (*Km*INUt) sequence. For the complete CRISPR system, the hammerhead (HH) ribozyme, CRISPR site, gRNA scaffold and a HDV ribozyme was placed under the regulation and termination of the *Y. lipolytica* TEF1 promoter. This region was inserted in the pKM180::CAS9 plasmid between the *Km*INUt and the 1rRNA region as indicated in Figure 6D.

For the final part of the study the wide range CRISPR-Cas9 system was validated by targeting disruption of the *ADE2* gene in each of the selected yeast strains. When the *ADE2* gene is disrupted, the product P-ribosylaminoimidazole accumulates, due to the disruption in the adenine pathway (Figure 7) (Ugolini & Bruschi 1996). This leads to a phenotypic change in the colony's color changing from creamy to red. P-ribosylaminoimidazole itself is not red, the color emerges when this intermediate product is oxidized by the cells resulting in the accumulation of a red pigment.
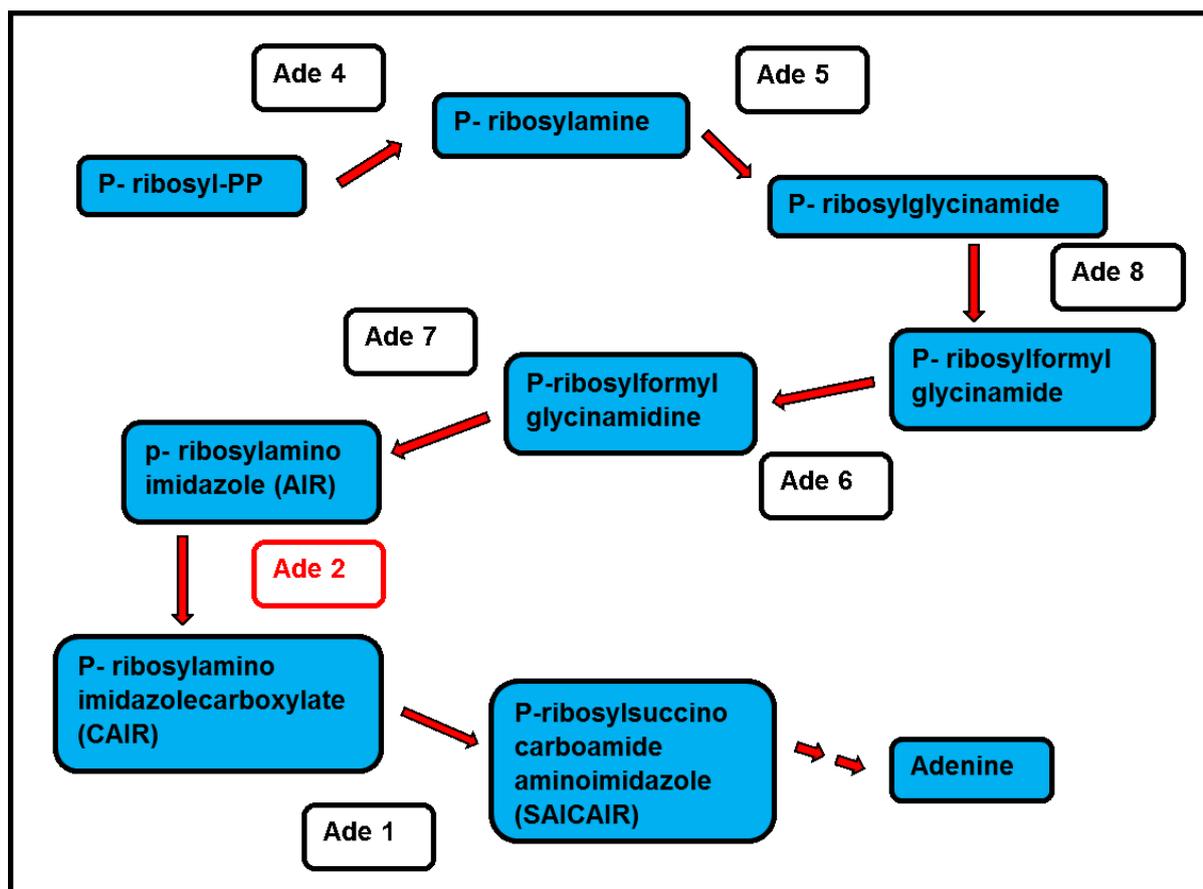
Figure 7: Displayed here is the adenine biosynthesis pathway (Ugolini & Bruschi 1996). The red Ade2 enzyme indicates where the pathway is targeted, meaning the *ADE2* gene is going to be disrupted. This will allow for the accumulation of P-ribosylaminoimidazole, which forms a red pigment when oxidised by the cells.

To insert the *ADE2* CRISPR site for each of the seven yeast species, two 50 bp oligonucleotides were designed (see Figure 2B). The forward oligonucleotide has a 20 bp overlap with the *Yarrowia lipolytica TEF1* followed by a part of the HH ribozyme sequence. In between these two region are an additional 6 bp that is the reverse complement of the first six base pairs of the *ADE2* CRISPR site. The reverse oligonucleotide has a 20 bp overlap with the gRNA scaffold part, the *ADE2* CRISPR site and 3'-side of the HH ribozyme sequence. The forward oligonucleotide and the reverse oligonucleotide contain a 16 bp overlap with each other and when combined in a PCR reaction will result in a 100bp double strand DNA fragment. The pKM::CCAS9+CRISPR vector were linearized with primers ylTEF1-3R and AHO1099. The oligonucleotides have an overlap of approximately 20 bp with the ends of the linear plasmid allowing the use of NEBuilder to connect the two fragments. For each

yeast a specific target site for disruption of the *ADE2* was selected, therefore the final CRISPR-Cas9 construct were made for each of the selected yeast species designated pKM::CCAS9+*ADE2*.

4.2. Transformation

To validate the developed CRISPR-Cas9 gene editing system the 12 selected yeasts were transformed with the specific pKM::CCAS9+*ADE2* containing the corresponding CRISPR site for each of the specific yeast strains. For every yeast that was transformed donor DNA was either included or omitted. If the CRISPR-Cas9 system is effective, the Cas9 protein will cleave the genomic DNA in the genome of the specific yeast. These double-stranded breaks (DSB) must be repaired in the genome as it is lethal to the cell and this can be done via non-homologous end joining (NHEJ) or homology-directed repair (HDR). NHEJ are error prone and happens spontaneously when a DSBs occurs in the genome of the organism and can therefore lead to a disrupted open reading frame. HDR on the other hand can be used to insert modifications (DNA region of heterologous DNA, i.e. donor DNA) at the target locus thus modifying the genome of the organism. Thus, if the specific yeast has a higher affinity to correct a double strand break through HDR, it will replace the full ORF of the *ADE2* gene with the donor DNA region (thus resulting in a complete deletion). This can easily be confirmed through PCR analysis. Alternatively, if repair is via NHEJ, the *ADE2* ORF will only be disrupted as a result of indels and not deleted, however this cannot be confirmed via PCR. Fortunately, the red phenotype will indicate a non-functional *ADE2* gene whether the *ADE2* gene was deleted or only disrupted.

Before transformation each of the plasmids were linearized by *Not*I digestion to remove the bacterial moiety and therefore exposing the rDNA regions to allow homologous recombination to take place. For *Saccharomyces cerevisiae CENPK42, Kluyveromyces lactis UOFS-Y1167, Yarrowia lipolytica Po1f, Debaryomyces hansenii UOFS-Y0610, Arxula adeninivorans UOFS-Y1219, Arxula adeninivorans Ls3* and *Ogataea polymorpha UOFS Y1507* the one-step transformation method was used to maximize transformation efficiency (Chen *et al.* 1992). For *Arxula adeninivorans UOFS-Y1220, Komagataella phaffii UOFS-Y1552, Komagataella phaffii SMD 1168, Komagataella phaffii Km.71 and Komagataella phaffii GS115* transformation using the bicine method was used to maximize transformation efficiency (Klebe *et al.* 1983). The

concentration of hygromycin was reduced for transformation for a number of yeast to accommodate for the reduction in growth due to the introduction of the *CAS9* vector as well as the gene disruption. This reduction in growth may have led to increased susceptibility to hygromycin for some yeast, even when the yeast strains possessed the necessary hygromycin resistance gene carried on the pKM::CCAS9+ADE2 vector.

The results of the transformation with or without donor DNA are listed in Table 10. In this table the number of colonies obtained are indicated. For each yeast, 20 colonies were randomly selected (except for *A. adeninivorans* UOFS-Y1220, where 10 were selected) and streaked on YNB dropout plates, lacking adenine to observe how many colonies turns red. This indicated if the CRISPR system were effective as well as the level of efficiency in terms of deletion or disruption of the *ADE2* gene. The number of colonies that turned red *versus* those that stayed white are also indicated in Table 17. As can be observed from Table 10, a moderate to high (i.e. 50% to 100%) efficiency was obtained for most of the yeasts tested.

Table 10: Results for the total colonies obtained after transformation (with or without donor DNA) and the results for the screening of *ade2* mutants when streaked out on the YNB dropout plates, lacking adenine.

Yeast	Transformed Colonies			
	With donor DNA		Without donor DNA	
	Total	Red/White	Total	Red/White
Saccharomyces cerevisiae CENPK42	27	15/20	64	18/20
Kluyveromyces lactis UOFS-Y1167	26	14/20	23	12/20
Yarrowia lipolytica Po1f	23	20/20	43	20/20
Debaryomyces hansenii UOFS-Y0610	27	20/20	206	20/20
Arxula adeninivorans UOFS-Y1220	13	4/10	15	1/10
Arxula adeninivorans UOFS-Y1219	126	20/20	TMTC	20/20
Arxula adeninivorans Ls3	TMTC*	18/20	TMTC	18/20

Ogataea polymorpha UOFS Y1507	TMTC	10/20	TMTC	13/20
Komagataella phaffii UOFS-Y1552	64	4/20	154	6/20
Komagataella phaffii SMD 1168	55	10/20	165	11/20
Komagataella phaffii Km.71	56	14/20	104	11/20
Komagataella phaffii GS115	49	20/20	80	20/20

*Too many to count

Table 11 represents the HDR and NHEJ disruption efficiencies of the red colonies obtained, for each yeast, when transformation with donor DNA was performed.

Table 11: HDR and NHEJ disruption results for the red colonies obtained after transformation with donor DNA.

Yeast	Transformed Colonies With donor DNA	
	HR (%)	NHEJ (%)
Saccharomyces cerevisiae CENPK42	40	60
Kluyveromyces lactis UOFS-Y1167	36	64
Yarrowia lipolytica Po1f	50	50
Debaryomyces hansenii UOFS-Y0610	60	40
Arxula adeninivorans UOFS-Y1220	25	75
Arxula adeninivorans UOFS-Y1219	40	60
Arxula adeninivorans Ls3	50	50
Ogataea polymorpha UOFS Y1507	20	80
Komagataella phaffii UOFS-Y1552	0	100
Komagataella phaffii SMD 1168	0	100
Komagataella phaffii Km.71	0	100
Komagataella phaffii GS115	30	70

Colonies selected from the transformations where donor DNA was added and that turned red, were also further analysed via PCR analysis (see below results obtained for each yeast strain). This indicated if the *ADE2* ORF were replaced by the donor DNA region, i.e. repair of the Cas9 induced double strand break via HDR resulting in a complete deletion, or if the double stranded break was repaired via NHEJ.

4.3. Saccharomyces cerevisiae

In *Saccharomyces cerevisiae*, the addition of the donor DNA resulted in a reduction of transformation efficiency and of the 20 colonies randomly selected, 15 had a red phenotype, indicating a 75% disruption efficiency. Displayed in Figure.8 (B) are the different colonies streaked out to indicate the ratio of successful vs unsuccessful gene deleted mutants. The red colonies were then analysed through PCR to determine how many of the 15 colonies contained the donor fragment, meaning the gene were repaired through HR, and how many was repaired through NHEJ. A fragment size of 449 bp (the size of the donor fragment) will confirm the integration of the donor fragment. A size of 1 716 bp (size of the original gene) will indicate that the gene was disrupted through NHEJ. From Figure 8 (A) the total number of donor fragment repair colonies obtained were 6 out of 15 red colonies (green arrow and box), resulting in a HR-disruption efficiency of 40% when the yeast is transformed with a donor fragment. Thus, giving the NHEJ a repair efficiency of 60%, 9 out of the 15, which are indicated by the red arrow and box.
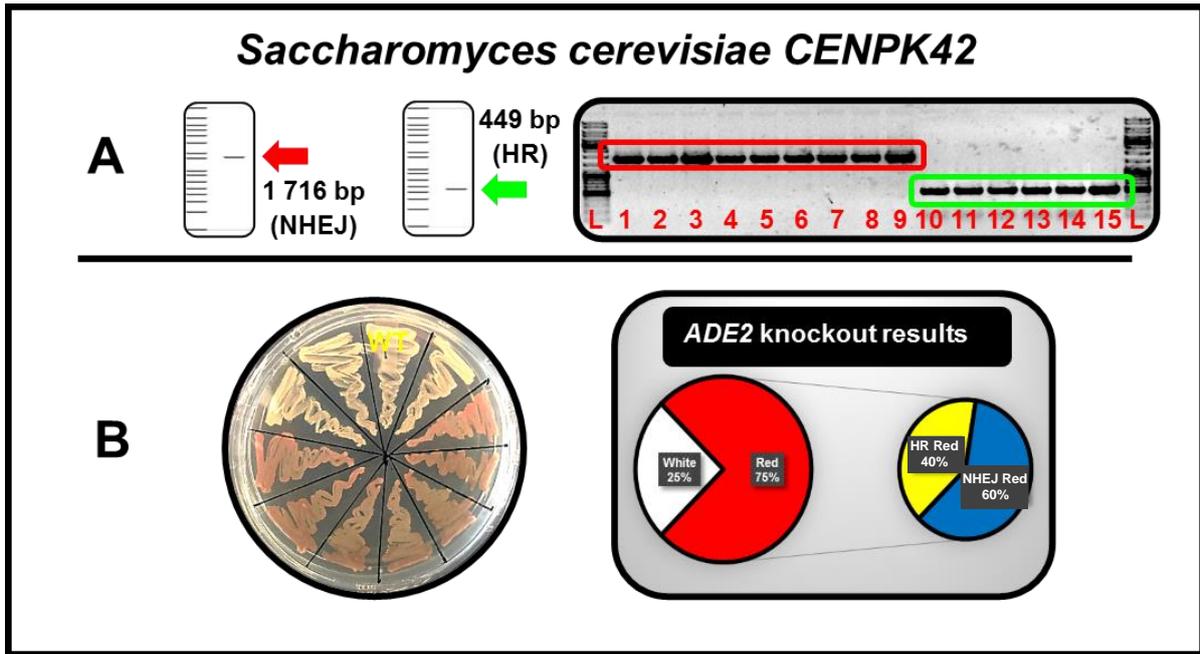
Figure 8: A) *ADE2* disruption results for *Saccharomyces cerevisiae CENPK42* when transformed with donor DNA. For the NHEJ repair mechanism, the PCR fragment size will be 1 716 bp (red arrow and box) and for the HR repair mechanism, the fragment size will be 449 bp (green arrow and box). B) On the plate are the wild type (WT) as well as the different colonies streaked out to display the ratio of successful colonies obtained vs unsuccessful. The pie chart also indicates the successful vs unsuccessful, as well as the HR- vs NHEJ repair mechanisms regarding the disruption of the *ADE2* gene.

4.4. *Kluyveromyces lactis*

In *Kluyveromyces lactis* the editing efficiency yielded was 70% when a donor fragment was provided (14/20), which are indicated by the pie chart as well as the plate where the ratio of successful vs unsuccessful *ADE2* mutants have been streaked out (Figure 9 (B)). From Figure 9 (A) it can be seen that from the 14 red colonies analysed, only 5 contained the donor fragment (421 bp = green arrow and box) and the other 9 red colonies was repaired through NHEJ (64%) due to their fragment sizes being the original gene size (1 710 bp = red arrow and box). This indicates that when a donor fragment was provided for editing, gene disruption via HR occurred in 36% of the red colonies.
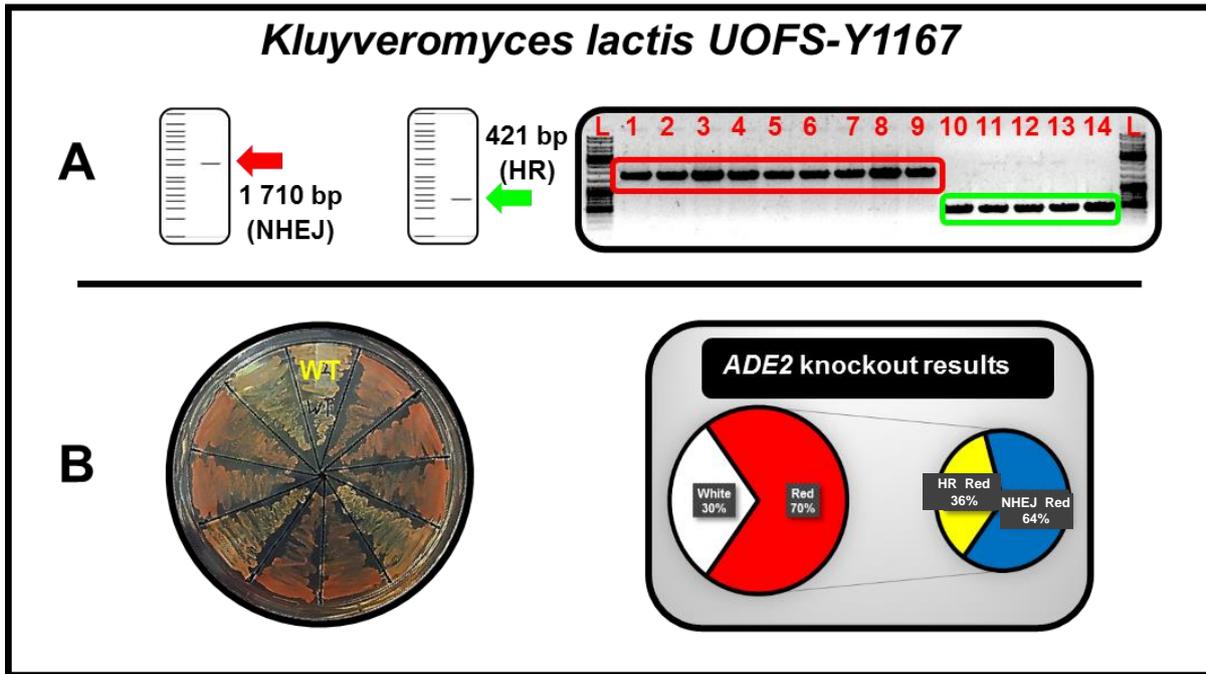
Figure 9: A) *ADE2* disruption results for *Kluyveromyces lactis UOFS-Y1167* when transformed with donor DNA. For the NHEJ repair mechanism, the PCR fragment size will be 1 710 bp (red arrow and box) and for the HR repair mechanism, the fragment size will be 449 bp (green arrow and box). B) On the plate are the wild type (WT) as well as the different colonies streaked out to display the ratio of successful vs unsuccessful colonies obtained. The pie chart also indicates the successful vs unsuccessful, as well as the HR- vs NHEJ repair mechanisms regarding the disruption of the *ADE2* gene.

4.5. *Yarrowia lipolytica*

In *Yarrowia lipolytica* an editing efficiency of 100% were obtained when a donor fragment was included with transformation. This is also indicated in Figure 10 (B), by the plated colonies in ratio of successful vs unsuccessful colonies. Figure 10(A) indicates the gene disruption through HR (green arrow and box) occurred in 50% of the colonies screened. The donor fragment size was 375 bp which were obtained in 9 out 18 of the red colonies screened. The other 50% were repaired via NHEJ (red arrow and box), which was observed by PCR analysis when the other 10 colonies retained the original gene size (1 698 bp).
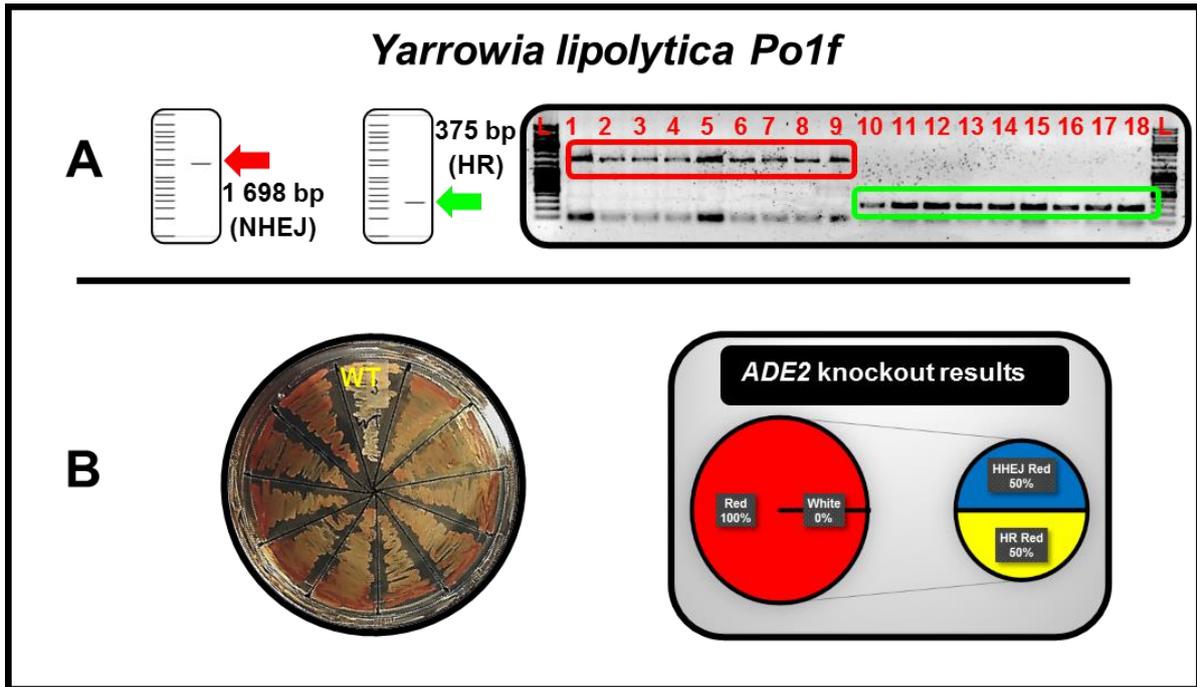
Figure 10: A) *ADE2* disruption results for *Yarrowia lipolytica Po1f* when transformed with donor DNA. For the NHEJ repair mechanism, the PCR fragment size will be 1 716 bp (red arrow and box) and for the HR repair mechanism, the fragment size will be 449 bp (green arrow and box). B) On the plate are the wild type (WT) as well as the different colonies streaked out to display the ratio of successful colonies obtained vs unsuccessful. The pie chart also indicates the successful vs unsuccessful, as well as the HR- vs NHEJ repair mechanisms regarding the disruption of the *ADE2* gene.

4.6. *Debaryomyces hansenii*

For *Debaryomyces hansenii* an editing efficiency of 100% were obtained when a donor fragment was included with transformation. From Figure 11(A) can be seen that the donor fragment has a size of 401 bp (green arrow and box), this fragment size was obtained in 6 out of 15 of the red colonies when screened with PCR analysis. The original gene size is 1 686 bp (red arrow and box) and it was obtained in the other 9 colonies screened. In Figure 11 (B) are the different colonies streaked out to indicate the ratio of successful vs unsuccessful gene deleted mutants.
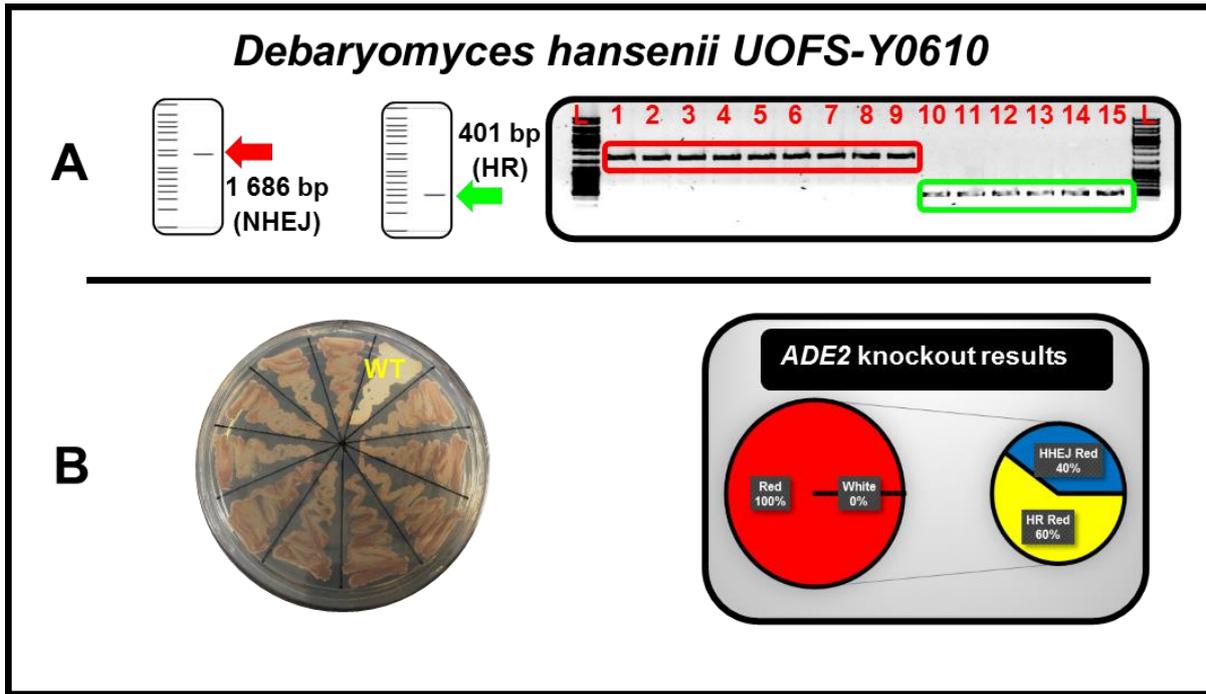
Figure 11: A) *ADE2* disruption results for *Debaryomyces hansenii UOFS-Y0610* when transformed with donor DNA. For the NHEJ repair mechanism, the PCR fragment size will be 1 716 bp (red arrow and box) and for the HR repair mechanism, the fragment size will be 449 bp (green arrow and box). B) On the plate are the wild type (WT) as well as the different colonies streaked out to display the ratio of successful colonies obtained vs unsuccessful. The pie chart also indicates the successful vs unsuccessful, as well as the HR- vs NHEJ repair mechanisms regarding the disruption of the *ADE2* gene.

4.7. *Arxula adeninivorans*

The donor fragment in *Aruxla adeninivorans* is 436 bp (red arrows in Figure 12(A), 13(A) and 14(A)) in size and when transformed with the pKM::CCAS9+ADE2 vector, an editing efficiency of 40% were obtained in the UOFS-Y1220 strain (Figure 12(A)). When screened, via PCR analysis, a HR editing efficiency of 25% were obtained, meaning of the 8 red colonies screened two (green arrow and box) of them tested successful for the donor fragment (436 bp) and the other 6 still showed the original gene size (1 710 bp), indicating a 75% editing efficiency for NHEJ (red arrow and box). In Figure 12(B), 13(B) and 14(B) are the different colonies streaked out to indicate the

ratio of successful vs unsuccessful gene deleted mutants. The UOFS-1219 strain [Figure 13(A)] yielded in a 40% HR editing efficiency (4 out of 10 indicated by the green arrow and box) and a NHEJ efficiency of 60% (6 out of 10 indicated by the red arrow and box).Lastly the Ls3 strain [Figure 14(A)] yielded 50% HR editing efficiency (9 out of 18, indicated by the green arrow boxes) when fragment sizes of 436 bp was observed. The other 9 out of 18 (50%) red colonies screened tested positive for NHEJ, due to the original gene size (1 710 bp) that was obtained through PCR analysis.
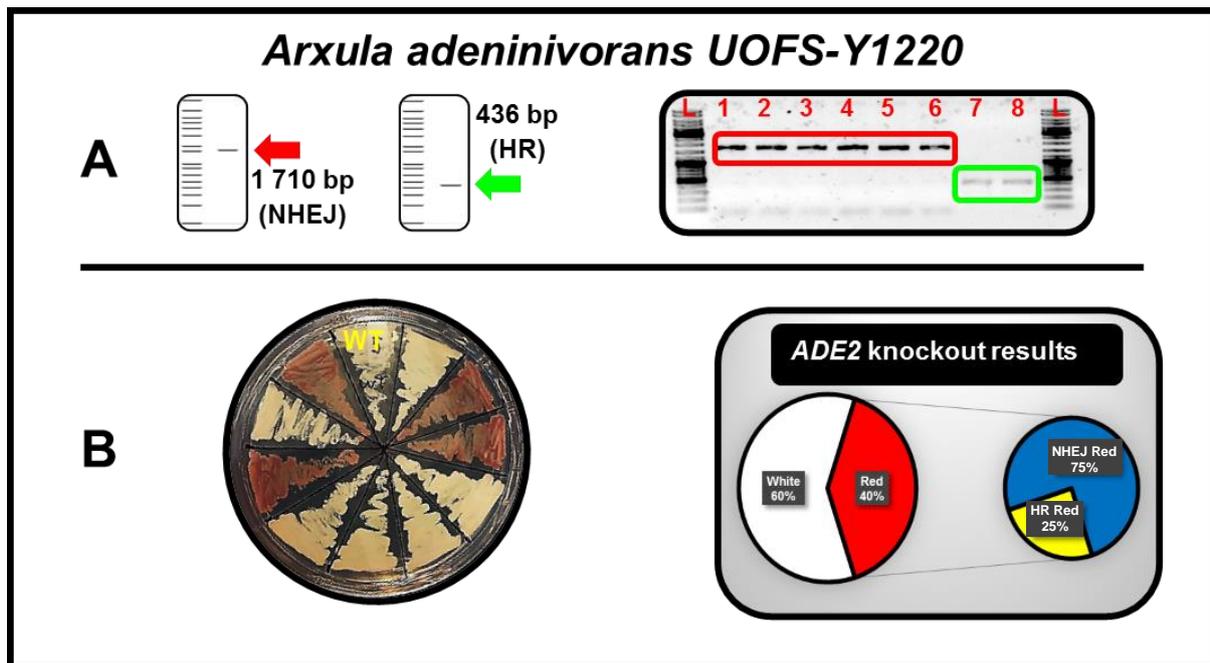

Figure 12: A) *ADE2* disruption results for *Arxula adeninivorans UOFS-Y1220* when transformed with donor DNA. For the NHEJ repair mechanism, the PCR fragment size will be 1 716 bp (red arrow and box) and for the HR repair mechanism, the fragment size will be 449 bp (green arrow and box). (B) On the plate are the wild type (WT) as well as the different colonies streaked out to display the ratio of successful colonies obtained vs unsuccessful. The pie chart also indicates the successful vs unsuccessful, as well as the HR- vs NHEJ repair mechanisms regarding the disruption of the *ADE2* gene.
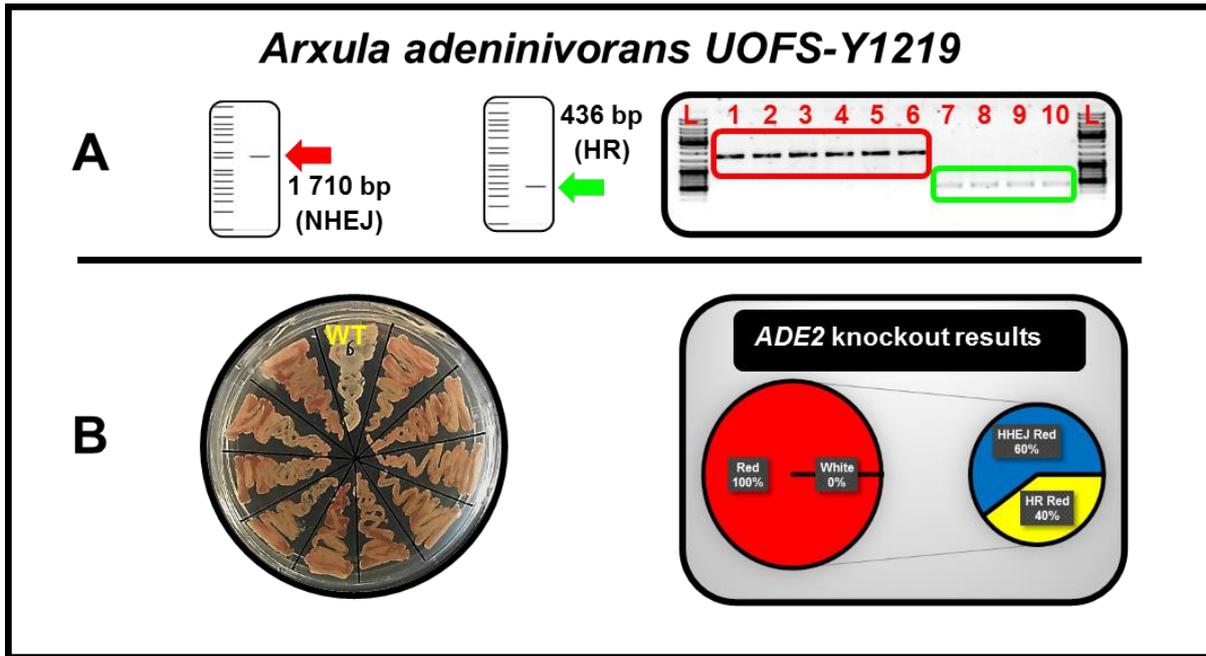
Figure 13: A) *ADE2* disruption results for *Arxula adeninivorans UOFS-Y1219* when transformed with donor DNA. For the NHEJ repair mechanism, the PCR fragment size will be 1 716 bp (red arrow and box) and for the HR repair mechanism, the fragment size will be 449 bp (green arrow and box). B) On the plate are the wild type (WT) as well as the different colonies streaked out to display the ratio of successful colonies obtained vs unsuccessful. The pie chart also indicates the successful vs unsuccessful, as well as the HR- vs NHEJ repair mechanisms regarding the disruption of the *ADE2* gene.
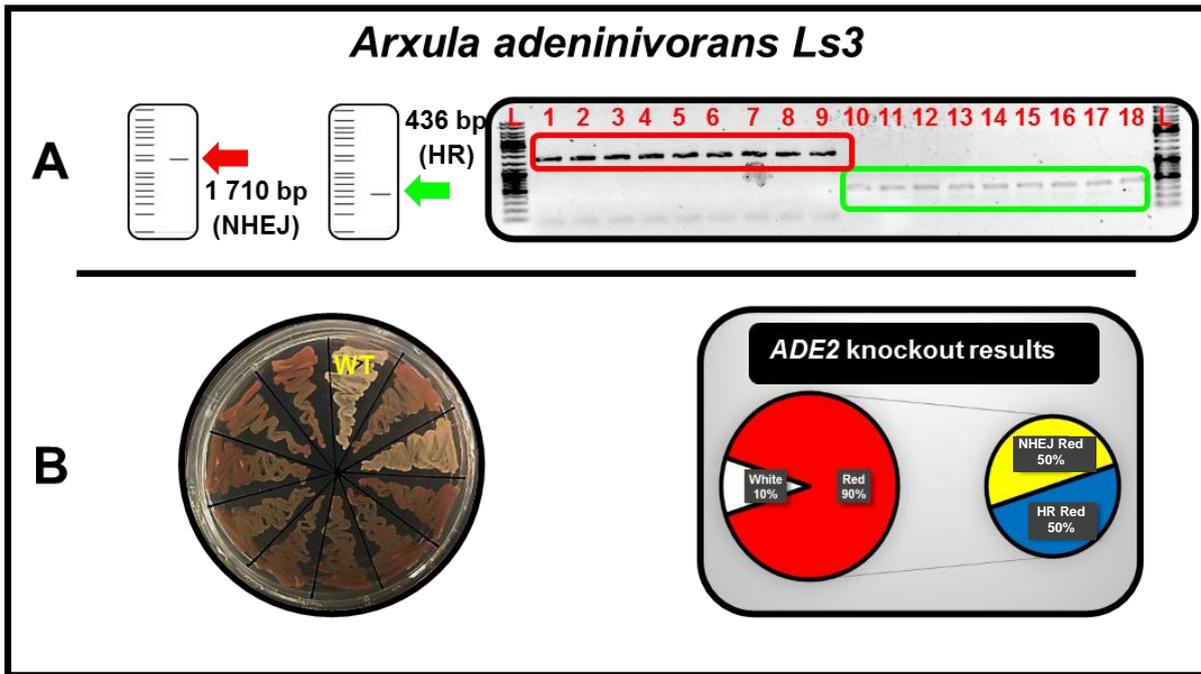
Figure 14: A) *ADE2* disruption results for *Arxula adeninivorans Ls3* when transformed with donor DNA. For the NHEJ repair mechanism, the PCR fragment size will be 1 716 bp (red arrow and box) and for the HR repair mechanism, the fragment size will be 449 bp (green arrow and box). B) On the plate are the wild type (WT) as well as the different colonies streaked out to display the ratio of successful colonies obtained vs unsuccessful. The pie chart also indicates the successful vs unsuccessful, as well as the HR- vs NHEJ repair mechanisms regarding the disruption of the *ADE2* gene.

4.8. *Ogataea polymorpha*

Ogataea polymorpha displayed a 90% disruption efficiency, indicated by Figure 15(B) with the streaked-out colonies in ratio of successful vs unsuccessful colonies. In regards to HR, the editing efficiency obtained (green arrow and box) was 20%, due to only 2 out of the 10 red colonies containing the donor fragment (443 bp) and the other 8 the original gene size (1 704 bp). Meaning the NHEJ efficiency was at 80%. In Figure 15 (B) are the different colonies streaked out to indicate the ratio of successful vs unsuccessful gene deleted mutants, when donor DNA is included with the transformation.
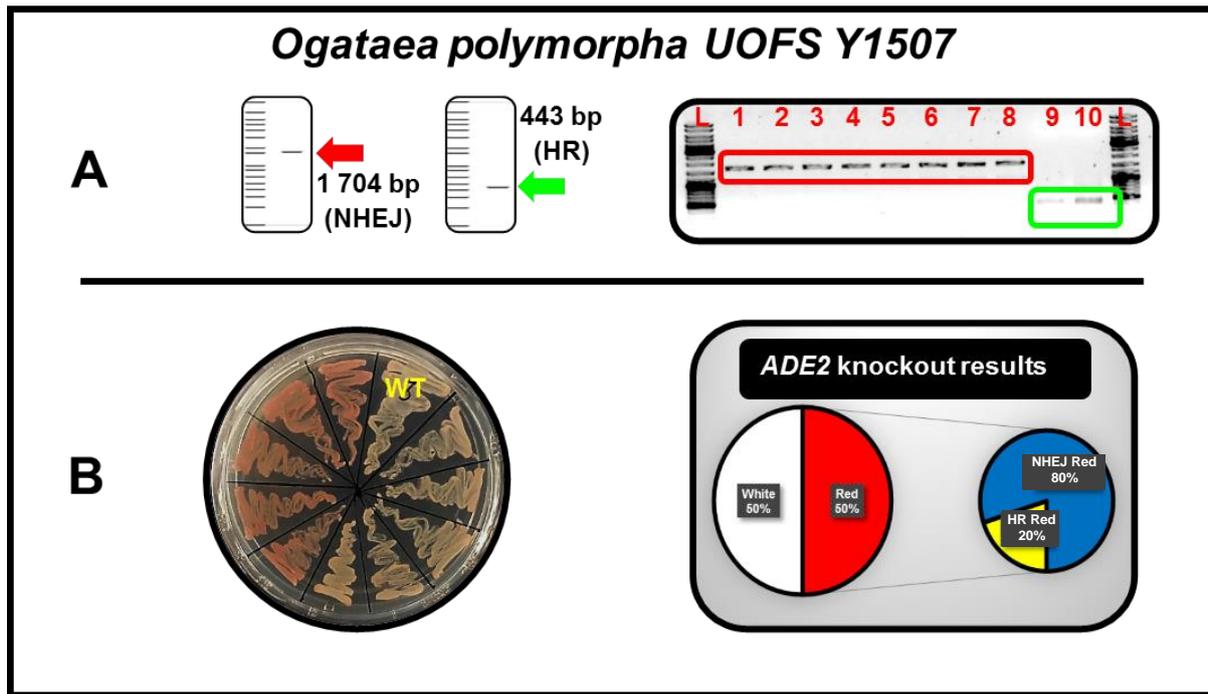
Figure 15: A) *ADE2* disruption results for *Ogataea polymorpha UOFS Y1507* when transformed with donor DNA. For the NHEJ repair mechanism, the PCR fragment size will be 1 716 bp (red arrow and box) and for the HR repair mechanism, the fragment size will be 449 bp (green arrow and box). B) On the plate are the wild type (WT) as well as the different colonies streaked out to display the ratio of successful colonies obtained vs unsuccessful. The pie chart also indicates the successful vs unsuccessful, as well as the HR- vs NHEJ repair mechanisms regarding the disruption of the *ADE2* gene.

4.9. *Komagataella phaffii*

The *Komagataella phaffii* donor fragment is 396bp in size. However, in Figure 16(A), strain UOFS-Y1552 yielded no donor fragments and only displayed the original size (1 698 bp) of the gene. Meaning the 10% disruption efficiency obtained were through NHEJ (red arrow and box). The same trend was observed in the SMD 1168 strain (Figure 17(A), indicated by the red arrow and box), where all of the 10 red colonies obtained portrayed the original size (1 698 bp) of the gene. This also happened in the Km 71 strain (Figure 18 (A) indicated by the red arrow and box), the 14 red colonies all had the original gene size.

However, successful HR disruption efficiency (30%) was obtained in the Gs 115 strain where 10 red colonies were tested and 3 (indicated in Figure 19 (A) by the green arrow and box) of the 10 displayed the donor fragment (396 bp), after PCR analysis. Also, the other 7 colonies showed the original gene size (1 698 bp) resulting in a 70% NHEJ efficiency.

In Figure 16 (B), 17(B), 18(B) and 19(B) are the different colonies streaked out to indicate the ratio of successful vs unsuccessful gene deleted mutants, when the donor DNA is included with the transformation.
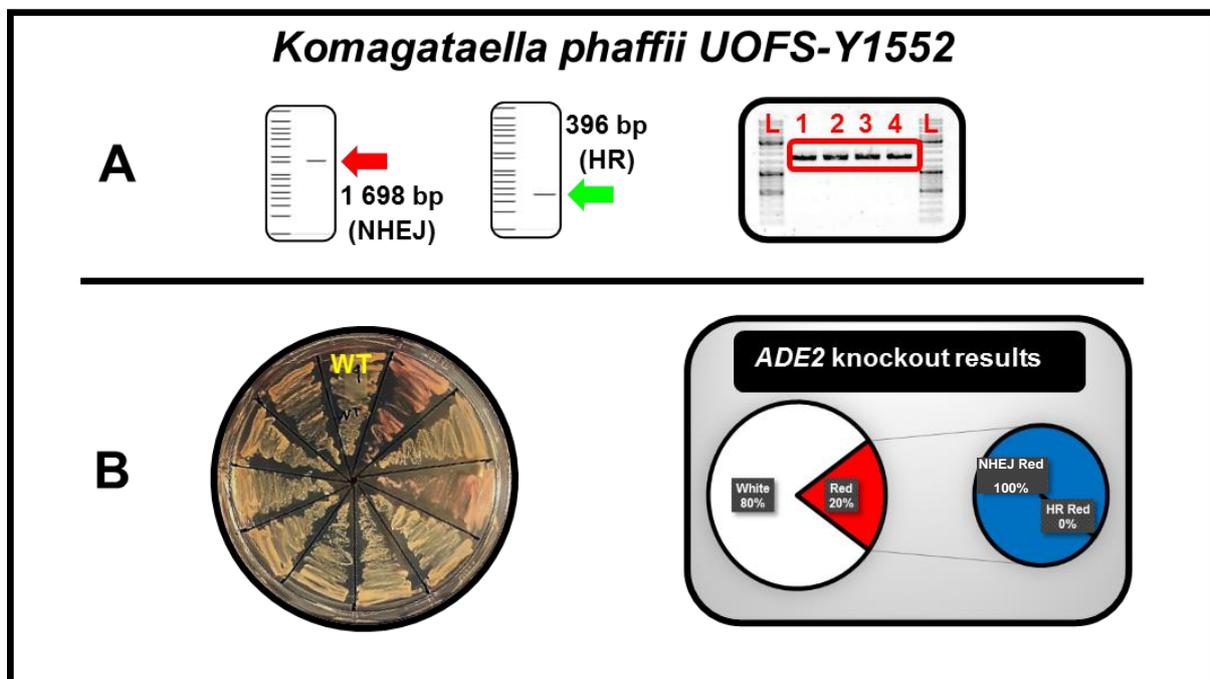

Figure 16: A) *ADE2* disruption results for *Komagataella phaffii UOFS-Y1552* when transformed with donor DNA. For the NHEJ repair mechanism, the PCR fragment size will be 1 716 bp (red arrow and box) and for the HR repair mechanism, the fragment size will be 449 bp (green arrow). B) On the plate are the wild type (WT) as well as the different colonies streaked out to display the ratio of successful colonies obtained vs unsuccessful. The pie chart also indicates the successful vs unsuccessful, as well as the HR- vs NHEJ repair mechanisms regarding the disruption of the *ADE2* gene.

Figure 17: A) *ADE2* disruption results for *Komagataella phaffii SMD 1168* when transformed with donor DNA. For the NHEJ repair mechanism, the PCR fragment size will be 1 716 bp (red arrow and box) and for the HR repair mechanism, the fragment size will be 449 bp (green arrow). B) On the plate are the wild type (WT) as well as the different colonies streaked out to display the ratio of successful colonies obtained vs unsuccessful. The pie chart also indicates the successful vs unsuccessful, as well as the HR- vs NHEJ repair mechanisms regarding the disruption of the *ADE2* gene.
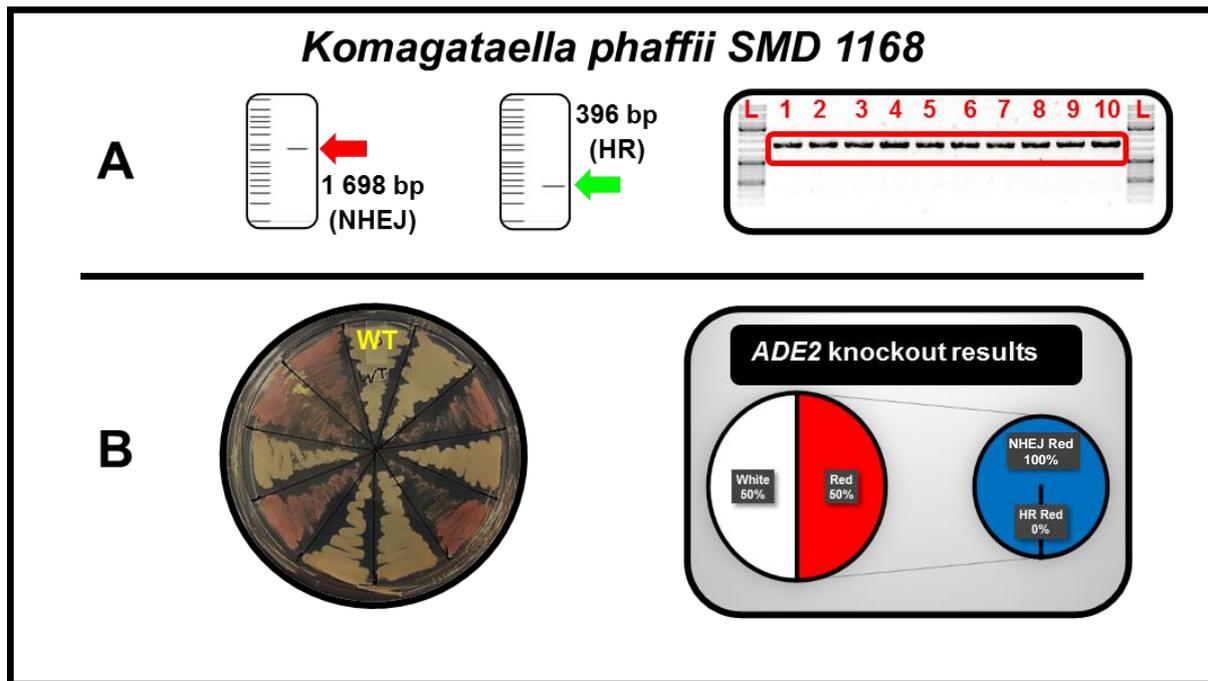
Figure 18: A) *ADE2* disruption results for *Komagataella phaffii Km.71* when transformed with donor DNA. For the NHEJ repair mechanism, the PCR fragment size will be 1 716 bp (red arrow and box) and for the HR repair mechanism, the fragment size will be 449 bp (green arrow). B) On the plate are the wild type (WT) as well as the different colonies streaked out to display the ratio of successful colonies obtained vs unsuccessful. The pie chart also indicates the successful vs unsuccessful, as well as the HR- vs NHEJ repair mechanisms regarding the disruption of the *ADE2* gene.
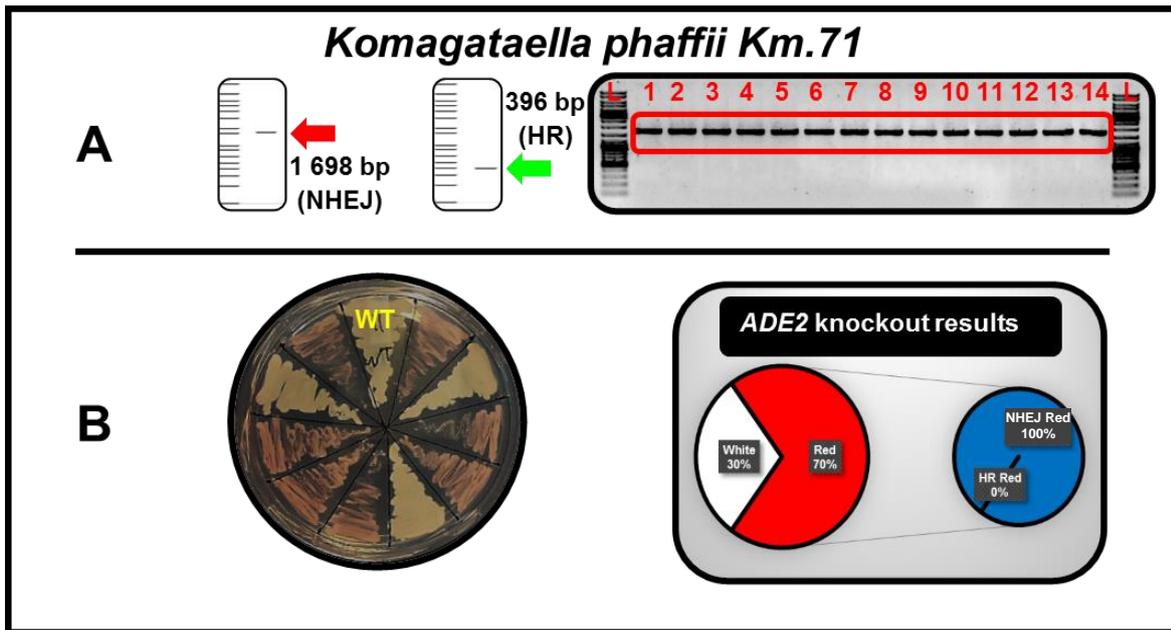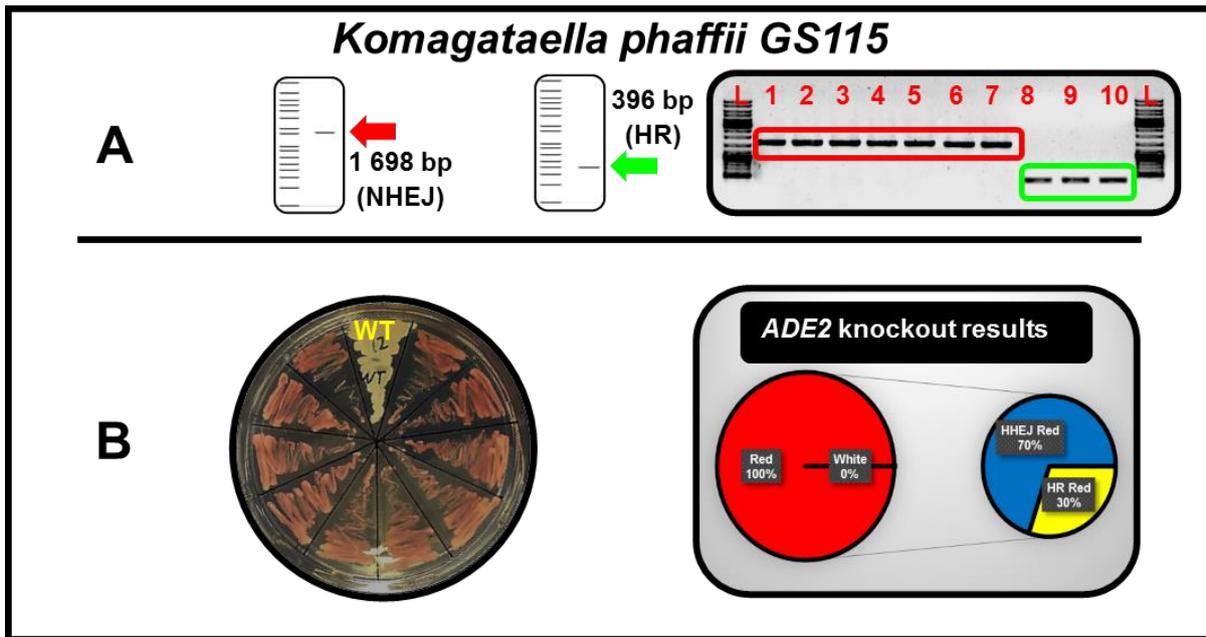
Figure 19: A) *ADE2* disruption results for *Komagataella phaffii GS115* when transformed with donor DNA. For the NHEJ repair mechanism, the PCR fragment size will be 1 716 bp (red arrow and box) and for the HR repair mechanism, the fragment size will be 449 bp (green arrow and box). B) On the plate are the wild type (WT) as well as the different colonies streaked out to display the ratio of successful colonies obtained vs unsuccessful. The pie chart also indicates the successful vs unsuccessful, as well as the HR- vs NHEJ repair mechanisms regarding the disruption of the *ADE2* gene.

5. Conclusion

Although gRNA constructs, as part of an effective CRISPR-CAS9 gene editing system, are designed to satisfy the special demands of the specific yeast, they nevertheless follow similar construction principles. For the expression of the gRNA, Pol III promoters have to be used, due to the extensive modification and processing that is applied when Pol II promoters are used for transcription. However, Pol III promoters are very specific for each organism. Thus, for the system to be classified as wide range, the expression of the gRNA needs to be expressed using the same promoter in all of the yeast, hence the development of a gRNA construct containing ribozymes on both the ends of the gRNA. The ribozymes promote self-cleavage, thus removing any post transcriptional modifications. In this study a gRNA construct were developed, which consisted of a *Yarrowia lipolytica TEF1* promoter and terminator, the gRNA (which consist of the

Chapter 3

CRISPR site followed by the gRNA scaffold) and ribozymes (HH and HDV) flanking the gRNA.

After the development of the TEF1-RGR-gRNA construct, it was incorporated into the pKM::CCAS9 vector to yield the pKM::CCAS9+CRISPR plasmid. Each of the *ADE2* target sites of each yeast were then ligated into the pKM::CCAS9+CRISPR to yield the pKM::CCAS9+ADE2. The vectors were successfully transformed into the different yeasts either in the presence or absence of donor DNA. The transformed colonies were tested for *ADE2* inactivation using YNB dropout plates lacking adenine. If successful for gene disruption/deletion the colonies obtained a red phenotype due to the cells that oxidizes the accrued P-ribosylaminoimidazole. The results indicated that the wide range CRISPR-Cas9 system were successful in disrupting the *ADE2* gene in all of the yeasts tested. When the donor DNA were included in the transformation, it was shown that NHEJ were the preferred repair mechanism. However, the HR efficiency can be increased in most yeast by modifying and optimizing the donor fragment. A more direct approach can be taken and the NHEJ mechanism can be disrupted through inactivation of the KU70 and/or KU80 gene(s) (Boulton 1996; Daley *et al.* 2005; Verbeke *et al.* 2013; Kretzschmar *et al.* 2013; Juergens *et al.* 2018). These genes are crucial in the NHEJ repair pathway and inactivation of these genes results in a decrease in NHEJ repair mechanism.

This proven efficacy of the developed wide range CRISPR-CAS9 gene editing system provides an efficient and precise method to target any gene. This is the first reported system to be tested in such a broad range of yeasts, which include seven biotechnologically relevant yeast, with some of the yeasts tested consisting of more than one strain. For the yeasts included in this study, moderate to high (50% to 100%) disruption or deletion efficiencies were obtained, with the exception of one strain of *A. adeninivorans* (40%) and one strain of *K. phaffii* (30%).

To conclude, this system's gene editing efficiency indicated comparable results to any developed CRISPR-Cas9 system, proving that this system serves as a valuable, cost efficient tool for easy and fast gene disruption in a number of yeast species.

6. References

Addgene., (2015) 'Addgene'. https://www.addgene.org/. Accessed, January 30, (2019)

Bétermier, M.,Bertrand, P.,& Lopez, B.S., (2014) 'Is Non-Homologous End-Joining Really an Inherently Error-Prone Process?'. *PLoS Genet* **10**, e1004086. doi: 10.1371/journal.pgen.1004086

Boulton, S., (1996) 'Identification of a *Saccharomyces cerevisiae Ku80* homologue: roles in DNA double strand break rejoining and in telomeric maintenance'. *Nucleic Acids Res* **24**, 4639–4648. doi: 10.1093/nar/24.23.4639

Chen, D.C.,Yang, B.C.,& Kuo, T.T., (1992) 'One-step transformation of yeast in stationary phase'. *Curr Genet* **21**, 83–84. doi: 10.1007/BF00318659

Chung, C.T.,Niemela, S.L.,& Miller, R.H., (1989) 'One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution.'. *Proc Natl Acad Sci* **86**, 2172–2175. doi: 10.1073/pnas.86.7.2172

Daley, J.M.,Palmbos, P.L.,Wu, D.,& Wilson, T.E., (2005) 'Nonhomologous End Joining in Yeast'. *Annu Rev Genet* **39**, 431–451. doi: 10.1146/annurev.genet.39.073003.113340

Devi, S.,Kedlaya, R.,Maddodi, N.,*et al.*, (2009) 'Calcium homeostasis in human melanocytes: role of transient receptor potential melastatin 1 (TRPM1) and its regulation by ultraviolet light'. *Am J Physiol Physiol* **297**, C679–C687. doi: 10.1152/ajpcell.00092.2009

DiCarlo, J.E.,Norville, J.E.,Mali, P.,*et al.*, (2013) 'Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems'. *Nucleic Acids Res* **41**, 4336–4343. doi: 10.1093/nar/gkt135

Friedland, A.E.,Tzur, Y.B.,Esvelt, K.M.,*et al.*, (2013) 'Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system'. *Nat Methods* **10**, 741–743. doi: 10.1038/nmeth.2532

Gao, Y.,& Zhao, Y., (2014) 'Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing'. *J Integr Plant*

Biol **56**, 343–349. doi: 10.1111/jipb.12152

Geneious., (2017) 'Geneious R10'. http://www.geneious.com/. Accessed, January 30, (2019)

Gratz, S.J.,Cummings, A.M.,Nguyen, J.N.,*et al.*, (2013) 'Genome Engineering of *Drosophila* with the CRISPR RNA-Guided Cas9 Nuclease'. *Genetics* **194**, 1029–1035. doi: 10.1534/genetics.113.152710

Haft, D.H.,Selengut, J.,Mongodin, E.F.,& Nelson, K.E., (2005) 'A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes.'. *PLoS Comput Biol* **1**, e60. doi: 10.1371/journal.pcbi.0010060

Haurwitz, R.E.,Jinek, M.,Wiedenheft, B.,*et al.*, (2010) 'Sequence- and Structure-Specific RNA Processing by a CRISPR Endonuclease'. *Science (80-)* **329**, 1355–1358. doi: 10.1126/science.1192272

Horwitz, A.A.,Walter, J.M.,Schubert, M.G.,*et al.*, (2015) 'Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas'. *Cell Syst* **1**, 88–96. doi: 10.1016/j.cels.2015.02.001

Hwang, W.Y.,Fu, Y.,Reyon, D.,*et al.*, (2013) 'Efficient genome editing in zebrafish using a CRISPR-Cas system'. *Nat Biotechnol* **31**, 227–229. doi: 10.1038/nbt.2501

Jiang, W.,Zhou, H.,Bi, H.,*et al.*, (2013) 'Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice'. *Nucleic Acids Res* **41**, e188–e188. doi: 10.1093/nar/gkt780

Jinek, M.,Chylinski, K.,Fonfara, I.,*et al.*, (2012) 'A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity'. *Science (80-)* **337**, 816–821. doi: 10.1126/science.1225829

Juergens, H.,Varela, J.A.,Gorter de Vries, A.R.,*et al.*, (2018) 'Genome editing in Kluyveromyces and Ogataea yeasts using a broad-host-range Cas9/gRNA co-expression plasmid'. *FEMS Yeast Res* **18**, . doi: 10.1093/femsyr/foy012

KAPA Biosystems Inc., (2014) 'KAPA HiFi HotStart ReadyMix PCR Kit Technical Data Sheet'. https://www.kapabiosystems.com/product-applications/products/pcr-2/kapa-hifi-pcr-kits/. Accessed, January 28, (2019)

Chapter 3

Kistler, K.E.,Vosshall, L.B.,& Matthews, B.J., (2015) 'Genome Engineering with CRISPR-Cas9 in the Mosquito *Aedes aegypti*'. *Cell Rep* **11**, 51–60. doi: 10.1016/j.celrep.2015.03.009

Klebe, R.J.,Harriss, J. V.,Sharp, Z.D.,& Douglas, M.G., (1983) 'A general method for polyethylene-glycol-induced genetic transformation of bacteria and yeast.'. *Gene* **25**, 333–341

Kretzschmar, A.,Otto, C.,Holz, M.,*et al.*, (2013) 'Increased homologous integration frequency in *Yarrowia lipolytica* strains defective in non-homologous end-joining'. *Curr Genet* **59**, 63–72. doi: 10.1007/s00294-013-0389-7

Makarova, K.S.,Grishin, N. V.,Shabalina, S.A.,*et al.*, (2006) 'A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action'. *Biol Direct* **1**, 1–7. doi: 10.1186/1745-6150-1-7

New England Biolabs., (2017) 'NEBuilder HiFi'. https://www.neb.com/-/media/catalog/datacards-or-manuals/manuale2621.pdf. Accessed, January 25, (2019)

New England Biolabs Inc., (2014) 'NEBcloner'. http://nebcloner.neb.com/#!/redigest. Accessed, January 07, (2019)

New England BioLabs Inc., (2017) 'NEB PCR Cloning Kit Instruction Manual'. https://www.neb.com/-/media/catalog/datacards-or-manuals/manuale1202.pdf. Accessed, January 25, (2019)

Novagen., (2011) 'KOD Hot Start DNA Polymerase protocol'. http://goo.gl/UWwdYi. Accessed, January 25, (2019)

Pâques, F.,& Haber, J.E., (1999) 'Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*.'. *Microbiol Mol Biol Rev* **63**, 349–404

Pitcher, R.S.,Brissett, N.C.,& Doherty, A.J., (2007) 'Nonhomologous End-Joining in Bacteria: A Microbial Perspective'. *Annu Rev Microbiol* **61**, 259–282. doi: 10.1146/annurev.micro.61.080706.093354

Chapter 3

Sung, P.,& Klein, H., (2006) 'Mechanism of homologous recombination: mediators and helicases take on regulatory functions'. *Nat Rev Mol Cell Biol* **7**, 739–750. doi: 10.1038/nrm2008

Thermo Fisher Scientific Inc., (2015a) 'Thermo Scientific GeneJET Gel Extraction Kit'. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0012661_GeneJET_Gel_Extraction_UG.pdf. Accessed, January 07, (2019)

Thermo Fisher Scientific Inc., (2015b) 'Restriction Enzymes–Thermo Scientific - ZA'. https://www.thermofisher.com/za/en/home/brands/thermo-scientific/molecular-biology/thermo-scientific-restriction-modifying-enzymes/restriction-enzymes-thermo-scientific.html. Accessed, January 07, (2019)

Ugolini, S.,& Bruschi, C. V., (1996) 'The red/white colony color assay in the yeast Saccharomyces cerevisiae: Epistatic growth advantage of white *ade8*-18, ade2 cells over red ade2 cells'. *Curr Genet* **30**, 485–492. doi: 10.1007/s002940050160

Verbeke, J.,Beopoulos, A.,& Nicaud, J.-M., (2013) 'Efficient homologous recombination with short length flanking fragments in *Ku70* deficient *Yarrowia lipolytica* strains'. *Biotechnol Lett* **35**, 571–576. doi: 10.1007/s10529-012-1107-0

Vyas, V.K.,Barrasa, M.I.,& Fink, G.R., (2015) 'A *Candida albicans* CRISPR system permits genetic engineering of essential genes and gene families'. *Sci Adv* **1**, e1500248. doi: 10.1126/sciadv.1500248

Wang, H.,Yang, H.,Shivalila, C.S.,*et al*., (2013) 'One-step generation of mice carrying mutations in multiple genes by CRISPR/cas-mediated genome engineering'. *Cell* **153**, 910–918. doi: 10.1016/j.cell.2013.04.025

Weninger, A.,Hatzl, A.-M.,Schmid, C.,*et al*., (2016) 'Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast *Pichia pastoris*'. *J Biotechnol* **235**, 139–149. doi: 10.1016/j.jbiotec.2016.03.027

Yan, H.,Opachaloemphan, C.,Mancini, G.,*et al*., (2017) 'An Engineered orco Mutation Produces Aberrant Social Behavior and Defective Neural Development in Ants'. *Cell* **170**, 736–747.e9. doi: 10.1016/j.cell.2017.06.051

Chapter 3

Chapter 4

Summary and concluding remarks

1. Summary and concluding remarks

The CRISPR-Cas9 tool consists of the Cas9 enzyme and a gRNA (Jinek *et al.* 2012) that binds to the Cas9 enzyme and guides it to the target DNA. This is then followed by the binding of the Cas9-gRNA complex to the target DNA, where the Cas9 enzyme then cleaves the DNA and creates double stranded breaks.

Application of the CRISPR-Cas9 system relies on the effective regulation (i.e. homozygous promoter regions) and translation (i.e. codon optimisation) of the *CAS9* gene in each yeast host (DiCarlo *et al.* 2013; Gao & Zhao 2014; Ryan *et al.* 2014; Zhang *et al.* 2014; Bao *et al.* 2015; Horwitz *et al.* 2015; Generoso *et al.* 2016). For the gRNA, expression is very important to obtain the correct functionality and therefore the choice of the correct promoter is essential (Gao & Zhao 2014). If any RNA Polymerase II promoters are used, in for example yeasts, the RNA produced will contain elements added during post-translational modifications because these promoters produces mRNA. Thus, for functional expression, a non-mRNA producing promoter needs to be used. Fortunately, the use of the self-catalysing characteristic of ribozymes can circumvent these problems. Therefore, any promoter can be used as long as the gRNA is flanked by self-splicing ribozymes to remove any post translational modifications, resulting in a functional gRNA.

This study describes new adaptations to CRISPR-Cas9 systems through the development of a new CRISPR-Cas9 system for easy and fast gene editing in a wide variety of different yeast species. Chapter 2 discussed the construction of three different pKM::CAS9 expression constructs, each with a different codon optimized *CAS9* [*P. pastoris* (PCAS9)*, C. albicans* (CCAS9) and *H. sapiens* (HCAS9)] (Gao & Zhao 2014; Vyas *et al.* 2015; Weninger *et al.* 2016). Each of the *CAS9* genes contained one or more nuclear localization sequence to ensure localization to the

nucleus. These *CAS9* expression constructs consisted of the pKM180 backbone, which makes use of a part of the rRNA locus from *Kluyveromyces marxianus* for integration into the yeasts and selection of positive transformants is facilitated by the dominant selection of Hygromycin via the constitutively driven Hygromycin resistant (*HPH*) gene. The three different *CAS9* constructs were transformed into seven different biotechnologically relevant yeasts using two different chemical transformation protocols (Klebe *et al.* 1983; Chen *et al.* 1992) to test for Cas9 expression and the effect that each of the Cas9 proteins have on each of the yeasts. Cas9 expression was subsequently confirmed through western blot analysis. It was observed that expression of the *P. pastoris* optimised *CAS9* resulted in multiple protein bands when analysed using a western blot. Expression of the *H. sapiens* lead to slower growth of transformed yeasts as well as increased sensitivity toward Hygromycin. It was therefore apparent that the pKM::CCAS9 vector (i.e. the pKM180 containing the *C. albicans* optimised CAS9 gene) was the best candidate to proceed with the testing of complete CRISPR-CAS9 gene editing system.

For the construction of the complete wide rage CRISPR-CAS9 gene editing system, where a single system could be used in a range of different yeasts, it was essential to use a promoter for the expression of the gRNA that could be recognised by a range of different yeast species. For this purpose we made use of the RGR-gRNA strategy, where the gRNA is flanked by self-splicing ribozymes (Gao & Zhao 2014). The RGR-gRNA was incorporated into the vector which resulted in a vector that consist of a CCAS9 gene and a gRNA construct which is flanked by ribozymes.

To target the *ADE2* gene in each of the yeast species, a specific CRISPR target sequence was designed for each yeast and inserted in the completed pKM::CCAS9+CRISPR vector. This vector was then transformed into the different yeasts to determine the efficiency of inactivation (either through disruption or deletion) of the *ADE2* gene. Inactivation of the *ADE2* gene results in an observable red phenotype if plated on media that lack adenine. When the vector was transformed with or without donor DNA, all yeast species (depending on the strain selected) showed moderate to high deletion or disruption efficiency of the *ADE2* gene. It should be noted that when donor DNA was included in the transformation, most of the yeast displayed gene disruption through homologous recombination (HR). Better HR disruption

efficiency can possibly be obtained by modifying the donor fragment to obtain optimal integration (via HR) of the donor fragment. Enhanced HR is also likely by disrupting the *KU70/80* genes (Boulton 1996; Daley *et al.* 2005; Verbeke *et al.* 2013; Kretzschmar *et al.* 2013; Juergens *et al.* 2018b). These genes are crucial in the NHEJ repair pathway. Thus, by disrupting these genes a decrease in NHEJ and an increase in HR disruption efficiencies will be observed.

In conclusion, the developed wide range CRISPR-CAS9 gene editing system proved to be successful by disrupting the *ADE2* gene in all of the yeast tested. This system therefore allows rapid, cost-effective genetic manipulation of biotechnologically relevant yeast.

2. References

Bao, Z.,Xiao, H.,Liang, J.,*et al.*, (2015) 'Homology-Integrated CRISPR–Cas (HI-CRISPR) System for One-Step Multigene Disruption in *Saccharomyces cerevisiae*'. *ACS Synth Biol* **4**, 585–594. doi: 10.1021/sb500255k

Boulton, S., (1996) 'Identification of a *Saccharomyces cerevisiae Ku80* homologue: roles in DNA double strand break rejoining and in telomeric maintenance'. *Nucleic Acids Res* **24**, 4639–4648. doi: 10.1093/nar/24.23.4639

Chen, D.C.,Yang, B.C.,& Kuo, T.T., (1992) 'One-step transformation of yeast in stationary phase'. *Curr Genet* **21**, 83–84. doi: 10.1007/BF00318659

Daley, J.M.,Palmbos, P.L.,Wu, D.,& Wilson, T.E., (2005) 'Nonhomologous End Joining in Yeast'. *Annu Rev Genet* **39**, 431–451. doi: 10.1146/annurev.genet.39.073003.113340

DiCarlo, J.E.,Norville, J.E.,Mali, P.,*et al.*, (2013) 'Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems'. *Nucleic Acids Res* **41**, 4336–4343. doi: 10.1093/nar/gkt135

Gao, Y.,& Zhao, Y., (2014) 'Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing'. *J Integr Plant Biol* **56**, 343–349. doi: 10.1111/jipb.12152

Generoso, W.C.,Gottardi, M.,Oreb, M.,& Boles, E., (2016) 'Simplified CRISPR-Cas genome editing for *Saccharomyces cerevisiae*'. *J Microbiol Methods* **127**, 203–

205. doi: 10.1016/j.mimet.2016.06.020

Horwitz, A.A.,Walter, J.M.,Schubert, M.G.,*et al.*, (2015) 'Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas'. *Cell Syst* **1**, 88–96. doi: 10.1016/j.cels.2015.02.001

Jinek, M.,Chylinski, K.,Fonfara, I.,*et al.*, (2012) 'A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity'. *Science (80-)* **337**, 816–821. doi: 10.1126/science.1225829

Juergens, H.,Varela, J.A.,Gorter de Vries, A.R.,*et al.*, (2018) 'Genome editing in Kluyveromyces and Ogataea yeasts using a broad-host-range Cas9/gRNA co-expression plasmid'. *FEMS Yeast Res* **18**, . doi: 10.1093/femsyr/foy012

Klebe, R.J.,Harriss, J. V.,Sharp, Z.D.,& Douglas, M.G., (1983) 'A general method for polyethylene-glycol-induced genetic transformation of bacteria and yeast.'. *Gene* **25**, 333–341

Kretzschmar, A.,Otto, C.,Holz, M.,*et al.*, (2013) 'Increased homologous integration frequency in *Yarrowia lipolytica* strains defective in non-homologous end-joining'. *Curr Genet* **59**, 63–72. doi: 10.1007/s00294-013-0389-7

Ryan, O.W.,Skerker, J.M.,Maurer, M.J.,*et al.*, (2014) 'Selection of chromosomal DNA libraries using a multiplex CRISPR system.'. *Elife* **3**, 1–15. doi: 10.7554/eLife.03703

Verbeke, J.,Beopoulos, A.,& Nicaud, J.-M., (2013) 'Efficient homologous recombination with short length flanking fragments in *Ku70* deficient *Yarrowia lipolytica* strains'. *Biotechnol Lett* **35**, 571–576. doi: 10.1007/s10529-012-1107-0

Vyas, V.K.,Barrasa, M.I.,& Fink, G.R., (2015) 'A *Candida albicans* CRISPR system permits genetic engineering of essential genes and gene families'. *Sci Adv* **1**, e1500248. doi: 10.1126/sciadv.1500248

Weninger, A.,Hatzl, A.-M.,Schmid, C.,*et al.*, (2016) 'Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast *Pichia pastoris*'. *J Biotechnol* **235**, 139–149. doi: 10.1016/j.jbiotec.2016.03.027

Zhang, G.-C.,Kong, I.I.,Kim, H.,*et al.*, (2014) 'Construction of a Quadruple Auxotrophic Mutant of an Industrial Polyploid *Saccharomyces cerevisiae* Strain

by Using RNA-Guided Cas9 Nuclease'. *Appl Environ Microbiol* **80**, 7694–7701. doi: 10.1128/AEM.02310-14