MACQUARIE
University
SYDNEY·AUSTRALIA

Exploring the Genome Rearrangement System, SCRaMbLE, to Introduce DNA into *Saccharomyces cerevisiae*

Elizabeth Louise Izolde Wightman

A thesis submitted in partial fulfilment of the degree of Master of Research

Supervisor: Professor Helena Nevalainen

Department of Chemistry and Biomolecular Sciences,
Macquarie University

9 October 2015

Table of Contents

Abstract

A novel genome rearrangement tool, SCRaMbLE (synthetic chromosome rearrangement and modification by loxP mediated evolution), has been designed for the synthetic genome of *S. cerevisiae*. Collaborators of the 'yeast 2.0' project initially construct one version of the synthetic genome. However, SCRaMbLE can rapidly generate billions of unique genomes in even a small culture. Only one of the 16 chromosomes is currently complete, limiting full genome rearrangement by SCRaMbLE. As a contribution to the international project, I constructed and integrated a 'megachunk' of synthetic chromosome 14 into *S. cerevisiae*. Currently, SCRaMbLE is limited to *S. cerevisiae* DNA only. I aimed to explore SCRaMbLE in a novel application: as a tool to introduce foreign DNA into *S. cerevisiae*. To develop protocols, the *S. cerevisiae* marker gene *URA3* was used to SCRaMbLE into yeast. SCRaMbLEing *URA3* increasing the transformation efficiency significantly compared to non-SCRaMbLEd *URA3*. gDNA from the filamentous fungus *Trichoderma reesei* was used as an example of SCRaMbLEing large libraries of foreign DNA into yeast. Successful attachment of SCRaMbLE recombination sites to gDNA fragments was achieved using a cloning approach. gDNA fragments were SCRaMbLEd into yeast and cellulase activity was screened.

Declaration

The research presented in this thesis is original work by the author, unless otherwise references or acknowledged, conducted between January 2015 and October 2015 at Macquarie University, Australia. This material has not been submitted for qualification or assessment to any other institution

Elizabeth Wightman

Acknowledgements

First, I want to thank my supervisor, Professor Helena Nevalainen, for her support and guidance throughout this year. Thank you for always being willing to spend time with me. Your constant guidance and advice throughout the year has been invaluable.

A huge thank you to Tom Williams who was there at every step of the way. Thank you for all the time you spent with me, for always being happy to answer my never ending questions, and for helping to make my project the best it could be.

I want to thank Liisa Kautto, Anna Gryshyna, Heinrich Kroukamp and Elizabeth Daniel for their support and guidance in and outside the lab.

Last but not least, I thank my Mum, Dad and brother for their love and support. I couldn't have got through this year without them. I also thank my boyfriend Alex for being more supportive and encouraging than I could have asked for.

Abbreviations

The following abbreviations are used within the text:

CIP	Calf Intestinal alkaline phosphatase
CRISPR	Clustered regularly interspaced short palindromic repeats
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
EBD	Estrogen binding domain
gDNA	Genomic deoxyribonucleic acid
h	hour
K.O.	Knock out
LB	Lysogeny broth
LiOAc	Lithium acetate
ml	mililitre
µl	microlitre
min	minute
mRNA	Messenger ribonucleic acid
O.D.	Optical density
PEG	Polyethylene glycol
PCR	Polymerase chain reaction
PDA	Potato dextrose agar
RNA	Ribonucleic acid
Sc2.0	Saccharomyces cerevisiae 2.0
SCRaMbLE	Synthetic chromosome rearrangement and modification by loxP mediated evolution
SOC	Super optimal broth with catabolite repression
ss	Single stranded
Syn	Synthetic chromosome
TAR	Transformation associated recombination
qPCR	Quantitative polymerase chain reaction
YPD	Yeast extract peptone dextrose
YPG	Yeast extract-peptone-glycerol
V	Volts

Chapter 1

Introduction

1.1 Whole genome engineering

Successful whole genome engineering would provide an unparalleled level of control over the manipulation of genomes and biological processes. The motivations of genome engineering are diverse; genome engineering could be used as a platform to design and engineer organisms with almost any process imaginable with genomes optimised for their intended purpose. In medicine, microbes could be engineered to hunt and destroy cancer cells, in industry, yeast could be engineered to efficiently produce biofuels from raw cellulosic material, and in environmental management, algae could be manipulated to turn industrial CO_2 and wastewater nutrients into renewable chemicals.

Engineering at the whole genome scale is a fundamentally difficult task. Predictable engineering at this scale is close to impossible due to the complexity of biological networks and systems. Rational engineering approaches at this scale often fail due to the prerequisite of complete knowledge required for designing the desired changes. Instead, non-rational approaches, such as traditional evolutionary engineering have often been more successful. New approaches are required to fulfil the ambitious goals of whole genome engineering. Specifically, a novel tool has been developed for synthetic biology for utilisation in the yeast *Saccharomyces cerevisiae*. The system, called SCRaMbLE, is a powerful approach for whole genome engineering and has the potential to be exploited for novel applications.

1.2 *S. cerevisiae* as a model for novel genetic engineering technologies

Novel and revolutionary genetic engineering techniques for developing new phenotypes are constantly emerging for use in *S. cerevisiae*; amongst these are tools for whole genome editing (*e.g.* the CRISP-Cas system (DiCarlo *et al.*, 2013)) and synthetic biology approaches (*e.g.* Gibson assembly (Gibson *et al.*, 2009)) which allow for engineering at the whole genome scale. *S. cerevisiae,* also known as baker's or brewer's yeast, is one of the most widely researched organisms in existence and is the model organism for higher eukaryotic life.

There are a number of reasons why *S. cerevisiae* has been so successful in molecular biology studies. Firstly, there is the availability of highly efficient transformation methods, such as the LiAc/SS carrier DNA/PEG method (Gietz and Schiestl, 2007). The organism is also

Generally Regarded as Safe (GRAS), and the first eukaryotic organism to have its full genome published (Goffeau *et al*., 1996). Finally, there are many tools facilitating genetic engineering significantly, including an array auxotrophic strains, specialised expression vectors and selectable markers (discussed in (Ostergaard *et al*., 2000).

A high rate of homologous recombination makes *S. cerevisiae* highly desirable for targeted genetic engineering. The unique capability is exploited in transformation procedures, as foreign DNA can be effectively guided to a target location. Prior to transformation associated recombination (TAR), exogenous DNA is flanked with fragments homologous to the target locus (Larionov *et al*., 1994). The fragment is inserted into the chromosomes by homologous recombination reactions. The phenomenon is used to insert, delete or disrupt genetic material with accuracy and efficiency (figure 1) (discussed in (David and Siewers, 2014)).

Figure 1: Transformation associated recombination (TAR) can be used to replace a target gene with a genetic marker. Red boxes represent homologous sequences, as do blue boxes. Black crosses show homologous recombination between homologous sequences.

1.3 Rational genetic engineering of *S. cerevisiae*

Rational engineering is the knowledge-based modification of genetic material to bring about pre-defined changes in cellular functioning. The approach requires comprehensive knowledge concerning the pathways and networks of interest in order to predict the responses of the planned genetic manipulation. Though successful in many instances, the technique has been limited in engineering at the genome scale in *S. cerevisiae*, due to limited available knowledge, resources and technologies (Reviewed in (Heinemann and Panke, 2006)).

Metabolic engineering as an example of rational engineering

The goal of metabolic engineering is to re-build metabolic pathways through the deletion or insertion of one or several genes by TAR into the host chromosome (Raab *et al*., 2010, Ostergaard *et al*., 2000, Runguphan and Keasling, 2014). As an example, four genes were systematically deleted in an attempt to increase the production of succinic acid in *S.*

cerevisiae (Raab *et al*., 2010). The genes *SDH1*, *SDH2*, *IDH1* and *IDP1* are involved in succinate dehydrogenase activity. It was predicted that with their deletion, the carbon flux would be redirected into the glyoxylate cycle, and would allow succinate to accumulate as an end product. As predicted, the quadruple-deletion strain showed a 4.8-fold increase in succinic acid compared to the wild-type strain.

Genome editing as an example of rational engineering

The clustered regularly interspaced short palindromic repeat (CRISPR) and CRISPR-associated (Cas) systems are innovative tools that are simple, targeted and efficient. The CRISPR/Cas system originates from the bacterial immune system (Garneau *et al*., 2010). Following the invasion of foreign DNA, short endogenous guide-RNA (gRNA) sequences form a complex with the endonuclease Cas9 to seek out and destroy invading DNA (Jinek *et al*., 2012). The Cas9 cleavage of target DNA is site specific, determined by base pairing between the 5' end of the gRNA and target. The system became of great interest as a targeted genome-editing tool due to its effective RNA-guided endonuclease activity. CRISPR/Cas9 has been applied to genome engineering in *S. cerevisiae*. DiCarlo *et al*. (2013) used the system for donor-DNA homologous recombination. In this technique, the CRISPR-Cas system induces double stranded breaks that are repaired by homologous recombination. There was a simultaneous introduction of 'donor DNA' where an exogenous, homologous sequencing carrying the desired genetic sequence recombines with the target. In this proof of concept study, CRISPR-Cas was used for the targeted repair of a mutated *ADE2* marker gene.

Limitations of rational genetic engineering

The complexity of biological networks and systems is what limits rational engineering and leads to: failures, results less successful than predicted, or unwanted side effects. Firstly, the approach requires comprehensive knowledge of cellular pathways involved in the process of interest. This includes a complete understanding of genetics, regulatory factors and enzymes involved, along with their activity and interactions within networks. Even intricate, well-researched metabolic maps can fail to predict the effect of genetic modifications. Systems biology is the discipline which aims to understand and model the cell as a whole and is used to predict cellular responses. However, systems biology has its own drawback; the models are only as good as the experimental data they are based off. Furthermore, they often rely on assumptions, due to lack of sufficient experimental information. As an example, a sophisticated genome-scale model was used to design metabolic engineering strategies in *S. cerevisiae* (Bro *et al*., 2006). The model scored a number of strategies for the reengineering

of the redox metabolism. However, the engineered strains did not perform as expected. While the model predicted a complete elimination of the formation of glycerol using a particular approach, the resulting strain showed only 40% lower formation of the product, demonstrating the limitations of even intricate, well-researched models.

1.4 Non-rational genetic engineering

Non-rational (or 'non-targeted') genetic engineering is not limited by available knowledge; it requires little to no information of the genetic basis of desired phenotypes. The technique usually involves mutagenesis to generate a mutant library which is screened to isolate organisms with desirable traits. The approach can be used for engineering at the whole genome scale, to develop strains with complex phenotypes that are often difficult to predict rationally (Sonderegger and Sauer, 2003, Wisselink *et al*., 2009).

Evolutionary engineering to enhance *S. cerevisiae*

Evolutionary engineering (also 'adaptive laboratory evolution' or 'directed evolution') is a non-rational method of genetic engineering involving multiple cycles of mutagenesis and screening. The design is based on natural evolution, where advantageous traits are developed and maintained in a population. The population is first subjected to a mutagen, followed by large-scale phenotypic screening, usually involving a selective pressure. The selection of chosen organisms will possess genetic components required to survive in the screening environment. Evolutionary engineering has been utilised to generate novel *S. cerevisiae* strains. One study developed yeast with higher tolerance to ethanol, which is beneficial in fermentation procedures (Stanley *et al*., 2010). In the study, the parent yeast strain was either mutated chemically, or allowed to mutate spontaneously, before being subjected to ethanol stress. Following multiple rounds, the procedure achieved two strains which demonstrated increased growth rates in sub-lethal ethanol concentrations, and improved survivability in lethal ethanol concentrations, compared to the parent strain.

The use of non-rational genetic engineering reveals novel genotype-phenotype links

Non-rational genetic engineering approaches are powerful because they reveal potential links between complex phenotypes and the corresponding genotype. This contrasts to rational approaches where the genotype/phenotype relationship is predicted, and the results of the study either confirm or contradict the prediction. Non-rational approaches open the possibility to uncover new information about the genetic basis of specific traits. Therefore, it is important that following the isolation of desirable mutants, a rigorous analysis of the strains is

performed to identify genetic mutations. Genome sequencing analysis is invaluable in uncovering the genetic basis of strain improvement. Recently, genome sequencing of a UV-mutated *S. cerevisiae* strain revealed that a single-nucleotide point mutation improved amylase production by 35% (Liu *et al.*, 2014). The authors were then able to theorise the function of the gene involved, and the effect of the mutated gene of amylase production. Overall, the analysis of newly developed strains can reveal the genetic basis of observed phenotypes and uncover novel targets for future rational genetic engineering studies.

1.5 Synthetic Biology

Synthetic biology is the application of genetic engineering principles on a larger scale, focusing on whole systems of genes. The goal of synthetic biology is to re-design and re-engineer life by modifying and extending the behaviour of organisms to perform novel processes. The field has developed powerful tools for engineering at the whole genome scale. These are used to develop complex artificial devices, pathways and networks.

There are three defining principles of synthetic biology (Heinemann and Panke, 2006). Firstly, there is a standardisation of biological components. Lego-like 'biobricks' are genetic sequences coding for one part of a network, which are standardised and shared. This means pathways can be engineered without the need for an understanding of the complex underlying genetics. A good example of this is in the context of the International Genetically Engineered Machine Competition (iGEM). The global competition sees teams of undergraduate students build synthetic networks and systems brick by brick, using these standardised components. In the 2009 competition, the Cambridge team won with '*E. chromi*', an engineered *E. coli* which responds to different concentrations of pollutants in water and generates corresponding colours visible to the naked eye (Discussed in (Porcar and Peretó, 2014). The second principle of synthetic biology states that pathways and networks should be simple; complex and unnecessary genetic components should be removed or simplified. Finally, synthetic biology demands the development of technologies for rapid and effective engineering. SCRaMbLE (discussed later) is a novel synthetic biology tool developed which fulfil this principle.

An innovative molecular tool that was developed for synthetic biology is 'Gibson assembly' which can be used to assemble large biosynthetic pathways. The *in vitro* technique efficiently assembles numerous molecules of overlapping DNA using three enzymes: an exonuclease, a polymerase and a ligase (Gibson *et al.*, 2009). The method works in a rapid, isothermal, single-step reaction, a feature distinguishing it from traditional cloning methods (figure 2).

The technique has demonstrated its potential for future biosynthetic pathway engineering by rapidly constructing very large DNA molecules. An early version of the procedure was used to assemble the entire *Mycoplasma genitalium* genome (Gibson *et al*., 2008).

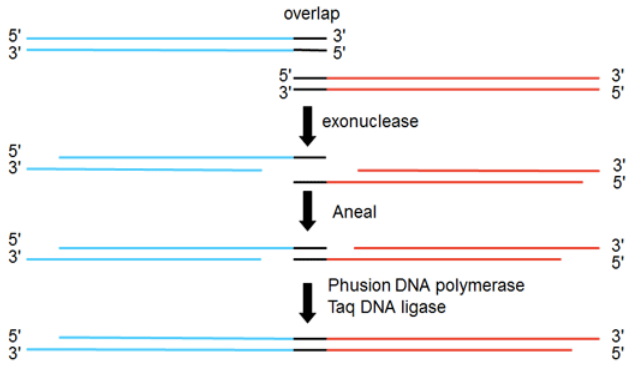

Figure 2: During Gibson assembly, an exonuclease removes nucleotides from double stranded DNA at the 5' end, complementary single strand overhangs anneal, a polymerase fills the gaps and a ligase seals the sequence.

1.5.1 Yeast 2.0

Collaborators from around the world are currently working to build the very first synthetic eukaryotic genome. The goal of the 'yeast 2.0' project is to re-design and re-engineering the entire genome of *S. cerevisiae*. The synthetic yeast, 'Sc2.0', will have a number of genetic modifications compared to the wild type, and contain features to facilitate innovative future studies, including a novel non-rational genetic engineering tool, SCRaMbLE.

The synthetic genome was designed in accordance to specific design principles (Dymond *et al*., 2011). Firstly, the genome will result in fitness and phenotype as close as possible to wild type *S. cerevisiae*; synthetic yeast should be at wild type standards of growth, morphology and viability. Secondly, the genome will lack destabilising elements and will be streamlined. To achieve this, many elements deemed unnecessary were deleted or moved. Finally, it will comprise genetic features facilitating a multitude of future studies. This principle is satisfied by the replacement of stop codons, and the SCRaMbLE system (discussed later).

A number of institutions around the world are building and integrating the 16 synthetic chromosomes of *S. cerevisiae* piece by piece, until all wild-type DNA is replaced with synthetic DNA. Each synthetic chromosome is comprised of ~30 kb 'megachunks' which are made up of four 'chunks'. The chromosomes are built one megachunk at a time, which are integrated into the chromosome by homologous recombination (figure 3). At the 3' end of each megachunk is a selective marker, either the *LEU2* or *URA3* gene, which allows for selection of the integration of each megachunk on a medium lacking leucine or uracil respectively. The presence of each marker is temporary and is replaced upon the integration

of the next sequential megachunk. Only two markers are required for the building of synthetic chromosomes, as the genes are alternated with each megachunk integration.

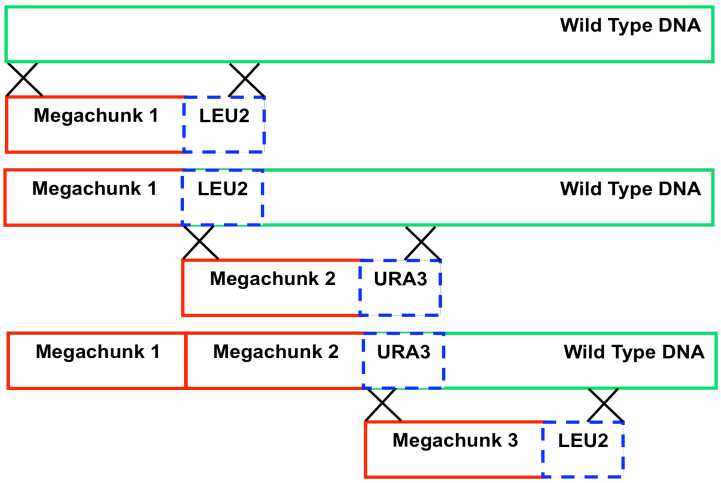

Figure 3: Construction of each chromosome is performed one megachunk at a time. Each megachunk is integrated by homologous recombination (black crosses). After each recombination, the marker from the previous megachunk is lost and the marker from the new megachunk is gained.

Modifications in the synthetic sequence of Sc2.0

Compared to the wild type sequence, elements which are identical in the synthetic genome include the gene order and the sequence of non-coding regions. This was to avoid disrupting cell fitness. Four types of genetic elements were removed or relocated (figure 4) (Dymond *et al*., 2011). Firstly, many introns were removed, as they were deemed unnecessary. Secondly, retrotransposons were removed to eliminate highly repetitive sequences, known to increase genome instability. Thirdly, the removal of subtelomeric repeats at chromosome ends was performed to generate more streamlined chromosomes. Specifically, subtelomeric Y' elements have unknown functions and were deleted. Subtelomeric X elements function in telomeric silencing and possibly in chromosome segregation. These highly divergent X elements were replaced with the core X element sequence to preserve function. Finally, tRNA sequences were relocated to a new 'neochromosome' to reduce genome instability.

A number of features were added to the synthetic genome (figure 4). Firstly, the TAG stop codon was replaced by TAA. This facilitates future studies by allowing for the expansion of the genetic code. This can be achieved by the addition of a twenty-first, unnatural amino acid, coded by TAG. Secondly, strings of code were replaced with 'PCRTags' specific to either the synthetic or wild type sequence. By utilising these convenient genetic markers, sequences from synthetic or wild type origin can be rapidly verified using PCR. This feature is crucial in the building of the genome, as it confirms the incorporation of each megachunk of synthetic DNA, paired with the loss of the equivalent wild type sequence. Finally, loxPSym

recombination sequences were designed to reside 3 bp after the stop codon of all non-essential genes. These sequences are essential for the SCRaMbLE system, discussed later.

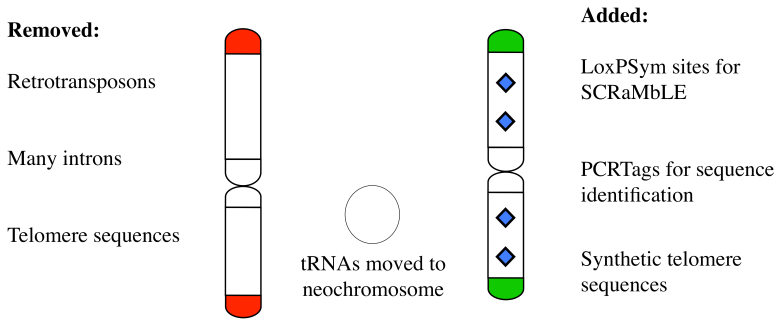

Figure 4: Elements added and removed in Sc2.0. Removed from the wild type DNA were retrotransposons, many introns and telomere sequences. tRNA sequences were moved to a neochromosome. Elements added were loxPsym sites (blue diamonds), PCRTags and synthetic telomere sequences. The synthetic chromosome is usually 10-20% shorter than the wild type (*e.g.* synthetic chromosome III is ~13% shorter than the wild type).

1.5.1.1 SCRaMbLE: a powerful genome rearrangement system

SCRaMbLE (synthetic chromosome rearrangement and modification by loxP mediated evolution) is the defining feature of the synthetic yeast genome, facilitating powerful and inducible evolution. Initially the yeast 2.0 project engineers only one synthetic genome, however SCRaMbLE can generate billions of unique genomes on demand in a small cell culture. These unique genomes have highly variable structures and contents. The system induces the rearrangement of genetic material, involving gene insertions, deletions, translocation or inversion of DNA flanked with recombination sites called loxP (figure 5).

The SCRaMbLE mechanism relies on the Cre/lox recombination system. LoxP recombination sites are 34 bp sequences, taken from the P1 bacteriophage, and are the substrate for Cre recombinase activity. LoxP sequences were designed to reside 3 bp after the stop codon of ~5000 non-essential genes throughout the 16 synthetic chromosomes (figure 5). LoxP sites, also called loxPsym, are symmetrical, which enables huge combinatorial diversity, as they are able to recombine in either direction (Dymond *et al.*, 2011). Importantly, the sequences are shorter than what is required for homologous recombination in yeast, so Cre recombinase needs to be expressed in order for recombination to occur (Dymond *et al.*, 2011). SCRaMbLE is induced with the controlled expression of Cre recombinase. To achieve inducible expression, an engineered version of Cre recombinase was fused to the murine estrogen-binding domain (EBD). The estrodiol inducible Cre-EBD variant has low basal activity and is

controlled by the daughter-cell-specific promoter *SCW11*. Therefore, there is a short and strong wave of recombinase production occurring only once in a cell's lifetime.

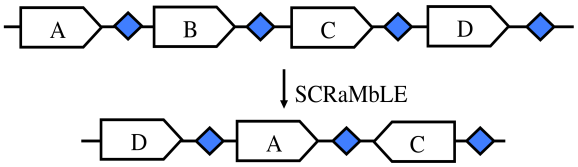

Figure 5: LoxP recombination sequences (blue diamonds) reside throughout the synthetic genome, after non-essential genes (A-D). Following chemical induction, a heterologous recombinase causes translocation, duplication, deletion or inversion of non-essential genes.

In accordance to the third design principle, the SCRaMbLE mechanism allows for genetic flexibility to facilitate future studies. On demand, SCRaMbLE can generate billions of unique genomes which can be screened for desirable traits (figure 6). Through the sequencing of strains, analysis can reveal common subsets of genes required for certain phenotypes. This approach could uncover new directions for rational engineering approaches.

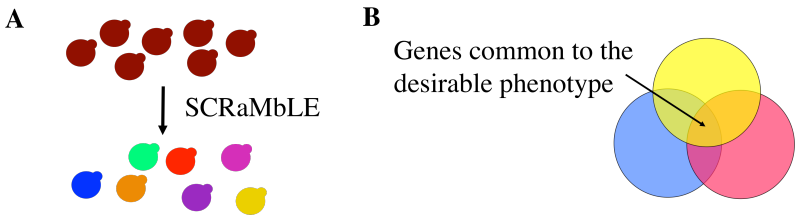

Figure 6: SCRaMbLE produces huge genetic diversity, resulting in billions of unique genomes which differ in size and configuration (**a**). After the application of appropriate selection pressures, desirable yeast colonies can be isolated and their genomes sequenced. (**b**) Analysis may reveal common subsets of genes responsible for the desired phenotype.

Currently, only synthetic chromosome III (SynIII) has been constructed and is available for SCRaMbLEing in the laboratory. SCRaMbLE can only be used genome wide once all 16 chromosomes are complete. Upon completion, thousands of loxP sites will reside throughout the chromosomes and SCRaMbLE will produce a level of genetic diversity never before seen.

The completed yeast 2.0 project will be one of the greatest technical achievements in molecular biology of all time, being the first synthetic eukaryotic genome. Even the synthesis of the 583 kb *Mycoplasma genitalium* genome was vastly simpler (Gibson *et al*., 2008); it was orders of magnitude smaller, compared with the 12 Mb yeast genome. The yeast 2.0 project is not only an increase in scale, but involves the complete redesign of the genome. In contrast,

M. genitalium genome did not incorporate significant modifications. Other than the addition of 'watermarks', no synthetic sequences were incorporated, the genome was simply rebuilt.

1.6 SCRaMbLE for introducing foreign DNA into *S. cerevisiae*

Until now, SCRaMbLE has been limited to rearranging *S. cerevisiae* DNA only. Therefore, only existing *S. cerevisiae* phenotypes are possible to generate using SCRaMbLE. There are endless foreign phenotypes that would be useful to engineer in yeast, but without the development of new protocols, they are impossible to introduce with SCRaMbLE.

In the current work SCRaMbLE was explored as a novel tool to introduce foreign DNA into *S. cerevisiae*. In this application, SCRaMbLE was considered a non-rational approach that has significant advantages compared to rational engineering approaches. The engineering of complex phenotypes, such as those requiring whole pathways, is very difficult to engineer rationally. However, whole libraries of loxP-flanked genes could potentially be SCRaMbLEd into yeast allowing for the selection of entire foreign metabolic pathways.

SCRaMbLEing *Trichoderma reesei* gDNA as a proof-of-concept

DNA from *Trichoderma reesei* was chosen to SCRaMbLE into yeast in order to explore the system in a novel application. *T. reesei* is a filamentous fungus known for its ability to break down cellulose to glucose, and its exceptional protein production ability (Mandels and Reese, 1957). Cellulose is a linear polymer of glucose units joined by (beta-1,4)-glycosidic linkages, utilised by *T. reesei* as an energy and carbon source. The three classes of cellulases produced by *T. reesei* are cellobiohydrolases (CBH), endoglucanases (EG), and B-glucosidase (Montenecourt, 1983). These enzymes, for which the genes are located in different chromosomes (Carter *et al*., 1992) act synergistically to hydrolyse cellulosic substrates. The *T. reesei* genome codes for 10 cellulases and 16 hemicellulases (Martinez *et al*., 2008).

Cellulose is a cheap product which can be converted to sugars for fermentation into ethanol by *S. cerevisiae,* which is highly valuable as a biofuel. Therefore, it would be beneficial to engineer yeast that can hydrolyse cellulosic biomass and produce ethanol at a high rate. If the hydrolysis and fermentation were economically combined in a consolidated bioprocess (CBP), it would reduce the costs of biomass conversion (van Zyl et al., 2007).

In addition to show a proof-of-concept for SCRaMbLEing *T. reesei* DNA into yeast, the recombinant yeast strains were screened for cellulase activity. Screening was performed

following the SCRaMbLEing into *S. cerevisiae* of a *T. reesei* DNA library representative of the whole genome into (chapter 5). Cellulase activity was screened on Avicel, an insoluble cellulose microcrystalline degraded by exocellulases, and also screened on Carboxymethyl cellulose (CMC), a derivative of cellulose with carboxymethyl groups (-CH_2-COOH) bound to hydroxyl groups of glucopyranose monomers on the cellulose backbone. CMC is a substrate for endoglucanases. CMC and Avicel are polymeric substrates which would need to be hydrolysed by cellulases before they can be taken into cell and used for growth and energy.

1.6.1 The Challenges of SCRaMbLEing in foreign DNA

In the application of SCRaMbLE to introduce foreign DNA into yeast, there are issues that must be acknowledged. From previous literature involving more traditional methods of DNA introduction into heterologous expression hosts, the most apparent issues were the number of introns contained in foreign genes and expression of DNA under non-homologous promoters.

Intron splicing in yeast

It is important to consider issues relating to introns and their splicing. Introns are the non-coding sequences of DNA, whereas exons are the protein coding sequences. Prior to synthesis of mature mRNA, introns are cleaved and the final mature mRNA holds the information required for protein synthesis. The issue of intron splicing of foreign eukaryotic genes came to light shortly after the first transformation procedures became available. For example, in an early study there was abnormal expression of the heterologous rabbit β-globin gene in *S. cerevisiae* (Beggs *et al*., 1980). After analysis of the transcripts, no splicing of the primary β-globin transcript could be detected and a mutated protein was expressed.

Compared to other eukaryotes, including filamentous fungi, *S. cerevisiae* has a limited number of introns. Only 283 of the ~6000 yeast genes contain introns, accounting for only 5% of the genome (Davis *et al*., 2000, Juneau *et al*., 2007, Spingola *et al*., 1999, Kellis *et al*., 2003). In contrast, 68% of *T. reesei* genes contain introns (Gurr *et al*., 1987). Additionally, the vast majority of intron-containing genes in yeast have only one intron and fewer than 10 genes have more than one intron (Spingola *et al*., 1999). However, in terms of *T. reesei* cellulase genes, *cbhI*, *egI* and *egIII* contain two introns, and *cbhII* contains three introns (Teeri *et al*., 1987). *S. cerevisiae* is capable of splicing, however, the issue is if the splicing machinery will be efficient enough to synthesise good quality mRNA requires for high volume protein production (Spingola *et al*., 1999).

Foreign promoters

An important consideration for the expression of heterologous genes is the promoter. Promoters initiate transcription of mRNAs by providing a sequence that is recognised by RNA polymerase. They also require various regulatory proteins to respond to signals for switching the promoter on and off (Lee *et al*., 2002). An early study of heterologous gene expression demonstrated the importance of a functional transcriptional promoter in the host organism and showed the result of promoter/host incompatibility (McNeil and Friesen, 1981). In this study, the herpes simplex virus (HSV) thymidine kinase gene (*tk*) along with its native promoter was cloned into yeast plasmids and transformed into *S. cerevisiae*. These plasmids propagated as autonomously replicating plasmids. However, no RNA specific to the *tk* coding sequence was detected. The authors concluded that the HSV promoter was unrecognised by *S. cerevisiae* RNA polymerases.

In a previous research, it was reported that a region of the *T. reesei* cellobiohydrolase *cbhI* promoter drives expression in *S. cerevisiae*. Within the *T. reesei chb1* promoter there are two regulatory regions. One region controls basal expression, while the other controls induced expression. In the study by (Carraro *et al*., 1998), the *T. reesei cbh1* regulatory region, *UARcb1*, was inserted in a yeast multicopy plasmid. It was engineered as a "promoter-fusion" with the *S. cerevisiae CYC1* promoter, upstream of a *lacZ* reporter gene in a yeast multicopy plasmid. After transformation into *S. cerevisiae*, yeast were grown on a glucose medium, then a glycerol medium. In this instance, the glucose or glycerol was the sole carbon source. Overall, the effect of the *cbh1* promoter region in yeast was comparable to that in *T. reesei*; the expression of *lacZ* in *S. cerevisiae* was repressed by glucose (Ilmen *et al*., 1997). Specifically, *lacZ* expression increased at least 10-fold after a shift to a glycerol medium.

It is important to note that SCRaMbLE has the potential to surmount the issues concerning heterologous protein production in *S. cerevisiae*. Along with genes directly involved in cellulose hydrolysis, other *T. reesei* DNA may be imported that can encode a selection of useful functions, for example, the regulatory factors for *T. reesei* promoters have potential to be SCRaMbLEd and expressed in *S. cerevisiae*.

1.7 Engineering cellulolytic yeast

There have been many attempts to develop cellulolytic yeasts through the introduction of foreign DNA. Specifically *T. reesei* genes have been rationally engineered into *S. cerevisiae* under homologous promoters. Below are some examples.

Although heterologous cellobiohydrolases can be produced in *S. cerevisiae* the titers of functionally secreted enzymes are relatively low. For example, T. reesei genes encoding CBHI and CBHII have been expressed in *S. cerevisiae* under the native *ENO1* promoter and terminator and the *ADH2* promoter and terminator respectively (Den Haan *et al*., 2007a). However, protein expression was too low for growth on Avicel cellulose.

T. reesei genes encoding endoglucanase I and II have been cloned and expressed in *S. cerevisiae* under the transcriptional control of the yeast *ENOI* promoter (Du Plessis *et al*., 2010). The genes were co-expressed with the synthetic, codon-optimised *T. reesei* cellobiohydrolase gene and the *Saccharomycopsis fibuligera* β-glucosidase gene. The recombinant strains hydrolysed phosphoric-acid-swollen cellulose into mainly cellobiose, cellotriose and glucose. The accumulation of cellobiose suggested the poor enzyme activity of β-glucosidase. A successful cellulolytic *S. cerevisiae* strain would secrete at least one endoglucanase, cellobiohydrolase and a β- glucosidase. The expression of these enzymes needs to be sufficient, and at a ratio that would prevent the accumulation of one product and also produce enough glucose to support growth on cellulose.

The *T. reesei* cellulase, endoglucanase I (EGI), has been previously expressed in *S. cerevisiae* (Knowles *et al*., 1998). A cDNA coding for endoglucanase I (EGI) was first isolated from a *T. reesei* cDNA library. Expression vectors were assembled with EGI cDNA which was under control of the yeast ADH1 promoter. After cultivation of hybrid yeasts, activity of the cellulolytic enzyme was investigated. Enzyme activity was highest in the cultivation medium consisting 1-5% of total cell protein (Knowles *et al*., 1998) demonstrating that the protein could be produced under the yeast promoter.

1.8 Aims of the study

The overall aim of this work was to explore SCRaMbLE as a novel tool to introduce foreign DNA into *S. cerevisiae*. Currently, only synthetic chromosome III is ready for SCRaMbLEing. In order to broaden the scope for SCRaMbLEing, a synthetic megachunk O was introduced into synthetic chromosome XIV. To develop protocols for SCRaMbLEing eukaryotic DNA into *S. cerevisiae,* the homologous *URA3* was introduced into yeast. As a proof of concept for the ability to SCRaMbLE heterologous eukaryotic DNA into yeast, a *T. reesei* gDNA library was SCRaMbLEd into yeast which were screened for cellulase activity.

Chapter 2

Construction of Megachunk O of Yeast Synthetic Chromosome XIV

Before SCRaMbLEing of the entire yeast genome can take place, all synthetic chromosomes must be complete; SCRaMbLE relies on loxP recombination sequences that reside throughout the synthetic yeast genome. Currently, only synthetic chromosome III (SynIII) has been constructed and is available for SCRaMbLEing. In order to induce SCRaMbLE over the whole genome, the construction and integration of all chromosomes is required. To contribute to the building of Sc2.0, megachunk O (the 15th sequential megachunk) of yeast synthetic chromosome XIV (SynXIV) was constructed. The megachunk, containing the loxP sites required for SCRaMbLE, was integrated into *S. cerevisiae*, replacing equivalent wild type DNA. Screening techniques including replica plating, qPCR analysis and fitness assays were performed to confirm the integration of megachunk O, and the fitness of the resulting *S. cerevisiae* strain.

2.1 METHODS

Extracting chunks O1-O4 from four different plasmids

The four chunks of megachunk O in SynXIV were provided by Genscript on pUC57 plasmids. Chunks were isolated from plasmids using appropriate restriction enzymes (table 1). The conditions of digestion were as suggested by NEB (New England Biolabs).

Table 1: Length of each chunk in megachunk O and restriction enzymes used to extract them from pUC57 plasmids.

Chunk	Length of chunk	Restriction enzymes required	
O1	9631 bp	*Bso*BI	*Bst*EII
O2	8115 bp	*Bst*EII	*Bgl*I
O3	7659 bp	*Bgl*I	*Bgl*I
O4	9248 bp	*Bgl*I	*Bae*I

Following digestion, DNA fragments were separated on a gel (gels in this thesis are 1% agarose run in TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8)) at 80V for 1 h. Bands containing each chunk were excised from the agarose gel. DNA was purified using the Zymo Research Zymoclean™ Gel DNA Recovery Kit (all kits mentioned in this thesis were used according to the manufacturers instructions unless specified otherwise).

Ligation of chunk O1-O4 to form megachunk O

Chunks O1-O4 were pooled in diminishing quantities; 4 µg O1, 2 µg O2, 1 µg O3, 0.5 µg O4. DNA was purified using the QIAGEN QIAquick® PCR Purification Kit. Chunks were ligated (Figure 7) in 20 U/µL NEB T4 ligase and 1X NEB T4 DNA Ligase Reaction Buffer at 16 °C for 18 h. DNA ligase was deactivated at 65 °C for 20 min.

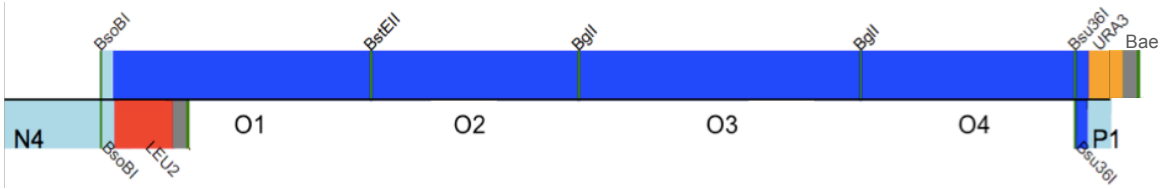

Figure 7: Megachunk O of SynXIV. Chunks O1-4 are represented by dark blue boxes. The lines between each chunk represent restriction enzyme sites. The N megachunk *LEU2* marker, shown in red, is lost with O megachunk integration. The O megachunk marker is *URA3* shown in yellow.

Transformation of megachunk O into *S. cerevisiae*

Megachunk O was transformed into a *S. cerevisiae* strain containing the previous megachunk (megachunk N) using the LiAc/SS carrier DNA/PEG method (Gietz and Schiestl, 2007). Four transformations were set up; two using megachunk O DNA, one negative control and one positive control (pRS416 plasmid containing *URA3*). Following transformation, cells were plated onto –ura plates (1X Sigma Yeast Nitrogen Base Without Amino Acids, 1X Sigma Yeast Synthetic Drop-out Medium Supplements without uracil, 20 g/L agar, 20 g/L glucose) and incubated for ≥ 2 days at 30 °C. (drop-out plates in this thesis are made as above with appropriate Sigma drop-out supplements)

Replica plating of –ura plates and gDNA extraction from colonies

To screen for the gain of the megachunk O *URA3* marker, and the loss of the megachunk N *LEU2* marker, transformation plates were replica plated onto –leu, then –ura plates and incubated overnight at 30 °C. The gDNA of colonies that grew on –ura plates and not on –leu plates was extracted using the Epicenter Master Pure Yeast DNA purification kit.

qPCR of gDNA to screen for gain of synthetic DNA and loss of wild type DNA

A qPCR analysis was set up to screen PCRTags in megachunk O; 21 primer pairs annealed to wild type PCRTags and 21 primer pairs annealed to synthetic PCRTags (table 7, appendix). Reactions were set up with: 1X Roche LightCycler® 480 SYBR Green I Master, 20 ng/µl

gDNA, 250 nM primer pair. Controls were set up (negative control, wild-type control with BY4742 gDNA, synthetic positive containing megachunk O DNA). Reactions were carried out in the Roche LightCycler® 480 instrument II using the program detailed in table 2.

Table 2: qPCR program used for megachunk O synthetic and wild type primers.

Program	Cycles	Target temperature (°C)	Hold (sec)
Incubation	1	95	5
Anneal	40	95	30
		64	30

Fitness assay of *S. cerevisiae* to identify growth defects resulting from the integration of megachunk O

Following the integration of each megachunk, the fitness of the resulting *S. cerevisiae* strain/s is compared to the wild type and a knockout strain to check for growth defects. Four yeast strains were grown in 5 ml of YPD overnight at 30 °C, shaking at 200rpm; two colonies which amplified all synthetic primers and no wild type primers in the qPCR analysis, the wild type strain BY4742, and a knockout strain (BY4741 APJ1Δ::KanMX). The knockout strain had the same gene knocked out that was disrupted by *URA3* in the two O-megachunk strains. YPD aliquots were inoculated with each overnight culture to an O.D. of 0.125, then incubated at 30 °C, shaking at 200 rpm until samples reached at least O.D. 0.5. A 100 μl sample of the lowest O.D. sample, and equivalent cell counts of the remaining three samples were spun down at 13,000 RCF. Samples were resuspended in water and serially diluted to 10^{-5}. A spot assay was performed on two YPD agar plates and two YPG agar plates, with one of each incubated at 30 °C and one incubated at 37 °C. After sufficient growth, the size and abundance of the colonies was observed.

2.2 RESULTS

Currently, it is only possible to SCRaMbLE SynIII. In order for SCRaMbLE to work through the whole genome, all 16 synthetic chromosomes must be constructed and integrated into *S. cerevisiae*. Construction into SynXIV is beneficial, as once this chromosome is complete, it can also be SCRaMbLEd. Megachunk O of SynXIV was built by ligating chunks O1-O4. The megachunk was transformed into a *S. cerevisiae* strain containing the previous megachunk N. Replica plating suggested which colonies had lost the N megachunk marker *LEU2* and gained the O megachunk marker *URA3*. qPCR screening confirmed the gain of synthetic DNA and

the loss of all equivalent wild type DNA. Two colonies gave positive results for this screening. Through fitness testing, a potential defect in these two colonies was discovered, however the defect may be due to gene disruption by the *URA3* marker which is reversed with the addition of the next megachunk.

Figure 8 shows the agarose gel electrophoresis of chunks O1-O4 following plasmid digestion. Bands at the top of lanes 2-5 represent chunks. The remaining plasmid fragment migrated further down the gel. The bands at the bottom of lanes 3, 4 and 5 (below 0.5 kb) may be undigested plasmid.

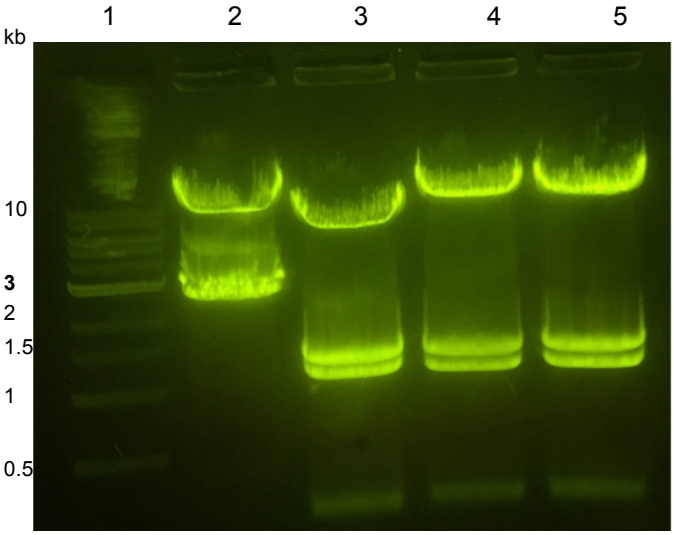

Figure 8: Agarose gel DNA electrophoresis of chunks O1-O4 after isolation from plasmids. Lane 1: NEB 1 kb DNA ladder, lane 2: chunk O1 digested with *Bso*BI and *Bst*EII, lane 3: chunk O2 digested with *Bst*EII and *Bgl*I, lane 4: chunk O3 digested with *Bgl*I, lane 5: chunk O4 digested with *Bgl*I and *Bae*I.

Following the transformation of megachunk O into *S. cerevisiae*, thousands of colonies grew on –ura plates. After replica plating, ~50 colonies were observed growing on –ura and not on –leu plates, suggesting the loss of *LEU2* and the gain of *URA3*. gDNA from twelve colonies was extracted for qPCR analysis. Through the qPCR analysis of gDNA from two colonies, all synthetic primers were amplified and none of the wild type primers were amplified by cycle 30, suggesting the replacement of wild type DNA with synthetic DNA.

After the integration of each megachunk, the fitness of the resulting *S. cerevisiae* strain must be screened. The fitness of two O megachunk colonies, the wild type strain BY4742 and a knockout strain BY4741 APJ1Δ::KanMX were assessed by growth on YPD (agar 20 g/l, BactoPeptone 20 g/l, BactoYeast extract 10 g/l, 20 g/l glucose) and YPG plates (agar 20 g/l, BactoPeptone 20 g/l, BactoYeast extract 10 g/l, 30 ml/l glycerol) incubated at 30 °C and 37 °C (figure 9). The size and number of colonies across the four strains on YPD plates showed no significant differences. On YPG plates, there was a significant difference in colony size and viability when comparing the four strains; there were less colonies and

colonies were smaller in O megachunk strains compared to the wild type and the knockout strains.

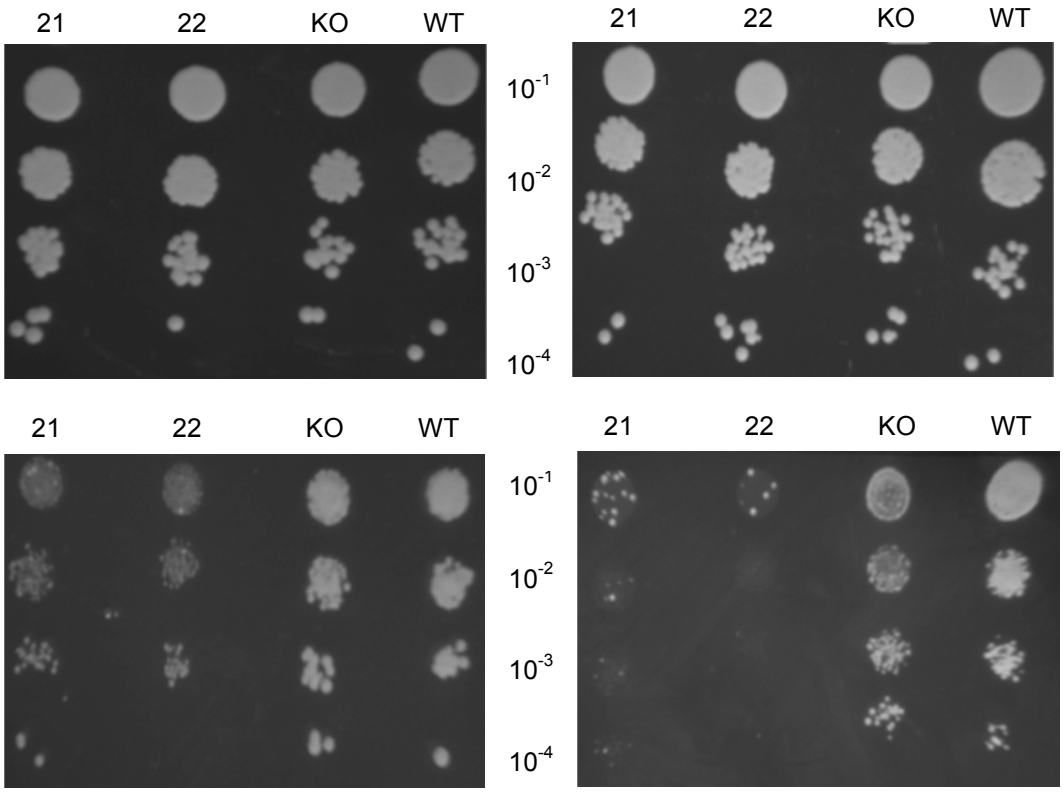

Figure 9: Series of fitness assays of four *S. cerevisiae* strains on YPD and YPG agar plates. In each image, column 1: colony 21, column 2: colony 22, column 3: knock out strain BY4741 APJ1Δ:KanMX (KO), column 4: wild type strain BY4742 (WT). Each row is a 10x dilution from the previous. Top left: YPD 30 °C, top right: YPD 37 °C, bottom left: YPG 30 °C, bottom right: YPG 37 °C.

Despite the two O megachunk colonies demonstrating an apparent growth defect compared to the knock out strain, the growth defect may still be due to gene disruption by the *URA3* marker cassette. There may, therefore, be no growth defect compared to the wild type once the disruption of the gene is reversed when *URA3* is removed with integration of the next consecutive megachunk. A test can be performed to confirm this theory, discussed below.

2.3 DISCUSSION

Construction of SynXIV is essential to provide another synthetic yeast chromosome for SCRaMbLEing. Megachunk O of SynXIV was successfully transformed into *S. cerevisiae*, replacing the equivalent wild type sequence. Replica plating suggested a number of colonies had gained the *URA3* marker from megachunk O and lost the *LEU2* marker from megachunk

N. Two colonies were isolated that passed qPCR analysis, suggesting that all wild type DNA was replaced with the synthetic sequence. Through fitness screening, no difference was found in viability or growth rate between these two colonies and the knockout or wild types strains on YPD plates, however on YPG plates a significant growth defect was observed.

The protocol prior to transformation was optimised to increase the probability of isolating colonies containing all chunks. Chunks O1-O4 were pooled for ligation in a ratio of approximately 8:4:2:1, so that O1 was most abundant, and O4 was least abundant. This protocol was designed as O4 contained the *URA3* selection marker and was therefore used for initial screening on –ura plates. Therefore, if O4 was least abundant, the probability that growing colonies also contained O1-O3 was relatively high compared to a protocol where all four chunks were pooled in equal quantities.

Genomic DNA was extracted from colonies that were positive for the *URA3* marker and negative for the *LEU2* marker. qPCR was carried out with two sets of primers, one which was specific to synthetic DNA sequences and one specific to the native chromosome. For two colonies, amplification using all synthetic primers occurred before 35 cycles. Using wild type primers, there was amplification of a few pairs, however the amplification did not begin until after ~35 cycles. Using SYBR Green, non-specific background can show up in late cycles (Pfaffl, 2001, Pfaffl *et al*., 2002), so these amplifications were disregarded.

One of the design principles of yeast 2.0 is that the synthetic yeast should demonstrate equal fitness to the wild type. It is important that fitness testing takes place regularly throughout the building of the synthetic genome, following the incorporation of each megachunk. Firstly, this allows for the rapid identification of any defects. If fitness testing occurred only at the completion of each chromosome, it would be difficult to determine the location of DNA causing the defect. Secondly, fitness testing after each cycle allows for more rapid correction of the issue. Since each megachunk is added sequentially, it is relatively easy to redesign the offending sequence, have the DNA re-synthesised, and simply repeat the integration.

The fitness of two megachunk O colonies was compared to the wild type strain BY4742 and the knockout strain BY4741 APJ1Δ::KanMX. Ideally, there should be no fitness defect compared to the wild type, however, fitness defects may occur due to gene disruption by the *URA3* marker cassette. Marker cassettes were designed to reside at the 3' end of each megachunk, often disrupting a gene. Importantly, with the integration of the next sequential

megachunk, the marker is removed and a different marker is integrated at the 3' end of the new megachunk (Dymond and Boeke, 2012). Therefore, fitness defects caused by marker disruption are eliminated with the next megachunk integration. For this reason, fitness testing requires comparison with a specific knockout strain from the yeast gene deletion project, (Winzeler *et al*., 1999). The knockout strains have a gene deletion corresponding to the gene that is disrupted by the marker cassette. The strain BY4741 APJ1Δ::KanMX has had the *YNL077w* gene knocked out and was used for comparison with megachunk O colonies, as *YNL077w* was disrupted by *URA3*.

There was an apparent fitness defect in the two O megachunk colonies compared to the wild type and knockout strains. Despite the knock out strain showing no obvious growth defects, the defects of O megachunk colonies may still be due disruption of *YNL077w*. Importantly, the knockout strain has the complete removal of the *YNL077w* gene, while the O megachunk colonies only has a disruption in the gene (Winzeler *et al*., 1999). The disrupted *YNL077w* encodes a 528 amino acid protein of the Hsp40 (DnaJ) family, APJ1 (Anti-Prion DnaJ). The introduction of the *URA3* marker resulted in the last 57 amino acids of APJ1 to be deleted and replaced by 16 amino acids encoded in the *URA3* cassette (figure 27 in appendix for alignment). The growth defect in megachunk O colonies may therefore be a result of a partly functional, mutant protein. It is possible that this mutated protein causes a growth defect, but complete loss of the protein, as found in the knock out strain, does not.

There was a significant growth defect on YPG but not YPD. Interestingly, the APJ1 protein functions as a chaperone and forms one of the components of the 70 kilodalton heat shock protein (Hsp70), and is implicated to have function in protein degradation and import of proteins into mitochondria (Sahi *et al*., 2013). When growing on glucose *S. cerevisiae* carries out fermentation with no oxidative phosphorylation and little TCA cycle activity, whereas on glycerol aerobic respiration is used (Wills, 1990). Since aerobic respiration is dependent on processes in the mitochondria, this may explain the growth defect on YPG but not YPD.

In the O megachunk, there were little changes made to the native chromosome sequence; no tRNA genes, introns or transposons removed. Changes to the megachunk O chromosome region include the additions of the PCR tags, loxP sites and stop codon swaps. While possible, it is unlikely that the problem is caused by these changes, as they are not significant changes and occur in every megachunk. It is more likely that the change in phenotype is due to the altered APJ1 protein. Indeed, the deletion or replacement of a small percentage of amino acids

in yeast proteins can disrupt the structure and function significantly. For example, Vamvaca *et al*. (2009) demonstrated that the deletion of just 10 N-terminal amino acids in the lipid-membrane-binding protein α-Synuclein caused significant disruption to the native α-helical structure. Additionally, the propensity of membrane binding dropped dramatically, suggesting that the 10 deleted amino acids were essential for protein folding and function. This demonstrates that the function of proteins can be significantly hindered with the disruption of a small percentage of total amino acids.

There is simple and effective way to check if the *URA3* marker integration, and subsequent *YNL077w* gene disruption, caused the growth defect. The native *YNL077w* gene could be transformed back into one of the O megachunk strains and simultaneously remove the *URA3* marker. Following transformation, colonies would be grown on the *URA3* counter selection, 5-Fluoroorotic Acid (5-FOA), to select for loss of *URA3*. The fitness assay on YPD and YPG plates could be repeated, and if there is no longer a fitness defect, it would prove that the defect was a result of *YNL077w* disruption.

An important design principle of yeast 2.0 specifies that the genome should be highly extensible and able to facilitate a multitude of future experiments. The design of the genome has enabled this principle using two features; the replacement of TAG stop codons, and the SCRaMbLE system. The elements for these two features are found in every megachunk, including the recently integrated, megachunk O of SynXIV. The replacement of TAG stop codons with TAA stop codons facilitates the expansion of the genetic code in yeast 2.0 by the potential to incorporate a twenty-first, unnatural amino acid. The addition of unnatural amino acids to the *S. cerevisiae* genetic code has been previously achieved by engineering tRNA synthetase/tRNA pairs. Variants of the pyrrolysyl-tRNA synthetase/$tRNA_{CUA}^{Pyl}$ pair were used to introduce five unnatural amino acids in *S. cerevisiae* (Hancock *et al*., 2010). Pyrrolysyl-tRNA synthetase is used charge tRNA molecules bound to the amino acid pyrrolysine, preparing for the transfer of pyrrolysine to a growing polypeptide. Pyrrolysyl-tRNA synthetases (PyIRS) were used to introduce unnatural amino acids because of their ability to activate an assortment of pyrrolysine and lysine analogues (Polycarpo *et al*., 2006). The mRNA codon utilised in this system on is UAG, usually the stop codon on eukaryotic organisms which can be re-engineered. Using this approach, the five unnatural amino acids were incorporated into proteins in yeast. Yeast 2.0 facilitates the introduction of unnatural amino acids using methods similar to this and allows for the expression of novel polypeptides.

SCRaMbLE is the defining feature of the yeast 2.0 project. Although the yeast 2.0 project initially builds one version of the synthetic genome, the SCRaMbLE system will enable the generation of many different versions of the genome that vary in structure and gene content (Dymond *et al*., 2011). Even in a small culture containing a billion cells, SCRaMbLE can generate a billion unique genomes. With a simple selection pressure such as elevated temperature, or exposure to alcohol, genomes can be selected that have arrangements which allow them to grow more optimally in these conditions. If a handful of selected colonies are isolated and sequenced, it can give valuable insight into the gene sets that are required for these phenotypes. With this knowledge, it may be possible to rationally design genomes and genetic modules that encode complex phenotypes such as heat tolerance.

The limitation of SCRaMbLE is that it does not induce point mutations in protein encoding genes (Dymond and Boeke, 2012). Promoter-gene cassettes are deleted or shuffled, however proteins cannot not undergo gain of function mutations with structural changes. However, in the future SCRaMbLE could be used in conjunction with traditional mutagenesis techniques such as UV or EMS to induce these mutations. Mutated genes can then be SCRaMbLEd.

When all synthetic chromosomes are complete, it will be the first synthetic eukaryotic organism. The achievement goes far beyond the 582 kb *Mycoplasma genitalium* genome, which was chemically synthesised, assembly and cloned (Gibson *et al*., 2008). By contrast, the changes to the *M. genitalium* genome were minimal, with only the addition of a few 'watermarks'. Furthermore, the synthetically synthesised bacterial genome is simpler and smaller, and lacks any feature to facilitate future novel experiments.

Currently, the SCRaMbLE mechanism is limited to SynIII, the only synthetic chromosome that is complete and available, containing loxP recombination sites that are essential for the SCRaMbLE system. However, SCRaMbLE was designed to be used throughout the entire genome. When SynXIV is complete, it will be another chromosome that can be SCRaMbLEd simultaneously with SynIII.

Chapter 3

Proof of Concept: SCRaMbLEing the *URA3* Marker into Yeast

The overall aim for this thesis was to explore the use of SCRaMbLE to introduce foreign DNA into synthetic chromosomes of *S. cerevisiae*. Before attempting to SCRaMbLE in a large amount of heterologous *T. reesei* DNA, it was essential to first develop and optimise a protocol using homologous DNA. The *S. cerevisiae* marker gene, *URA3*, was chosen to SCRaMbLE into yeast since there would be no issues with intron splicing or promoter compatibility. Overall, SCRaMbLE was successfully utilised to integrate loxP-flanked *URA3* in yeast, while *URA3* without loxP was not SCRaMbLEd.

3.1 METHODS

A significant portion of the work presented from here on is PCR based. Tables 3 and 4 detail the reagents and program used for standard PCR protocols.

Table 3: Reagents used for PCR protocols in this thesis.

Reagent	**Final Concentration**
NEB Phusion® High-Fidelity DNA Polymerase	1.0 U/50 μl (0.02U/ μl)
Phusion® HF Buffer	1X
dNTPs	200 μM
Forward primer	0.5 μM
Reverse primer	0.5 μM
DNA Template	< 250 ng

Table4: Touch down PCR program

Number of cycles	**Temperature (°C)**	**Time**
1	95	4 min
15	95	30 s
	65 (-1 per cycle)	30 s
	72	Variable
10	95	30 s
	50	30 s
	72	Variable
1	4	Hold

Extraction of pRS416 plasmid containing *URA3*

The pRS416 plasmid was used as a control in this study, as it contains the *URA3* marker cassette which is not flanked by loxP. pRS416 was extracted from an *E. coli* DH5α strain. *E. coli* was grown overnight in LB ampicillin, shaking at 37 °C. pRS416 was extracted using the Invitrogen Quick Plasmid Miniprep Kit.

Construction of loxP-*URA3*-loxP DNA fragments

In order to SCRaMbLE *URA3* into SynIII of yeast, the gene requires flanking by loxP (figure 10). The first loxP site of SynXIV (in chunk A1) was amplified using primers that anneal to outside the sequence (figure 10). The PCR reaction was set up as described in table 3, with primers 'loxP A1 FWD' and 'loxP A1 REV' and DNA from megachunk A as template (sequences of all primers referred to in this thesis are listed in table 8, appendix). A touchdown PCR was performed as described in table 4, with an extension time of 15 s. The PCR product was purified using the QIAGEN QIAquick® PCR Purification Kit.

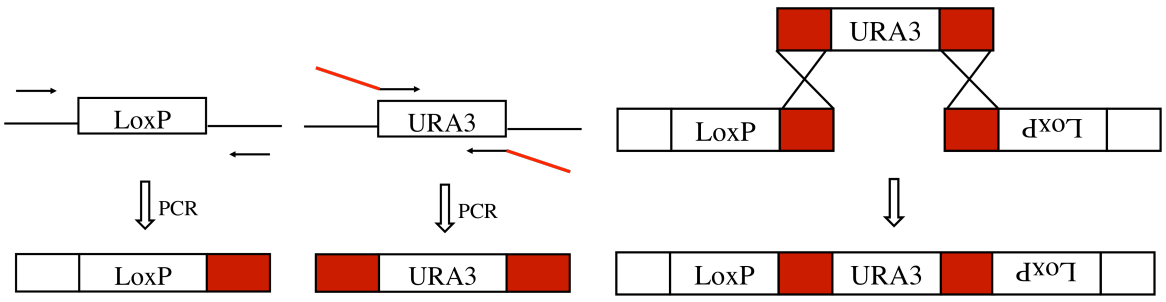

Figure 10: Construction of the loxP-*URA3*-loxP fragment. LoxP was amplified from SynXIV using primers that annealed outside the sequence. *URA3* was amplified from the pRS416 plasmid using primers that had 5' extensions homologous to the 3' flanking region of loxP. loxP-*URA3*-loxP was constructed using Gibson assembly which recombined homologous sequences (represented in red).

To amplify *URA3* from the pRS416 plasmid, primers were designed so that the 5' extensions were complementary to the 3' flanking sequence of the amplified loxP fragment (figure 10). The PCR reaction was set up as described in table 3, with primers 'URA3 FWD' and 'URA3 REV' and pRS416 as template. A touchdown PCR was performed as described in table 4, with an extension time of 30 s. The product was purified using the QIAGEN QIAquick® PCR Purification Kit. *URA3* was further amplified by using a primer that anneals to both the 5' and 3' ends. The PCR reaction was set up as described in table 3, with 'WGA primer' and the

PCR product containing *URA3* as template. A touchdown PCR was performed as described in table 4, with an extension time of 30 s. The PCR product was purified as above.

Construction of loxP-*URA3*-loxP fragments was performed using a 2:1 ratio of loxP and *URA3* fragments respectively. Fragments were incubated with 1X NEBuilder® HiFi DNA Assembly Master Mix at 50 °C for 1 h. LoxP-*URA3*-loxP fragments were PCR amplified using the loxP forward primer, which anneals to both ends of loxP-*URA3*-loxP, as the 3' loxP is inverted (figure 10). The reaction was set up as described in table 3, with primer 'loxP A1 FWD' and the Gibson assembly mixture as template. A touchdown PCR was performed as described in table 4, with an extension time of 30 s. The PCR product was purified using the QIAGEN QIAquick® PCR Purification Kit.

Construction of a diploid yeast strain for importing DNA using SCRaMbLE

A diploid yeast strain was constructed with two copies of each wild type chromosome, besides chromosome III, of which it contained one wild type copy and one synthetic copy (SynIII). The strain was constructed to decrease viability loss generated by SCRaMbLEing SynIII. In order to induce SCRaMbLE, the strain requires the Cre-EBD plasmid. The plasmid contains an engineered Cre recombinase fused to the murine estrogen-binding domain (EBD) (Dymond *et al*., 2011). This facilitates Cre recombinase expression upon estradiol exposure.

YPD media were inoculated with yeast haploids BY4742 (MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0) and BY4741 (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 containing SynIII chromosome), and incubated overnight shaking at 30 °C. An equal number of cells from each medium were inoculated into one YPD aliquot and incubated overnight shaking at 30 °C. The BY4742 and BY4741 yeast alone, as well as the mixed culture, were streaked onto –lys –met plates to select for cells with double auxotrophies originating from both haploid strains (his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 met15Δ0).

The Cre-EBD plasmid containing a *HIS3* marker was transformed into the diploid strain to allow for SCRaMbLEing through expression of Cre recombinase. A diploid colony was re-streaked on a –lys –met plate in 2x2 cm square and incubated at 30 °C for 18 hours. Cells were harvested into 1 ml of sterile water, spun down and resuspended in a transformation mix containing 30% PEG, 0.1M LiOAc, 0.28mg/ml ss carrier DNA, and 840 ng/ml of Cre-EBD plasmid. Cells were incubated at 42 °C for 1.5 hours, then spun down and resuspended in sterile water. Cells were plated onto –lys –met –his plates and incubated for 48 hours at 30 °C.

A PCR was performed to confirm the diploid genotype. DNA from the diploid strain and both BY4742 and BY24741 strains was extracted using the Epicenter Master Pure Yeast DNA Purification Kit. DNA samples were PCR amplified with primers that annealed to BY4742 or BY4741 sequences. Three primers were designed; one binding to sequences in both BY4741 and 4742 (MAT-locus REV), one specific to the Mat A locus (MAT-A FWD) and one specific to the Mat α locus (MAT-α FWD). Each of the two primer pairs was tested against the three DNA samples. The PCR reactions were set up using 1X Promega GoTaq® DNA Polymerase Master Mix, 0.5 μM of each primer and <250ng of purified yeast DNA. A touchdown PCR was performed as described in table 4, with an extension time of 45 s.

Transformation of loxP-*URA3*-loxP and pRS416 into diploids and haploids

The transformation performed was a variation of one previously described (Gietz and Schiestl, 2007). 768 ng loxP-*URA3*-loxP DNA or 5.06 ng pRS416 plasmid (containing *URA3* not flanked by loxP) were used in each transformation. The constructed diploid strain and a haploid strain (BY4741 containing SynIII and the Cre-EBD plasmid) were used for transformation. Following heat shock for 30 min, transformation media were replaced with YPD. SCRaMbLE was induced immediately following transformation by exposure to 1 μM estradiol (added to YPD), and cells were incubated at 30 °C, shaking for 3 hours. Media were replaced with sterile water and cells were plated onto two –ura plates, incubating at 30 °C for 3 days.

3.2 RESULTS

The potential of SCRaMbLE to be used to integrate DNA into synthetic yeast chromosomes was explored using the *S. cerevisiae URA3* marker gene. The *URA3* marker was flanked by loxP to generate loxP-*URA3*-loxP fragments. These fragments were SCRaMbLEd into diploid and haploid yeast by inducing the system with estradiol immediately following transformation. The pRS416 plasmid was also transformed and SCRaMbLEd as a control, as it contains the *URA3* gene and is not flanked by loxP. Using the diploid strain, the non-SCRaMbLEd loxP-*URA3*-loxP samples had significantly lower transformation efficiency than the SCRaMbLEd samples. In contrast, the pRS416 plasmid showed no significant difference in the SCRaMbLE and no-SCRaMbLE conditions using haploids or diploids. The haploid strain showed a small difference between SCRaMbLEd and non-SCRaMbLEd loxP-*URA3*-loxP.

PCR confirmation of diploid genotype

The genotype of the constructed diploid strain was confirmed by PCR. As shown in figure 11, BY4741 DNA was amplified with the primer pair specific to the MAT A locus while BY4742 DNA was amplified with the primer pair specific to the MAT α locus. Both pairs of primers amplified diploid DNA, confirming that it was a diploid strain.

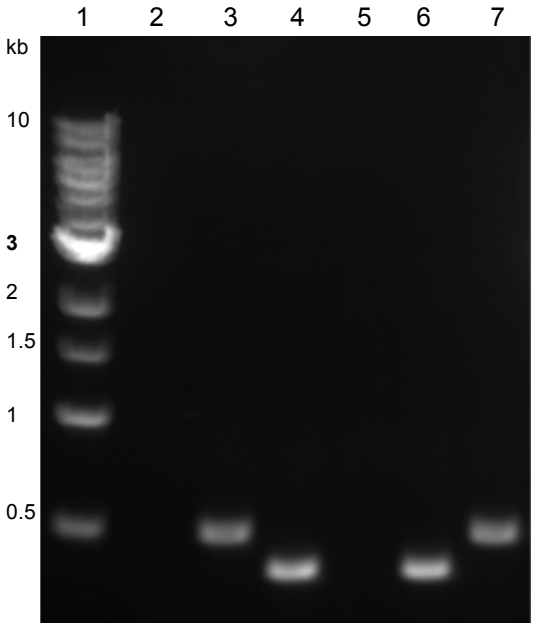

Figure 11: Agarose gel electrophoresis showing PCR confirmation of diploid genotype. Lane 1: NEB 1 kb DNA ladder, lane 2: BY4741 DNA amplified with MAT α primer pair, lane 3: BY4741 DNA amplified with MAT A primer pair, lane 4: BY4742 DNA amplified with MAT α primer pair, lane 5: BY4742 DNA amplified with MAT A primer pair, lane 6: diploid DNA amplified with MAT α primer pair, lane 7: diploid DNA amplified with MAT A primer pair.

LoxP attachment to the *URA3* marker gene

The PCR amplification of *URA3* and loxP-*URA3*-loxP is shown in figure 12. Lane 2 of image (a) shows the loxP fragment at just over 1 kb in size. The *URA3* gene is 1.1 kb, which suggested successful amplification. Lane 2 of image (b) shows the loxP-*URA3*-loxP fragment at almost 1.5 kb in size. The fragment was expected to be 1.4 kb (1.1 kb *URA3* + 2 x 150 bp loxP fragments).

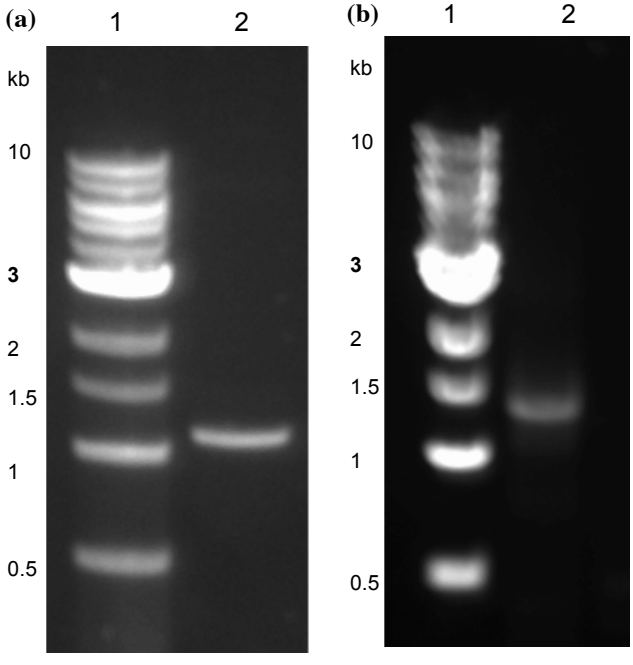

Figure 12: (a) Agarose gel electrophoresis of *URA3* DNA after second round of PCR amplification. Lane 1: NEB 1 kb DNA ladder, lane 2: PCR amplified *URA3* using the 'WGA primer'. **(b)** Agarose gel DNA electrophoresis of loxP-*URA3*-loxP DNA following Gibson Assembly and PCR amplification. Lane 1: NEB 1 kb DNA ladder, lane 2: PCR amplified loxP-*URA3*-loxP using primers 'loxP A1 FWD' and 'loxP A1 REV'.

SCRaMbLE-mediated integration of loxP-*URA3*-loxP DNA into yeast

Following transformation and SCRaMbLEing of DNA into diploid yeast, there were significantly more colonies on –ura plates in the loxP-*URA3*-loxP SCRaMbLE condition compared with the no-SCRaMbLE condition (figure 13). The transformation efficiency was on average 44-fold higher in the loxP-*URA3*-loxP SCRaMbLE condition (table 5). In contrast, the difference between the pRS416 (*URA3* not flanked by loxP) SCRaMbLE and no SCRaMbLE conditions was negligible, with a transformation efficiency 1.07-fold higher in the pRS416 no-SCRaMbLE group (figure 14, table 5). Results using haploid yeast showed a much smaller difference between the loxP-*URA3*-loxP conditions. The transformation efficiency was 2.14-fold higher in the SCRaMbLE condition than the no-SCRaMbLE condition. Using pRS416 DNA, the transformation efficiency was 1.12-fold higher in the no-SCRaMbLE condition than the SCRaMbLE condition.

Table 5: Transformation efficiency of *URA3* integration into *S. cerevisiae* using SCRaMbLE

	loxP-*URA3*-loxP SCRaMbLE	loxP-*URA3*-loxP no SCRaMbLE	pRS416 SCRaMbLE	pRS416 no SCRaMbLE
Diploids	1.55×10^3	3.5×10	7.11×10^4	7.63×10^4
Haploids	1.31×10^2	6.1×10	3.52×10^4	3.94×10^4

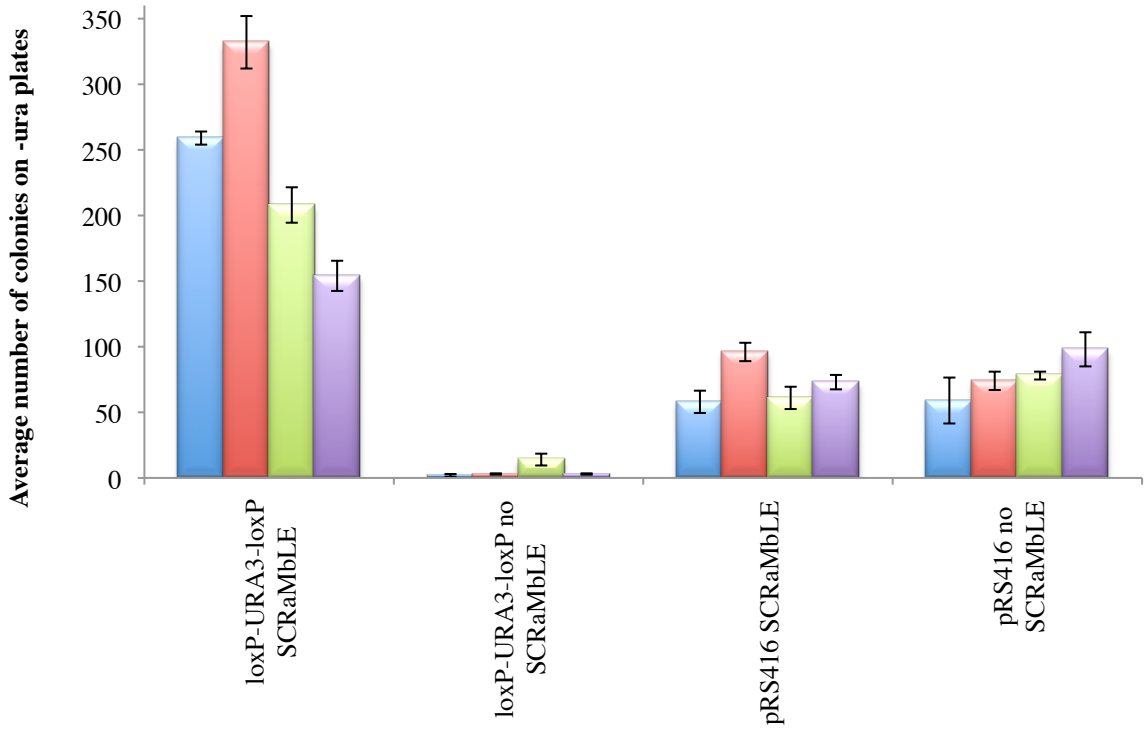

Figure 13: Average number of diploid yeast colonies growing over two –ura plates. Yeast were transformed with either loxP-*URA3*-loxP DNA or pRS416 plasmid DNA and were SCRaMbLEd for 3 hours or not SCRaMbLEd. Blue, red, green and purple columns represent four biological replicates performed as separate experiments.

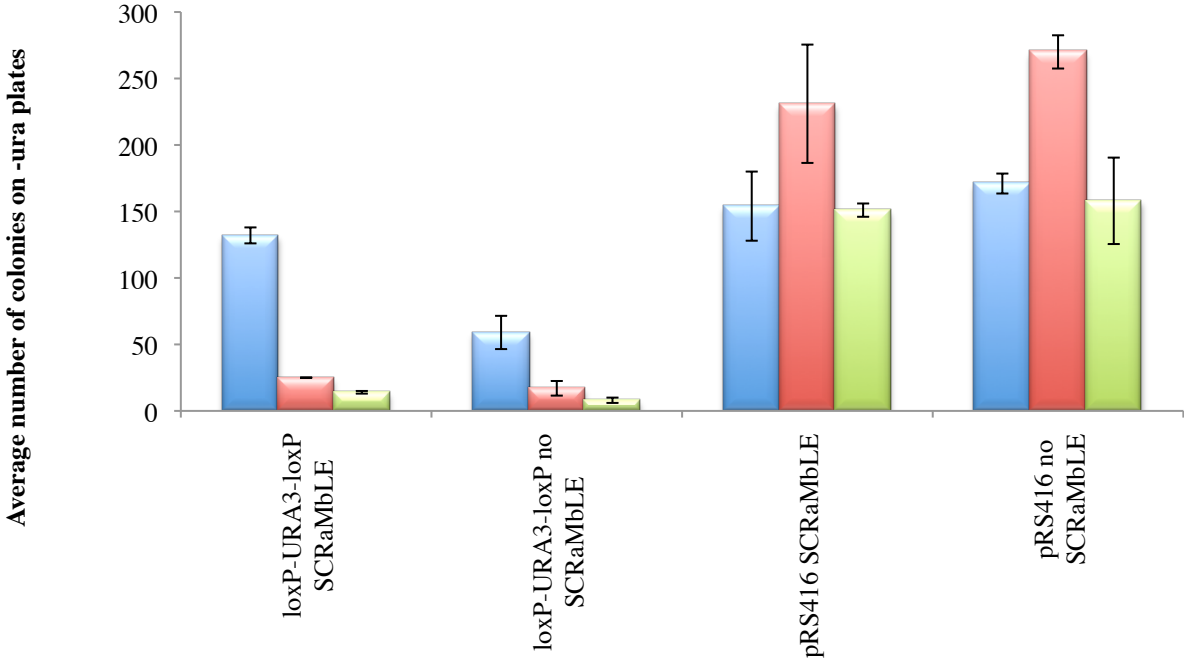

Figure 14: Average number of haploid yeast colonies growing over two –ura plates. Yeast were transformed with either loxP-*URA3*-loxP DNA or pRS416 plasmid DNA and were SCRaMbLEd for three hours or not SCRaMbLEd. Blue, red and green columns represent three biological replicates performed as separate experiments.

Overall loxP-flanked DNA was successfully SCRaMbLEd into *S. cerevisiae*, and resulted in a higher transformation efficiency compared to when DNA was not SCRaMbLEd. The results of this chapter demonstrate that DNA must be flanked by loxP for SCRaMbLEing to occur.

DISCUSSION

SCRaMbLE is the novel genome rearrangement system designed for the synthetic genome of *S. cerevisiae*. SCRaMbLE here was used in a new application; to introduce DNA into yeast chromosomes. Before attempting to SCRaMbLE in heterologous DNA, it was essential to develop protocols for the SCRaMbLEing of a single homologous gene. SCRaMbLE was used to integrate loxP-flanked *URA3* into SynIII of yeast. Using diploids, a significant increase in transformation efficiency occurred following the SCRaMbLEing of loxP-flanked *URA3*, compared to the non-SCRaMbLEd equivalent. This difference was absent using the non-flanked *URA3* of pRS416. Furthermore, there was a significant difference between SCRaMbLEing diploid and haploids; the increase in transformation efficiency for SCRaMbLEd loxP-flanked *URA3* was significant when using diploids, but was much smaller with haploids.

SCRaMbLE causes loss of viability due to the generation of lethal genotypes (Dymond *et al*., 2011). A diploid strain was constructed to minimise loss of viability, as diploid strains can maintain recessive lethal mutations (Wloch *et al*., 2001, Eyre-Walker and Keightley, 2007). The strain was constructed by mating the haploid strains BY4741 and BY4742, and had two of each wild type chromosome besides chromosome III, of which it had one wild type and one synthetic. During SCRaMbLEing, wild type chromosome III remains unSCRaMbLEd. Therefore, it was hypothesised that in the case that lethal genotypes were generated by SCRaMbLEing SynIII, the wild type chromosome III would remain in its original configuration, allowing cells to remain viable. It was predicted that there would be greater loss of viability in experiments using haploids compared to using diploids.

When using diploids in this experiment, results were as expected. The transformation efficiency was 44-fold higher in the loxP-*URA3*-loxP SCRaMbLE condition compared to the non-SCRaMbLE condition (table 5). In contrast, there was a 1.07-fold increase in transformation efficiency in the pRS416 no-SCRaMbLE group compared to the SCRaMbLE group, demonstrating that loxP is required for SCRaMbLEing.

Using haploids, it was expected that there would be a higher loss of viability due to SCRaMbLE in haploid cells compared to the diploid cells. Results were as expected in using loxP-*URA3*-loxP; the increase in transformation efficiency between SCRaMbLEd and non-SCRaMbLEd samples were much higher in diploids than in haploids and this is assumed to be due to loss of viability. However, it was interesting that when using pRS416, the same trend was absent. If loss of viability were the sole cause, there would have been many less colonies in the pRS416 SCRaMbLE condition compared to the non-SCRaMbLE condition, since SCRaMbLEing still occurs in SynIII. An explanation for this is that the transformation of loxP-*URA3*-loxP prior to SCRaMbLE provided many more loxP sites in SynIII, and therefore facilitated more SynIII rearrangements than without these extra loxP sites. These extra rearrangements can explain the loss of viability with loxP-*URA3*-loxP transformed haploids, and not the pRS416 haploids, since the plasmids did not contain any loxP sites.

The theory that extra loxP sites (introduced via transformation) may facilitate more potential genome arrangements could be explored by testing the viability of yeast on non-selective media (*e.g.* YPD). Since SCRaMbLE produces lethal phenotypes, the more rearrangements that are generated, the more lethal the system becomes. Therefore, a study could be performed where various numbers of loxP-*URA3*-loxP molecules are transformed and subsequently SCRaMbLEd into yeast. A dot assay on YPD could reveal if more transformed loxP-flanked fragments facilitate more possible chromosome rearrangements and subsequent, a higher loss of viability.

In the protocol optimised for SCRaMbLEing in DNA, SCRaMbLE was induced for three hours immediately following a 30 minute heat shock transformation. In a previous protocol, transformation and SCRaMbLE were performed simultaneously in diploid cells; estradiol was added to the transformation mixture and cells were heat shocked and simultaneously SCRaMbLEd for three hours. The results for simultaneous transformation and SCRaMbLE were similar to the results using haploid cells, as SCRaMbLEd yeast did not show a large increase in transformation efficiency compared to non-SCRaMbLEd yeast (data not shown). The reason that this protocol produced suboptimal results is that the promoter regulating Cre-EBD expression, the SCW11 promoter, is expressed only in daughter cells. During heat-shock, cells are stressed and may not divide as they would in YPD media. Therefore, the generation of daughter cells may be at a much lower rate and Cre-EBD expression (and therefore SCRaMbLE induction) would be much lower throughout the population. Therefore, loxP-

URA3-loxP is SCRaMbLEd at a much lower rate resulting in a lower transformation efficiency.

To explain the mechanism resulting in increased transformation efficiency for SCRaMbLEd loxP-flanked *URA3* compared to non-SCRaMbLEd, I propose that immediately following transformation, DNA has entered the nucleus but has not yet been integrated into the chromosomes. It is possible that many fragments never integrate into the chromosome and are ultimately degraded by nucleases. During SCRaMbLE, Cre recombinase integrates loxP-flanked DNA fragments into synthetic chromosomes that perhaps would otherwise be degraded. This theory explains how SCRaMbLEing after transformation produces many more colonies compared to transformation alone. It is important to also note that since the loxP sequence is too short for homologous recombination, SCRaMbLE does not work simply by providing more places for transforming DNA to integrate (Dymond *et al*., 2011).

It would be valuable to determine the effectiveness of SCRaMbLE when the system is used at different time points following transformation. In the protocol developed here, SCRaMbLE was induced immediately after transformation. In the future, cells could be incubated in YPD in 30 minute increments following transformation (prior to SCRaMbLE induction) to investigate the effect. It was also hypothesised that transformed DNA resides in the nucleus before either integration into the chromosomes or degradation. This method can be used to investigate this theory by suggesting how long DNA resides in the nucleus, prior to integration or degradation.

Overall, this study was a proof of concept, demonstrating that the novel evolution system SCRaMbLE can be utilised to import DNA into synthetic yeast chromosomes. Specifically, the *S. cerevisiae URA3* gene was flanked with loxP sequences and was integrated at a much higher rate than *URA3* fragment that were not flanked with loxP. In the future, the technique developed here could be adapted to insert multiple copies of PCR amplified genes for specific enzymes, for example those involved in a particular metabolic pathway.

Chapter 4

PCR Approaches to Attach Recombination Sequences to *T. reesei* DNA

In chapter 3, SCRaMbLE was used to integrate the homologous *URA3* marker into diploid yeast containing one copy of SynIII. In order to further explore the potential of SCRaMbLE for the integration of DNA into yeast, protocols were developed for the integration of large amounts of heterologous DNA. In order to SCRaMbLE random fragments of DNA from the *T. reesei* genome (or any other genome) into yeast, they first must be flanked by loxP recombination sequences. A number of PCR-based protocols were developed to achieve this. PCR-based methods were chosen to use over traditional cloning methods, as DNA fragments can be easily and rapidly engineered with the right primers and PCR program. If successful, loxP sites could be added to random fragments of *T. reesei* gDNA in a single step. This chapter describes in detail three protocols in the methods and results section. Many more protocols were developed, some of which are discussed later.

4.0 METHODS

Isolation of *Trichoderma reesei* gDNA used for each of the three methods

Each of the below three protocols require pure *T. reesei* gDNA. 10^7 *T. reesei* (RUT-C30 strain) spores were plated onto PDA plates (30 g/l potato dextrose agar) covered with sterile cellophane disks and incubated for 5 days at 28 °C. Spores were harvested, and crushed with a mortar and pestle with liquid nitrogen to obtain a homogenous powder. gDNA was extracted using the Sigma-Aldrich GenElute™ Plant Genomic DNA Miniprep Kit.

<u>Protocol 1</u>: Addition of loxP sequences to random fragments of *T. reesei* gDNA using a degenerate primer

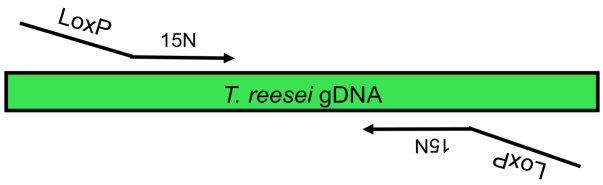

Figure 15: Random amplification of *T. reesei* gDNA using a degenerate primer with loxP sequence as the 5' extension.

This method used a PCR primer which annealed randomly to *T. reesei* gDNA by a 15 nucleotide degenerate sequence where each position could be A, T, C or G (figure 15). The 5' primer extension contained the loxP sequence, as well as the primer-binding site 'M13F' for re-amplification. The PCR reaction was set up as described in table 3, with primer 'M13-

loxP-15N', and *T. reesei* gDNA as template. A touchdown PCR was performed as described in table 4, with an extension time of 5 min.

<u>Protocol 2</u>: Amplification of loxP from SynXIV, and preparation of random *T. reesei* gDNA fragments for Gibson assembly with loxP

In this method loxP fragments were amplified from SynXIV (figure 16). The PCR reaction was set up as described in table 3, with primers 'loxP A1 FWD' and 'loxP A1 REV', and DNA from SynXIV megachunk A as template. A touchdown PCR was performed as described in table 4, with an extension time of 15 sec. Random fragments of *T. reesei* gDNA were amplified using a primer with a 15 nucleotide degenerate sequence (figure 16). The 5' extension was homologous to the 3' flanking region of loxP from chunk A1. The PCR reaction was set up as described in table 3, with primer 'chXIV loxP 5' flank-15N', and *T. reesei* gDNA as template. A touchdown PCR was performed as described in table 4, with an extension time of 5 min.

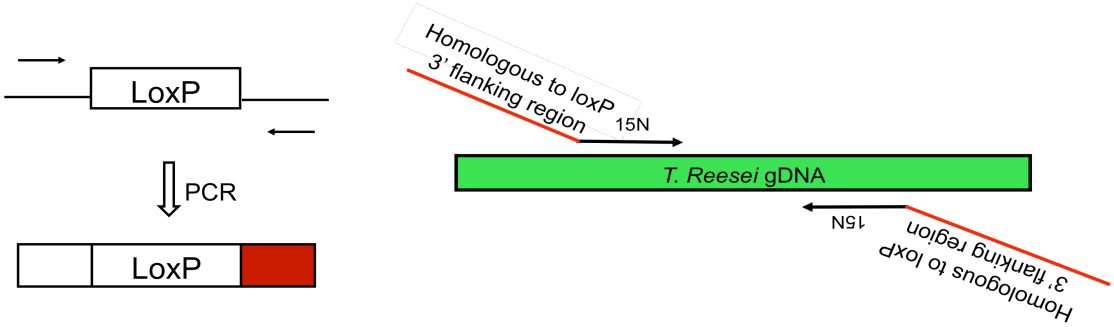

Figure 16: LoxP was amplified from SynXIV using primers that annealed ~150bps outside the loxP sequence, creating flanking fragments on each side. *T. reesei* DNA was amplified using degenerate primers. The 5' extension was a sequence homologous to the 3' flanking region of loxP. Homologous sequences are represented in red.

LoxP-gDNA-loxP were Gibson assembled using a 2:1 ratio of loxP and *T. reesei* gDNA fragments respectively. DNA fragments were incubated with 1X NEBuilder® HiFi DNA Assembly Master Mix at 50 °C for 1 h. LoxP-gDNA-loxP fragments were PCR amplified as described in table 3, with primer 'loxP A1 FWD', and the Gibson assembly mixture as template. A touchdown PCR was performed as described in table 4, with an extension time of 5 min.

<u>Protocol 3</u>: Restriction digestion of *T. reesei* gDNA and loxP and subsequent ligation

***T. reesei* gDNA *Hind*III incomplete digestion**

gDNA from *T. reesei* was partially digested as follows; <1 ng *T. reesei* gDNA, 0.4 U/ 50 µl NEB *Hind*III, 1X NEBuffer 2.1, incubated at 37 °C for 10 min. DNA was purified using the QIAGEN QIAquick® PCR Purification Kit

Digestion of loxP fragments with *Hind*III recognition sites at the 3' end

In order to attach *Hind*III restriction sites to the 3' ends of loxP, a primer pair was used where the reverse primer had a 5' extension which included the *Hind*III restriction site (figure 17). These primers annealed to ~150 bp either side of the first loxP site in SynXIV. The PCR reaction was set up as described in table 3, with primers 'loxP A1 FWD' and 'loxP A1 REV *Hind*III extension', and SynXIV megachunk A DNA as template. A touchdown PCR was performed as described in table 4, with an extension time of 15 s. A PEP-PCR was also set up as described by Arneson *et al*. (2008b) (table 9, appendix for PCR program). The PCR product was purified using the QIAGEN QIAquick® PCR Purification Kit.

The loxP fragments, with 3' *Hind*III restriction sites, were digested by 0.4 U/ 50 µl NEB HindIII in 1X NEBuffer 2.1, and incubated at 37 °C for 1 h. DNA was purified using the QIAGEN QIAquick® PCR Purification Kit. LoxP fragments were dephosphorylated using 0.2 U/µl NEB CIP in 1X CutSmart® Buffer and incubated for 5 min at 37 °C. DNA was purified using the QIAGEN QIAquick® PCR Purification Kit.

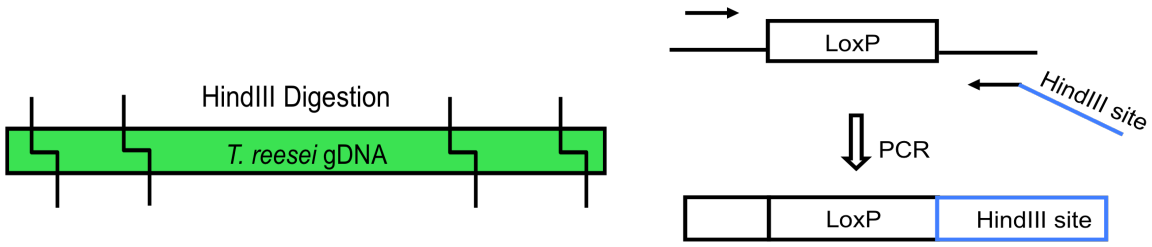

Figure 17: Incomplete *Hind*III digestion of *T. reesei* gDNA and the addition of *Hind*III restriction recognition sites to 3' ends of loxP. *Hind*III recognition sites were added using a primer with the sequence in the 5' extension. LoxP sequences were then digested.

Preparation of loxP-gDNA-loxP fragments

A ligation was carried out with *Hind*III digested loxP and *Hind*III incompletely digested *T. reesei* gDNA using 1 U/µl NEB T4 DNA Ligase, 1X NEB T4 DNA Ligase Reaction Buffer, <1 µg *Hind*III incompletely digested *T. reesei* gDNA and <1 µg *Hind*III digested loxP. The

reaction was incubated at 16 °C overnight. In order to amplify gDNA fragments that had been flanked with loxP fragments, the forward primer only was used to selectively amplify only these fragments. The PCR reaction was set up as described in table 3, with primer 'loxP A1 FWD' and the loxP/gDNA ligation mix as template. A touchdown PCR was performed as described in table 4, with an extension time of 8 min.

The PCR products were run on a 1% agarose gel at 80V for 1 hour. A band at >3 kb was excised and DNA was purified using the Zymo Research Zymoclean™ Gel DNA Recovery Kit. A PCR was performed to further amplify loxP-gDNA-loxP fragments using the forward primer of loxP only. The PCR reaction was set up as described in table 3, with primer 'loxP A1 FWD' and the loxP-gDNA-loxP PCR product as template. A touchdown PCR was performed as described in table 4, with an extension time of 8 min.

RESULTS

In order to explore SCRaMbLE to introduce heterologous DNA into yeast, foreign DNA must first be flanked with loxP. A series of PCR-based approaches were developed in this chapter in order to attach loxP sequences to random fragments of *T. reesei* gDNA.

The addition of loxP sequences to random fragments of *T. reesei* gDNA using a degenerate primer was unsuccessful due to the primer not amplifying *T. reesei* gDNA. Figure 18 shows the 'M13-loxP-15N' degenerate primer amplification of *T. reesei* gDNA. Lanes 2 and 3 show *T. reesei* gDNA amplification and the negative control respectively, with both lanes showing primer dimers below the 0.5 kb ladder band. The protocol was attempted with a range of modifications (discussed later); each time primer dimers appeared.

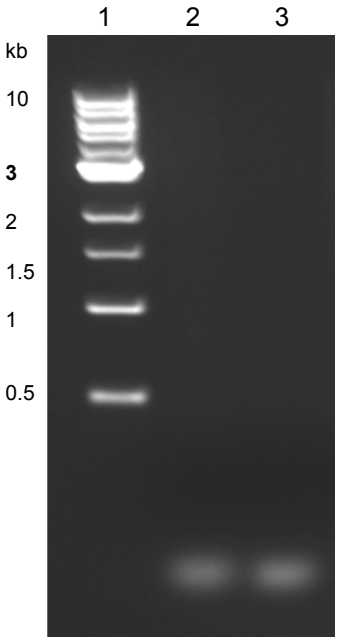

Figure 18: Agarose gel electrophoresis of M13-loxP-15N primer PCR amplified *T. reesei* gDNA. Lane 1: NEB 1 kb DNA Ladder, lane 2: *T. reesei* gDNA amplified using M13-loxP-15N primer, lane 3: negative control using no template DNA.

In the second protocol, loxP was successfully amplified from SynXIV with flanking regions on either side, but the preparation of random fragments of *T. reesei* gDNA for Gibson assembly was unsuccessful. Figure 19 lanes 2 and 3 show the PCR amplification of loxP from SynXIV using primers that annealed outside the loxP sequence. As a control, the primers were used to amplify the equivalent sequence from the wild type chromosome XIV. Since this sequence lacks the 34 bp loxP sequence, it ran slightly lower down the gel, seen in lane 4.

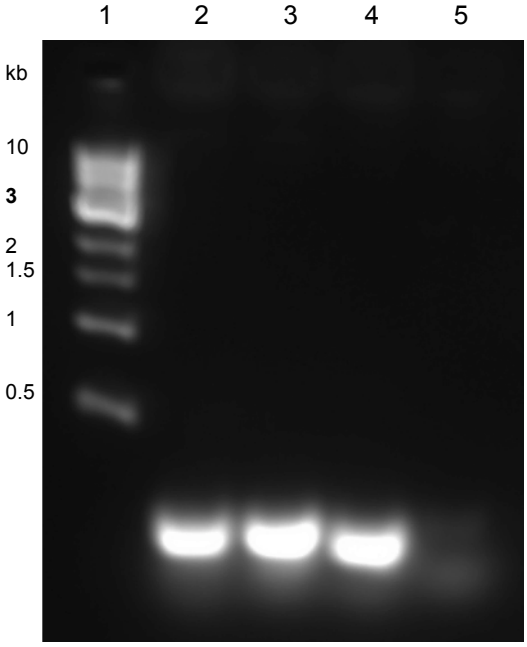

Figure 19: Aragorse gel electrophoresis of PCR amplified loxP. Lane 1: NEB 1 kb ladder, lanes 2 and 3: loxP amplified from SynXIV using primers 'loxP A1 FWD' and 'loxP A1 REV', lane 4: DNA fragment amplified from wild type DNA using primers 'loxP A1 FWD' and 'loxP A1 REV'.

As part of the second protocol, figure 20 shows PCR amplification using the 'chXIV loxP 5' flank-15N' primer, which was designed to add sequences to random fragments *T. reesei* gDNA. These sequences were homologous to the 3' flanking region of loxP which was amplified from SynXIV. Image (a) shows primer dimers at the ends of each lane, including the negative controls. Lanes 5 and 6 are PCR products using the PEP PCR protocol smear from ~1 kb to the primer dimer bands. Image (b) shows re-amplification of DNA which was cut out and purified from the gel in image (a). Lanes 2 and 3 show 'WGA primer' amplification of bands ~4-8 kb excised from lanes 5 and 6 respectively. Lane 4 shows 'WGA primer' amplification of a ~1 kb band excised from lane 5. Lane 5 is the no DNA control. All lanes show amplification at ~0.5 kb and down. The smear in lane 5 starts and finishes earlier.

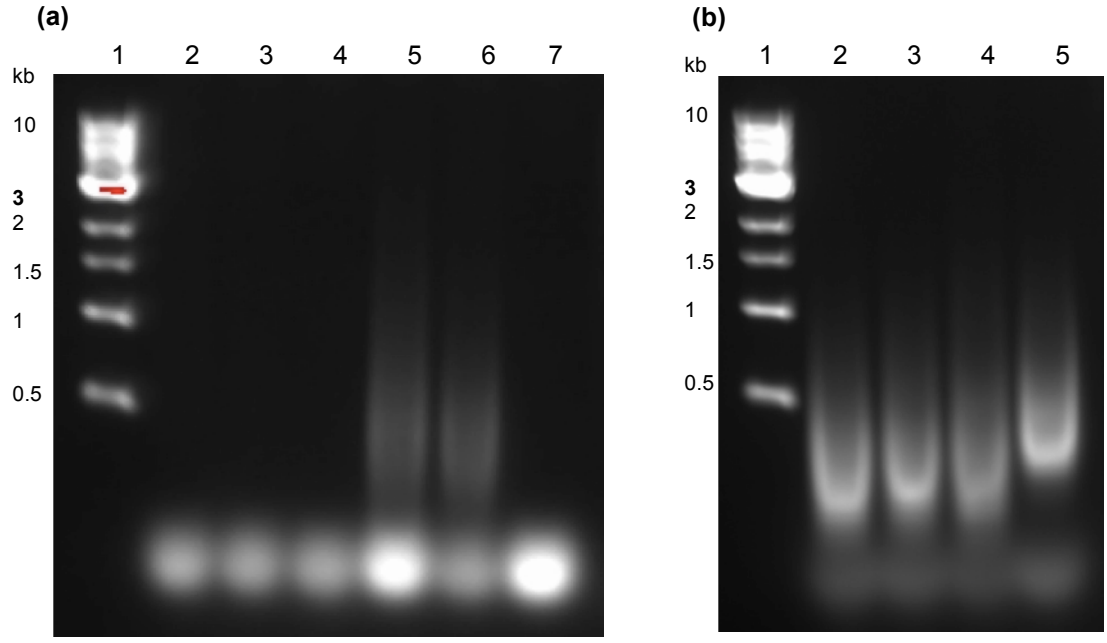

Figure 20: (a) Agarose gel electrophoresis of 'chXIV loxP 5' flank-15N' primer PCR amplified *T. reesei* gDNA. Lane 1: NEB 1 kb DNA Ladder, lane 2: *T. reesei* gDNA with DMSO (dimethyl sulfoxide) (touch down PCR), lane 3: *T. reesei* gDNA (touch down PCR), lane 4: no template (touch down PCR), lane 5: *T. reesei* gDNA with DMSO (PEP PCR), lane 6: *T. reesei* gDNA (PEP PCR,) lane 7; no template (PEP PCR). **(b)** re-amplification of DNA which was cut out and purified from the gel in image. Lane 1: NEB 1 kb DNA Ladder, lanes 2 and 3: 'WGA primer' amplification of bands ~4-8 kb, lane 4: 'WGA primer' amplification of a ~1 kb band, lane 5: no DNA control.

Using the third protocol involving the restriction digestion of both *T. reesei* gDNA and loxP, and subsequent ligation, the fragments were successfully prepared, but ligation was unsuccessful. A partial digestion of *T. reesei* gDNA was achieved using a 10 minute incubation with *Hind*III, shown in figure 21 (a) lane 2. This sample was used for ligation as a band of intact gDNA was still visible, suggesting incomplete digestion. *Hind*III recognition sites were successfully attached to the 3' ends of loxP fragments shown in figure 21 (b) lanes 2 and 3. The bands representing loxP with 3' *Hind*III recognition site are slightly higher on the gel compared to loxP alone, due to being slightly longer in length.

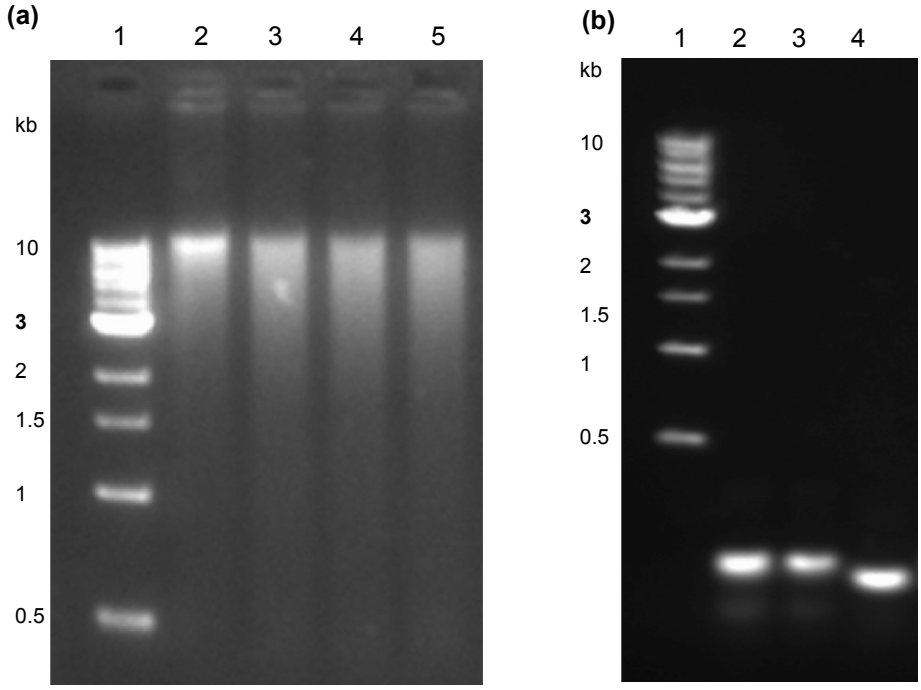

Figure 21: (a) Agarose gel electrophoresis showing partial digest of *T. reesei* gDNA using *Hind*III. Lane 1: NEB 1 kb DNA Ladder, lane 2: 10 min *Hind*III digestion of *T. reesei* gDNA, lane 3: 30 min *Hind*III digestion of *T. reesei* gDNA, lane 4: 1 hour *Hind*III digestion of *T. reesei* gDNA, lane 5: 2 hour *Hind*III digestion of *T. reesei* gDNA. **(b)** Agarose gel electrophoresis showing addition of *Hind*III site to loxP. Lane 1: NEB 1 kb DNA Ladder, lane 2 and 3: loxP with *Hind*III site amplified using 'loxP A1 FWD' and 'loxP A1 REV *Hind*III extension' primers, lane 4: loxP without *Hind*III site amplified using 'loxP A1 FWD' and 'loxP A1 REV' primers.

After ligation of *Hind*III digested *T. reesei* gDNA and loxP, loxP-gDNA-loxP fragments were amplified using a primer than annealed to only side of loxP. Figure 22 (a) shows bright bands below 0.5 kb in lanes 2 and 3 and also smears that extend from these bands to ~3 kb. Gel bands were excised from these lanes at ~1 kb and ~3-8 kb and were PCR re-amplified using the same primer. The product of PCR is shown in image (b) and the expected product sizes (~1 kb and ~3 kb) are absent.

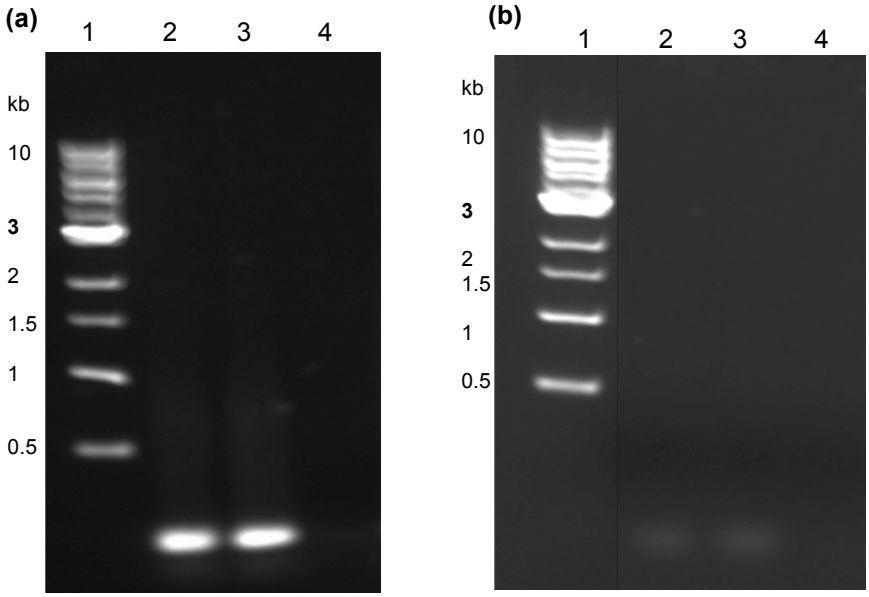

Figure 22: (a) Agarose gel electrophoresis showing PCR amplification of loxP-gDNA-loxP using 'loxP A1 FWD' primer. Lane 1: NEB 1 kb DNA Ladder, lane 2 and 3: PCR amplification of loxP-gDNA-loxP, lane 4: PCR negative control. **(b)** Agarose gel electrophoresis showing PCR re-amplification of loxP-gDNA-loxP purified from gel slices in image (a) using primer 'loxP A1 FWD'. Lane 1: NEB 1 kb DNA Ladder, lane 2: PCR amplification of DNA purified from ~3-8 kb gel slice, lane 3: PCR amplification of DNA purified from ~1 kb gel slice, lane 4: no DNA control.

The third protocol was promising as there were smears in the gel after the first round of loxP-gDNA-loxP amplification. However, after a second round of amplification, the protocol was ultimately unsuccessful as the expected bands were absent.

DISCUSSION

The SCRaMbLEing of heterologous DNA first requires that fragments are flanked with loxP recombination sequences. This chapter details a number of PCR-based methods that were implemented in order to attach loxP sequences to random fragments of *T. reesei* gDNA. Thus far, the developed protocols remain unsuccessful.

The use of degenerate primers is a simple and inexpensive way to amplify random fragments of DNA that has been used for decades. In an early example, Bohlander *et al.* (1992) used degenerate primers for the successful sequence-independent amplification of chromosomal material. The primers used in this chapter have a relatively long stretch of random nucleotide sequence. The sequences used were at least 15 bp long to increase the probability that longer *T. reesei* gDNA fragments would be amplified. If the random sequence was short, it would

bind to more locations on the genome and produce fragments of shorter length. The ideal length of the randomly amplified fragments was at or above 2-3 kb, to conserve protein coding genes, *e.g.* those involved in cellulase activity.

The first method of loxP attachment to random fragments of *T. reesei* gDNA using a degenerate primer was unsuccessful because of primer dimers. Primer dimers formed because the loxP sequence is palindromic, and primers are able to bind each other. In order to avoid this, there were multiple modifications made to the protocol. A higher concentration of gDNA was used to increase the chance of primer annealing to *T. reesei* gDNA instead of to other primers. The NEB Phusion® GC Buffer and DMSO were utilised at appropriate concentrations as they facilitate amplification of G/C rich regions. To confirm if *T. reesei* fragments were flanked with loxP, the initial PCR products were amplified with the M13F primer, however a smear was not seen on the gel. All attempts resulted in primer dimers and it was concluded that a new protocol must be developed. LoxP sequences cannot be added in the form of primer extensions, because single stranded loxP sequences (generated in the denaturation phase of PCR) form dimers.

Primer dimers were unavoidable when loxP sequences were part of the primer sequence. Therefore the in second method, degenerate primers were utilised with 5' extensions that were homologous to the 5' region of loxP. The loxP sequence was successfully amplified from SynXIV; primers annealed to outside of the loxP sequence to provide buffer regions to prevent the possibility of dimer formation. The issue here was optimising the PCR method for amplification of *T. reesei* gDNA. The touch down method resulted in DNA bands under 0.5 kb on the gel. The primer extension preamplification (PEP) PCR method was also utilised which is designed for use with a completely degenerate primer (Arneson *et al.*, 2008b). It was interesting that this PCR protocol resulted in a smear (although, below 3 kb). After cutting out from the smear at 4-8 kb and 1 kb and PCR re-amplification using the WGA primer (anneals to the constant 5' region generated in the initial PCR), the smears appeared identical, at below 3 kb. The expected bands at 4-8 kb and 1 kb were absent suggesting that the smear is something other than loxP-gDNA-loxP. The products may have been generated from only one primer-binding event. This means that each fragment can only undergo linear PCR using one priming site rather than two per molecule.

Another related method that was developed was a modified version of the degenerate-oligonucleotide-primed (DOP) PCR (Carter *et al.*, 1992). In this technique, a primer is

designed with defined regions at the 5' and 3' ends, with a random hexamer sequence in the middle. The 3' fixed sequence was designed to anchor the primer to common *T. reesei* sequences. This approach was used in a study where primers were designed with arbitrary but defined 3′ four-base sites (Primers, 2007). These sequences had moderate frequency in the target genomes and assisted in primer annealing. The primer 'chXIV loxP 5' flank-15N' was modified to include the sequence ATGTGG added to the 3' end. The new primer 'chXIV loxP 5' flank-15N-ATGTGG' was designed to encourage annealing to the *T. reesei* genome. The PCR protocol was set up as previously described for whole genome amplification (Arneson *et al*., 2008a). Results showed bands < 0.5 kb on an agarose gel and no smear (data not shown)

The method using restriction enzymes to digest loxP and *T. reesei* gDNA and subsequent ligation was unsuccessful. The step of adding *Hind*III recognition sites to the 3' ends of loxP worked well, as did the partial digestion of *T. reesei* gDNA. However, following ligation and PCR amplification, the expected smear (similar to that seen in the *Hind*III *T. reesei* gDNA partial digestion), was absent. The bands at the bottom of figure 22 (a) appear similar in size to the loxP amplification (with *Hind*III recognition site) in figure 21 (b). By using the primer 'loxP A1 FWD' alone, a single loxP fragment can only undergo linear amplification as the primer only anneals to one end. Fragments may therefore be one or more loxP fragments. Gel slices at 1 kb and ~3-8 kb were excised and re-amplified using the same primer. A nanodrop analysis confirmed the presence of DNA after purification from agarose, however, following re-amplification, the expected bands at 1 and >3 kb were missing. It is unclear why PCR amplification was unsuccessful.

With all protocols, NEB Phusion® High-Fidelity DNA Polymerase was used first due to its high fidelity and speed. Following unsuccessful attempts, NEB Taq DNA polymerase was also used since it has been shown to tolerate minor mismatches between degenerate primers and templates (Wang *et al*., 2000). Wang *et al*. (2000) found that Taq polymerase amplified fragments which failed to be amplified by the higher-fidelity enzyme Vent® DNA Polymerase. In this work however, Taq polymerase was also unsuccessful in amplification.

In chapter 3, it was demonstrated that DNA requires flanking by loxP recombination sequences to be SCRaMbLEd into *S. cerevisiae*. Overall, the PCR-based protocol developed in this chapter to attach loxP to *T. reesei* gDNA fragments were promising but unsuccessful. A new approach was therefore developed (chapter 5) which allowed for more control and screening at each stage of the protocol.

Chapter 5

A cloning approach of loxP attachment to *T. reesei* DNA and subsequent SCRaMbLEing into *S. cerevisiae*

The approaches used in the previous chapter to attach loxP sites to *T. reesei* gDNA were PCR based involving multiple degenerate primers and an approach based on ligation of digested gDNA and loxP. The approach presented here was cloning-based, which allowed for more control and screening at each stage. In order to attach loxP sites to *T. reesei* gDNA fragments, a plasmid was constructed with two consecutive loxP sequences on a pRS416 plasmid backbone (figure 29, appendix). Two loxP sites and a linear pRS416 fragment were PCR amplified; each fragment had homologous overlapping fragments to the other two fragments, allowing for construction by Gibson assembly. Between the two loxP sequences were the *Xho*I and *Bst*BI restriction sites for the digestion of plasmids and the cloning of digested *T. reesei* DNA. Fragments of loxP-gDNA-loxP were amplified from library of plasmids and SCRaMbLEd into *S. cerevisiae*. Cellulase activity was screened on Avicel and CMC.

METHODS

In order to Gibson assemble a plasmid containing a restriction site flanked by two loxP sequences, three DNA fragments were prepared. Primers were designed which generated overlapping regions between the three fragments, and also added the restriction sites.

Preparation of two loxP fragments and the pRS416 fragment

To amplify the 'loxP3' fragment, a PCR reaction was set up as described in table 3, with primers 'loxP3 FWD' and 'loxP3 REV' and DNA from SynXIV megachunk A as template. A touchdown PCR was performed as described in table 4, with an extension time of 15 s. The 'loxP4' fragment was amplified as above, with primers 'loxP4 FWD' and 'loxP4 REV'. PCR products were purified using the QIAGEN QIAquick® PCR Purification Kit.

To amplify pRS416, a PCR reaction was set up as described in table 3, with primers 'pRS416 FWD' and 'pRS416 REV' and pRS416 plasmid as template. A touchdown PCR was performed as described in table 4, with an extension time of 2 min. The product was purified using the QIAGEN QIAquick® PCR Purification Kit. The DNA was digested with DpnI restriction enzyme to remove methylated DNA as follows; 0.4 U/µl NEB DpnI, 1X NEB CutSmart® Buffer, <1μg PCR product. The reaction was incubated at 37 °C for 1 h. DNA was

dephosphorylated adding NEB CIP (0.2 U/µl) and incubated for 5 min at 37 °C. DNA was purified using the QIAGEN QIAquick® PCR Purification Kit.

Gibson assembly loxP3-loxP4-pRS416 plasimd

As with the construction of loxP-*URA3*-loxP in chapter 3, construction of the loxP3-loxP4-pRS416 plasmid was performed by Gibson assembly which recombined overlapping sequences (figure 23). The reaction was set up using a 6:6:1 ratio of loxP3, loxP4 and pRS416 respectively. DNA fragments were incubated with 1X NEBuilder® HiFi DNA Assembly Master Mix at 50 °C for 1 h.

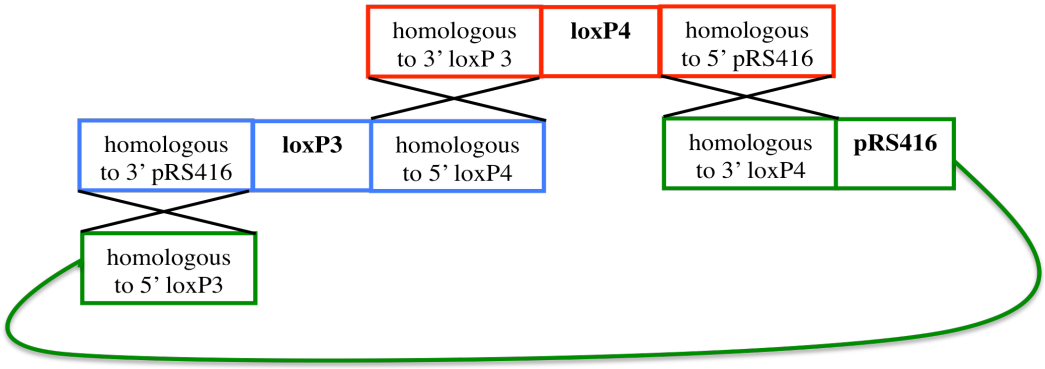

Figure 23: Construction of the loxP3-loxP4-pRS416 plasmid. Three fragments were prepared (represented in blue, red and green) using PCR with overlapping regions to facilitate construction by Gibson assembly.

Transformation of loxP3-loxP4-pRS416 plasmid into *E. coli*

Bioline DH5α™ *E. coli* competent cells and loxP3-loxP4-pRS416 were heat shocked at 42 °C for 45 sec. Cells were incubated in SOC for 1 h, shaking at 37 °C. Cells were plated on LB ampicillin and incubated at 37 °C for 18 h. To identify which colonies contain the correct loxP3-loxP4-pRS416 plasmid, a PCR was performed using primers that annealed outside the loxP3-loxP4 region. The PCR reaction was set up as described in table 3, with primers 'M13 F' and 'M13 R', and a sample of cells as template. A touchdown PCR was performed as described in table 4, with an extension time of 15 sec. Colonies which showed positive results in the colony PCR were harvested and cultured in liquid LB ampicillin overnight, shaking at 37 °C. Plasmid DNA was extracted using the Invitrogen Quick Plasmid Miniprep Kit.

Confirmations of correct loxP3-loxP4-pRS416 plasmid sequence

Plasmid DNA was digested to check the correct assembly of the loxP3-loxP4-pRS416 plasmid. The reaction was set up as follows; 0.5 U/µl Promega PvuII, 1X Promega buffer B,

<1μg loxP3-loxP4-pRS416 plasmid. The digestion mixture was incubated at 37 °C for 15 min. Also, the plasmids were sent to Macrogen (Korea) for Sanger sequencing using 'M13 F' and 'M13 R' primers. The primers anneal to outside of the two loxP sequences and the restriction recognition sites (the correct sequence of the remainder of the plasmid was not crucial).

_Xho_I digestion of loxP3-loxP4-pRS416 plasmid and _T. reesei_ gDNA

The loxP3-loxP4-pRS416 plasmid was digested as follows; 0.4 U/µl NEB *Xho*I, 1X NEB CutSmart® Buffer, <1μg loxP3-loxP4-pRS416 plasmid, incubated at 37 °C for 15 min. The DNA was dephosphorylated by adding NEB calf intestinal phosphatase (CIP) (0.2 U/µl) and incubated for 5 min at 37 °C. DNA was purified using the QIAGEN QIAquick® PCR Purification Kit. *T. reesei* gDNA was extracted as described in chapter 3. gDNA was partially digested as follows; 0.008 U/µl NEB *Xho*I, 1X NEB CutSmart® Buffer, <1μg *T. reesei* gDNA, incubated at 37 °C for 40 min.

Ligation of _T. reesei_ gDNA and the loxP3-loxP4-pRS416 plasmid

*Xho*I digested *T. reesei* gDNA and loxP3-loxP4-pRS416 were ligated using a 6:1 ratio of insert to plasmid respectively; 1 U/µl NEB T4 DNA Ligase, 1X NEB T4 DNA Ligase Reaction Buffer, <1 µg *T. reesei* gDNA and <1 µg loxP3-loxP4-pRS416 plasmid, incubated at 16 °C overnight. A control was set up with no *T. reesei* gDNA.

E. coli transformation of plasmids containing _T. reesei_ DNA fragments

Bioline DH5α™ *E. coli* competent cells and the ligation mixtures from above were separately heat shocked at 42 °C for 45 sec. Cells were incubated in SOC for 1 h, at 37 °C. A sample of cells from each sample were plated onto LB ampicillin and incubated overnight at 37 °C for colony counting. Remaining cells were added to 10 volumes of LB ampicillin and incubated at 37 °C overnight. *E. coli* plasmid DNA containing *T. reesei* DNA fragments was extracted from the overnight culture using the Invitrogen Quick Plasmid Miniprep Kit.

Amplification of loxP-gDNA-loxP fragments from heterologous plasmids

To amplify loxP-gDNA-loxP, a PCR reaction was set up as described in table 3, with primers 'loxP3 FWD' and 'loxP4 REV' and *E. coli* plasmid DNA as template. A touchdown PCR was performed as described in table 4, with an extension time of 5 min. The product was run on an agarose gel, and bands were excised above 3 kb. DNA was purified using the Zymo Research Zymoclean™ Gel DNA Recovery Kit. A PCR was performed to amplify fragments from the gel slices, and was set up as above.

SCRaMbLE and transformation of loxP flanked gDNA into *S. cerevisiae*

DNA was transformed and SCRaMbLEd for 3 hours using the diploid *S. cerevisiae* strain constructed in chapter 3 using the same protocol as described in chapter 3. Yeast cells were plated onto 20g/L avicel and 20g/L CMC agar plates and incubated for 5 days at 30 °C.

RESULTS

In order to SCRaMbLE foreign fragments of DNA into synthetic chromosomes of yeast, they must be flanked with loxP recombination sequences. PCR-based approaches developed in the previous chapter were unsuccessful to achieve this, so a cloning-based protocol was developed. In this method, there was more control at each stage, and the success of each step could be confirmed. The protocol was used successfully to attach loxP sites onto fragments of partially digested *T. reesei* gDNA. Following transformation into diploid *S. cerevisiae*, no colonies grew on CMC or Avicel media.

Figure 24 shows the partial digestion of *T. reesei* gDNA. The sample shown in lane 3 was utilised for cloning. It shows a smear and also maintains a band above 10 kb signifying partially digested gDNA using *Xho*I.

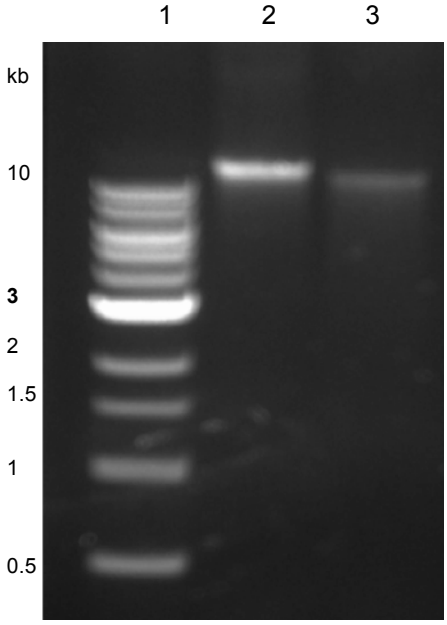

Figure 24: Agarose gel electrophoresis showing intact *T. reesei* gDNA and partially digested *T. reesei* gDNA using *Xho*I. Lane 1: NEB 1 kb DNA ladder, lane 2: undigested *T. reesei* gDNA, lane 3: *Xho*I partially digested *T. reesei* gDNA.

Following transformation of the constructed loxP3-loxP4-pRS416 plasmid into *E. coli* and plating onto LB ampicillin, a number of colonies grew and were screened by a colony PCR. Figure 25 (a) shows the PCR amplicons. Lanes 2, 3, 5 and 7-11 show a band at ~200bp suggesting incorrect plasmid assembly, and no amplification in lanes 4 and 11 suggests the priming site was absent. The ~400bp band in lane 6 is the expected size for correct loxP3-

loxP4-pRS416 plasmid assembly. Figure 25 (b) shows the PvuII endonuclease digest of *E. coli* purified plasmid. In lane 3, the faint band at ~600bp is the expected size if the loxP3-loxP4-pRS416 plasmid was correctly assembled. As a final confirmation of plasmid construction, the plasmid was sequenced by Macrogen. Plasmid sequencing results aligned with the designed sequence confirming correct construction (Figure 28, appendix)

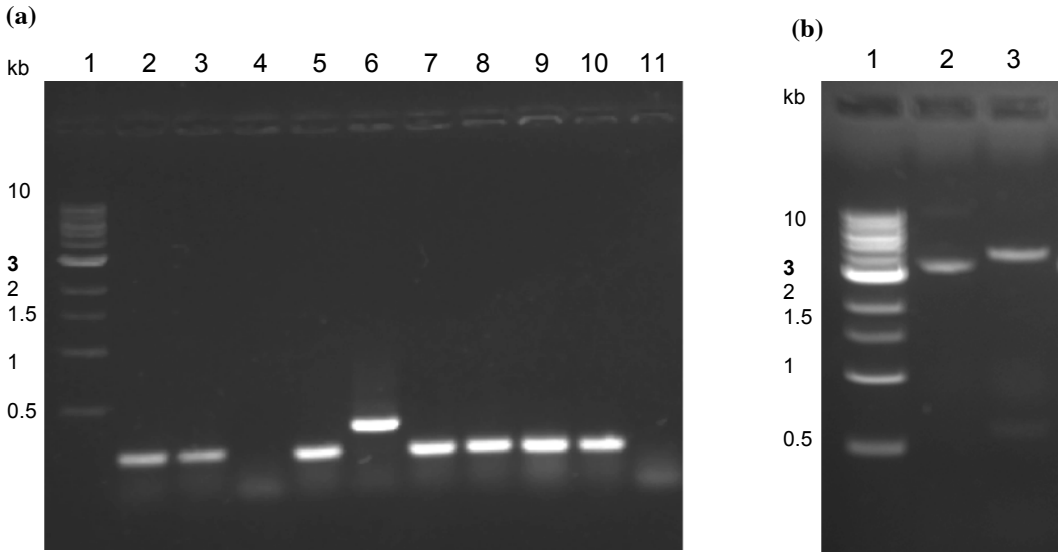

Figure 25: (a) Agarose gel electrophoresis showing PCR products of the colony PCR of *E. coli*. Lane 1: NEB 1 kb DNA ladder, lanes 2-11: ten different colony PCRs. **(b)** Agarose gel electrophoresis showing digestion of purified *E. coli* plasmid DNA using PvuII endonuclease. Lane 1: NEB 1 kb DNA ladder, lane 2: purified *E. coli* plasmid DNA, lane 3: purified *E. coli* plasmid DNA cut with PvuII endonuclease.

Following cloning of *T. reesei* gDNA into the plasmid, a transformation into *E. coli* was performed using both the *T. reesei*/plasmid ligation and the plasmid-only ligation. This was to determine if any growth on LB ampicillin was due to re-ligation of linear loxP3-loxP4-pRS416 molecules. Table 6 shows that after transformation of the plasmid and gDNA ligation, there were 442 colonies, compared to 90 colonies for the plasmid only ligation (representing plasmids that have re-ligated).

Table 6: Number of *E. coli* colonies on LB ampicillin after transformation with ligations of plasmid and *T. reesei* gDNA or plasmid only.

	Number of colonies on LB ampicillin plates
loxP3-loxP4-pRS416 plasmid + *T. reesei* gDNA ligation	442
loxP3-loxP4-pRS416 plasmid-only ligation	90

Following transformation of the ligation between the loxP3-loxP4-pRS416 plasmid and *T. reesei* gDNA, a liquid culture of *E. coli* cells with plasmids containing *T. reesei* DNA was grown. Plasmids were extracted and loxP-gDNA-loxP fragments were amplified using the 'loxP3 FWD' and 'loxP4 REV' primers. Figure 26 (a) lane 2 shows a smear of PCR product following the first round of loxP-gDNA-loxP amplification. The presence of the smear indicated that fragments of different sizes had been ligated into the *Xho*I site of the loxP3-loxP4-pRS416 plasmid. The sizes of the fragments ranged from ~10 kb to <0.5. DNA was extracted from a gel slice excised at >2 kb and reamplified using the same primers. The expected smear at >2 kb was absent (figure 26).

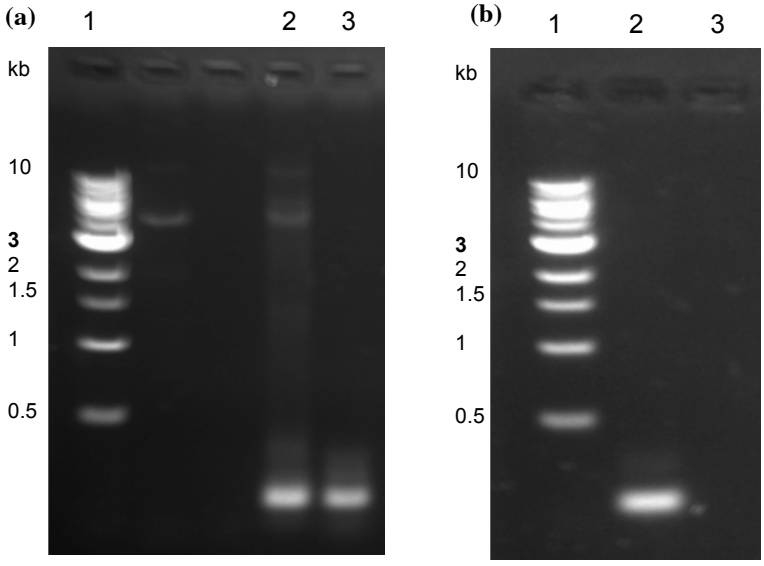

Figure 26 (a) Gel electrophoresis of loxP-gDNA-loxP DNA after first round of PCR amplification. Lane 1: NEB 1 kb DNA ladder, lane 2: PCR amplification of loxP-gDNA-loxP using primers 'loxP3 FWD' and 'loxP4 REV' following plasmid purification from heterologous *E. coli* colony, lane 3: negative control. **(b)** Agarose gel electrophoresis of loxP-gDNA-loxP DNA after second round of PCR amplification. Lane 1: NEB 1 kb DNA ladder, lane 2: PCR amplification of loxP-gDNA-loxP DNA using primers 'loxP3 FWD' and 'loxP4 REV' following purification from >2 kb gel slice, lane 3: negative control.

DNA samples from two stages of the protocol were transformed into yeast; DNA from the after the first loxP-gDNA-loxP PCR amplification, and DNA from after the second loxP-gDNA-loxP PCR re-amplification (following DNA purification of gel slice). No growth was observed on Avicel or CMC after >5 days incubation at 30 °C.

DISCUSSION

To utilise SCRaMbLE for integration of foreign DNA into synthetic chromosomes of yeast, DNA fragments require flanking by loxP recombination sequences. *T. reesei* gDNA fragments were flanked with loxP using a cloning approach. A plasmid was constructed, using the pRS416 plasmid as a backbone, which included two restriction sites immediately flanked by two loxP sites. Digested gDNA fragments were cloned into the *Xho*I site. loxP-gDNA-loxP fragments were PCR amplified. LoxP-gDNA-loxP fragments were transformed and SCRaMbLEd into *S. cerevisiae*. Cellulase activity was screened on avicel and CMC, however no colonies grew.

The success of flanking gDNA with loxP sites is largely attributed to the design and construction of the loxP3-loxP4-pRS416 plasmid (figure 29, appendix). The incorporated *Xho*I and *Bst*BI restriction sites were immediately flanked by the two loxP fragments, allowing for digested *T. reesei* gDNA fragments to be cloned into. These enzymes were suitable for *T. reesei* digestion as they cut frequently and fairly uniformly throughout the genome. *Xho*I cleaves 15,471 times in the genome with an average fragment length of 2159 bps while *Bst*BI cleaves 6,100 times with an average fragment length of 5475 bps (information generated by an *in silico* digest using Geneious software). To improve the likelihood of obtaining representative genomic samples of fragments that varied in length and composition, a partial *Xho*I digestion of gDNA was performed to ensure that cleavage did not occur at all available restriction sites,

The constructed diploid *S. cerevisiae* strain was used for the SCRaMbLEing of loxP-flanked *T. reesei* gDNA. This strain showed optimal results when introducing the *URA3* marker with loxP-mediated recombination (chapter 3). In future SCRaMbLE experiments, it would be beneficial to incorporate multiple copies of cellulase genes to increase cellulase activity increases. Yamada *et al*. (2011) constructed diploid strains of *S. cerevisiae* from haploids containing single copies of EG, CBH and BGL genes. The diploids displayed six-fold higher phosphoric acid swollen cellulose (PASC) degradation activity than the parent haploid strains. In another example, the copy number of *egl* genes was increased from one to two; two copies

of *egl2* increased the endoglucanase activity by 3-fold, while two copies of *egl1* increased the endoglucanase activity by 2.2-fold (Miettinen-Oinonen and Suominen, 2002). To achieve results similar to these, a diploid could be constructed that contains two copies of each available synthetic chromosome. This will increase the frequency of loxP-mediated insertion and SCRaMbLEing of cellulase genes into the genome. Another way to increase the copy number of cellulase genes would be to attach loxP sequences specifically to cellulase genes and SCRaMbLE into yeast to optimise copy number.

As an example, a cellulolytic yeast has been engineered by importing a selection of heterologous genes into yeast. A recombinant *S. cerevisiae* strain was developed capable of the hydrolysis and fermentation of amorphous cellulose. After transformation of an endoglucanase of *T. reesei* (EGI) and the b-glucosidase of *Saccharomycopsis fibuligera* (BGL1) the *S. cerevisiae* strain successfully grew on phosphoric acid swollen cellulose (PASC) (Den Haan *et al*., 2007b). These genes could be flanked with loxP and SCRaMbLEd into yeast in order to achieve a strain with similar, or even more enhanced cellulase capability.

The limitation of the current protocol is the suboptimal *T. reesei* gDNA library. Ideally, the library would include fragments representative of the full genome, should be highly variable, and have a high copy number. To determine the genome coverage in the gDNA library, a plasmid pRS416-only ligation, along with the plasmid and gDNA ligation was set up. Following transformation of these ligations into *E. coli*, colonies growing on LB ampicillin may be due to plasmid/gDNA ligation, or re-ligation of the digested plasmid. The prior dephosphorylation of the plasmid minimised re-ligation, however colonies still grew in the plasmid-only control; 90 colonies grew as a result of re-ligation compared to 442 from the plasmid and gDNA ligation. Therefore, 80% of plasmids in the library have a gDNA insert. Since 100 µl of 2000 µl of of *E. coli* cells were plated, the remainder contains ~6,700 fragments in total. There were 15,471 possible fragments for complete *Xho*I digestion of *T. reesei* gDNA, and ~5500 fragments that are the ideal fragment length (between 2 and 15 kb). Since the average fragment length in a complete digestion is ~2 kb, if half of all restriction sites were cleaved, there would be many more fragments falling in the 2-15 kb category, perhaps double (11,000). Ideally, the library would contain at least 5-fold this many fragments to account for different fragment compositions. Therefore a library of 55,000 would be required to optimise this experiment, many more than the current library of ~6,700 fragments.

In order to develop cellulolytic yeast from the SCRaMbLEing in of random fragments of *T. reesei* gDNA requires a huge library. This is because *S. cerevisiae* not only requires the genes involved directly in cellulase activity, but it also requires gene for other processes such as promoter regulation. In the future, a gDNA library of closely related organisms could be developed and used to transform and SCRaMbLE into *S. cerevisiae*. A much smaller library would be required if SCRaMbLEing DNA from more closely related species such as the methylotrophic yeast, *Pichia pastoris*. *S. cerevisiae* is more likely to be able to express genes from this organism. This theory could be tested in future studies.

When importing heterologous DNA into a host, there are certain challenges to account for. These include issues such as protein secretion and ER stress. These issues are a legitimate but distant concern. Firstly, the process is not optimised yet and if it were optimised, then genes that facilitate better secretion, cellulase translation, intron splicing, and promoter regulation can also be SCRaMbLEd in with genes directly involved in the desirable phenotype. Therefore, once the protocol is optimised, SCRaMbLE has the potential to overcome many issues usually associated with heterologous protein production.

The approach of using SCRaMbLE to integrate random fragments of the entire *T. reesei* genome into *S. cerevisiae* is similar to a modified genome shuffling protocol developed recently. In genome shuffling, the genomes of two organisms are recombined to make a hybrid, having traits of each parent. This method was used to generate a *S. cerevisiae* strain with increased ethanol production on xylose (Zhang and Geng, 2012). The protocol involved two steps. Firstly, the whole genome of *Pichia stipitis* was transferred into *S. cerevisiae*. Following screening on xylose media, the best transformant was again transformed with the *S. cerevisiae* genome. The resulting recombinant strain had desirable traits from both of the parental strains. From *P. stipitis,* the strain was able to utilise xylose, and from *S. cerevisiae*, the strain was a good producer of ethanol. In fact, the new strain produced ethanol more efficiently, than *P. stipitis* did when fermenting xylose.

This genome shuffling technique is similar in some ways to introducing foreign DNA with SCRaMbLE, however SCRaMbLE is ultimately much more efficient for integrating heterologous DNA and rearranging the genome. Both are non-rational approaches relying on the random integration of sequences from a whole genome, followed by screening. As with genome shuffling, SCRaMbLE has been designed to rapidly generate simultaneous changes throughout the entire genome. Neither approach requires genome sequence data or knowledge

of pathways or networks of interest. Aside from the similarities, there are many advantages of SCRaMbLE. Firstly, the entire genome of the parent strain (*S. cerevisiae*) does not have to be completely shuffled. For example, in SCRaMbLE, there is the choice of which chromosomes are SCRaMbLEd by developing strains with different numbers of synthetic chromosomes. Also, the level of SCRaMbLEing can easily be controlled with different lengths of induction with estradiol. The parent can be shuffled and only a selection of genes can be introduced.

Overall, this chapter has presented a protocol to develop a library of *T. reesei* gDNA flanked with loxP recombination sequences. The screening of cellulase resulted in no isolates. The protocol requires optimisation and has the potential to be a powerful tool to integrate large libraries of DNA into yeast.

CHAPTER 6

Summary, Conclusions and Future Direction

The yeast 2.0 project is an international collaboration which is building the first eukaryotic synthetic genome. The defining feature of the genome is an inducible and powerful evolution system called SCRaMbLE. Though the project will initially build only one synthetic genome, SCRaMbLE can rapidly generate billions of unique genomes on demand. With simple screening conditions, such as high temperature, only cells with desirable genetic arrangements will survive. The sequencing of a few strains may shed light to the gene sets corresponding to the selected phenotype. This information can be applied to future rational engineering studies. The system was previously limited to SCRaMbLEing yeast DNA only. Therefore, the possible phenotypes that can be generated were finite. However, through the introduction of foreign DNA with SCRaMbLE, the possible genotypes and phenotypes are limitless. The overall aim of the current work was to explore SCRaMbLE for this novel application of introducing foreign DNA.

Currently, only one synthetic chromosome is complete and available for SCRaMbLEing. In chapter 2, a megachunk of SynXIV, megachunk O, was constructed and integrated into *S. cerevisiae*. Analysis confirmed that megachunk O had replaced the wild type equivalent DNA. Fitness screening of the final two strains suggested a potential growth defect, however, the defect may be due to integration of the *URA3* marker cassette which would be reversed with the integration of the next marker. The construction of this megachunk was a contribution to the international yeast 2.0 project.

Prior to SCRaMbLEing in heterologous DNA, it was essential to first develop protocols for the approach using a single homologous gene. A protocol was developed to SCRaMbLE in the loxP-flanked *S. cerevisiae* marker gene, *URA3*. The loxP-*URA3*-loxP fragment was transformed SCRaMbLEd into yeast. The transformation efficiency of SCRaMbLEd loxP-flanked *URA3* was significantly higher than the non-SCRaMbLEd equivalent. The pRS416 plasmid containing *URA3* (not flanked by loxP) was also SCRaMbLEd and not SCRaMbLEd as a control, showing only a negligible difference. This experiment demonstrated that foreign DNA requires flanking by loxP sequences to be SCRaMbLEd, and that loxP-flanked DNA is integrated into yeast chromosomes with high efficiency using SCRaMbLE.

To SCRaMbLE fragments of *T. reesei* gDNA into yeast, they need to be flanked by loxP. A number of PCR-based approaches were developed to achieve this, as they had the potential to be a simple and efficient. In two protocols, degenerate primers were used that had 5' extension sequences used to add loxP sites. One protocol digested *T. reesei* gDNA as well as modified loxP fragments, which were then ligated. The PCR-protocols were promising but ultimately unsuccessful to attach loxP sites to random fragments of *T. reesei* gDNA.

A cloning approach was successful in the attachment of loxP sites to *T. reesei* gDNA. The protocol had multiple checkpoints, which conferred a high level of control that the previous PCR-based approaches lacked. A plasmid was engineered with restriction sites immediately flanked by two loxP sequences. Digested *T. reesei* gDNA was cloned into the restriction sites, and loxP-gDNA-loxP fragments were amplified using PCR. Fragments were transformed and SCRaMbLEd into yeast. Cellulase activity was screened, however no colonies grew on selection media. The gDNA library was the limiting factor for the approach; a much larger *T. reesei* gDNA library could be developed in the future to would the chances of desirable gene integration.

Conclusions

In conclusion, SCRaMbLE was explored as a novel tool to integrate foreign DNA into *S. cerevisiae*. Protocols were developed to flank DNA with loxP; Gibson assembly was utilised to flank the single *URA3* genes, while a cloning approach was utilised to flank partially digested fragments of *T. reesei* gDNA with loxP. These loxP sequences were shown to essential to SCRaMbLEing. Overall, SCRaMbLE was used to efficiently integrate DNA into *S. cerevisiae* synthetic chromosome III. The approach is a novel method integrate large libraries of DNA in yeast while simultaneously shuffling endogenous and exogenous DNA.

Future direction

A good approach that could optimise foreign DNA SCRaMbLEing in the future, would be to utilise cDNA libraries. cDNA libraries are extremely useful for heterologous gene expression as they do not contain introns. Therefore, there would be no issues of intron splicing after integration into *S. cerevisiae* (Saloheimo *et al.*, 1997). In harnessing cDNA libraries, genetic promoters needs to be taken into account.

S. cerevisiae promoter libraries could be developed to moderate the expression of *T. reesei* genes involved in *e.g.* cellulase activity. Foreign promoters can be incompatible with a host

organism, so a well-characterised promoter library is invaluable for fine-tuned transcriptional control. A common strategy currently used to develop promoter libraries is the use of error-prone PCR to generate promoter mutants which are then screened. The strong translation and elongation factor 1 (TEF1) promoter was randomised by error prone PCR to produce promoter mutants (Alper *et al*., 2005). Differential expression was then screened using green fluorescent protein (GFP) fluorescence and expression levels monitored by qPCR. The resulting promoter library had a range of strengths which can be selected for different applications. More recently, a method was developed for constructing *S. cerevisiae* synthetic promoter libraries (Blazeck *et al*., 2012). Using a synthetic hybrid promoter approach, the strongest characterised promoters were developed. The approach is based on the combination of upstream activating sequences (UAS) with core promoters. The combination of UAS elements with *S. cerevisiae* core promoters generated a promoter library with a dynamic range of constitutive promoter activity, expanding the transcriptional capacity beyond that of the strong constitutive glyceraldehyde-3-phosphate dehydrogenase (P_{GPD}) promoter by over 2.5-fold. Following this, the inducible regulation of constitutive promoter expression was engineered through combinations with a galactose-inducible UAS element. Thus, a library of inducible promoters was established which were tunable by galactose induced gene expression. The promoters here demonstrated a huge dynamic range; an increase in range of almost 50-fold compared to the galactose inducible promoter $P_{GAL.}$ This approach could be harnessed to generate a pool of promoters to use with *T. reesei* cellulase genes, or any other genes to be imported into *S. cerevisiae* for SCRaMbLEing.

References

ALPER, H., FISCHER, C., NEVOIGT, E. & STEPHANOPOULOS, G. 2005. Tuning genetic control through promoter engineering. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 12678-12683.

ARNESON, N., HUGHES, S., HOULSTON, R. & DONE, S. 2008a. Whole-genome amplification by degenerate oligonucleotide primed PCR (DOP-PCR). *Cold Spring Harbor Protocols***,** 249(2):584-91

ARNESON, N., HUGHES, S., HOULSTON, R. & DONE, S. 2008b. Whole-genome amplification by improved primer extension preamplification PCR (I-PEP-PCR). *Cold Spring Harbor Protocols,* 249(2):592-9.

BEGGS, J. D., VAN DEN BERG, J., VAN OOYEN, A. & WEISSMANN, C. 1980. Abnormal expression of chromosomal rabbit β-globin gene in *Saccharomyces cerevisiae*. *Nature*, 283:835-40.

BLAZECK, J., GARG, R., REED, B. & ALPER, H. S. 2012. Controlling promoter strength and regulation in *Saccharomyces cerevisiae* using synthetic hybrid promoters. *Biotechnology and Bioengineering,* 109**,** 2884-2895.

BOHLANDER, S. K., ESPINOSA, R., LE BEAU, M. M., ROWLEY, J. D. & DÍAZ, M. O. 1992. A method for the rapid sequence-independent amplification of microdissected chromosomal material. *Genomics,* 13**,** 1322-1324.

BRO, C., REGENBERG, B., FÖRSTER, J. & NIELSEN, J. 2006. *In silico* aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. *Metabolic Engineering,* 8**,** 102-111.

CARRARO, D. M., JÚNIOR, J. R. F., SCHUMACHER, R., PEREIRA, G. G., HOLLENBERG, C. P. & EL-DORRY, H. 1998. A Region of the Cellobiohydrolase I Promoter from the Filamentous Fungus *Trichoderma reesei* Mediates Glucose Repression in *Saccharomyces cerevisiae*, Dependent on Mitochondrial Activity. *Biochemical and Biophysical Research Communications,* 253**,** 407-414.

CARTER, G. L., ALLISON, D., REY, M. W. & DUNN‐COLEMAN, N. S. 1992. Chromosomal and genetic analysis of the electrophoretic karyotype of *Trichoderma reesei*: mapping of the cellulase and xylanase genes. *Molecular Microbiology,* 6**,** 2167-2174.

DAVID, F. & SIEWERS, V. 2014. Advances in yeast genome engineering. *FEMS yeast research*. 15(1),1-14.

DAVIS, C. A., GRATE, L., SPINGOLA, M. & ARES, M., JR. 2000. Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. *Nucleic Acids Research,* 28**,** 1700-6.

DEN HAAN, R., MCBRIDE, J. E., LA GRANGE, D. C., LYND, L. R. & VAN ZYL, W. H. 2007a. Functional expression of cellobiohydrolases in *Saccharomyces cerevisiae* towards one-step conversion of cellulose to ethanol. *Enzyme and Microbial Technology,* 40**,** 1291-1299.

DEN HAAN, R., ROSE, S. H., LYND, L. R. & VAN ZYL, W. H. 2007b. Hydrolysis and fermentation of amorphous cellulose by recombinant *Saccharomyces cerevisiae*. *Metabolic Engineering,* 9**,** 87-94.

DICARLO, J. E., NORVILLE, J. E., MALI, P., RIOS, X., AACH, J. & CHURCH, G. M. 2013. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Research***,** 41(7), 4336-4343

DU PLESSIS, L., ROSE, S. H. & VAN ZYL, W. H. 2010. Exploring improved endoglucanase expression in *Saccharomyces cerevisiae* strains. *Applied Microbiology and Biotechnology,* 86**,** 1503-1511.

DYMOND, J. & BOEKE, J. 2012. The *Saccharomyces cerevisiae* SCRaMbLE system and genome minimization. *Bioengineered Bugs,* 3**,** 168-71.

DYMOND, J. S., RICHARDSON, S. M., COOMBES, C. E., BABATZ, T., MULLER, H., ANNALURU, N., BLAKE, W. J., SCHWERZMANN, J. W., DAI, J., LINDSTROM, D. L., BOEKE, A. C., GOTTSCHLING, D. E., CHANDRASEGARAN, S., BADER, J. S. & BOEKE, J. D. 2011. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature,* 477**,** 471-6.

EYRE-WALKER, A. & KEIGHTLEY, P. D. 2007. The distribution of fitness effects of new mutations. *Nature Reviews Genetics,* 8**,** 610-618.

GARNEAU, J. E., DUPUIS, M.-È., VILLION, M., ROMERO, D. A., BARRANGOU, R., BOYAVAL, P., FREMAUX, C., HORVATH, P., MAGADÁN, A. H. & MOINEAU, S. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature,* 468**,** 67-71.

GIBSON, D. G., BENDERS, G. A., ANDREWS-PFANNKOCH, C., DENISOVA, E. A., BADEN-TILLSON, H., ZAVERI, J., STOCKWELL, T. B., BROWNLEY, A., THOMAS, D. W. & ALGIRE, M. A. 2008. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science,* 319**,** 1215-1220.

GIBSON, D. G., YOUNG, L., CHUANG, R.-Y., VENTER, J. C., HUTCHISON, C. A. & SMITH, H. O. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods,* 6**,** 343-345.

GIETZ, R. D. & SCHIESTL, R. H. 2007. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature protocols,* 2**,** 31-34.

GOFFEAU, A., BARRELL, B., BUSSEY, H., DAVIS, R., DUJON, B., FELDMANN, H., GALIBERT, F., HOHEISEL, J., JACQ, C. & JOHNSTON, M. 1996. Life with 6000 genes. *Science,* 274**,** 546-567.

GURR, S. J., UNKLES, S. E. & KINGHORN, J. R. 1987. The Structure and Organisation of Nuclear Genes of Filamentous Fungi. *Gene Structure in Lower Eukaryotes,* 22**,** 93-193.

HANCOCK, S. M., UPRETY, R., DEITERS, A. & CHIN, J. W. 2010. Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair. *Journal of the American Chemical Society,* 132**,** 14819-14824.

HEINEMANN, M. & PANKE, S. 2006. Synthetic biology—putting engineering into biology. *Bioinformatics,* 22**,** 2790-2799.

ILMEN, M., SALOHEIMO, A., ONNELA, M.-L. & PENTTILÄ, M. E. 1997. Regulation of cellulase gene expression in the filamentous fungus *Trichoderma reesei*. *Applied and Environmental Microbiology,* 63**,** 1298-1306.

JINEK, M., CHYLINSKI, K., FONFARA, I., HAUER, M., DOUDNA, J. A. & CHARPENTIER, E. 2012. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science,* 337**,** 816-821.

JUNEAU, K., PALM, C., MIRANDA, M. & DAVIS, R. W. 2007. High-density yeast-tiling array reveals previously undiscovered introns and extensive regulation of meiotic splicing. *Proceedings of the National Academy of Sciences,* 104**,** 1522-1527.

KELLIS, M., PATTERSON, N., ENDRIZZI, M., BIRREN, B. & LANDER, E. S. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature,* 423**,** 241-254.

KNOWLES, J., PENTTILII, M., TEERI, T., NEVALAINEN, H., LEHTOVAARA-HELENIUS, P., AHO, S., NITISINPRASERT, S., PALOLIEIMO, M. & KERIINEN, S. 1998. Yeast Strains Producing cellulolytic enzymes and methods and means for constructing them. United States patent number CA 1305931 C.

LARIONOV, V., KOUPRINA, N., ELDAROV, M., PERKINS, E., PORTER, G. & RESNICK, M. 1994. Transformation - associated recombination between diverged and homologous DNA repeats is induced by strand breaks. *Yeast,* 10**,** 93-104.

LEE, T. I., RINALDI, N. J., ROBERT, F., ODOM, D. T., BAR-JOSEPH, Z., GERBER, G. K., HANNETT, N. M., HARBISON, C. T., THOMPSON, C. M. & SIMON, I. 2002. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science,* 298**,** 799-804.

LIU, Z., LIU, L., ÖSTERLUND, T., HOU, J., HUANG, M., FAGERBERG, L., PETRANOVIC, D., UHLÉN, M. & NIELSEN, J. 2014. Improved production of a heterologous amylase in *Saccharomyces cerevisiae* by inverse metabolic engineering. *Applied and Environmental Microbiology,* 80**,** 5542-5550.

MANDELS, M. & REESE, E. T. 1957. Induction of cellulase in *Trichoderma viride* as influenced by carbon sources and metals. *Journal of Bacteriology,* 73**,** 269.

MARTINEZ, D., BERKA, R. M., HENRISSAT, B., SALOHEIMO, M., ARVAS, M., BAKER, S. E., CHAPMAN, J., CHERTKOV, O., COUTINHO, P. M. & CULLEN, D. 2008. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nature Biotechnology,* 26**,** 553-560.

MCNEIL, J. B. & FRIESEN, J. D. 1981. Expression of the Herpes simplex virus thymidine kinase gene in *Saccharomyces cerevisiae*. *Molecular and General Genetics MGG,* 184**,** 386-393.

MIETTINEN-OINONEN, A. & SUOMINEN, P. 2002. Enhanced production of *Trichoderma reesei* endoglucanases and use of the new cellulase preparations in producing the stonewashed effect on denim fabric. *Applied and Environmental Microbiology,* 68**,** 3956-3964.

MONTENECOURT, B. S. 1983. *Trichoderma reesei* cellulases. *Trends in Biotechnology,* 1**,** 156-161.

OSTERGAARD, S., OLSSON, L. & NIELSEN, J. 2000. Metabolic engineering of *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology Reviews,* 64**,** 34-50.

PFAFFL, M. 2001. Development and validation of an externally standardised quantitative insulin-like growth factor-1 RT-PCR using LightCycler SYBR Green I technology. *Rapid Cycle Real-Time PCR*. 3, 13-16.

PFAFFL, M., GEORGIEVA, T. M., GEORGIEV, I. P., ONTSOUKA, E., HAGELEIT, M. & BLUM, J. 2002. Real-time RT-PCR quantification of insulin-like growth factor (IGF)-1, IGF-1 receptor, IGF-2, IGF-2 receptor, insulin receptor, growth hormone receptor, IGF-binding proteins 1, 2 and 3 in the bovine species. *Domestic Animal Endocrinology,* 22**,** 91-102.

POLYCARPO, C. R., HERRING, S., BÉRUBÉ, A., WOOD, J. L., SÖLL, D. & AMBROGELLY, A. 2006. Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *FEBS Letters,* 580**,** 6695-6700.

PORCAR, M. & PERETÓ, J. 2014. *Synthetic Biology*: The iGEM Competition. *Springer Netherlands*, 55-62.

PRIMERS, P. 2007. High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences. *Biotechniques,* 43**,** 649-656.

RAAB, A. M., GEBHARDT, G., BOLOTINA, N., WEUSTER-BOTZ, D. & LANG, C. 2010. Metabolic engineering of *Saccharomyces cerevisiae* for the biotechnological production of succinic acid. *Metabolic Engineering,* 12**,** 518-525.

RUNGUPHAN, W. & KEASLING, J. D. 2014. Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid-derived biofuels and chemicals. *Metabolic Engineering,* 21**,** 103-113.

SAHI, C., KOMINEK, J., ZIEGELHOFFER, T., YU, H. Y., BARANOWSKI, M., MARSZALEK, J. & CRAIG, E. A. 2013. Sequential duplications of an ancient member of the DnaJ-family expanded the functional chaperone network in the eukaryotic cytosol. *Molecular Biology and Evolution***,** 30 (5):985-98.

SALOHEIMO, M., NAKARI-SETAELAE, T., TENKANEN, M. & PENTTILÄ, M. 1997. cDNA cloning of a *Trichoderma reesei* cellulase and demonstration of endoglucanase activity by expression in yeast. *European Journal of Biochemistry,* 249**,** 584-591.

SONDEREGGER, M. & SAUER, U. 2003. Evolutionary engineering of *Saccharomyces cerevisiae* for anaerobic growth on xylose. *Applied and environmental microbiology,* 69**,** 1990-1998.

SPINGOLA, M., GRATE, L., HAUSSLER, D. & ARES, M. 1999. Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA,* 5**,** 221-234.

STANLEY, D., FRASER, S., CHAMBERS, P. J., ROGERS, P. & STANLEY, G. A. 2010. Generation and characterisation of stable ethanol-tolerant mutants of *Saccharomyces cerevisiae*. *Journal of Industrial Microbiology & Biotechnology,* 37**,** 139-149.

TEERI, T. T., LEHTOVAARA, P., KAUPPINEN, S., SALOVUORI, I. & KNOWLES, J. 1987. Homologous domains in *Trichoderma reesei* cellulolytic enzymes: gene sequence and expression of cellobiohydrolase II. *Gene,* 51**,** 43-52.

VAMVACA, K., VOLLES, M. J. & LANSBURY, P. T. 2009. The first N-terminal amino acids of α-synuclein are essential for α-helical structure formation in vitro and membrane binding in yeast. *Journal of Molecular Biology,* 389**,** 413-424.

VAN ZYL, W. H., LYND, L. R., DEN HAAN, R. & MCBRIDE, J. E. 2007. Consolidated bioprocessing for bioethanol production using *Saccharomyces cerevisiae*. *Biofuels*. 108, 205-35.

WANG, Z., RAIFU, M., HOWARD, M., SMITH, L., HANSEN, D., GOLDSBY, R. & RATNER, D. 2000. Universal PCR amplification of mouse immunoglobulin gene variable regions: the design of degenerate primers and an assessment of the effect of DNA polymerase 3′ to 5′ exonuclease activity. *Journal of Immunological Methods,* 233**,** 167-177.

WILLS, C. 1990. Regulation of sugar and ethanol metabolism in *Saccharomyces cerevisiae*. *Critical Reviews in Biochemistry and Molecular Biology,* 25**,** 245-280.

WINZELER, E. A., SHOEMAKER, D. D., ASTROMOFF, A., LIANG, H., ANDERSON, K., ANDRE, B., BANGHAM, R., BENITO, R., BOEKE, J. D. & BUSSEY, H. 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science,* 285**,** 901-906.

WISSELINK, H. W., TOIRKENS, M. J., WU, Q., PRONK, J. T. & VAN MARIS, A. J. 2009. Novel evolutionary engineering approach for accelerated utilization of glucose, xylose, and arabinose mixtures by engineered *Saccharomyces cerevisiae* strains. *Applied and Environmental Microbiology,* 75**,** 907-914.

WLOCH, D. M., SZAFRANIEC, K., BORTS, R. H. & KORONA, R. 2001. Direct estimate of the mutation rate and the distribution of fitness effects in the yeast *Saccharomyces cerevisiae*. *Genetics,* 159**,** 441-452.

YAMADA, R., TANIGUCHI, N., TANAKA, T., OGINO, C., FUKUDA, H. & KONDO, A. 2011. Direct ethanol production from cellulosic materials using a diploid strain of *Saccharomyces cerevisiae* with optimized cellulase expression. *Biotechnol Biofuels,* 4(8), 102-109.

ZHANG, W. & GENG, A. 2012. Improved ethanol production by a xylose-fermenting recombinant yeast strain constructed through a modified genome shuffling method. *Biotechnol Biofuels,* 5**,** 46-51.

Appendix

Table 7: Primer pairs used for qPCR analysis of yeast DNA isolated from yeast colonies transformed with O megachunk.

Primer Pair Name	Forward Primer Sequence	Reverse Primer Sequence
YNL091W_3_Syn	CTTGGCTGCTACCTTGTCAAGCTGTTTC	TCTGCCTTCTTCCAATCTATCGGTTTCG
YNL091W_4_Syn	CAGCGCTAATATCTTGAGTGCTAAGCCA	AGGCTGAGGTAAATGGCTTTGTTCCATG
YNL091W_1_Syn	CTCAATGTTGAGCAGCAGTGGTTTTGCT	TGAGGTGATGTTTCTGCCACTCAACTCT
YNL088W_1_Syn	CTTCGGTAGTAGATGCGAGATCCCATTG	GCTTAAAGCACTGTCACCTTCGGTTAAG
YNL088W_2_Syn	CGGTTTGGCTCAAAATTTCGTCGGTAGT	TGGGATATAGGTTGACCAACCGGTACCG
YNL088W_3_Syn	CATTACCTTGAGCCCAGAGGAAATGGCC	CTTCTCAACCTCCCACTGCAATCTTTCA
YNL088W_4_Syn	CTTACAAAGAGACGCTGAAGCCAGAGGC	CTTGGTAGCAACGTCATCCTGTTCAGGG
YNL087W_2_Syn	TACCCCTCATGATATGAGCGACTTGGAC	GCCGCTCAACTCGAAGGTTAAATAAGGA
YNL087W_1_Syn	CGGTGACGACCACAACTTGGGTTCATTG	AGCTGAGGTGATAGCAGCATAGCCGCAT
YNL087W_3_Syn	CAGAACTATCAAGACCGGTTGGAGCGGT	ATCGCTTTGTGGTAATTCGGTAGCGGTA
YNL087W_4_Syn	AACCATCGACCCTGAGTCAGACACTACT	ACCGCCGAAAACACCAGCTTTAACCTTG
YNL085W_1_Syn	AGCTAGCATCAGCAGCACTACCACCAGT	GAACCAGCTACTATATGGGGCAACTTGA
YNL085W_2_Syn	TGCTTTGAGCCAAGCTACCTTGAGAAGC	AGCGGCGTTCAACTCCTGCATAACCTTA
YNL084C_1_Syn	TCTTTTGCTCAACTTGCTACCACCTTGG	CGACTTGAGAAGTAGCCAACCACCAACC
YNL083W_1_Syn	CTTGATTGCTCGTACCGACTTGAGTAGC	GCCACCACCCTCTCTGAACATGTCCTTA
YNL082W_1_Syn	TTTCCGTGGTGAGGCTTTGAGCAGCTTG	GACGGTCAAGCACTTAGTGAACTGTCTT
YNL082W_2_Syn	CTTACAGGCTGTTACCGTCTTCAAGAGC	ATTATCACGTCTCAAACCGCCATCTTCC
YNL080C_1_Syn	TGGACGCTTTGGTGATTGGCTGTGGCTA	ATATAGCACCGTTGCTGCTATCTTGCGT
YNL078W_1_Syn	CGCTTTGGTTGACAGCACCAGCAATAGC	ACGTCTTTGGCTGGTATGTCTGATAGGT
YNL077W_1_Syn	CACTCCAAGCAGCAACGGTAGTAAGTCA	ACCCAAGCCTTGACACTGTTGACAGATG
YNL077W_2_Syn	CTTGCAAGTTACTGTCCAACCAGGTAGC	GTACAAGTGGCCAAAGCCTGATCTGACG
YNL091W_3_Wt	TCTAGCAGCCACTTTAAGCTCGTGCTTT	CCGACCTTCTTCTAAACGGTCAGTTTCA
YNL091W_4_Wt	ATCTGCGAACATACTTTCCGCCAAACCT	TGGCTGTGGCAAGTGTGATTGTTCCATA
YNL091W_1_Wt	TAGCATGCTTTCTTCTTCCGGGTTCGCA	ACTAGTTATATTGCGACCGGAAAGCTCG
YNL088W_1_Wt	TTTTGGGTCCCGTTGTGAGATTCCTCTT	TGACAAGGCGGAATCCCCTTCTGTCAGA
YNL088W_2_Wt	TGGGCTAGCCCAAAACTTTGTTGGGTCC	AGGAATGTAAGTACTCCAGCCAGTGCCA
YNL088W_3_Wt	AATCACGCTATCACCTGAGGAAATGGCT	TTTCTCTACCTCCCACTGTAACCTTTCG
YNL088W_4_Wt	TTTGCAACGAGATGCAGAAGCTCGCGGT	TTTAGTGGCTACATCGTCCTGTTCTGGT
YNL087W_2_Wt	CACTCCACACGACATGTCTGATCTTGAT	ACCAGAGAGCTCAAATGTCAAGTATGGG
YNL087W_1_Wt	TGGCGATGATCATAATCTCGGGAGCCTT	TGCACTAGTTATCGCTGCGTAACCACAG
YNL087W_3_Wt	TAGGACCATTAAAACAGGCTGGTCGGGC	GTCTGATTGAGGCAATTCCGTGGCAGTT
YNL087W_4_Wt	TACTATTGATCCAGAGAGCGATACCACC	GCCACCAAAGACTCCGGCCTTTACTTTA
YNL085W_1_Wt	TGCGTCGATTTCGTCAACCACTACTTCC	AAACCAAGAGGAGTAAGGAGCCACTTGG
YNL085W_2_Wt	AGCCCTATCTCAAGCCACTCTTCGTTCA	CGCAGCATTTAGCTCCTGCATGACTTTT
YNL084C_1_Wt	CCGCTTAGAAAGTTTTGAGCCTCCTTGA	TGATTTACGGTCCTCTCAACCTCCTACA
YNL083W_1_Wt	TCTCATCGCAAGAACGGATCTATCCTCG	ACCCCCCCCCTCTCGAAACATATCTTTT
YNL082W_1_Wt	GTTTAGAGGGGAGGCCCTATCTTCTTTA	AACTGTGAGACATTTGGTAAACTGGCGC
YNL082W_2_Wt	ACTGCAGGCCGTGACAGTTTTTAAATCG	GTTGTCTCTTCGTAACCCACCGTCTTCT
YNL080C_1_Wt	GGGTCTTTTAGGGCTTTGTGAATGAGAG	GTACTCTACTGTGGCCGCCATTTTAAGA
YNL078W_1_Wt	TGCGCTTGTGGATTCTACATCGAACTCG	TCTTCGTTGTGAAGTGTGTCGAATTGGC
YNL077W_1_Wt	TACCCCCTCTTCTAATGGCTCCAAAAGC	GCCAAGACCTTGGCACTGTTGGCAAATA
YNL077W_2_Wt	TCTACAAGTCACCGTTCAACCGGGATCG	ATATAGATGACCGAAACCGCTCCGAACA

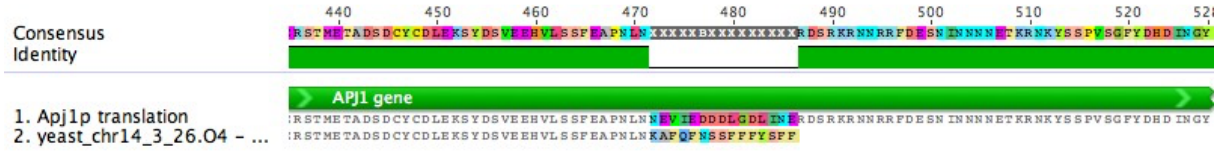

Figure 27: Sequence alignment between the APJ1 protein amino acid sequence (sequence number 1) and the megachunk O sequence (sequence number 2) using Geneious software. The introduction of the *URA3* marker resulted in the last 57 amino acids of APJ1 to be deleted and replaced by 16 amino acids encoded in the *URA3* cassette.

Table 8: Sequences of primers used throughout this thesis

Name	Sequence
chXIV loxP 5' flank-15N	GTTTTAGTTTAGTGCAGCCCACATACTACTNNNNNNNNNNNNNN
chXIV loxP 5' flank-15N-ATGTGG	GTTTTAGTTTAGTGCAGCCCACATACTACTNNNNNNNNNNNNNNNATGTGG
loxPsym sense random 15mer	ATAACTTCGTATAATGTACATTATACGAAGTTATNNNNNNNNNNNNNN
loxP 3 REV (extension for loxP 4)	CGTATAGGTTTCGAACTCGAGTTAGCGTTGTGTGAGCATCG
loxP 4 FWD (extension for loxP 3)	CAACGCTAACTCGAGTTCGAAACCTATACGCTCTGAGTTGA
loxP 3 FWD (extension for pRS)	ATTGGGTACAGGCAAAAACGGTGCAAGAC
pRS REV (extension for loxP 3)	TTTTTGCCTGTACCCAATTCGCCCTATAGTGA
loxP 4 REV (extension for pRS)	TTCGATATCCCACTACGAAGCGGTGAGAG
pRS FWD (extension for loxP 4)	TCGTAGTGGGATATCGAATTCCTGCAGCCC
URA3 FWD	GTTTTAGTTTAGTGCAGCCCACATACTACTAAGCTTTTCAATTCAATTCATCAT
URA3 REV	GTTTTAGTTTAGTGCAGCCCACATACTACTCCCGGGTAATAACTGATATAATTAA
WGA primer	GTTTTAGTTTAGTGCAGCCCA
loxP A1 FWD	ACCACAATCATCTCACAGTATGT
loxP A1 REV	AGTAGTATGTGGGCTGCACT
M13_F	GTAAAACGACGGCCAGT
M13_R	CATGGTCATAGCTGTTTCC
M13-loxP-15N	GTAAAACGACGGCCAGTATAACTTCGTATAATGTACATTATACGAAGTTATNNNNNNNNNNNNN
loxP A1 REV HindIII Extension	TATTATAAGCTTAGTAGTATGTGGGCTGCACT
MAT-A FWD:	ACTCCACTTCAAGTAAGAGTTTG
MAT-α FWD:	GCACGGAATATGGGACTACTTCG
MAT-locus REV:	AGTCACATCAAGATCGTTTATGG

Table 9: PEP PCR program (Arneson *et al*., 2008b)

Number of Cycles	Temperature °C	Time
1	94	2 min
50	94	1 min
	28 ramping to 55	>2 min (0.1 °C/sec)
	55	4 min
	68	30 sec
1	68	8 min

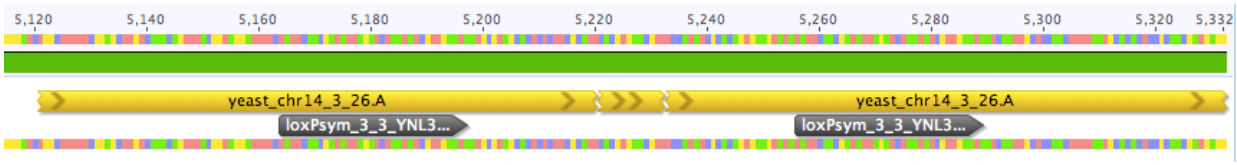

Figure 28: Alignment of loxP3-loxP4-pRS416 plasmid sequences. The top multicoloured line represents the sequence of *in silico* designed plasmid. The bottom multicoloured line represents the sequence of the same plasmid purified from *E. coli* and sequenced by Macrogen. The two long yellow bars are annotations showing the two loxP fragments with flanking sequences from SynXIV, while the short yellow bar shows the *Xho*I and *BstB*I. The two dark grey bars are annotations showing the exact location of the 34 bp loxP sequences (without flanking fragments). The green bar signifies 100% sequence alignment. This image represents a small portion of the total plasmid.

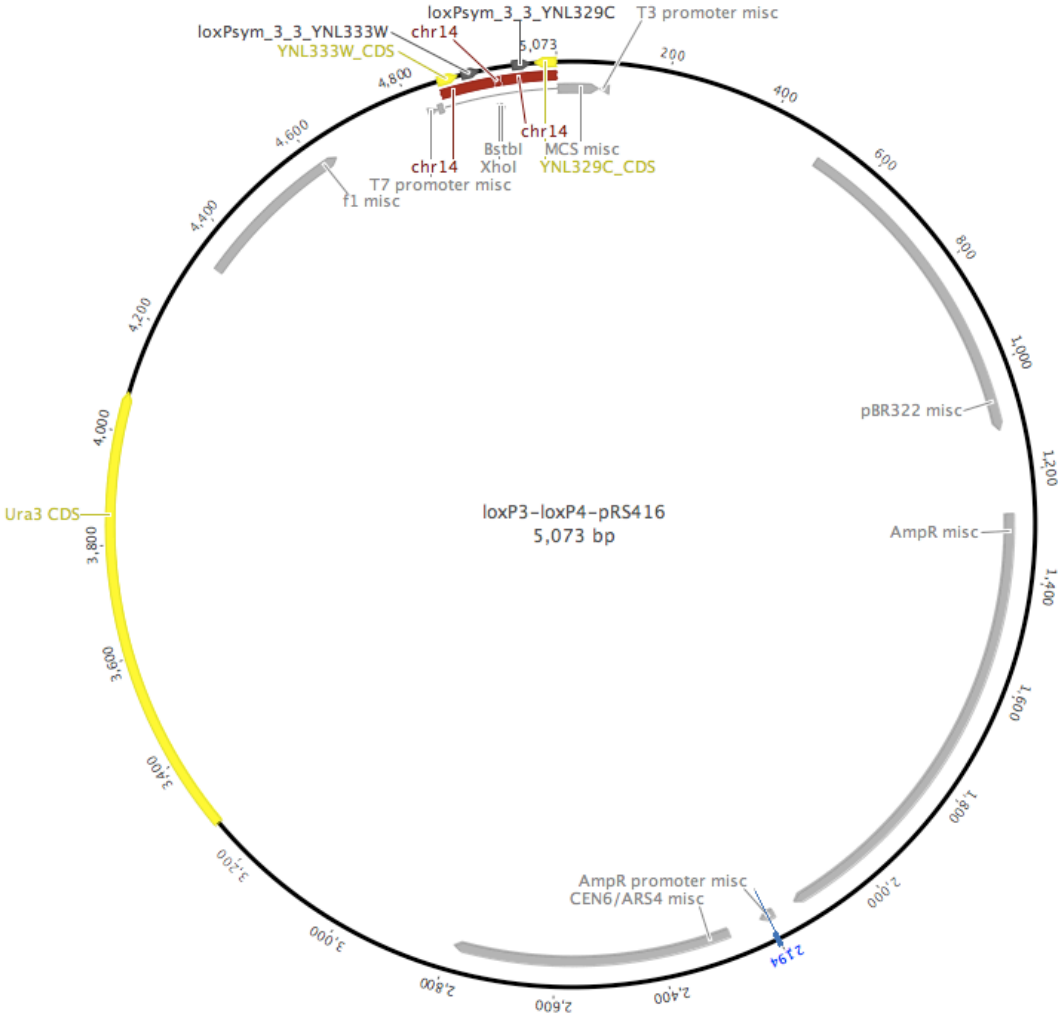

Figure 29: Plamid map of loxP3-loxP4-pRS416 showing major annotations.