Developing a Synthetic Yeast for the Expression of Heterologous Genes using SCRaMbLE

Dejana Jovicevic

Submitted for the Degree of Doctor of Philosophy

Supervised by Dr Tom Ellis
Imperial College London
Department of Bioengineering
Centre for Synthetic Biology and Innovation

March 11, 2016

ABSTRACT

Synthetic genomics is a new and fast emerging multi-disciplinary field of research, representing the largest scale of work underway in synthetic biology. The *Saccharomyces cerevisiae* version 2 (Sc2.0) project is currently the leading example of synthetic genomics research, and is an attempt to perform the first redesign, synthesis and assembly of a complete synthetic genome for an eukaryotic organism. Major changes are being made to the redesigned DNA sequence of the new yeast genome and these include a variety of different deletions, sequence recodings at every gene and also insertions of DNA motifs. The most significant insertion is the placement of a recombinase recognition site called loxPsym throughout the genome, placed downstream of all non-essential genes. These recombinase sites act as recombination hotspots for Cre recombinase, to bring about recombination-mediated genomic rearrangements of the synthetic chromosomes. Together the Cre recombinase and loxPsym inserts make up the inducible Synthetic Chromosome Rearrangement and Modification by LoxPsym Evolution (SCRaMbLE) system.

This thesis describes the design, synthesis, hierarchical assembly and *in vivo* integration of synthetic DNA for the construction of synthetic chromosome XI for the Sc2.0 project. With the first 90 kb of the synthetic chromosome complete, the SCRaMbLE toolkit was then examined. It was hypothesised that along with causing gene deletions, inversions and duplications, this Cre-lox system could also be implemented to insert heterologous DNA into the synthetic chromosomes. This thesis shows that with the correct formatting of heterologous DNA, SCRaMbLE can be further developed to generate a new synthetic biology method called 'SCRaMbLE-in' suitable for the insertion of heterologous genes into synthetic chromosomes as they are rearranging to produce diverse synthetic yeast strains with novel functions.

Having successfully developed and investigated SCRaMbLE-in, this method was then used for the simultaneous introduction of multiple genes that can confer a selective benefit to yeast. By providing three heterologous genes encoding enzymes that together reconstitute the oxidoreductase xylose-utilisation pathway, a synthetic yeast strain capable of growth on the lignocellulosic sugar xylose was produced by SCRaMbLE-in. This work thus demonstrates a new approach to constructing strains for metabolic engineering projects, where incorporation of heterologous genes and rapid evolution of the yeast genome can be done simultaneously in one pot.

DECLARATION

I herewith certify that all the material in this thesis is my own work, except for quotations from published and unpublished sources which are clearly indicated and acknowledged as such. The source of any picture, diagram or other figure that is not my own work is also indicated.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr Tom Ellis, for giving me a chance to be part of his lab and work on such a fantastic project. I had a very impromptu start to my PhD course and was naïve and generally clueless about it. He willingly accepted me as a member of his group, and has spent many hours showing me the basics in the office and in the lab. His continuous support, advice, understanding and above all, patience, have been invaluable, and I wouldn't have been able to do this PhD without him. I am truly grateful for both the opportunity to be part of your lab, as well as to have been able to experience and learn over the last four years of my life about my subject, and myself.

Secondly, I must give mention to the post-docs in my group who have had to put up with my incessant questioning, comments, conversation (not always scientific), and the general noise that I always generated. To Dr Ben Blount, who was tasked with the most jejune of tasks, showing me how to pipette, how to miniprep, and generally everything in the lab. I owe pretty much all my laboratory skills and training to you. You've clarified so many issues and misunderstandings I have had in my head over the years, spent hours and hours on end helping me set up experiments, checking my work, helping me with my figures, and dealing with my random questions. To Dr Ali Awan, who is always ready to give advice, to listen, to talk, to joke, to help, to laugh, to proof read this thesis. Thank you to you both.

To everyone in the Ellis group, you have all contributed in some way or another. To my fellow PhD, George, it has been a hilarious day-to-day rollercoaster ride, and I am honoured to have shared it with you, and to call you my friend. To all my friends out there, especially the non-scientific ones who hadn't a clue about all the jargon I've thrown at you, but you all received it so well. There are so many of you to mention, you know who you are, you have all helped me on this journey.

To my five amazing sisters, Maja, Jovana, Tamara, Tanja and Angelina for always being there for me, throughout my whole journey, always ready for a giggle and always offering advice. I love you all very much. To my aunt Vesna who has guided me from the beginning and made all this possible. To Christina, Howard and Maria who have been a tremendous help and support.

To the best parents possible, Zoka and Bata, for you have never stopped listening and being there. Although you haven't always understood what I was talking about, it was never an issue, and you never stopped believing in me. You've been my driving force in the past four years, financially, emotionally and mentally. I am so very lucky to be your daughter. I dedicate this thesis to you.

Finally to Alex. Words simply aren't enough but know
that I couldn't have done it without you.
I could never thank you enough.

1.0 CHAPTER 1: Introduction

1.1 Synthetic Biology and Synthetic Genomics

1.1.1. The Rise of Genetic Manipulation and Biotechnology

Scientists have been manipulating DNA for over half a century in a bid to explore and exploit the nature, structure, function and interaction of all domains of life. There is a long history of humans engineering organisms, the most primal example perhaps being found in agriculture from selective breeding, for the regeneration of particular traits or phenotypes. Pivotal progress was made with the discovery of DNA by Franklin, Wilkins, Crick and Watson in 1953[1]. This led to significant genetic engineering feats, especially after the chemical synthesis of DNA oligonucleotides was first demonstrated in the 1960s[2, 3], and the chemical manufacture of the first synthetic gene was achieved in the 1970s[4]. Along with these advances in molecular biology, our ability to use DNA to create and transform life has also been aided by great progress in techniques from cell biology. The development and application of IVF in 1978 by Robert G Edwards[5], whose research was awarded the Nobel Prize for Medicine was a forerunner for nuclear transfer techniques that have since been used for cloning animals, such as the famous Dolly the Sheep[6]. Combining the transfer of genomes, chromosomes and DNA in and out of cells with the molecular cloning techniques developed in the past five decades has led to a biotechnology revolution. This has especially been aided by research such as the discovery of restriction enzymes[7, 8], allowing precision engineering of DNA so that genes and other DNA sequences can be cut-and-pasted into both prokaryotic and eukaryotic organisms.

Recombinant DNA applications are now widely dispersed in medical, chemical, agricultural and food industries. The first medically licensed recombinant DNA therapy was the production of insulin[9] which entered the clinic in the 1980s. Since then there has been a step change in the scale of gene manipulation achievable in both industry and research, and in the last 15 years this has given rise to new interdisciplinary field of synthetic biology.

1.1.2. Synthetic Biology

There are several definitions suggested for synthetic biology, but the most widely-accepted is: "*the design and construction of new biological parts, devices and systems, and redesign of existing, natural biological systems for useful purposes*"[10]. Synthetic biology is a diverse field incorporating elements from computer science,

genetic engineering, bioinformatics, modelling, cell biology and systems biology in order to engineer novel biological systems for specific applications and purposes[11] (Figure 1.1). The emergence of this new field has led to the characterisation and redesign of genes, promoters, terminators and other genetic regulatory elements and their use for the combinatorial construction of pathways, regulatory networks, chromosomes, and even genomes.

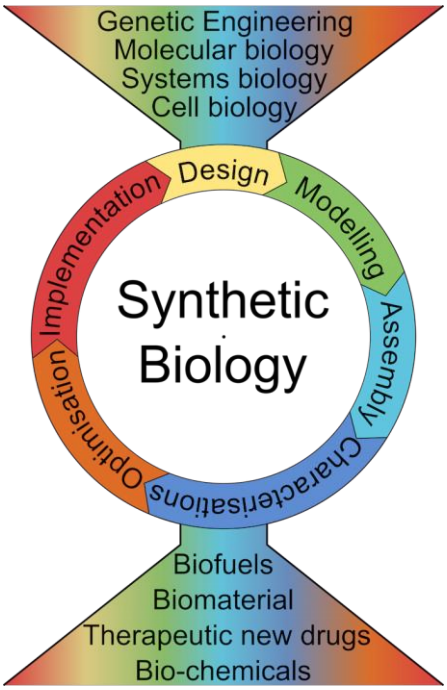


Figure 1.1 Synthetic Biology. Schematic illustration of the key processes and implementations associated with synthetic biology. The combination of genetic engineering, systems, molecular and cell biology to engineer genes, metabolic pathways and systems has evolved into the new discipline of synthetic biology. This new approach of rational design, DNA assembly, followed by the characterising, optimising and implementing has the potential to give rise to efficient, useful and functional products such as biomaterials, novel therapeutic drugs, biomaterial and bio-chemicals.

The expansion of synthetic biology as a field has been made possible due to the decreasing costs of synthetic DNA manufacture[12] as well as many other biotechnological developments. Examples of prokaryotic and eukaryotic systems engineered for new biological purposes include genetic networks designed to change the function of cells to behave as oscillators[13] toggle switches[14], or to perform multicellular pattern formation[15]. Other achievements from the field that have potential global impacts are the reprogramming of microbes to sense and detect biofilms[16], the production of whole-cell biosensors[17] that report the detection of environmentally harmful chemicals, and cells that produce artemisinic acid[18], the precursor of the important antimalarial drug artemisin.

1.1.3. Synthetic Biology Limitations and Challenges

Despite many advances in this field, there are still practical limitations to what synthetic biology can do. Roberta Kwok reviewed the challenges faced by synthetic biologists in 2010, and stated that the first and foremost issue was that there is a lack of characterisation and standardisation of all biological parts (such as promoters, terminators, or genes)[19] which impedes the construction of ever more complex context-independent genetic circuitry. For example, promoter A can be characterised and shown to function under a certain set of conditions, but i) as there is a lack of standardisation across the field[20, 21] this part may not be suitable for others to use and ii) whilst the promoter may show functionality in host organism X with gene Y to produce Z, it may not exhibit the same properties and interactions in a different host cell or when paired with a different gene. The Registry of Standard Biological Parts (partsregistry.org) is an online catalogue of biological parts that directly aims to tackle the first of these issues, and facilitates the building of genetic circuits from a large database of modularised and interchangeable parts. It does not answer the second issue; that parts do not exhibit true modularity across biology.

Another challenge reported by Kwok was the lack of predictability in the dynamic behaviour of genetic circuits. This invariably limits the complexity of biological systems constructed, and was previously discussed by Purnick *et al* (2009). In that review the authors pointed out that a greater understanding of the biological environment needed to be integrated into synthetic biology in order for greater progress. In particular it is important to consider the inherent intra- and extracellular processes, along with traits such as noise, crosstalk, and cell death[22]. Tuning the expression of (multiple) gene circuits and changing cell behaviour is not an easy task due to the complexity of such processes inside the cell and because of the trade-off between the cell's intentions and the engineer's intentions. A proposed way to get around this is to expand work on genetic circuits to "cell-free" systems based on the extracted intracellular machinery[23]. To this end, Forster and Church (2007) reported the *in vitro* synthesis, replication and evolution of DNA[24], RNA[25], protein[26] and small molecules[27] that bypass dependence on cellular machinery. These are not restricted by membrane compartmentalisation and so increase the levels of flexibility for engineers[28]. Another possibility is to consider and account for the unwanted but unavoidable interactions between genetic circuits and the natural process inherent to cells. For example, researchers have begun to measure and design around the burden on gene expression that effects a cell when it is used a 'chassis' for running a synthetic genetic circuit[29].

1.1.4. Genome Editing Methods

Other areas of synthetic biology have diverged away from the engineering of genetic circuits and new biological systems, and instead have focused on developing methodologies for whole genome manipulation. Zinc finger nucleases (ZFNs) were one of the primary genome editing systems to be reported and combine the use of a restriction enzyme domain (such as FokI) with a programmable DNA binding domain, called zinc finger proteins, which bind a target DNA sequence for a tethered FokI enzyme to cleave nearby and cause double stranded breaks (DSBs) in the DNA sequence. Double stranded breaks are then repaired using non homologous end joining (NHEJ) which can lead to insertion or deletion mutations (indels) at the selected location, causing frameshifts in the coding region. The same technology can even cleave large segments of DNA by utilising multiple ZFNs in parallel[30].

TALENS, or transcription activator-like effectors nucleases, are similar to zinc finger domains wherein a nuclease domain has been paired with a DNA-binding domain, in this case from a transcription activator like effector (TALEs) protein. TALEs were discovered in the plant-infecting *Xanthomonas* bacteria species, as a mechanism to alter the plant's native gene expression and bring on bacterial infection[31]. TALENs are now used as customisable genome editing tools that also causes DSBs at specific loci. These are used to study the effects of gene knockouts, and can be modified to also provide transcriptional control over gene expression rather than cutting the DNA[32].

The most recent and significant targeted genome editing technology to be reported is the clustered regulatory, interspaced, short palindromic repeat (CRISPR) and Cas9 coupled RNA-guided DNA cutting technique. This comes from a prokaryotic defence system used to recognise phage sequences and break them down prior to infection of bacterial cells. CRISPR-Cas9 facilitates targeted *in vivo* genome editing by providing the Cas9 nuclease with a guide RNA (gRNA) that directs the Cas9 nuclease to the particular nucleotide sequence within the genome that a user wishes to cleave. Since its discovery and particularly due to the easy manipulation of guide RNAs to target different specific DNA sequences, CRISPR-Cas9 has been implemented in a wide variety of organisms including bacteria[33], zebrafish[34], the silkworm[35], yeast[36], mouse[37] and even in human embryos[38] in order to cleave genomic DNA generating double stranded DNA breaks[39] or even just generating single-stranded nicks[40].

1.1.5. Synthetic Genomics

The above described genome editing technologies are a top-down approach, which are used to bring about changes to existing genomes. In the past few years, scientists from synthetic biology have also taken a different bottom-up approach, with the aim of designing and assembling synthetic genomes that work in cells. This field, more commonly referred to as synthetic genomics, applies gene synthesis, DNA assembly and genome engineering tools together to construct synthesised chromosomes or drastically modify the DNA content throughout existing chromosomes to give new functions[41].

The team from the J. Craig Venter Institute was the first to report the synthesis and assembly of a complete bacterial genome, constructing the 582 kb *Mycoplasma genitalium* chromosome, in 2008[42]. They followed this up in 2010 with a 1 Mb synthetic chromosome encoding the *Mycoplasma mycoides* genome, which was shown to be functional when transplanted into a host cell whose own genomic DNA had been ablated. This yielded the world's first synthetic cell, a *Mycoplasma* bacterium programmed by a synthesised genome[43]. This landmark research demonstrated that synthetic DNA could be assembled and be functional up to the genome scale; but the DNA and genetic content encoded by the synthetic genome was little changed. The genome sequence was intentionally kept almost identical to the natural genome sequence, with only a few silent watermark motifs incorporated to distinguish the synthetic genome from the natural DNA. The next major challenge for synthetic genomes, therefore, would be to build upon this achievement but to do so with redesign.

In 2013, a successfully redesigned genome arrived, when researchers reported completion of a long term project to systematically edit the 4.6 Mb chromosome of *Escherichia coli* to remove all instances of the TAG stop codon and replace these with the TAA stop codon, freeing-up the TAG codon to be repurposed for the intentional incorporation of a non-natural amino acids in desired positions of any *E. coli* expressed proteins[44]. Rather than synthesising and assembling a synthetic *E. coli* chromosome containing the DNA sequence changes required, the approach taken here was to use genome engineering tools[45, 46] to edit the natural chromosome at hundreds of positions in parallel. In contrast to the previous projects of the J. Craig Venter Institute, the new *E. coli* chromosome was not synthetically assembled, but did endow the bacterial cells with an impressive new ability of great value to biotechnology and bioscience research: the ability to encode amino acids beyond the canonical 20 at desired positions within proteins.

The works presented by Isaacs and Gibson in these two cases were based on prokaryotic organisms, which have small genomes and less than 5000 genes. The next step in synthetic genomics is to apply the bottom-up approach of genome synthesis to the next level of complexity in life. This is now being done by demonstrating genome and chromosome synthesis in a eukaryote, the model organism *Saccharomyces cerevisiae*.

1.2 The Synthetic Yeast Project: *Saccharomyces cerevisiae* Sc2.0

1.2.1 *Saccharomyces cerevisiae* Yeast

The eukaryotic organism, *S. cerevisiae*, known as Baker's or Brewer's yeast, has long been exploited by humankind, and is arguably our oldest example of biotechnology. It is known for being straightforward to culture, generally regarded as safe (GRAS) to work with by the FDA[47] and in the past three decades research of its well-characterised genome[48] has revealed that is has a relatively compact and stable genome, lacking much of the complexity of genomes seen in multicellular organisms. In addition, it has high tolerance to ethanol[49], making it very suitable for industrial scale fermentation. The prevalent use of *Saccharomyces cerevisiae* in the biotech industry remains the use of this yeast for the production of bioethanol[50]. More biotechnological uses relating to metabolic engineering are given in Section 1.3.1 of this chapter.

S. cerevisiae yeast is a unicellular organism and one of the simplest examples of a eukaryote, the kingdom that incorporates animals, fungi and plants. It is also widely-used, both commercially and in research, where it is particularly relevant as a model organism for how eukaryotic genomes operate and evolve. Its partial orthogonality with much of the processes of multicellular eukaryotes has helped yeast achieve the production of over 70% of recombinant mammalian proteins that have previously failed to be generated in *E. coli*, thus making yeast the second-most important organism for biologics production. Its suitability for this is a result of having much of the post-translational machinery necessary for the proper folding and structure of functional mammalian proteins[51]. While not as complex as a mammalian cell, yeast can at least be a model for many of the process which we wish to study.

The 16 chromosomes that make up the genome of yeast were some of the first large scale DNA sequences to be revealed[52], and the *S. cerevisiae* genome was the first completed eukaryotic genome sequencing project, finished in 1996[53]. Given our in-depth understanding of this organism and its genome, it is an ideal candidate to extend synthetic genomics beyond bacteria. From this point on, unless

explicitly stated otherwise, the use of the term yeast will exclusively apply to *Saccharomyces cerevisiae*.

1.2.2. Design Principles for a Synthetic Yeast Genome

The international synthetic yeast genome project aims to synthesise and assemble the DNA of all 16 yeast chromosomes, and to incorporate into this sequence a significant number of design changes intended to test biological hypotheses and provide new ways to exploit yeast as a valuable technology. The project was conceived by Professors Jef Boeke and Srinivasan Chandrasegaran from Johns Hopkins University and as it began to take shape, they sought help from computer science expert and colleague Dr Joel Bader to guide the design tools and data management such a large project would require. By 2007, the idea had been implemented into an active venture to make a *S. cerevisiae* genome version 2.0; giving the project its now widely used label 'Sc2.0'.

Before any synthesis and assembly could be done, the design changes for a synthetic genome needed to be agreed upon. Based on existing knowledge of yeast and eukaryotic genetics and how DNA sequences encode function, decisions were made as to what could be removed, relocated and added to the synthetic genome. The genome redesign needed to strike the balance between making enough changes for the new yeast to be easily exploited and yet being conservative enough that the resultant yeast still grows and behaves in the lab like the strains of *S. cerevisiae* from which it is derived. The redesigned chromosomes of Sc2.0 may differ from their natural counterparts by only a few percent of the DNA bases, but across the scale of the genome this represents a large number of modifications and a significant increase compared to the minor changes made in the previously described bacterial synthetic genomics project.

The changes incorporated can be categorised into four types; deletions, replacements, translocations and insertions (summarised in Figure 1.2). Firstly, in accordance with the design principle that the synthetic yeast genome must exhibit equal or better fitness of the parent strain, all genome destabilising elements were removed, including retrotransposons or Ty elements that are highly repetitive and unstable[54] regions of DNA that contribute to chromosome loss. The telomeric regions of *S. cerevisiae* chromosomes share structural and functional similarity with all eukaryotic chromosomes conserving chromosomal length, and preserving the segregation, and completing replication[55]. However, the subtelomeric regions, which have an unknown function and lack of genes, are made up of a widely conserved X core and Y' elements, contain highly repetitive sequences and are classed as

destabilising. The Y' elements were therefore removed from the design of the Sc2.0 genome. Out of the 6000 genes, 285 contain introns of unknown function and are regarded as non-essential in the majority of cases. Deletion of all non-essential introns is thus another design feature of the Sc2.0 yeast genome[56].

Secondly, as a result of the research demonstrated by Isaacs *et al*, the Sc2.0

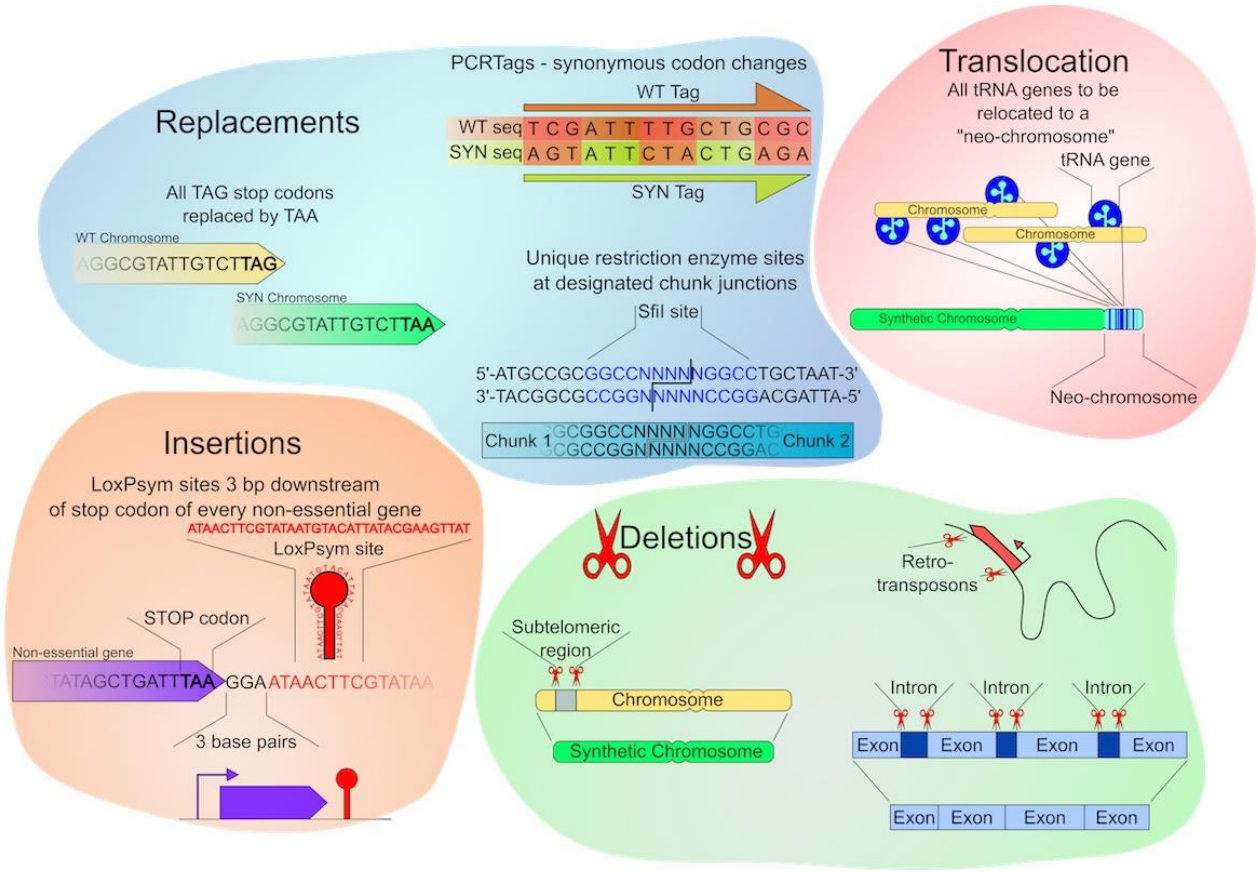


Figure 1.2 Sc2.0 Design Principles. Summary of the changes implemented into the Sc2.0 genome. The three types of insertions (blue) include the encoding of loxPsym sites three base pairs downstream of all non-essential genes, the synonymous codon changes to facilitate distinction between the wild type (WT) and synthetic (SYN) genomes, and the inclusion of unique restriction sites at all chunk or mini-chunk borders in order to enable the building of the synthetic chromosome. All tRNA genes are excised from their native chromosomal loci onto a neo-chromosome (red), and replacements include the recoding of all TAG stop codons to TAA (tan). Lastly, the deletions shown above include the excision of all retrotransposons, introns and subtelomeric regions from the novel genome as a result of their intrinsic destabilising effects on genome stability.

team replaced all TAG stop codons with TAA in the genome design. Again this was in order to free up a codon for potential future use in encoding non-canonical amino acids into proteins as and where is desired. Other elements were recoded to facilitate screening and construction. Unique restriction enzyme sites that are used for the hierarchical construction of the synthetic genome from small fragments of DNA to much larger sections up to 50 kb. The unique restriction sites were encoded by

introducing small synonymous recoded codons within protein-coding sequences in order to define and cleave the borders of the synthetic fragments being assembled. A second type of short synonymously recoded region in the protein-coding sequences of the synthetic genome are codon changes named “PCRtags”. These codon changes act as orthogonal primer-binding site sequences in the genomic DNA, designed specifically for primer pairs to anneal to distinguish between the wild type DNA or the synthetic DNA. They are used as a verifying tool for researchers to distinguish between the original and synthetic DNA using a simple Polymerase Chain Reaction (PCR) process. Following integration of the synthetic DNA, the oligonucleotides designed to anneal only to the synthetic DNA encoded into each protein coding sequence will be amplified, making a PCRtag. If amplification of genomic DNA occurs with the oligonucleotides matching synthetic DNA, but not with the corresponding oligonucleotides that match the natural DNA, then synthetic DNA incorporation and replacement of wild type DNA is successful.

The third alteration was the relocation of all tRNA genes from their natural loci to a novel “neo-chromosome”. Moving tRNAs was proposed as they are known to encode unstable areas into chromosomes and because they have inherent redundancy[57]. However, unlike subtolemeric regions due to their essential function in cells, their sequences cannot be deleted entirely[58]. By relocating the tRNA genes onto the neo-chromosome, researchers are in some respects mirroring the genomic architecture of other yeasts, such as the well-studied yeast *Schizosaccharomyces pombe*, which organises its tRNA genes into clusters on chromosomes[59].

The final design alteration, and perhaps the most significant, is the insertion of loxPsym sites embedded into the synthetic genome to increase the flexibility of the chromosomes and allow massive evolution on demand. This insertion is a 34 bp palindromic sequence recombinase substrate site and is a key component to the Sc2.0 project. These loxPsyms make up part of the Synthetic Chromosome Rearrangement and Modification by LoxPsym Evolution (‘SCRaMbLE’) system, and they are designed to cause large genomic rearrangements. SCRaMbLE is an important component of this thesis and is reviewed in detailed in Section 1.25 of this Chapter.

1.2.3 Synthetic Chromosome III

Following the earlier work by others in bacteria, in 2014 a synthetic yeast chromosome became the third example of a synthetic chromosome that had been completed and shown to be functional in a cell[60]. This was also the first of its kind in a eukaryote. In constructing the synthetic version of the third chromosome of the

yeast *S. cerevisiae* (SynIII) the Sc2.0 team [60]incorporated both the synthesis and assembly approaches from the J. Craig Venter Institute projects with the redesign and novel function concept of the recoded *E. coli* project. The resulting synthetic chromosome was the impressive first step in the largest synthetic biology project currently underway globally.

As with the historic yeast genome sequencing project that ran two decades before, the first chromosome to be entirely synthesised was one of the shortest, chromosome III. It was originally 317 kb in size, but the new computer-aided SynIII design resulted in a more compact 273 kb chromosome: smaller than the synthetic bacterial chromosomes previously described by Gibson *et al*, but still a major feat. With the decreasing costs of custom DNA synthesis[12] one way to do construction would be to outsource synthesis of 1-10 kb DNA chunks to a company, as previously done by the J. Craig Venter Institute[42]. These could then be stitched together using a variety of different DNA assembly protocols[61], such as by Gibson Isothermal Assembly[62], Golden Gate Cloning[63], In Fusion[64], SLIC[65] or others.

However, for the assembly of SynIII the team, then based at Johns Hopkins University, took on a more hands on approach. They founded the undergraduate class called Build-a-Genome (BAG)[66, 67], where students were supervised whilst working to gradually assemble all the 60-80mer DNA oligonucleotides that encode an entire chromosome. Students initially used polymerase cycling assembly[68] to stitch together 1-20 of their oligos into ~750 bp building blocks (BB) and then worked in teams to ligate these together into “mini-chunks” of DNA, approximately 2–3 kb in size, sometimes using Gibson Assembly[62], but mainly using USER assembly[69]. USER assembly is based on the use of specially designed primers that contain a deoxyuridine (U) residue in the 5’ end of primers to amplify PCR products. The U-sites are cleaving substrate for deoxyuridine excision reagent for the digestion of the PCR products to form 3’ overhangs on the BBs. The building blocks were designed to be complimentary to each other, thus upon USER digestion, the BBs overhangs are ligated with each other to form ~3–4 kb mini-chunks (Figure 1.3). Issues with USER assembly include the necessity to create large libraries of specific primer pairs for the amplification of all the BBs with U-oligonucleotides, as well as being a time consuming process and the use of specific reagents. However, the reported efficiency rates of USER fusion are around 90% of correctly formed products[70], and can be manipulated for both *in vitro* and *in vivo* cloning.

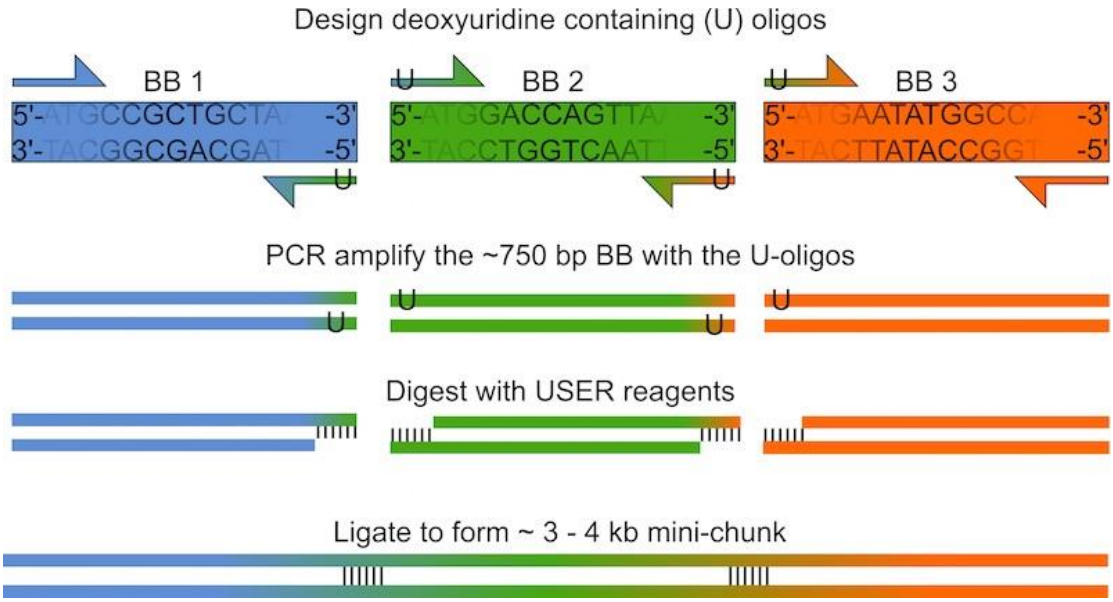


Figure 1.3 USER Assembly. The cloning strategy utilised by BAG students to assemble mini-chunks. Building blocks (BBs) were constructed from overlapping 60-80mer oligonucleotides cycled through Polymerase Chain Assembly to form the ~ 750 bp BBs (not shown), sharing regions of homology with each other. The BBs were amplified using specifically designed deoxyuridine (U) residue primers in the 5' ends, and the PCR products were cleaved at the U residues to form 5-9 bp of 3' overhangs. The 3' overhang of BB1 is complementary to the 3' overhang of BB2, and the building blocks are ligated to for a 3-4 kb mini-chunk which is directly transformed into yeast for homologous recombination for the building of synthetic chromosome III.

The student-made mini-chunks were sequence-verified and then transformed directly into yeast a linear DNA pieces. They were designed to have significant regions of homology with one another (a single mini-chunk shares one BB of homology with another mini-chunk) and also with the existing genome, and so directed by this homology, multiple recombination events occur inside the yeast cell, building the synthetic chromosome *in vivo*. The mini-chunks made by the BAG class were designed to be integrated into chromosome III and this was performed using yeast's native homologous recombination machinery, which efficiently assembles together DNA ends that share at least 40 bp of equivalent sequence.

The power of utilising yeast recombination to assemble DNA constructs *in vivo* was extensively reported previously by Gibson *et al* (2008) who used the intrinsic yeast machinery to assemble the *Mycoplasma genitalium* genome by providing 25 overlapping fragments into the nucleus of the yeast cell. The SynIII chromosome and its intermediates use the same recombination principles to hierarchically assemble the synthetic chromosome using iterative rounds of transformation and recombination.

The limitations of yeast recombination are currently not fully known and during the construction of SynIII, Annaluru and colleagues reported using up to 12

mini-chunks at a time to assemble ~30–40 kb of the chromosome in one go. Unlike the previous all-in-one synthetic *Mycoplasma* genome projects where a single circular chromosome needs to be constructed, in the case of SynIII construction of the synthetic genome requires recombining the mini-chunks directly into an existing chromosome to replace the native DNA. As this is not done in one go, the construction creates a series of hybrid part-natural-part-synthetic chromosomes along the way to completion, each of which is checked for sequence fidelity and growth in multiple conditions. This allows those doing the construction to track any detrimental phenotypes as and when they arise. Currently, the only drawbacks of using yeast recombination include basal error rates[71], improper joining of DNA by non-homologous end joining (NHEJ), and the necessity to culture yeast for several days on agar plates to obtain viable colonies.

Following the introduction of any synthetic yeast genome DNA into the yeast cell, integration is verified using PCRtag analysis that amplifies the synthetic DNA. The PCRtag verification steps following synthetic DNA integrations are the most time consuming steps of the process of constructing synthetic yeast chromosomes and they have to be done every round before the next integration can begin. The PCRtags are useful to confirm the presence of the synthetic DNA (and loss of the wild type DNA) but they cannot differentiate between small and undesired discrepancies, such as single nucleotide polymorphisms (SNPs), which may or may not be recognised in the ensuing fitness tests. Identification of SNPs in synthetic chromosomes is only carried out by genome sequencing.

The Sc2.0 SynIII chromosome took several years and nearly a hundred students and researchers to complete. For the Sc2.0 team to realise their dream of a completely synthetic genome, they realised they would need more help. Following the historic model of the international collaboration to sequence the yeast genome in the 1990s, Sc2.0 went global in 2011, with groups from around the world joining in the effort to synthesise, assembly and verify all the remaining chromosomes needed for a synthetic yeast. The International Synthetic Yeast Genome Project now includes groups from countries as diverse as China, Singapore, Australia, the United Kingdom and across the USA, and each team is being allocated work on individual chromosomes with the aim of bringing these together by yeast mating in order to finish the genome project before the end of the decade. By then, it is expected that the cost of synthetic DNA will have decreased by orders of magnitude, especially if innovations in constructing synthetic DNA directly from printed oligo arrays continue[12]. Considerably, lower costs will make it likely that yeast or bacteria chromosome redesign and synthesis will be within the reach of the average research

group by 2020. The Sc2.0 project therefore provides a framework to develop the methods and technologies for work that will become commonplace in the future. Innovations on parallelising the step-by-step integration (e.g. by exploiting meiotic recombination between multiple versions of semi-synthetic chromosomes) and using new technologies to accelerating verification steps (e.g. next generation sequencing) will certainly aid in making future projects not only cheaper, but also faster.

In the most recent publication on synthetic yeast, the full synthesis of SynIII[60] the authors also went to great lengths to proves that yeast with SynIII no longer contains the natural chromosome III or any of its natural genes, only the water-marked synthetic versions. Whole genome sequencing also showed that the chromosome DNA sequence was almost exactly as designed, and a variety of assays in the lab demonstrated that this new yeast with one chromosome entirely replaced by its synthetic counterpart can still grow in a variety of different conditions with equal efficiency as the standard lab yeast from which it was redesigned. The fact that so many design changes can be applied to the DNA sequence of a whole chromosome with such minimal impact on cell fitness makes this breakthrough truly impressive. The radical recoding inherent in the design of the synthetic yeast genome project is thus validated, even at the full-chromosome scale.

1.2.4 Synthetic Chromosomes VI.L and IX.R

The building of chromosomes VI (SynVI) and IX (SynIX) began after SynIII construction began but used different approaches. Successful replacement of regions of these two chromosomes with synthetic genome DNA was reported by Dymond *et al* (2011) three years before SynIII was completed and published. In this paper the details of construction are given and show different approaches. The right arm of SynIX (SynIX.R) was synthesised as a 91 kb bacterial artificial chromosome (BAC) by Codon Devices, whereas a region of the left arm of SynVI (SynVI.L) was outsourced for synthesis as chunks ranging in 2–10 kb in size from Epoch BioLabs. In this latter case, the chunks were ligated together to make 30 kb mega-chunks which were then integrated into yeast. SynIX.R construction was the most distinct when compared to that of the other chromosomes, in that it was introduced as a circular BAC into a truncated diploid strain that lacked a copy of the right arm of chromosome IX. Consequent sporulation experiments to generate haploids isolated the haploid strain containing the native left arm of chromosome IX only, as well as the BAC providing the synthetic copy of the right arm of chromosome IX[72]. Semi-synthetic chromosome VI used a digestion-ligation coupled method to assemble the

SynVI.L synthetic DNA *in vitro* first, before transforming into the yeast cell. This method is now the most commonly-used method in the global Sc2.0 project.

In this 'mega-chunk method' (Figure 1.4), the purchased chunks for synthetic chromosome VI were provided in bacterial vectors which were first isolated and linearised by digestion at the unique restriction sites embedded within the synthetic sequence. The 3' overhang generated at the first chunk anneals to the 5' overhang of the second chunk, and the 3' overhang of the second chunk anneals to the 5' overhang of the third chunk, and so on. Chunks can thus be stitched together by

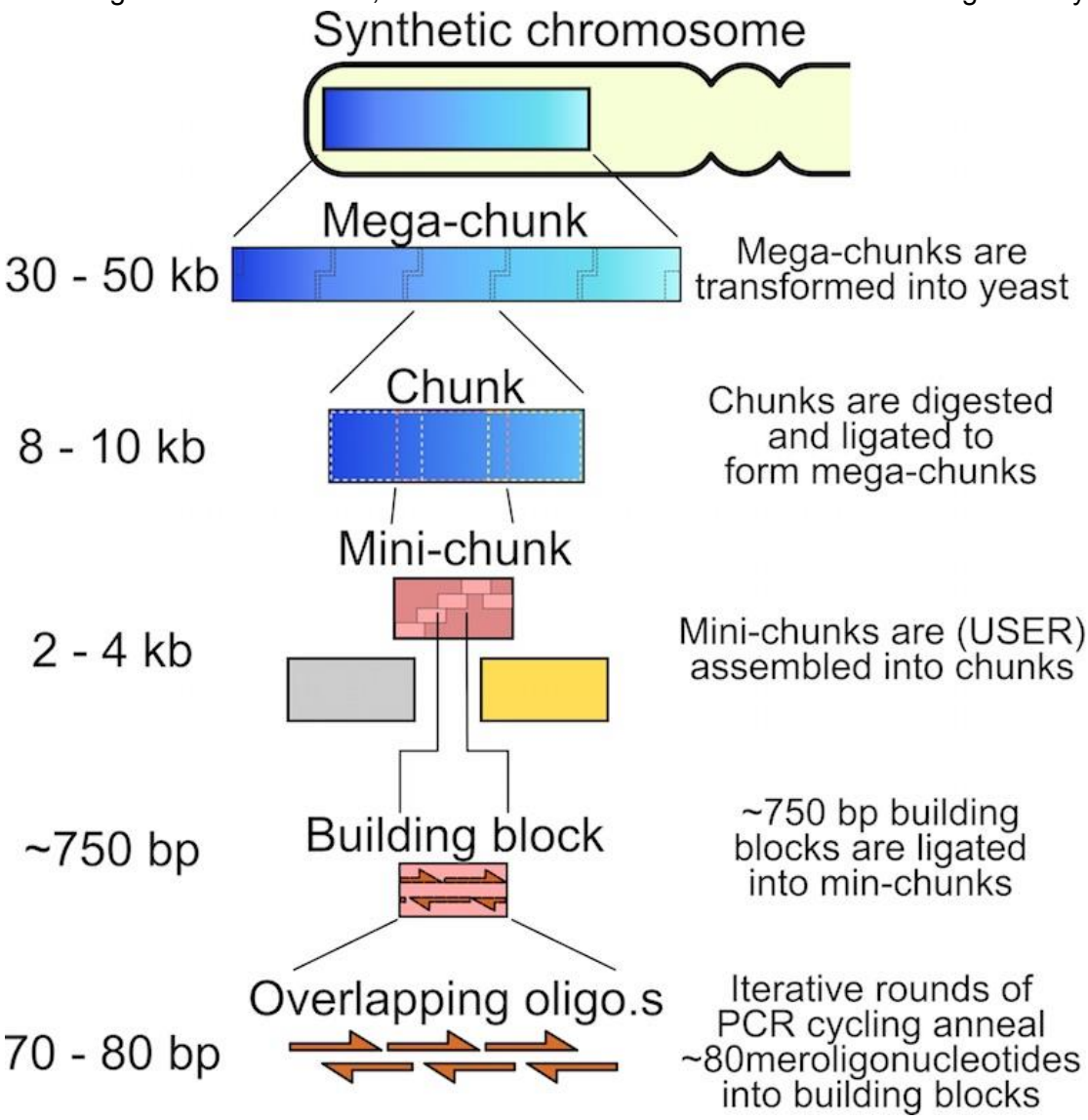


Figure 1.4 Hierarchical Assembly of Sc2.0. The hierarchy and strategy employed in the Sc2.0 project for the construction of synthetic yeast chromosomes. 70-80 bp oligonucleotides are ordered from DNA synthesising companies and assembled into ~750 bp building blocks (BB) by B-A-G students. BBs are ligated to form mini-chunks, which are the basis of chunks. In general, now most groups outsource synthetic DNA as chunks to optimise time and effort invested in building synthetic chromosomes. Chunks are digested using complimentary restriction enzymes to create 3 to 4 base pair overhangs, used to ligate chunks into a single 30 to 50 kb mega-chunk. Mega-chunks are transformed into yeast to replace the wild type chromosome with the new synthetic sequence.

classic restriction digestion and ligation to form a larger fragment of synthetic DNA, called a 'mega-chunk'. Each far-right chunk within a mega-chunk is also synthesised to encode an auxotrophic marker, such as genes that allow growth in the absence of leucine or uracil. This allows for selection of the mega-chunk following transformation and integration into the genome. The integration is verified using the same methodology as described earlier for SynIII: with the use of PCRtags to differentiate between the synthetic and the native DNA sequences. The next mega-chunk is then incorporated once the verified partially-synthetic strain is deemed to have no loss of fitness.

Following the construction, integration and verification of synthetic chromosome III, and parts of VI and IX, the original Sc2.0 team gathered enough evidence to support their aims and methods and the remaining 13 chromosomes were distributed to other global partners (Figure 1.5). Professor Jef Boeke, now at New York University, leads work on chromosomes I, IV, and VIII and Professor Srinivasan, Chandrasegaran is working on synthetic chromosome IV (jointly with NYU). China has two universities and a biotechnology company contributing to

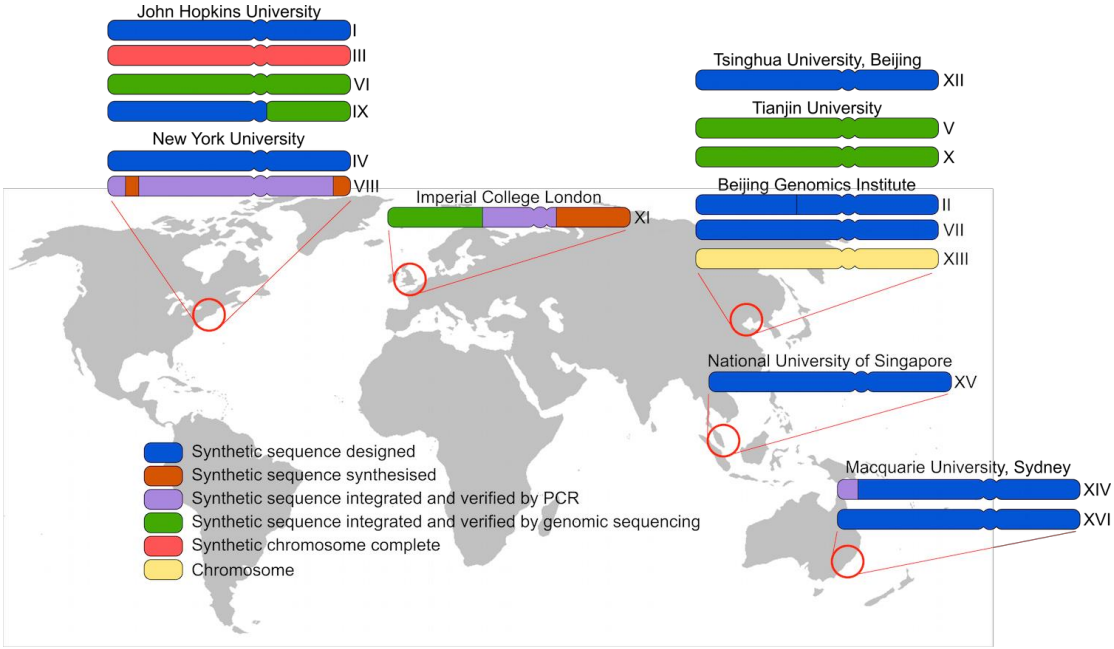


Figure 1.5 Sc2.0 Chromosomes Mapping. The Sc2.0 is now a global project, with groups from around the world contributing to the synthesis and assembly of a synthetic genome. The USA have two research teams from Johns Hopkins and New York Universities work on six chromosomes. China has three collaborators, Tsinghua and Tianjin Universities and BGI also working on six synthetic chromosomes. Singapore is the latest country to become a member of the Sc2.0 consortium, taking on the synthesis, assembly and integration of chromosome XV. Australia's Macquarie University has taken on two chromosomes. Lastly, from the UK are Imperial College London contributing to the completion of chromosome XI, and Edinburgh University is creating the neo-chromosome to house all the relocated tRNAs (not shown).

chromosomes II, VII, and XIII (BGI Genomics), V and X (Tianjin University) and XII (Tsinghua University). Singapore has undertaken work on chromosome XV (National University of Singapore) and most-recently the newest partners are Macquarie University of Australia who are contributing to the assembly and completion of chromosomes XIV and XVI. Lastly, the United Kingdom has two universities working with the Sc2.0 project: Patrick Cai's group at Edinburgh University is endeavouring to produce a neo-chromosome that will host all the tRNA genes, and the Ellis Lab at Imperial College London is undertaking the assembly of synthetic chromosome XI.

1.2.5 Synthetic Chromosome Rearrangement and Modification by LoxPsym Mediated Evolution (SCRaMbLE)

The most-interesting and potentially the most valuable design change added to the synthetic genome is the aforementioned SCRaMbLE system. This uses an inducible version of the recombinase enzyme to recombine regions of the genome where the loxPsym sites have been added as inducible recombination hotspots. The Cre-Lox recombination system naturally originates from the prokaryotic bacteriophage P1[73], and is used by the phage to implement the circularisation and replication of its DNA during infection of a host bacterial cell. The Cre recombinase is a 38 kD chimeric protein that catalyses the recombination between two loxP sites, and belongs to the integrase family of site-specific recombination enzymes[74, 75]. Shortly after its discovery it was swiftly repurposed as a new site-specific recombination tool, allowing genes to be integrated into the chromosomes of many organisms including eukaryotes like yeast[76]. Even now, Cre-Lox recombinases are still the primary tool for inserting gene expression cassettes into the genomes of transgenic mice[77].

The chromosomes of the Sc2.0 genome, including SynIII, SynIX.R and SynVI.L, all incorporate symmetrical 34 bp loxPsym almost always placed three base pairs downstream of the stop codon of non-essential genes. They differ from standard loxP sites in that the sequence has been recoded to be palindromic[78, 79]. This size of these sequences is thought to be below the threshold for yeast to perform its own homologous recombination between multiple copies of this sequence, and instead their recombination is dependent on the Cre recombinase being expressed in the yeast nucleus. The Cre recombinase used for SCRaMbLE experiments is CreEBD, a version engineered to be a spatio-temporally controlled [80], by fusion of Cre to a murine oestrogen binding domain (EBD). Translocation of CreEBD to the yeast nucleus is induced by the addition of β-oestradiol. Upon the introduction of β-oestradiol, the Cre recombinase is released from the EBD domain,

allowing it to complete its required protein folding so that it can travel from the cytosol of the cell into the nucleus wherein it then binds and recombines loxPsym sites. For the SCRaMbLE system, CreEBD is provided on a YCp plasmid and is expressed under the control of a daughter specific promoter, SCW11, ensuring its expression is only active in dividing yeast cells[81].

The symmetrical nature of the chosen loxPsym site means that when two loxPsym sites are brought together by a Cre recombinase, the DNA between them can undergo either a deletion, a translocation or an inversion (so long as the DNA between the two sites is at least 82 base pairs)[78]. If, for example a non-essential gene was located between two loxPsym sites on the chromosome, then expression of the Cre recombinase, could lead to this gene being deleted, or having its orientation inverted or being translocated to another chromosomal region with loxPsym sites, or indeed all of the above (Figure 1.6). This is precisely what has been observed in SCRaMbLE verification experiments performed previously with the semi-synVI.L and synIX.R chromosomes described in 2011. Genes flanked by loxPsym sites that encode enzymes in the amino acid biosynthesis pathways, for example, are lost when the Cre-recombinase is expressed in growth media where the amino acids are provided. Cells that delete the biosynthesis genes can survive just as long as the amino acid is given[12] but when it is removed they can no longer make the essential molecule.

Further research conducted by Annaluru *et al* in 2014 showed that a diploid version of SynIII (generating SynIII/III and thus has both mating types MATa/α), consistently lost the loxPsym-flanked MATα gene found on SynIII with the induction of SCRaMbLE as it was no longer needed. However, they also found that SCRaMbLE induction in haploid cells led to large quantities of cell lethality, presumably as a result of essential gene loss. Shen and colleagues recently published extensive research on the effects SCRaMbLE on a synthetic chromosome strain, reporting their findings in late 2015[82]. They found that 64 rearranged versions of SynIX.R generated by inducing SCRaMbLE were unique in their layouts, with no two exhibiting the same genotype. This demonstrated the capability of SCRaMbLE to rapidly produce genome diversity. They sequenced the 64 rearranged chromosomes and found there were many complex rearrangements due to multiple recombination events. The fact that all 64 versions give healthy haploid yeast cells, demonstrates the capabilities of yeast to tolerate significant rearrangements of its genome. Initial findings in this paper and in Dymond *et al* (2011) showed that cell lethality is directly proportional to the SCRaMbLE induction times. A ten-fold decrease in cell viability is seen with six hours of induction, a 100-fold decrease is seen with 12 hours of

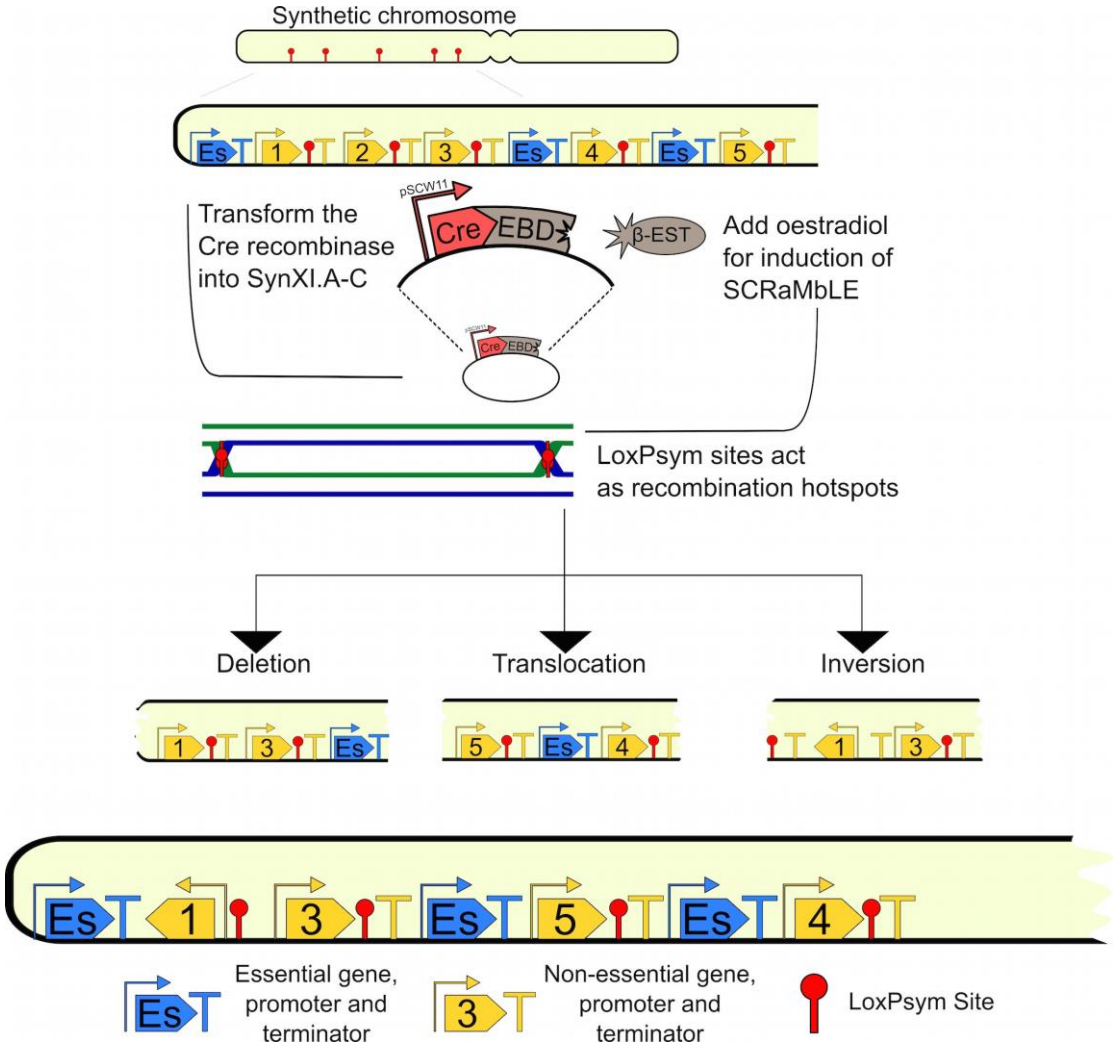


Figure 1.6 Sc2.0 SCRaMbLE schematic. Synthetic chromosomes contain loxPsym sites in their 3' UTR regions, three base pairs downstream of stop codons. Upon induction with β-oestradiol, the Cre recombinase enzyme is released from the oestrogen binding domain, and travels to the nucleus of cells whereby it catalyses the recombination between the site specific loxPsym sequences. Induction can cause a number of rearrangements, such as inversions, deletions, translocations, and most recently shown duplications (not shown), thus creating a new synthetic chromosome strain. Adapted from Jovicevic *et al* (2014)[124]

induction and more than a 1000-fold decline in cell numbers is seen after induction periods much longer than 12 hours.

Thus far SCRaMbLE is thought to cause cell death by essential gene loss, and potentially also by large genomic deletions and rearrangements that lead to lethality. However, large changes are still tolerated. The largest recorded deletion of in the sequenced SynIX.R SCRaMbLEd strains is 41,999 bp (which corresponds to almost half of the size of the synthetic right arm of chromosome IX) and the greatest inverted fragment included 38,925 base pairs[82]. Furthermore, apart from the

deletions and inversions mentioned, there was also a large number of duplications in these sequenced strains. Duplications occurred in almost half of all the induced strains, duplicating both essential and non-essential genes with no evidence of detrimental effects on cell fitness.

Interestingly, the generation of duplications was not initially expected and is hypothesised to be as a result of a "double rolling mechanism" whereby recombination of a loxPsym site occurs within a formed chromosome replication fork and travels with the DNA replicating bubble until it recombines with another loxPsym site[82]. However currently there is no direct evidence for this, so the exact cause for duplications is as of yet unknown.

The palindromic sequence of the loxPsym sites should theoretically cause deletions and inversions to occur at the same frequencies. The sequencing data of the SCRaMbLEd strains in Shen *et al* proved consistent with this prediction. However it should be noted that only deletions that do not negate the survival of the cell will be recorded and tolerated, and as such the excision of essential genes will result in cell lethality, and thus cannot be screened for. Several deletions that were associated with decreased cell fitness were identified in almost 50% of all the deletions recorded, but they are limited by the lengths tolerated as a result of essential gene loss. A particularly significant conclusion from this study showed that the new junctions generated as a result of SCRaMbLE from gene rearrangements, had no significant effects on gene expression and produced viable cells. The new junction types are not found naturally and in many cases cause the disruption of the 3' UTR of genes. Shen *et al* found that this did not decrease the fitness of cells.

SCRaMbLE is a tool for simultaneously studying the necessity for individual genes to be in the Sc2.0 genome, and study the effects caused by the loss or rearrangements of not just one gene, but many combinations of genes together. Previous efforts in the yeast community have investigated the role each gene plays within the genome by first deleting individual genes one-by-one[83], and then by deleting every pair of genes and measuring the effect on yeast fitness[84]. The SCRaMbLE system built into the synthetic genome allows more than billions of different combinations of gene deletions, inversions and translocations to be sampled in a single test-tube and only the genome rearrangements that will allow the cells to keep growing competitively in the conditions of the experiment will be seen after recovery of the yeast cells. The diversity that can be sampled when the entire Sc2.0 project is finished with thousands of loxPsym sites placed throughout the synthetic genome will be ground breaking for research. In theory, this strain could be used to

determine the minimal gene set required to have a functioning eukaryotic cell growing in the lab, although this has not been reported as of yet.

It is probable instead that there will be multiple minimalistic genomes generated from SCRaMbLE, as the likelihood is that there will be many combinations of minimal gene sets that cells can survive with. For example, consider a yeast cell with genes A and B; the yeast cell can survive with only gene A or with only gene B, but it cannot survive without genes A and B, thus two versions of the minimal cell are created. Now scale this up to the 6,000 genes that yeast actually contains – the predictability of all the possible combinations exceeds our current knowledge, capabilities and methodologies to efficiently map all the interactions. SCRaMbLE will allow researchers to develop and study diverse and novel rearrangements, by inducing laboratory-driven evolution to create minimal eukaryotic genomes[85]. Sc2.0 strains simply need to be subjected to continual rounds of Cre recombinase rearrangements in rich growth media, and theoretically over time all non-essential genes would be deleted.

1.3 SCRaMbLE and Yeast Metabolic Engineering

1.3.1 Yeast Metabolic Engineering

Yeast has long been exploited as the chassis for metabolic engineering studies, due to centuries of experience working with yeast in the wine and beer industries[86, 87]. Applications of yeast engineered for biotechnology are not limited to the food industry and include many examples such as the production of therapeutics like non-ribosomal peptides[88, 89] (a versatile group of precursors for antibiotics) or the production of chemicals for industry, such as the production of 1,2-propandiol[90]. The pharmaceutical industry in particular has gained significant output from the metabolic engineering of microorganisms, predominantly *S. cerevisiae*.

A recent paper from the Dueber Lab is an excellent example of yeast metabolic engineering showing the production of opioids from glucose[91]. Opioids, such as morphine and codeine, are analgesic compounds belonging to the diverse, plant-based family benzylisoquinoline alkaloids (BIA) metabolites. BIAs have thus far struggled to be industrially produced from microorganisms, and are extracted from plants such as the opium poppy. DeLoache *et al* (2015)[91] showed the expression of a seven-enzyme BIA pathway in yeast to facilitate S-reticuline production from glucose, a major metabolite that is necessary for the production of opiates. In their study, they reported the generation of an enzyme-coupled biosensor, with functionality and activity of the pathway detectable by eye from cell colour change. They developed an optical sensor by optimising activity of a tyrosine hydroxylase

through PCR mutagenesis, an enzyme necessary for the conversion of L-Tyrosine to dihydroxyphenylalanine (L-DOPA), which in turn is catabolised further to the produce betaxanthin (a yellow fluorescing compound).

The generation of this tyrosine hydroxylase was a major feat for this research group, as it was the first example of this enzyme produced using the methods of synthetic biology and exhibiting high functionality in yeast. They used this pathway and biosensor to identify a new tyrosinase hydroxylase enzyme (derived from the beetroot plant) called CYP76AD1, and implemented it within their pathway to produce a yellow-fluorescing compound, betaxanthin. They then mutated the original sequence of CYP76AD1 to generate 17 mutants by selecting for the most yellow colonies. Isolating the most beneficial sequence mutations, they engineered the new enzyme version to have a 2.7-fold increased activity when compared to the wild type.

The new CYP76AD1 mutant enzyme was expressed alongside the DODC enzyme (DOPA decarboxylate) and together allow the conversion of L-tyrosine to the S-reticuline precursor, dopamine. The combination of dopamine with 4-hydroxyphenylacetaldehyde (4-HPAA), another by-product of the catabolism of L-Tyrosine from the Ehrlich pathway found endogenously in yeast, generates S-norcoclaurine. This compound is four catabolic steps away from the producing S-reticuline. Through the combination of the expression of the tyrosine hydroxylising CYP76AD1 mutant that the Berkley team developed in their study, along with yeast's natural intracellular Ehrlich pathways, and the heterologous expression of the final enzymes for completing the pathway, DeLoache *et al* demonstrated an *S. cerevisiae* engineered to produce S-reticuline from L-tyrosine from growth on glucose media. Although the quantity of the S-reticuline produced was not high enough for commercial production, their work did show that *S. cerevisiae* could successfully express the range of enzymes required for the production of this compound, as well as the ability to run a biosynthesis pathway in yeast for the production of a pharmaceutically-important product.

Alongside work towards biosynthesis of medical compounds, *S. cerevisiae* has also been exploited at large scale for the production of biofuels[92-94]. Biofuels offer alternatives to the current fuel types (such as gasoline, jet fuel, diesel) and include a broad range of biologically-produced alcohols, such as ethanol, butanol, pentanol and propanol[95]. The scientific interest for the production of biofuels from agricultural products has seen a steady increase over the past two decades, leading to a variety of so-called 'second generation biofuels' (biofuels from biomass) being made. Our improvements in biotechnology coupled with the depletion of non-renewable fossil fuels such as coal and oil, mean that there is scope for generating

microbes that produce fuels such as bioethanol or biobutanol that can be widely used in the commercial, domestic and transport industries[96]. Secondly, the necessity to provide a sustainable source of "bioenergy" is becoming fast in-demand, and is seen as a green alternative to fossil fuel burning. The production and burning of biofuels, such as bioethanol, has a substantially smaller carbon emission levels, (although not eliminated entirely) compared to coal-based power, for instance.

Second generation biofuel and bioethanol fermentation has concentrated on the use of hexose-sugar rich foods as the biomass feedstock, such as maize[97] and wheat[98]. However these pose an issue as a biomass source as they can become strong competition with food markets[99]. Lignocellulosic biomass is a large agricultural by-product that does not compete with food markets. It is found abundantly on Earth as wood chippings, tree bark, and waste and has long been studied for its potential as an alternative green source for future advanced biofuel production.

A study by Wei *et al* (2015)[100] recently reported the metabolic engineering of *S. cerevisiae* to heterologously express three separate pathways with the aim of breaking down (pre-treated) lignocellulosic biomass to produce a high-tolerance, ethanol-generating yeast strain. The pathways in question tackled the breakdown of the two major sugars found in lignocellulose (glucose and xylose), as well as reducing the accumulation of acetic acid – which is toxic to yeast cells. This study employed a synthetic biology approach to create synergism between three separate pathways and the breakdown of multiple substrates to output a single desired product: bioethanol.

First the authors needed to establish hexose-sugar (glucose) metabolism from pre-treated cellobiose. *S. cerevisiae* cannot naturally metabolise xylose, and even when a recombinant xylose-utilising strain is produced, because xylose travels into the cells via glucose transporters, the presence of glucose completely inhibits the uptake of xylose. Thus the co-consumption of glucose and xylose at the same time, from substances like lignocellulose that contain both, is difficult to implement. A study by a related team in 2011[101] evaded this issue by expressing a β-glucosidase (for the hydrolysis of cellobiose into glucose in the yeast cell) and a cellodextrin transporter (for the uptake of cellobiose into the cell) to create a recombinant *S. cerevisiae* strain capable of catabolising cellobiose. Using this same strain, the authors co-expressed an optimised oxidoreductase xylose-metabolising pathway, to produce a strain that was capable of co-consumption of xylose and cellobiose. Furthermore, fermentation of this strain produced high enough ethanol yields that were significantly greater than the yields obtained from *S. stipitis* fermentation (see

the following section). Wei *et al* (2015)[100] reconstituted the same strain they developed previously, and optimised it further by adding a third pathway; the acetate reducing pathway that decreases the toxicity of acetic acid by also fermenting it into bioethanol. Their results generated significant amounts of ethanol (from a laboratory-based *S. cerevisiae* strain) and exemplified that co-consumption of several substrates simultaneously was possible, and could be applied industrially.

Thus, like the previous example of S-reticuline production in yeast by DeLoache *et al*[91], the production of bioethanol from the reports provided by Wei *et al*[100] and Ha *et al*[101] showed an example of extensive metabolic engineering in yeast, co-expressing eight heterologous genes encoding eight enzymes, of which none are found in yeast naturally. In conclusion, these studies, among a flurry of others, all exhibit the suitability and desire to keep exploiting the most manipulated eukaryotic organism in research, but now for the production of metabolic compounds of utility in healthcare, bioenergy and other sectors e.g. industrial chemistry.

1.3.2 Xylose as a Lignoceullosic Sugar for Engineered Yeast Metabolism

As briefly mentioned above, lignocellulose biomass is low cost, renewable carbohydrates and typically contains the three components lignin (10-40%), cellulose (40-55%) and hemicellulose (25-55%)[100, 102]. Within this mass the sugar xylose is the most abundant monosaccharide after glucose. Xylose is a five carbon pentose monosaccharide, whose polymer xylan is the main constituent of hemicellulose. There are a few organisms that possess the ability to metabolise xylose naturally, such as the bacteria *E. coli*,[103] or the fungi *Piromyces* sp. E2[104], both of which use the (bacterial) xylose utilisation pathway, called the Isomerase (IS) pathway. The IS pathway is formed of two genes, the xylA gene encoding for the xylose isomerase enzyme that catalyses the conversion of xylose to D-xylulose, and the XYL3 gene that catalyses the conversion of the xylulose to xylulose-5-phosphate, and encodes for the enzyme xylulokinase. Xylulose-5-phosphate is the prerequisite compound that enters the sugar-metabolising pentose phosphate pathway, which is a complex multistep metabolism platform used to break down and produce useful by-products such as ethanol or glycerol, in bacteria or yeast[105]. The other xylose metabolising pathway is called the oxidoreductase (OR) pathway and is predominantly found in yeasts such as *Scheffersomyces stipitis*. It facilitates the breakdown of xylose using genes XYL1, XYL2 and XYL3, encoding for the enzymes xylose reductase (XR), xylitol dehydrogenase (XDH) and xylulokinase (XKS1), respectively. The two pathways are summarised in Figure 1.7.

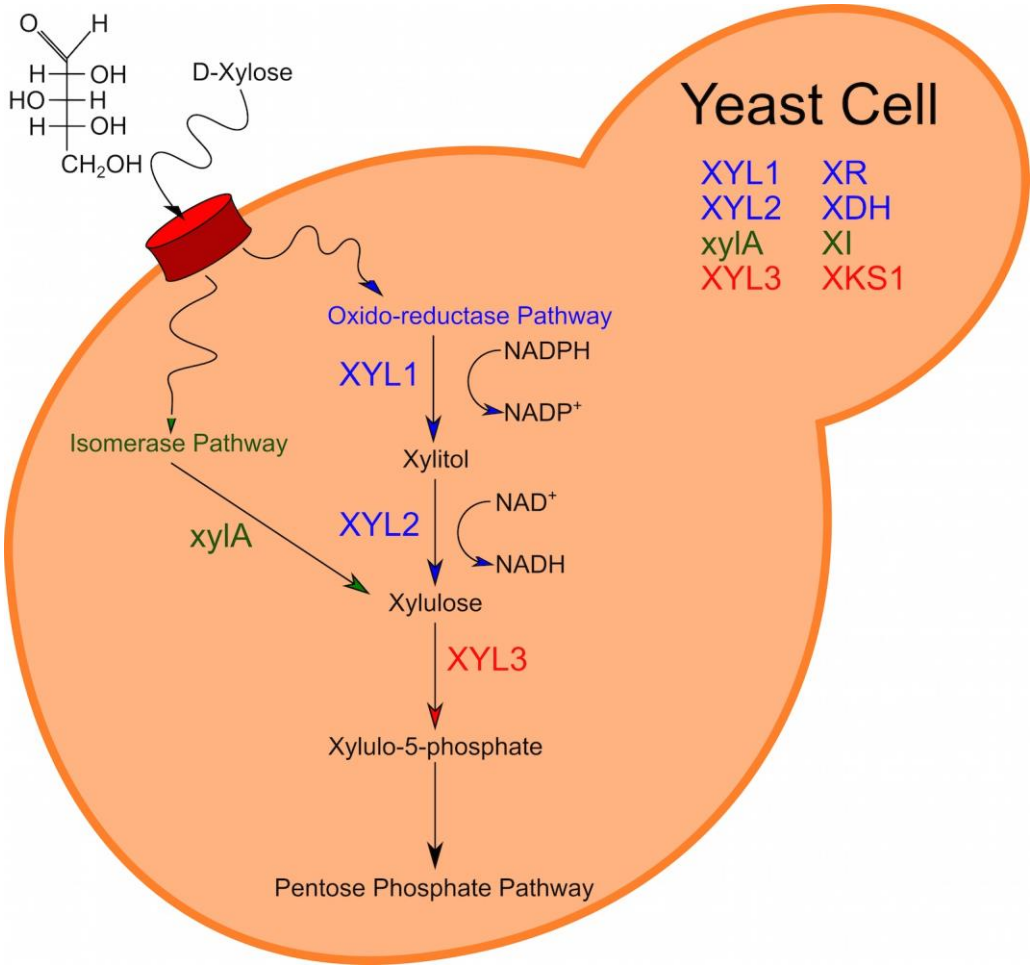


Figure 1.7 Xylose Utilisation Pathways. Schematic illustration of the xylose metabolic pathway in yeast. Xylose is secreted into the cell via glucose transporters, wherein it is metabolised via the oxidoreductase pathways (blue) or the isomerase pathway (green). Upon the production of xylulo-5-phosphate, this product enters the anaerobic phase of the pentose phosphate pathway.

The two pathways, bacterial and yeast, have both been extensively studied and expressed in *S. cerevisiae*[106-109], and have their equal associated benefits and drawbacks. The IS pathway allows easier implementation and optimisation due to being a single gene pathway (a version of the XYL3 gene is found intrinsically in *S. cerevisiae*[110]), however as its origins are derived from bacteria or fungi, it requires codon optimisation for functional expression in yeast. Furthermore, despite efforts to codon optimise, mutate or constitutively[106] express under the control of strong promoters[107, 111], xylose metabolism in *S. cerevisiae* via the IS pathway is poor and remains to be significantly enhanced. Reasons for the inefficient enzymatic activity of the xylose isomerase and therefore the low xylose utilisation are not known exactly, however suggestions have been made that this may be due to protein mis-folding or post-translational modification[112].

For improved xylose utilisation, the OR pathway has in the past been implemented for functional expression in *S. cerevisiae*, and being derived from the

closely related *Scheffersomyces stipitis* yeast, the two enzyme-encoding genes of this pathway do not suffer from the functional expression issues as experienced frequently with the Isomerase pathway. Codon-optimisation could still be applied to improve this engineering; however it is not the prevalent drawback of this xylose-metabolising pathway. Instead, the inherent disadvantage that impedes and limits efficient break down of xylose in recombinant *S. cerevisiae* is thought to be the co-factor imbalance that is associated with expressing the XYL1 and XYL2 genes. The xylose reductase enzyme has a preference for the nicotinamide adenine dinucleotide phosphate (NAD(P)H) and oxidises it to NADP+ during the initial conversion of xylose to xylitol. XDH reduces NAD+ to NADH, and thus this incomplete coenzyme recycling step is the fundamental issue of the oxidoreductase pathway[113], leading to the increment of xylitol – a substance that is toxic to cells at high concentrations. *S. stipitis* does not experience this issue as it possesses both NAD(P)H and NADH XR species for cofactor utilisation[114].

Both pathways also rely on the XYL3 (XKS1) enzyme to feed xylulose into the central carbon metabolism. There are contradicting reports surrounding the boost in xylose catabolism that can be achieved by overexpression of the XYL3 (XKS1) gene. However, most previous work favours XYL3 overexpression for optimised growth on xylose-containing media. Rodrigues-Peña *et al* (1998)[110], the same research team that discovered a version of XKS1 found endogenously in *S. cerevisiae*, have reported that the moderate expression of the XKS1 gene facilitates some growth on xylose media, but that overexpression of this gene can negatively influence growth. Similarly, Johansson *et al* (2002)[105] observed that the uncontrolled overexpression of this gene greatly reduced the consumption rate of xylose (by 30%) despite it increasing overall ethanol production. In contrast Jin *et al* (2003)[115] showed that multiple copies of the XKS1 gene facilitated improved xylose utilisation and improved growth of their recombinant strain on xylose media as a result of increased enzymatic activity. However, they also found evidence of substrate-accelerated death when the enzyme was overexpressed[115]. Others have reported that overexpression of XKS1 led to increased rates of xylose utilisation, and better conversion rates to ethanol[116]. Nonetheless, despite the few mentioned approaches to create an efficient recombinant *S. cerevisiae* that metabolises xylose, it is clear that there are likely to be improvements that can be made and that tuning of expression of both heterologous pathway genes and native genes is a suitable way to achieve this.

1.3.3 The Potential for Sc2.0 and SCRaMbLE to Aid Metabolic Engineering

The SCRaMbLE system has been designed to lead to gene deletions and gene rearrangements, in order to study how the gene layout and content of the yeast genome defines how it works. However, the SCRaMbLE system could also be put to task for more biotechnology-focused aims, not just by catalysing rearrangements or deletions but as a vehicle for incorporating new genes into synthetic genomes and at the same time creating strain diversity by combinatorial genome rearrangement. Consider the situation where synthetic genes encoding a desired and selectable property – the ability to grow on xylose as the carbon source, for example. Theoretically if these genes are flanked by loxPsym sites and made available to the yeast cells as the SCRaMbLE system is activated (e.g. by co-transforming their DNA into yeast or by having it already present), then these genes may get incorporated into the new genome layouts. Thus, when only xylose is present as a carbon source in the growth media, only yeast with genomes that have rearranged to include the provided xylose-utilisation genes will be able to grow. Therefore, by applying certain types of environmental pressures it becomes favourable for synthetic yeast to incorporate any provided genes that can give the cell an advantage during and/or after the SCRaMbLE period. Beyond just simply adding the heterologous genes to the synthetic yeast, SCRaMbLE also has the potential to at-the-same-time create millions of different genotypes, effectively creating diverse strain libraries. With suitable screening systems in place, the one-pot integration of heterologous genes whilst simultaneously generating diverse strain backgrounds has the potential to simplify and reduce the steps typically taken in metabolic engineering where heterologous genes are first tested for function in one strain, before then being tested in strain libraries or optimised by further engineering to knock-out or overexpress native genes.

One can imagine even further than the initial case of providing identified heterologous loxPsym-formatted genes such as those known to aid xylose utilisation. Potentially, whole libraries of genes (e.g. cDNA libraries) could be adapted into loxPsym format and taken from a diverse range of sources such as from plants, bacteria or fungi. These diverse gene libraries could be deliberately mixed into synthetic yeast strains undergoing SCRaMbLE in conditions where it would be desirable to evolve to outperform normal yeast, such as in high concentrations of ethanol. The “SCRaMbLE-in” approach could theoretically aid in producing new strains of yeast with enhanced gene content suitable for industrial uses.

Synthetic yeast and its SCRaMbLE system offer a new kind of evolution to be exploited: a “hyper-evolution” where gene content and genetic arrangement is

radically altered, rather than individual nucleotide bases. This fits very much with the ethos of synthetic biology, where organism design is modular, and new functions can be derived by rewriting genetic modules into new combinations and arrangements.[22, 117]. In this thesis, I investigate the potential for synthetic yeast and SCRaMbLE to be used for the first time to SCRaMbLE-in heterologous genes. I then apply this system in metabolic engineering, producing a new xylose-utilising stain by SCRaMbLE-in.

1.4 Aims and Objectives

The aim of this thesis is to determine the utility of SCRaMbLE as a method for introducing new genes suitable into synthetic yeast genome strains. The work presented here determines the first protocol for a new approach called 'SCRaMbLE-in' and demonstrates its use in metabolic engineering for generating a synthetic yeast strain that can grow on xylose as the sole carbon source. However, before SCRaMbLE can be assessed and investigated, a partially-synthetic yeast chromosome first needed to be produced. As the Ellis Lab at Imperial College London were one of the first international collaborators on the Sc2.0 synthetic yeast genome project, we became the first group beyond the USA to test out the methodology for mega-chunk assembly as outlined above. In Chapter 3 of this thesis, I report on the use of mega-chunk assembly and assessment of a synthetic yeast construction training kit produced by our collaborators. I then describe my work on starting the design, construction, integration and verification on the UK's first synthetic chromosome: yeast synthetic chromosome XI. My initial work in this chapter resulted in a partially synthetic chromosome (SynXI.A-C) suitable for SCRaMbLE experiments.

Following work on construction of the UK's synthetic yeast chromosome, in Chapter 4, I then report on my investigation of how the SCRaMbLE toolkit can be used to introduce heterologous genes into the synthetic chromosomes. My work in this chapter determines the layout required for heterologous genes to be compatible with SCRaMbLE and how a 'SCRaMbLE-in' protocol was finalised. Using novel fluorescent yeast strains constructed to investigate SCRaMbLE, I further determine the kinds of rearrangements and deletions associated with SCRaMbLE-in.

Finally, in Chapter 5, I build upon my work in the other two results chapters to test the hypothesis that the SCRaMbLE-in can be used to introduce multiple genes, at once, and that this can be used to accelerate metabolic engineering. In Chapter 6, we combine the work from the previous two chapter and the partially synthetic yeast strains I have helped construct are used for SCRaMbLE-in along with genes that I have first verified that enable growth of *S. cerevisiae* yeast with xylose as the sole carbon source, in fitness and functionality. This results in the production of the world's first xylose-utilising synthetic yeast stain constructed by SCRaMbLE-in.

2.0 CHAPTER 2: Materials and Methods

All strains, plasmids and primers are detailed in the tables below, and have been constructed in this study, unless otherwise stated. Commonly used strains (such as DH10B, or BY4741), plasmids (for example, pRS416) and primers (like the M13 F and R primer pair) are used for the expression, construction or amplification, respectively, of constructs and strains used in this study. Descriptions for primers ordered for this study are included in table 2.4. All other strains and plasmids are mentioned in the upcoming chapters. Strains, plasmids and primers mentioned in this study are stored and found in the Ellis Lab Culture Collection.

2.1 Strains and Plasmids used

Table 2.1 *E. coli* strains

Strain	Description	Source	Reference
DH10B	*Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- ϕ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL* (Str^R) *rph spoT1 Δ(mrr-hsdRMS-mcrBC)*	New England BioLabs	NEB Catalogue (No. C3019I)
DH10B EPI400	*F^- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ^- rpsL (Str^R) nupG trfA tonA pcnB4 dhfr*	GenScript/ Epicentre	Epicentre Catalogue (No. C400EL10)
Turbo Competent	F' *$proA^+B^+$ $lacI^q$ ΔlacZM15 / fhuA2 Δ(lac-proAB) glnV galK16 galE15 R(zgb-210::Tn10)* Tet^S *endA1 thi-1 Δ(hsdS-mcrB)5*	New England BioLabs	NEB Catalogue (No. C2984I)

Table 2.2 *Saccharomyces cerevisiae* strains

Strain	Description	Source	Reference
BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	John Hopkins University	[118]
BY4742	MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0	John Hopkins University	[118]
SynXI.A	BY4741 MATa met15Δ0 ura3Δ0 ::LEU2	This study	Unpublished
SynXI.A-C	BY4741 MATa met15Δ0 ura3Δ0 ::LEU2	This study	Unpublished

dSynXI.A-C	MATa/MATα his3Δ1 met15Δ0 lys2Δ0 ura3Δ0 ::LEU2	This study	Unpublished
dSynXI.A-L	MATa/MATα his3Δ1 met15Δ0 lys2Δ0 ura3Δ0 ::LEU2	This study	Unpublished
YFL054C	BY4741 MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ::KanMX4	John Hopkins University, USA	EUROSCARF catalogue; (acc. No. Y05675)
YKL220C	BY4741 MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ::KanMX4	EUROSCARF	EUROSCARF catalogue; (acc. No. Y07039)
YJR092W	BY4741; Mat a; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; YJL092w::kanMX4	EUROSCARF	EUROSCARF catalogue; (acc. No. Y01331)
yDJM1	dSynXI.A-C + pDJM1	This study	Unpublished
yDJM2	dSynXI.A-C + pDJM2	This study	Unpublished
yDJM3	dSynXI.A-C + pDJM3	This study	Unpublished
yRC1841	dSynXI.A-C + pRC1841	This study	Unpublished
yRC1842	dSynXI.A-C + pRC1842	This study	Unpublished
yRC1844	dSynXI.A-C + pRC1844	This study	Unpublished
yRC1845	dSynXI.A-C + pRC1845	This study	Unpublished
yDJGFP	dSynXI.A-C + pDJGFP	This Study	Unpublished

Table 2.3 List of plasmids used in this study

Plasmid	Description	Source	Reference
pChrXI_A1(v.2)	Fragment A1(version 2) in pUC57 MCS, Ori, AmpR	GenScript	Unpublished
pChrXI_A2	Fragment A2 in pUC57 MCS, ori, AmpR	GenScript	Unpublished
pChrXI_B1	Fragment B1 in pUC57 MCS, ori, AmpR	GenScript	Unpublished
pChrXI_B2	Fragment B2 in pUC57 MCS, ori, AmpR	GenScript	Unpublished
pChrXI_B3	Fragment B3 in pUC57 MCS, ori, AmpR	GenScript	Unpublished
pChrXI_B4	Fragment B4 in pUC57 MCS, ori, AmpR	GenScript	Unpublished
pChrXI_B5	Fragment B5 in pUC57 MCS, ori, AmpR	GenScript	Unpublished
pRS413	Common yeast expression vector; f1 origin, ori, lacZα/MCS, AmpR, Cen6/ARS(H4), HIS3	John Hopkins University, USA	[119]
pRS415	Common yeast expression vector; f1 origin, ori, lacZα/MCS, AmpR, Cen6/ARS(H4), LEU2	John Hopkins University, USA	[119]

pRS416	Common yeast expression vector; f1 origin, ori, lacZα/MCS, AmpR, Cen6/ARS(H4), URA3	John Hopkins University, USA	[119]
pUC19	Common bacterial vector; lacZα/MCS, ori, AmpR	ThermoFisher, SD0061	[120]
pCre_EBD1	pRS413 backbone vector; SCW11p, *Cre_EBD* gene, tCYC1, f1 origin, ori, AmpR, Cen6/ARS(H4), HIS3	Imperial College London	Tim Weenink; Unpublished
pCre_EBD2	pRS416 backbone vector; SCW11p, *Cre_EBD* gene, tCYC1, f1 ori, ori AmpR, Cen6/ARS(H4), URA3	John Hopkins University, USA	[81]
pS1A2	f1 origin, ori, GPDp, *xylA*2* gene, AmpR, 2micron-ori, URA3	Uni. of Texas, USA	[111]
pS1A3	f1 origin, ori, GPDp, *xylA*3* gene, AmpR, 2micron-ori, URA3	Uni. of Texas, USA	[111]
pDJ002	pRS413 backbone vector; **loxPsym**, TEF1p, *XYL1* gene, **loxPsym**, tCYC1, AmpR, f1 ori, ori, Cen6/ARS, HIS3	This study	Unpublished
pDJ008	pUC19 backbone vector; G-block DJ002, ori, AmpR	This study	Unpublished
pDJ014	pRS416 backbone vector; pKan, *KanMX* gene, tKan, AmpR, f1 ori, ori, Cen6/ARS, URA3	This study	Unpublished
pDJ015	pRS416 backbone vector; pKan, *KanMX* gene, **loxPsym**, tKan, AmpR, f1 ori, ori, Cen6/ARS, URA3	This study	Unpublished
pDJ016	pRS416 backbone vector; **loxPsym**, pKan, *KanMX* gene, **loxPsym**, tKan, AmpR, f1 ori, ori, Cen6/ARS, URA3	This study	Unpublished
pDJ017	pUC19 backbone vector; pKan, *KanMX* gene, tKan ori, AmpR	This study	Unpublished
pDJ018	pUC19 backbone vector; pKan, *KanMX* gene, **loxPsym**, tKan ori, AmpR	This study	Unpublished
pDJ019	pUC19 backbone vector; **loxPsym**, pKan, *KanMX* gene, **loxPsym**, tKan ori, AmpR	This study	Unpublished
pDJG1	**loxPsym**, TEF2p, *XYL1* gene, tADH1, ConS, Con1, AmpRColE1, Cen6/ARS(H4), URA3	This study	Unpublished
pDJG2	TEF1p, *XYL2* gene, tTDH1, Con1, Con2, AmpRColE1, Cen6/ARS(H4), URA3	This study	Unpublished

pDJG3	TDH3p, *XYL3* gene, **loxPsym**, tENO2, Con2, ConE, AmpRColE1, Cen6/ARS(H4), URA3	This study	Unpublished
pDJG4	**loxPsym**, TEF1p, *xylA*3* gene, tADH1, ConS, Con2, AmpRColE1, Cen6/ARS(H4), URA3	This study	Unpublished
pDJG5	**loxPsym**, TEF2p, *XYL1* gene, tADH1, **loxPsym**, tENO2, KanRColE1, Cen6/ARS(H4), URA3	This study	Unpublished
pDJG6	**loxPsym**, TEF1p, *XYL2* gene, tTDH1, **loxPsym**, tENO2, KanRColE1, Cen6/ARS(H4), URA3	This study	Unpublished
pDJG7	**loxPsym**, TDH3p, *XYL3* gene, **loxPsym**, tENO2, KanRColE1, Cen6/ARS(H4), URA3	This study	Unpublished
pDJM1	**loxPsym**, TEF2p, *XYL1* gene, tADH1, TEF1p, *XYL2* gene, tTDH1, TDH3p, *XYL3* gene, **loxPsym**, tENO2, KanRColE1, Cen6/ARS(H4), URA3	This study	Unpublished
pDJM2	**loxPsym**, TEF1p, *xylA*3* gene, tADH1, TDH3p, *XYL3* gene, **loxPsym**, tENO2, KanRColE1, Cen6/ARS(H4), URA3	This study	Unpublished
pDJM3	**loxPsym**, TEF2p, *XYL1* gene, tADH1, TEF1p, *XYL2* gene, tTDH1, TDH3p, *XYL3* gene, **loxPsym**, tENO2, URA3 3' hom., KanRColE1, URA3 5' hom, Cen6/ARS(H4), URA3	This study	Unpublished
pRC1841	ScRPL18Bp, mRuby2, **loxPsym**, tADH1, ScYRA1p, sfGFP, **loxPsym**, tScSSA1, ScALD6p, mTagBFP2, **loxPsym**, tScPGK1, URA3, URA3 3' hom., KanR, ColE1, URA3 5' hom.	This study	Unpublished
pRC1842	ScYRA1p, sfGFP, **loxPsym**, tScSSA1, ScALD6p, mTagBFP2, **loxPsym**, tScPGK1, ScRPL18Bp, mRuby2, **loxPsym**, tADH1, URA3, URA3 3' hom., KanR, ColE1, URA3 5' hom.	This study	Unpublished

pRC1844	ScRPL18Bp, mRuby2, **loxPsym**, tADH1, ScYRA1p, sfGFP, **loxPsym**, tScSSA1, ScALD6p, mTagBFP2, **loxPsym**, tScPGK1, URA3, URA3 3' hom., KanR, ColE1, URA3 5' hom.	This study	Unpublished
pRC1845	ScRPL18Bp, mRuby2, **loxPsym**, tADH1, ScYRA1p, sfGFP, **loxPsym**, tScSSA1, ScALD6p, mTagBFP2, **loxPsym**, tScPGK1, URA3, URA3 3' hom., KanR, ColE1, URA3 5' hom.	This study	Unpublished
pDJGFP	ConS, HHF2p, sfGFP, loxPsym, tTDH1, Con2, URA3, CEN6/ARS, AmpRColE1	This study	Unpublished

Table 2.4 List of primers used in this study

Primer Name	Sequence (5' → 3')	Source	Description
DJ033	gagatgagattgctgtgcgtgaggcctattcggccgaccaaaatgaattgcatttaaacg	IDT DNA	Annealing primer equally distributed between Chr11_A1 and Chr11_A2
DJ034	cgtttaaatgcaattcattttggtcggccgaataggcctcacgcacagcaatctcatctc	IDT DNA	Annealing primer equally distributed between Chr11_A1 and Chr11_A2
DJ037	attagaattcctgtttagcttgcctcgt	IDT DNA	Primer for amplifying out the genomic KanMX4 sequence from YJR092W, special features include restriction enzyme sites encoded within the oligonucleotide
DJ038	aaatggatccataacttcgtataatgtacattatacgaagttattgattagaaaaactcatcgagcat	IDT DNA	Primer for amplifying out the genomic KanMX4 sequence from YJR092W, special features include restriction enzyme sites and a loxPsym site encoded within the oligonucleotide
DJ039	tattgcggccgcgttttcgac	IDT DNA	Primer used for amplifying out the endogenous genomic KanMX4 cassette from YJR092W, with the inclusion of restriction enzyme sites
DJ040	ttatgaattcctgtttagcttgcctcgtcccc	IDT DNA	Primer used for amplifying out the endogenous genomic KanMX4 cassette from YJR092W, with the inclusion of restriction enzyme sites
DJ041	gcatcgtctcatcggtctcatatggctaaagaatatttccct	IDT DNA	Primer used for adding Type 3 prefix to xylA*3

DJ042	atgccgtctcaggtctcaggatccttattg atacatcgcgat	IDT DNA	Primer used for adding Type 3 suffix to xylA*3
DJ043	gcatcgtctcatcggtctcatatgccatcc atcaaattgaac	IDT DNA	Primer used for adding Type 3 prefix to XYL 1
DJ044	atgccgtctcaggtctcaggatccttaaa cgaagattggaat	IDT DNA	Primer used for adding Type 3 suffix to XYL1
DJ045	gcatcgtctcatcggtctcatatgactgct aacccatctttggt	IDT DNA	Primer used for adding Type 3 prefix to XYL2
DJ046	atgccgtctcaggtctcaggatccttattct ggaccatcaatca	IDT DNA	Primer used for adding Type 3 suffix to XYL 2
DJ047	ctcatctgtaacatcattggc	IDT DNA	Used for iPCR of KanMX
DJ048	gcataagcttttgccattct	IDT DNA	Used for iPCR of KanMX
DJ049	ttttgctcacatgttcttt	IDT DNA	For amplification of xylose gene(s) in multi-gene cassettes
DJ050	ataaagtgttctaaactatgatgaa	IDT DNA	For amplification of xylose gene(s) in multi-gene cassettes
DJ051	caagtttccaaaattggaacatcttcgtagt cgaa	IDT DNA	iPCR XYL1 R (divergent)
DJ052	catctgtcttgcaagttgaacatcacccata cttg	IDT DNA	iPCR XYL1 F (divergent)
DJ053	catctgtcttgcaagttgaacatcacccata cttgc	IDT DNA	iPCR XYL2 R (divergent)
DJ054	aattaccaaaacggtagagaaaacgcccc aatcg	IDT DNA	iPCR XYL2 F (divergent)
DJ055	acatggtgttagagacttctctggtttgtcttt gg	IDT DNA	iPCR XYL3 R (divergent)
DJ056	ctgatgttgataacgaaaactgggacagat acaac	IDT DNA	iPCR XYL3 F (divergent)
DJ057	ccatccatcaaattgaactccggtt	IDT DNA	PCR XYL1 F (convergent)
DJ058	ttaaacgaagattggaatcttatcccaatc	IDT DNA	PCR XYL1 R (convergent)
DJ059	actgctaacccatctttggtcttg	IDT DNA	PCR XYL2 F (convergent)
DJ060	ttattctggaccatcaatcaaacacttaac	IDT DNA	PCR XYL2 R (convergent)
DJ061	ggattcttactacttgggtttcgactt	IDT DNA	PCR XYL3 F (convergent)
DJ062	tcaaggtcttttccaattcggacaat	IDT DNA	PCR XYL3 R (convergent)
DJ063	ccaaagtctaacaccgttccaagattatt	IDT DNA	IPCR XYL1 F V2 (divergent)
DJ064	caaaaccaacagctggcatatcata	IDT DNA	IPCR XYL1 R V2 (divergent)

DJ065	gaaatgtcatcgatcttgttcaagacc	IDT DNA	IPCR XYL2 R V2 (divergent)
M13F	TGTAAAACGACGGCCAGT	IDT DNA	pRS vector compatible primer outside MCS
M13R	CAGGAAACAGCTATGACC	IDT DNA	pRS vector compatible primer outside MCS

2.2 Media Preparation

2.2.1 Luria-Bertani (LB) Media

Powdered broth comprised of 10 g/l Tryptone; 10g/l NaCl; 5 g/l Yeast Extract, supplied by Sigma-Aldrich (catalogue No. L3522). Dissolve 25 g/l powder in distilled water, and autoclave.

2.2.2 LB Agar (MILLER)

Dissolve 37 g/l of powder in distilled water and autoclave (Merck-Millipore; catalogue number 1102830500)

2.2.3 2xYT Rich Media

EZMix powdered bacterial media. Manufactured by Sigma-Aldrich (catalogue number Y2627) Comprised of: 16 g/ Tryptone (pancreatic digest of casein), 10 g/l Yeast extract, 5 g/l NaCl, 0.6 g/l Inert binder (EZMix only). Dissolve 31.6 g/l of powder into distilled water and autoclave.

2.2.4 Synthetic Complete Dropout Dextrose Media and Agar (SC D/O)

Synthetic complete dropout media is used for the auxotrophic selection of yeast, with the option of excluding one or more essential amino acids for screening purposes. SC D/O was made using the following reagents: 20 g/l Glucose (VWR AnalaR Nomapur; catalogue number 101176K), 1.4 g/l Yeast Synthetic Dropout Medium Supplements (Sigma-Aldrich; catalogue number Y2001-20G), 6.8 g/l Yeast Nitrogen Base without Amino Acids (Sigma-Aldrich; catalogue number Y0626-250G), and for agar an addition of 20 g/l agar powder (Merck Millipore; catalogue number 1119251000).

Addition of amino acids:

Stock solutions were made of the amino acids by dissolving the following weights in 100 mL of sterilised water:

- Uracil: 0.2 g
- Leucine: 1 g

- Histidine: 1 g
- Tryptophan: 1 g

All solutions were fully dissolved and filter sterilised. Tryptophan, Histidine and Leucine are stored at 4 °C, and Uracil is stored at room temperature. Sterilised amino acids can be added to SC Media in several ways:

1. To liquid media/agar (not exceeding a temperature of 50 °C): 2 ml/l Histidine and Tryptophan, respectively; 10 ml/l Uracil and Leucine, respectively.
2. Top down spreading onto SC agar plates: 100 µl of Histidine, Tryptophan and Leucine, respectively; 200 µl Uracil.

2.2.5 Yeast Peptone Dextrose (YPD)

The following reagents were dissolved in the appropriate amount of water and autoclaved: 10 g/l of Yeast Extract (Merck-Millipore; catalogue number 1119261000), 20 g/l Glucose anhydrous (VWR AnalaR Nomapur; catalogue number 101176K); 20 g/l Peptone from Casein (Merck-Millipore; catalogue number 1119311000), and the addition of 20 g/l of agar (Merck-Millipore; catalogue number 1119251000) for YPD agar.

2.2.6 Synthetic Complete Xylose Media (SC-X)

D-Xylose (Sigma-Aldrich, catalogue number X1500-500G) was dissolved in distilled water to make a 50 % w/v stock solution, and filter sterilised. Synthetic complete drop out media (section 2.2.4) was made without the inclusion of glucose, and autoclaved (Yeast Nitrogen Base without Amino acids and Yeast Synthetic Dropout Medium with Supplements). The sterilised xylose was added to the autoclaved synthetic complete media to make a final concentration of 4% xylose in synthetic complete media.

2.2.7 Mixed Carbon Synthetic Complete Drop-out Media

Mixed synthetic complete dropout media was made using the same protocol as synthetic complete dropout media, whereupon the Yeast Nitrogen Base without Amino Acids and the Yeast Synthetic Dropout Medium Supplements were dissolved in distilled water and autoclaved. Glucose and Xylose were dissolved separately in distilled water to make 50% w/v stock solutions, and were filter sterilised. The sugars were then added to the sterilised broth at appropriate concentrations, typically at 0.05% Glucose with 4% Xylose, or 0.01% Glucose with 4% Xylose.

2.2.8 Yeast Peptone Xylose (YP-X)

This media was made exactly the same as YPD media (2.2.5 of this chapter), except instead of Glucose as the main carbon source, and exchanged for D-Xylose. The Xylose

was dissolved separately in distilled water to make a 50% w/v solution, and filter sterilised. The xylose was added to the autoclaved media YP-media to make a final concentration of 4%.

2.2.9 Mixed Carbon Yeast Peptone Media

Mixed YP-media was made using the same protocol as YP-X media (previous section), whereupon the Yeast Extract and Peptone from Casein were dissolved in distilled water and autoclaved. Glucose and Xylose were dissolved separately in distilled water to make 50% w/v stock solutions, and were filter sterilised. The sugars were then added to the sterilised broth at appropriate concentrations, typically at 0.05% Glucose with 4% Xylose, or 0.01% Glucose with 4% Xylose.

2.3 Methods Protocols

2.3.1 Gel Electrophoresis

1 % (w/v) agarose gel electrophoresis was used to separate DNA based on band sizes. Agarose powder (ThermoFisher Scientific; catalogue number 75000-500) was heated in 1X TAE solution (Tris base, acetic acid and EDTA, pH 8.0) until fully dissolved, cooled to 60 °C and poured into a mould tray. GelRed (VWR, catalogue number 89139-140; for imaging only) or SYBRSafe (ThermoFisher Scientific, catalogue number s33102; for consequent gel extractions), both at 10,000x concentrations, were added to the agarose gel using a micropipette, and mixed in gently. Any air bubbles were removed and the gel was left to set. DNA was prepared by adding 6x purple loading dye (supplied from NEB) and the mixture loaded into the set gel wells. 1 µl of NEB's 2-Log Ladder (see Figure 2.1) was added to the gel as a reference for band sizes. Gel moulds were placed into BioRad Gel Electrophoresis tanks, which were subsequently connected to the complementary BioRad power supply, and all gels were run at 100 V for 60 minutes. Upon completion, the gels were viewed either under the BioRad UV Illuminator (for imaging) or under a BlueBox (Clare Chemical Research) for gel extraction.

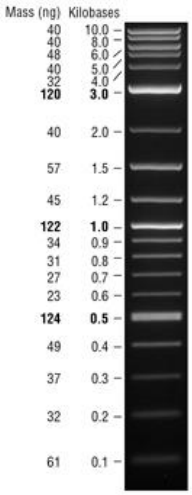


Figure 2.1 NEB 2-Log ladder. Figure showing the NEB 2-Log ladder used as a reference for DNA band sizes (visit https://www.neb.com/products/n3200-2-log-dna-ladder-01-100-kb for more information).

2.3.2 DNA Extraction and Purification

All plasmid DNA isolation was carried out using the Qiagen Spin Miniprep Kit. DNA purifications were completed using either the QIAquick Gel Extraction Kit or the QIAquick PCR Purification Kit. All protocols followed the provided manufacturer's instructions. DNA extractions are typically eluted in 50 µl of sterilised water, whereas DNA purifications are eluted at the lower volume of 30 µl in order to maintain a high DNA concentration. One variation added to the protocol was eluting using 100 µl of sterilised water (heated to 75 °C) when purifying DNA greater than 5 kb in size.

2.3.3 DNA Quantification

For average concentrations of isolated or extracted DNA verification was carried out using the NanoDrop 1000 Spectrophotometer, according to the manufacturers instructions. Prior to the measurement of each sample, a blank was made using the eluting liquid.

For sensitive experiments requiring more accurate estimations of DNA concentration, the Qubit 2.0 Fluorometer (ThermoFisher Scientific) was used, and DNA was prepared and measured according to the manufacturers instructions.

2.3.4 Restriction Digestions and Ligations

Restriction enzymes were provided by New England BioLabs (NEB) and digestion were carried out according to the manufacturers protocol, typically by incubation at 37 °C for one hour in CutSmart Buffer with a total reaction volume of 20 µl. Double digestions were set up in the same manner, unless otherwise required by the manufacturer.

Ligations were carried out using T4 DNA Ligase at room temperature (20 – 23 °C) for 1 – 3 hours, followed by heat inactivation at 65 °C for 20 minutes, all reagents supplied by New England BioLabs and according to the protocol provided by the manufacturer.

2.3.5 Polymerase Chain Reaction (PCR)

All primers were ordered from Integrated DNA Technologies (IDT), shown in Table 2.4. Upon delivery all primers were resuspended with sterile nuclease free water to a 100 mM concentration. These were then further diluted to 10 mM working stocks and utilised for all purposes thereon, adjusting the required volumes accordingly. All PCR amplification protocols were carried out using Phusion-HF DNA Polymerase (supplied from NEB) and reactions were set up with various volumes, depending on downstream reactions. The following reagents guideline was employed for all reactions: i) 10 – 30 ng template DNA; ii) 10 mM Primer 1; iii) 10 mM Primer 2; iv) 5x GC Buffer/5x HF Buffer; v) 0.02 U/µl Phusion-HF DNA Polymerase; vi) 2.5 mM dNTPs (NEB); vii) (optional) 2% DMSO; viii) nuclease free water. For analytical purposes a total volume of 20 µl was used, whereas for subsequent gel extractions larger volumes of 50 µl were utilised.

The thermocycler is programmed with the desired programme, typically a) an initial denaturation step at 98 °C for 3 minutes b) then a cycle of denaturation at 98 °C for 30 seconds, annealing stage for 30 seconds at 50 – 72 °C (according to primers), and an elongation at 72 °C for 30 seconds per kilobase of product. Repeat 25-30 times; c) a final elongation step at 72 °C for 10 minutes. NEB's online software tool "NEB Tm Calculator v1.7.2" was used for calculation of primer annealing temperature.

2.3.6 Touchdown PCR

For more complicated reactions, or for primers with annealing temperatures exceeding a difference of more than 5 °C, a touchdown PCR was used instead. The PCR reagents and protocol is the same, apart from the annealing temperature of the primers. An extra stage is added after the stage (a), whereby the starting annealing temperature is 2 – 3 °C higher than that of the highest annealing temperature of the two primers and is gradually reduced by 1 °C per cycle to 2 °C below the lowest annealing temperature. The protocol is then resumed as normal (as stated above from point (b)) with an annealing temperature of the lower of the two primers.

2.3.7 Preparation of *E. coli* Competent Cells

This protocol was adapted from Dr Ben Blount (ICL; unpublished data). In order to maintain maximum cell efficiency, cells should be kept at a low temperature at all times and in exponential growth phase. All steps should be carried out under sterile conditions. Day

One: Inoculate an *E. coli* colony in 5 ml of LB broth and incubate at 37 °C overnight, shaking. Prepare and autoclave 500 ml of LB broth in a conical flask and store overnight at 37 °C. Prepare and autoclave 500 ml of distilled water and store at 4 °C in a fridge overnight. Day Two: Re-suspend the 5 ml overnight culture in the heated LB conical flask at a dilution rate of 1:1000. Store at 37 °C shaking, and incubate for approximately 3-4 hours until $OD_{600\ nm}$ reaches 0.5. Prepare 20% (v/v) filter sterilised glycerol by incubating on ice, and place 70-80 sterilised microfuge tubes in a – 80 °C freezer. Following incubation, transfer all of the *E. coli* culture into 50 ml sterile falcon tubes (adding 30 ml of culture to each falcon tube), and place on ice for 30 minutes. During this period, pre-chill a bench top centrifuge rotor to 4 °C. Centrifuge the chilled falcon tube for 15 minutes at 4000 rpm. Carefully discard the supernatant ensuring not to disturb the bacterial pellet. Re-suspend in the equivalent amount of sterile chilled water, by on mixing the falcon tubes gently on ice until the pellet is fully re-suspended. Centrifuge as before. Discard the supernatant. Re-suspend the pellet in 2 ml of the pre-chilled 20% (v/v) glycerol (per falcon tube). Pool all the culture into one 50 ml falcon tube and centrifuge as before. Remove the supernatant. Re-suspend in a small volume of 20% (v/v) glycerol (with the aim of maintaining highly concentrated cells, a volume between 2 and 3 ml is appropriate), by gently shaking on ice. Transfer 50 µl of cells per chilled microfuge tube, and store at -80 °C immediately.

2.3.8 Golden Gate Assembly

The Golden gate assembly kit and protocol was used for cloning and plasmid construction purposes, courtesy of Robert Chen and the Dueber Lab[121]. For compatibility with the golden gate system, all “parts” (defined as individual promoters, genes, terminators, markers, etc.) must each have a “prefix” and a “suffix” sequence. These short flanking sequences determine the location and annealing to adjacent parts, which ensure correct assembly of “cassettes” (which are multiple parts assembled together into a single construct). Should any sequences need to be made compatible with the golden gate kit, this can be done easily through designing and ordering primers (with overhangs coding for the prefix and suffix sequences) which anneal to the desired part.

2.3.9 PCRtag Analysis

For each genomic DNA sample, create a master mix based on the following recipe for one reaction: 6.25 µl of GoTaq Green (Promega; catalogue number M7122), 0.25 µl of genomic DNA prep and 5.5 µl of sterilised water. Aliquot the 11.5 µl of the master mix into a PCR tube and add 1 µl of the pre-mixed primers (2.5 µM each). Thermocycle the mixture using the following PCR conditions: 95 °C for 3 minutes, (95 °C for 30 seconds, 55 °C for 30

seconds, 72 °C for 30 seconds) x 30, 72 °C for 7 minutes, 4 °C infinitely. Load the sample(s) directly onto a 1% agarose gel for band separation.

2.3.10 PEG Precipitation and Coupled Ligation (unpublished)

This protocol was developed by Dr Junbaio Dai (Tsingua University, China) formulated especially for the Sc2.0 project and the concentration of chunks, prior to the ligation into mega-chunk. Make a stock solution of 40% PEG6000/10mM $MgCl_2$, then mix with the chunks' DNA in a 2:1 ratio (chunk DNA: PEG mixture). Leave the mixture at room temperature for 15 minutes, and spin down at top speed for 10 minutes, using a cooled 4 °C centrifuge. Remove the supernatant, making an effort not to disturb the pelleted DNA. Resuspend using the following volumes of reagents: 14.5 µl of sterilized water, 1 µl T4 DNA Ligase (NEB), 2 µl 10X T4 ligase buffer (NEB), and 2.5u µl of the 40%PEG6000/10mM $MgCl_2$. Incubate overnight at 16 °C, and add the entire mixture to the transformation reaction.

2.3.11 Transformation Protocol used for Sc2.0

This protocol is courtesy to Dr Leslie Mitchell (NYU; unpublished protocol) and has been utilised in the exact way as instructed. Day one: The day before the scheduled transformation, inoculate the yeast strain of choice in 5 ml of appropriate media. Incubate at 30 °C for 14 – 18 hours until at full saturation. Day two: Resuspend the yeast cells in fresh appropriate media, so that starting OD 600 nm reads between 0.1 – 0.2. For logarithmic phase, grow for about four hours at 30 °C, or until OD 600 nm reads between 0.5 - 0.7. Prepare the yeast cells to be competent by harvesting the cells in 50 ml falcon tubes by centrifugation at 2000 rpm for 3 – 5 minutes in a clinical centrifuge. Discard the supernatant. Resuspend the cells by vortexing in 5 ml of 0.1 M Lithium acetate (Sigma-Aldrich; catalogue number 517992-100G). Adjust the volume to the original volume of the cells. Spin down the cells at 2000 rpm at room temperatures for 3 – 5 minutes, and discard the supernatant. Resuspend the cells in a small volume of 0.1 M Lithium acetate, so as to ensure 100 µl per transformation. Thus if you are doing four transformations, resuspend in 400 µl. Aliquot and denature the salmon sperm DNA (Trevigen; catalogue number 9610-5-D) in 100 °C heat block for 6-10 minutes, then place on ice for at least 5 minutes. For best efficiency, use carrier DNA that is freshly boiled and avoid repeated boiling/cooling cycles. Aliquot 100 µl of the concentrated competent cells to each microfuge tube, and add 10 µl of the salmon sperm DNA to each sample. Mix briefly by vortexing, then add up to 1 µg of ligated DNA, and mix again. Incubate for 30 minutes at RT or 30 °C. Vortex briefly and then add 600 µl of 50% PEG 3350 (Sigma Aldrich; catalogue number 202444-250G), 90 µl of 1 M Lithium

acetate, 100 µl DMSO (NEB) and top up with water so that the final volume is 1000 µl. Mix gently, but well and incubate the cells in PEG-LiOAc-DMSO mixture for 30 minutes at RT or 30 °C.

Heat shock stage: Place the tubes in a 42 °C water bath or heat block for 14 minutes, mixing gently halfway through. Spin down the cells in a microfuge tube at 2000 rpm for 3 - 5 minutes. Discard the liquid by decanting or aspirating. Wash the cells by gently resuspending in 1 ml of 5 mM Calcium chloride (Sigma Aldrich; catalogue number C1016-500G-D), and incubate for exactly 10 minutes at room temperature. Aliquot 250 µl onto the correct dropout plates, making sure to plate the entire transformation mix. Incubate the plates at 30 °C for 2 – 3 days face down.

2.3.12 Phenol-Chloroform Yeast Genomic DNA Prep

This protocol is adapted from Dr Leslie Mitchell's instructions (unpublished).

Day one: Inoculate colonies of choice in the appropriate media and grow overnight at 30 °C for 14 – 18 hours, until at full saturation. Day Two: Centrifuge 750 µl of the saturated overnight (2000 rpm at 2 minutes) in screw cap tubes (with O-rings) and discard the supernatant. Resuspend the pellet in 200 µl of breaking buffer by pipetting up and down or vortexing. [Breaking buffer: 50 mM Tris (8.0) (Fischer Scientific; catalogue number BP152-1), 100 mM NaCl (Sigma-Aldrich; catalogue number S9888-500G), 1% SDS (Sigma-Aldrich; catalogue number 436143-100G), 2% TX100 (Sigma-Aldrich; catalogue number T8787-50ML), and 1 mM EDTA (Sigma-Aldrich; catalogue number E6758-100G)].

Add 0.5 mm acid washed glass beads (Sigma-Aldrich; catalogue number G8772-100G) up to the top of the liquid/cell mixture. In a fume hood, add 200 µl of phenol – chloroform – isoamyl alcohol mixture (Sigma-Aldrich; catalogue number 77617-100ML). **WARNING: This chemical is highly toxic**. Screw the tubes shut, ensuring that the seal is tight. Agitate the cells for 5 – 10 minutes at room temperature using a vortex and a bead beater, or using a desktop shaker. Centrifuge for 10 minutes at top speed. In a fume hood, transfer 75 – 100 µl of the top, aqueous layer into labelled tube containing 1 ml 100% ethanol. Invert five times to mix and centrifuge for 20 minutes at 4 °C at top speed and discard the supernatant. Add 500 µl of 70% EtOH and invert to mix. Centrifuge for 5 minutes at room temperature and top speed and discard the supernatant. Air-dry the pellet at room temperature for 10 minutes. Resuspend in 50 µl of sterilized water. Store the genomic DNA prep indefinitely at -20 °C.

2.3.13 Bacterial Transformations

a) Electroporation:

Electrocompetent bacterial cells were transformed using Electroporation Cuvettes provided by Bio-Rad Laboratories (catalogue number 165-2088) and using the MicroPulser Electroporator, (also by Bio-Rad). Cells were made competent prior to electroporation.

Switch on the MicroPulser and adjust the settings to “Bacteria” and the Measurements to “Time ms”. Pipette 0.5 – 2 µl of DNA (using between 0.05 – 0.2 µg of DNA) between the metal plates of a pre-chilled electroporation cuvette. Transfer frozen competent cells onto ice and pipette 50 µl of the thawed cells as soon as possible on top of the DNA sample in the cuvette. Using tissue roll, wipe down the cuvette prior to placing it into the shocking chamber making sure it is entirely moisture-free. Place the cuvette into the slide and firmly push into the shocking chamber until contact has been made. Press the PULSE button once (accepted *E. coli* readings are between 4.00 – 6.00 ms) and immediately add 300 µl of LB broth, transfer the reaction into a 1.5 ml eppendorf tube and incubate at 37 °C for 30 – 45 mins. Plate 80 µl onto LB agar plates harbouring the appropriate antibiotic.

b) Heat-Shock:

Turbo Competent cells were the only strain used for heat-shock bacterial transformations and were made chemically competent prior to the procedure.

To 200 µl of frozen cells, add 50 µl of 5x KCM buffer, using the pipette to mix the reagents until fully resuspended and thawed. Combine 50 µl of the KCM-cells mixture and 10 µl of ~ 0.5 – 1 µg of DNA into a PCR tube. In a thermocycler, set up the following program: i) 10 minutes at 4 °C; ii) 1 minute at 42 °C; iii) 1 minute at 4 °C; iv) 30 minutes at 37 °C. Plate the entire transformation onto LB agar with the appropriate antibiotic.

2.3.14 Preparation and Transformation of DNA into Yeast Cells

This protocol is adapted from Gietz, R.D. *et al* (2002)[122]. Yeast transformations were always carried out using chemically competent cells and the cells were made competent during the procedure. Day One: From a plate or glycerol stock, inoculate the yeast strain of choice into 5 ml of appropriate media and incubate shaking at 30 °C overnight until full saturation. Day Two: Re-inoculate the saturated overnight culture into 10 ml of fresh media at OD 0.2 (at 600 nm) and set shaking at 30 °C. Harvest the cells until log phase or between 0.5 – 0.7 OD. Centrifuge the cells at room temperature at 2500 rpm for five minutes. During the centrifugation, boil 10 µl (per transformation) of 5 mg/ml salmon sperm DNA for five minutes at 100 °C and then place on ice. Discard the supernatant and gently re-suspend the cells in 5 ml of sterile water. Repeat centrifugation using the same conditions. Carefully remove the supernatant without disturbing the pellet. Using 100 µl of distilled water per

transformation, re-suspend the pellet and pipette into 1.5 ml eppendorf tubes. Using a microfuge, spin down at top speed for 20 seconds. Discard the supernatant and add the following reagents to the pellet in the listed order: i) 240 µl of 50% w/v PEG 3350; ii) 34 µl 1M Lithium acetate; iii) 10 µl of 5 mg/ml salmon sperm DNA; iv) 74 µl of DNA (0.1 – 10 µg) and water mix. Vortex rigorously until the cells and reagents are equally distributed in the tube.

Using a water bath or heat block, heat shock the mixture at 42 °C for 45 minutes. Spin down at top speed for 30 seconds and using a pipette, remove as much of the supernatant as possible. Without disturbing the pellet, gently pipette 1 ml of 5 mM Calcium chloride and gently invert the tube top-down on the bench. Invert the tube slowly after five minutes. Repeat three times. Using a pipette, re-suspend the pellet completely, and add between 80 – 200 µl of the transformation mix onto ready prepared SC agar plates. Incubate at 30 °C for 2 – 3 days.

2.3.15 Yeast Genomic DNA Chelex Preparation

This protocol is adapted from the Yeast Chelex Protocol as developed by Dr. Ben Blount[123]. Day One: Resuspend required yeast strain in 2 ml of YPD and incubate shaking overnight at 30 °C. Day Two: Transfer the overnight into a 2 ml eppendorf tube and spin down at 8000 rpm for 2 minutes. Remove the supernatant. Resuspend the pelleted cells in 300 µl of 5% v/v Chelex-100 (BioRad; catalogue number 1422822) and a small quantity of acid-washed glass beads. Vortex the tube(s) at top speed for 4 minutes, followed immediately by incubation at 100 °C for 10 minutes. Using a microfuge, spin down the reaction at 8000 rpm for 10 minutes. By carefully pipetting away, remove 30 – 100 µl of the supernatant making sure not to take up any of the glass beads-Chelex mix when doing so. Transfer the extracted genomic DNA into a clean tube and store at -20 °C.

2.3.16 Inverse PCR for Analysis of SCRaMbLE-d Yeast Strains

A multi-step inverse PCR reaction was carried out in order to determine the location of any integrated genes upon SCRaMbLE-in. Genomic DNA of strains of interest was isolated using a Phenol-Chloroform Genomic Prep (see section 2.3.13) and DNA was quantified using the QuBit 2.0 Fluorometer for accurate concentration determination (section 2.3.3).

(i) Regular digestion of the genomic DNA:

The genomic DNA was digested with XbaI (NEB) which cuts the SynXI.A-C regularly (35 times within the synthetic region), for 2 hours at 37 °C and according to the following recipe: (i) 900 ng genomic DNA; (ii) 1.5 µl XbaI; (iii) 5 µ 10x Cutsmart Buffer (NEB); (iv) n µl of sterilised water for final volume of 50 µl. The reaction was heat inactivated for 20 minutes

at 65 °C. Inactivation of the restriction enzyme was necessary in order to prevent any unwanted interaction or effects on downstream reactions. If heat inactivation is not possible, the digestion reaction needs to be PCR purified (section 2.3.2) in order to remove the presence of the enzyme. Restriction enzymes used should cut only the genomic DNA, but not within the known, SCRaMbLE-in gene (such as the KanMX sequence) in order to determine the integrated locus/loci of the heterologous gene.

(ii) Gradient Overnight Ligation:

A gradient overnight ligation was carried out in order to gradually circularise the digested DNA. The heat inactivated digestion reaction with the following reagents: (i) 30 µl digestion reaction; (ii) 2 µl of ATP; (iii) 1 µl T4 DNA Ligase (NEB); (iv) 17 µl sterilised water. The ligation reaction was placed in a medium-sized polystyrene icebox (supported by a float), and covered with the lid of the box. The gradient PCR encourages slow circularisation of the genomic DNA. Following the overnight incubation, the reaction could be stored in the fridge indefinitely.

(iii) Inverse PCR:

Divergent primers (DJ047 and DJ048) were designed to anneal within the KanMX sequence but amplifying outwards, to reveal the flanking genomic sequences (refer to Table 2.4 for primer sequences), thus the inversed PCR reaction. Without any purification following the ligation step, the PCR reaction was set up in the following way: (i) 3 µl ligation reaction; (ii) 2 µl of 10x dNTP mix; (iii) 1 µl of 10 mM forward primer; (iv) 1 µl of 10 mM reverse primer; (v) 0.5 µl Phusion Polymerase (NEB); (vi) 0.2 µl DMSO; (vii) 4 µl 5x GC Buffer; (viii) 8.3 µl sterilised water. A touchdown PCR was run on the thermocycler.

2.3.17 Glycerol Stocking of Plasmids and Strains

20% v/v distilled glycerol (VWR, East Grinstead, UK) was filter sterilised and used for storage of cells. Cell cultures (either bacterial or yeast) were resuspended in glycerol at a one-to-one ratio of volume into microfuge tubes (i.e. 600 µl of cell culture in 600 µl of 20 % glycerol. Tubes were briefly vortexed and immediately stored at -80 °C.

2.3.18 Yeast Growth Assays

All yeast strains were inoculated in appropriate media for overnight growth at 30 °C and harvested until full saturation. The next day cultures were spun down and resuspended in sterilised water three times, to wash away the media, leaving the cells only. Growth assays were carried out in 250 ml Erlenmeyer flasks, containing 50 ml of necessary media. The cells were resuspended at a starting OD_{600} of 0.03 AU (λ = 600 nm), and incubated for 80 hours at 30 °C. Should OD_{600} readings exceed 0.5 A.U., samples were diluted in the resuspension media, to ensure accuracy of measurements. OD_{600} measurements were

made using 1 ml disposable cuvettes (VWR; catalogue number 47744-644), and the Spectrophotometer apparatus (Jenway, model 6300).

2.3.19 Mating of Yeast cells

The two haploid strains were inoculated in 3 ml of YPD and harvested overnight at 30 °C. The following day, without taking OD measurements, 200 µl of each haploid culture was resuspended in 3.6 ml of fresh YPD and grown to exponential phase for four hours at 30 °C. The mating cultures were then streaked onto a clean YPD agar plate, and viewed under the Singer Spore Play Modular Microscope. Using the glass microscopic needle of the apparatus, zygotes were singled out, and transported to a clean part of the plate. These were then incubated and allowed to grow until pick-able colonies were observed, and then further testing was carried out to confirm diploidy of the newly formed cells.

2.3.20 Flow Cytometry and Gating Data

Flow Cytometry was carried out using the LSR BD Fortessa X-20. Prior to extracting statistics from the cell samples collected, several gates were applied to the flow cytometry data, with the aim of collecting cells of similar characteristics, shown in Figure 2.2. From the forward (FSC-A) and side scatter areas (SSC-A), a tight gate was added to try and include the area of the highest cell density, typically annotated in green or red. This first gate filters for cells of similar size and complexity, set at $3x10^3$ and $2x10^4$ arbitrary units (of the SSC-A scale), excluding any debris, contamination or noise. In all four control examples, the percentage of cells collected is above three quarters of the entire cell count (N = 10,000), thus exhibiting a well placed filter. Entering this subset population, we plot mRuby2-A vs sfGFP-A and apply a quadrilateral gate to expel fluorescing outliers situated on edges of the axes, in order to prevent skewed of future statistics. From gate two in Figure 2.2, yRC1841, yRC1842, yRC1844 and yRC1845 all have a catchment area of 99.9, 99, 99.9 and 99.5%, respectively. The same applies to gate three, whereby the filtering out of any mTagBFP2-A outliers is made, albeit all samples have 100% of the cells within the gate. These are control samples, however this is not the case for several of the SCRaMbLE-d samples, therefore we maintain this gate. The fourth and final gate was the hardest to apply, as it searched for yeast doublets, arising due to cell clumping during excitation and fluorescence measurement. Yeast single cells tend to have a similar area to height ratio, therefore by plotting FSC-A (area) vs. FSC-H (height), we can eliminate the population of cells (shown as the red polygon in Figure 2.2, yRC1841, gate four) that are exempt from the accepted sample group. We used FlowJo software (TreeStar) to manipulate all flow cytometry data.

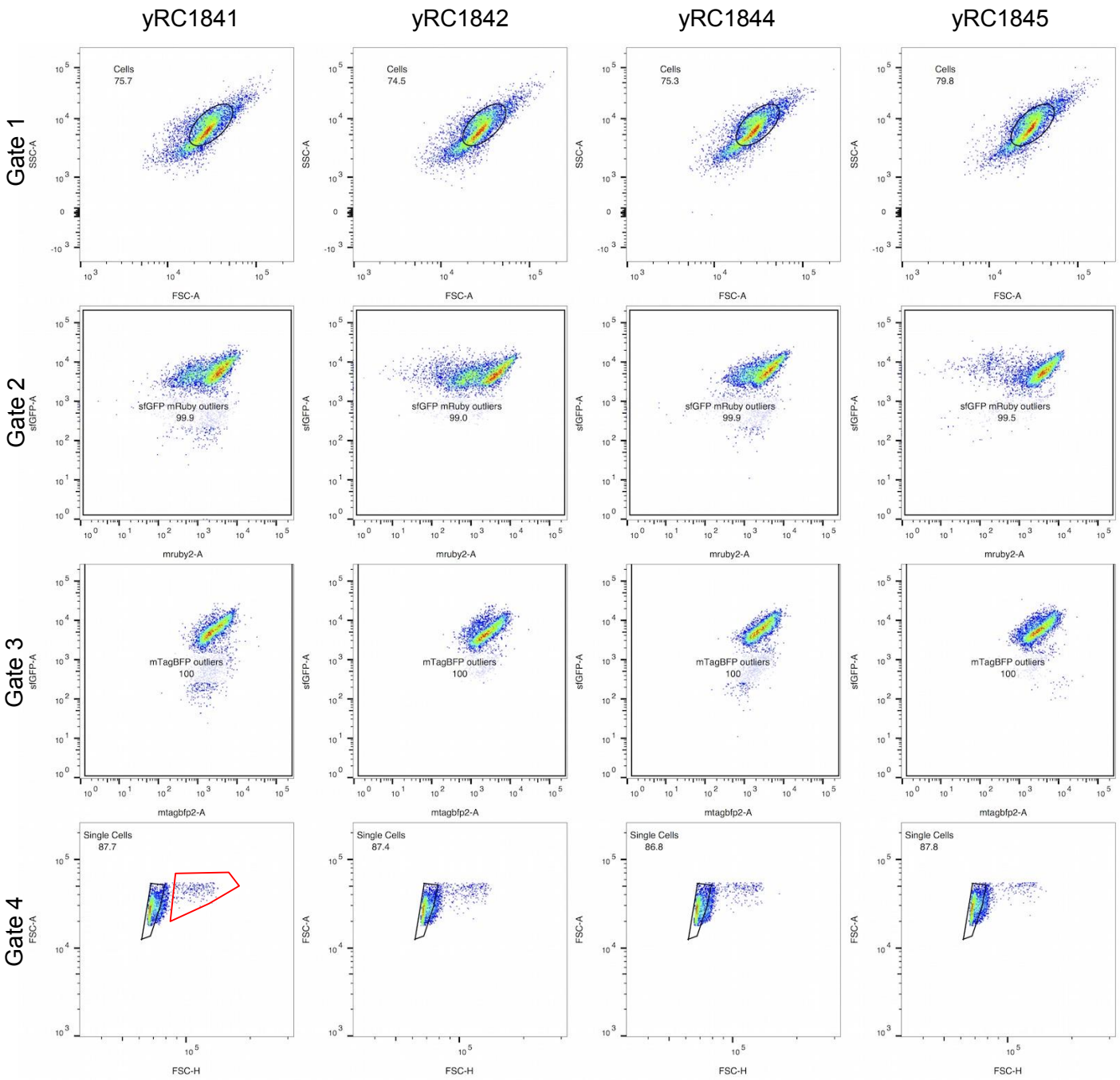


Figure 2.2 Gating methods applied to Cells. We used the four un-SCRaMbLE-d controls, yRC1841-5, to create gates on several levels of the flow cytometry analysis we obtained. In Gate 1, from the forward (FSC-A) and side area scatters (SSC-A), we applied an oval gate to filter for cells of similar size and complexity. We then entered into this cell population subset, to make Gate 2. Gates 2 and 3 are carried out using a square area, and both filter out samples based on the edges of the plot. These points have a tendency to skew data undesirably, therefore we isolated sfGFP-A vs. mRuby-A off the edges (gate 2), and repeated the same process for mTagBFP-A (vs sfGFP-A, gate 3). In the final gate, here we apply a polygon shaped gate to filter out doublets that can arise from cell clumping, especially as doublets can be misleading in this data and mistaken for duplications, or lead to bimodal populations. The red polygon shown in gate 4 of yRC1841 represents the population of cells that are excluded from the data set. Gates were tailored according to the control samples, and then applied to all samples.

3.0 CHAPTER 3: Building a Functional Synthetic Chromosome XI

In Chapter 1, the underlying principles surrounding the Sc2.0 project were introduced and the plethora of intended genomic sequence re-designs were detailed. These are summarised in Figure 1.2. In this Chapter, I will briefly describe the methods developed as the Sc2.0 project was established, and then describe the standardised hierarchical assembly approach used for SynXI construction. At the start of my PhD, I initially tested this assembly approach using a starter kit provided by our collaborators. I was then able to introduce this method as the main assembly approach used in the SynXI construction workflow now underway at Imperial College. The below aims summarise the intention of my work in this Chapter towards Synthetic Chromosome XI construction at Imperial College London.

3.1 Aims

- To ensure correct synthesis of the purchased synthetic DNA chunks and prepare these for “mega-chunk” assembly in order to make the first 90 kb of chromosome SynXI for the Sc2.0 project.
- To transform DNA mega-chunks in a specific order into the parent strain using iterative homologous recombination and auxotrophic selection.
- To determine by PCR and sequencing that synthetic DNA has entirely and accurately replaced the wild type DNA in yeast chromosome XI, converting the original sequence into desired synthetic DNA sequence.
- To investigate potential methods to increase the efficiency of mega-chunk assembly to aid in the further work of the Sc2.0 project.

3.2 Introduction

The Assembly Approaches used by the Sc2.0 Project and the Characteristic of *Saccharomyces cerevisiae* Chromosome XI

There were several methods used during the initial construction of synthetic chromosomes during the first five years of the Sc2.0 project. These were based on the contemporary research and the technologies available at the time. The most cost efficient method for the assembly of synthetic DNA involved the training and utilisation of "Build-a-Genome" (B-A-G) course undergraduates. As part of a teaching course module, the B-A-G students were tasked with building ~750 bp "building blocks" of double stranded DNA, which they assembled by annealing overlapping 70 – 80mer oligonucleotides using iterative rounds of PCR[68, 124]. They cloned these building blocks into bacterial vectors, and these were then used as the starting point of a second assembly stage. The 750 bp building blocks (BB) were digested from their vectors then ligated to one another to create a 2 to 4 kb "mini-chunks", with these cloned into an *E. coli*/yeast shuttle vectors. Overlapping sets of mini-chunks were subsequently transformed into yeast into the native chromosomes using homologous recombination to direct them to the desired locus. Although this method of training students to do the work was cheap and labour efficient, the associated drawbacks of setting up a course and having to do a lot of training made this a top-heavy investment that was time-consuming, and also error prone (both human error, and PCR-based errors). Despite these challenges, this was the method employed for the construction of SynIII, one of the smaller yeast chromosomes[60] and the first fully completed.

It was then recognised that in order for the synthetic yeast genome project to scale to constructing the complete genome there would need to be i) greater help and research expertise beyond B-A-G students, and ii) extensive funding for a near-future completion date. The global expansion of the project in the past five years, contributed both reinforcements in the form of the labour and in external funding provided by (professional) research groups. Assembly methods evolved as well, although all approaches still apply hierarchical assembly, albeit without many of the laborious and time-consuming cloning and PCR stages required to make building blocks. The majority of the research groups involved now outsource the assembly of chunks (or mini-chunks) using specialist DNA/gene manufacturing companies.

The outsourcing approach was first demonstrated by work described in Dymond *et al* (2011)[72]. In this work they enlisted the DNA synthesis company Codon Devices to produce a BAC (bacterial artificial chromosome) containing all the assembled synthetic DNA for the right arm of synthetic chromosome IX. Codon

Devices used their own proprietary DNA assembly methods to make this BAC. For the left arm of synthetic chromosome VI (SynVI.L), they developed and used an alternative method of "mega-chunk" assembly and transformation, which is has now become the generally accepted approach used for construction of all chromosomes, including our synthetic chromosome XI. For mega-chunk assembly, the starting point is "chunks", approximately 7 to 10 kb in size. These can be assembled together from mini-chunks by different DNA assembly methods or can simply be purchased. Mega-chunk assembly is carried out by digesting chunks with rare-cutter restriction enzymes and ligating these chunks together *in vitro* into linear 30 to 50 kb "mega-chunks" (as illustrated in Figure 3.1). The mega-chunk is built to contain an auxotrophic selection marker (e.g. Leu) and is transformed into yeast to replace the wild-type sequence for the equivalent chromosome region by homology-mediated integration. Only one 50 kb mega-chunk is added at a time, in one step. The reasoning behind the step-wise replacement of wild type chromosomal DNA by integration of one mega-chunk at a time is to ensure easy traceability of detrimental phenotypes. If the synthetic DNA introduced results in any morphological or fitness changes, the step-wise approach ensure that these changes can be easily ascribed to sequence changes in the latest integrated mega-chunk.

The step-wise approach to sequentially integrate multiple mega-chunks into the yeast genome to construct synthetic chromosomes is shown in Figure 3.2. In this method mega-chunks are transformed into the yeast genome using an iterative homologous recombination approach where the selection used to identify integration

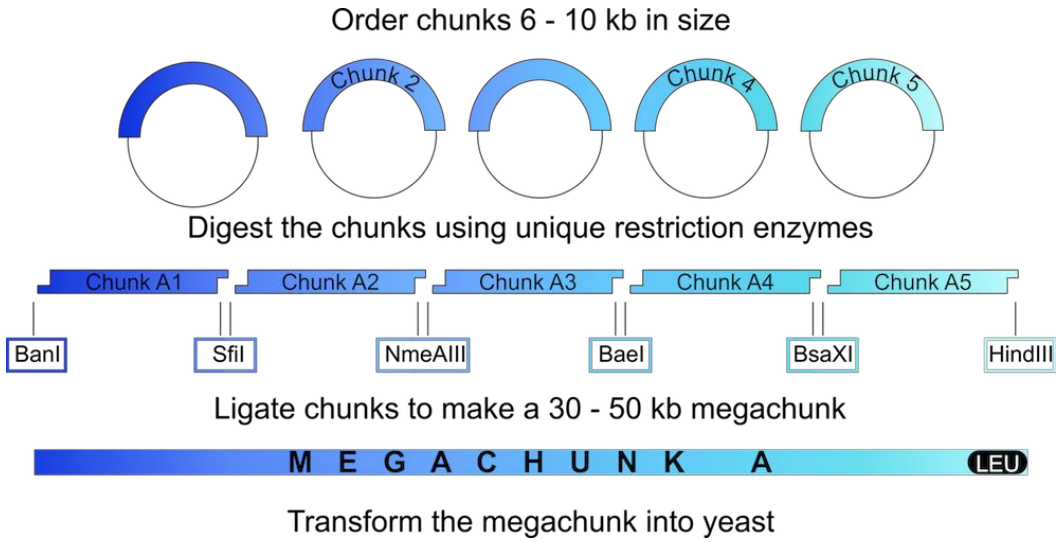


Figure 3.1 Mega-chunk Assembly. This illustration depicts the process of events from the delivery of synthetic chunk DNA to mega-chunk formation. The chunks encode unique restriction enzyme sites determining the sequential ligating to the ensuing chunk, to create a 30 – 50 kb synthetic mega-chunk. The mega-chunk is then transformed into yeast, recombining to replace the wild type DNA with the synthetic. Adapted from Jovicevic *et al* (2014)[125].

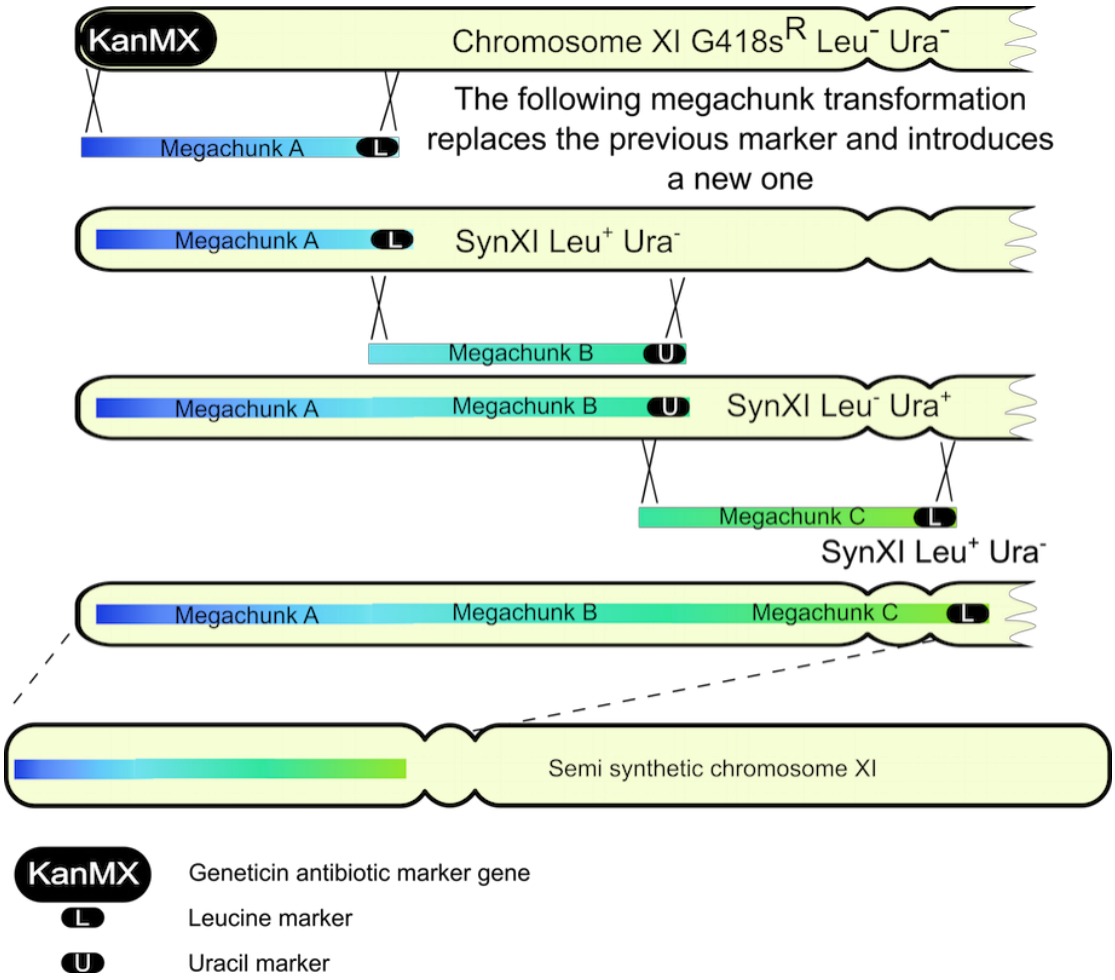


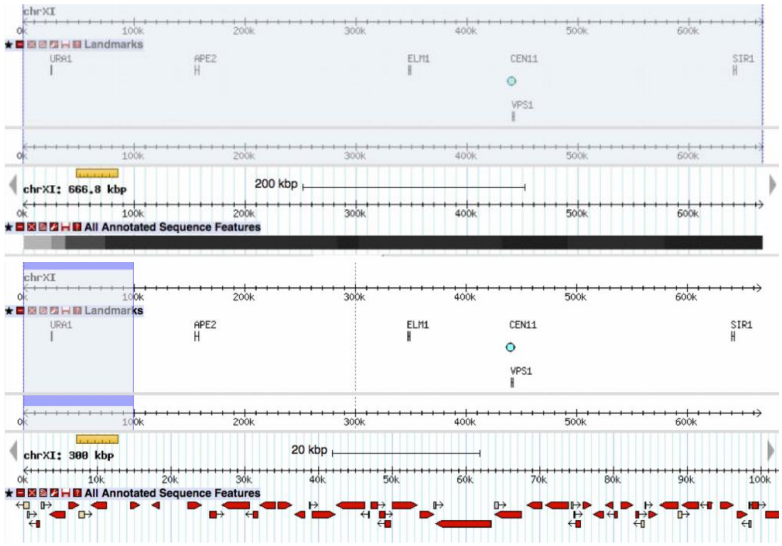
Figure 3.2 Iterative homologous recombination and marker swapping technique used for Sc2.0. The starting host strain contains a KanMX gene at the far left point of the chromosome (XI). The first mega-chunk, A, recombines the KanMX marker out and introduces the Leucine (L) marker, creating strain SynXI.A. This new partially-synthetic strain is consequently used for the next transformation round, of Mega-chunk B, which recombines before the integrated leucine marker, and swaps in the Uracil (U) marker, creating the new strain SynXI.A. Adapted from Jovicevic *et al* (2014)[125].

is iterated between two auxotrophic selection markers. For this approach, the far left part of every mega-chunk encodes a homologous region upstream of the marker gene already present as part of the previously integrated mega-chunk. The far right part contains a different selectable marker gene followed by homology to the wild type chromosome sequence. Thus with every transformation round, the existing marker gene is swapped out by the left region of the incoming mega-chunk, and this also swaps in its own marker about 50 kb downstream. This marker swapping technique is used in order to screen for genotypically correct integrants, where yeast colonies can be screened for the loss of the existing marker gene and addition of the new marker gene. However, as the first mega-chunk of every chromosome does not have a previously integrated marker to recombine out, starting strains are modified to include a KanMX gene (conferring resistance to Geneticin) at the chromosome's

starting transformation locus, usually the left arm telomere. This approach of mega-chunk assembly and iterative homology-directed integration, first developed for SynVI.L construction, is the approach we subsequently used for all assembly and integration steps for making synthetic chromosome XI.

Chromosome XI naturally consists of 666,816 bp of DNA, encoding 345 open reading frames (ORFs) and 16 tRNA genes. Figure 3.3a shows screenshots taken from SGD's (*Saccharomyces cerevisiae* Genome Database) website[126], a frequently used website that enables the genome browsing of the *S. cerevisiae* wild type genome. In 2012, our lab successfully secured £21,000 internal funding to initiate the synthesis, assembly and integration of the first 90 kb of DNA for synthetic chromosome XI. This synthetic chromosome was first designed using automated design software (Biostudio) by our lab in collaboration with researchers at Johns Hopkins University. In line with the rest of the Sc2.0 project, it was designed to have major modifications compared to the wild type chromosome sequence, including inserting 194 loxPsym sites into the 3'UTR of all non-essential genes, and deleting of the 16 introns that span over the wild type chromosome which have been previously shown to have no deleterious fitness effects[56]. The synthetic chromosome design was in total 7,139 bp shorter than the original sequence and has approximately 95% overall sequence identity compared to the wild type chromosome. The synthetic sequence also differs from the wild type sequence by the implementation of short synonymous codon changes in every protein-coding gene, allowing researchers to distinguish the DNA and mRNA sequence of the genes from the synthetic and wild type chromosomes. In this chapter, I briefly describe the design of SynXI, demonstrate the method of synthesis, assembly and verification of the first mega-chunks of chromosome SynXI, and finally assess alternative DNA preparation techniques that have the potential to increase assembly efficiencies for Sc2.0.

a)



b)

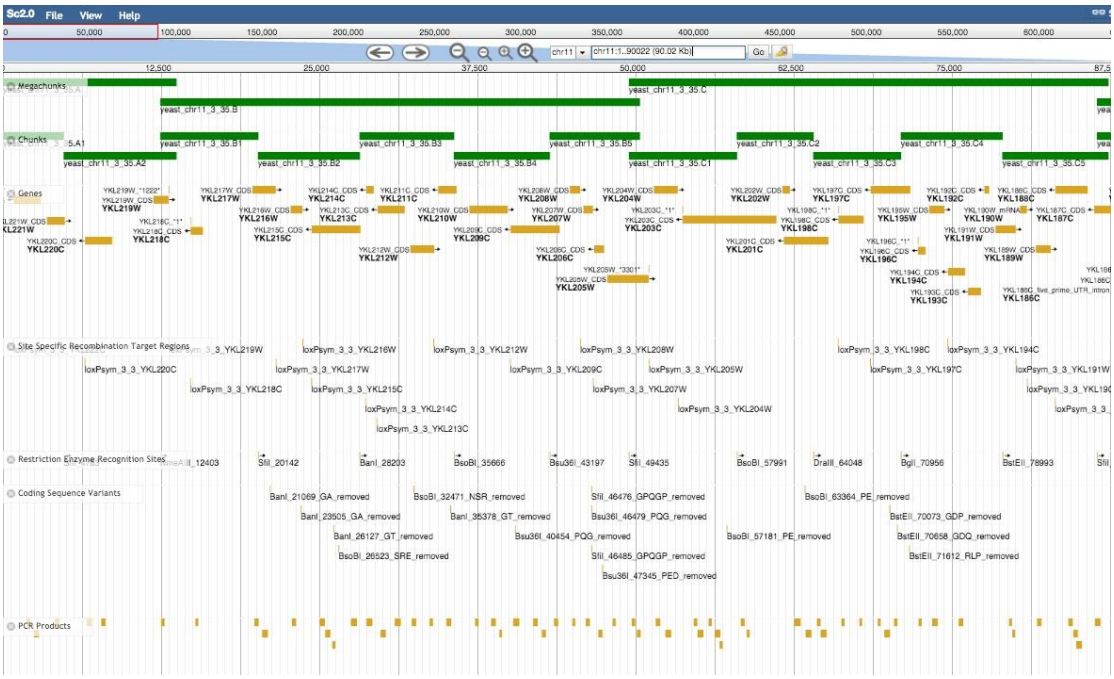


Figure 3.3 The Wild Type and Synthetic Chromosome XI Sequence. Screenshots of the publicly accessible SGD (*Saccharomyces cerevisiae* Genome Database; found at: http://www.yeastgenome.org/) website showing the whole wild type chromosome XI (a) and the annotated features and genes of the first 100 kb of the chromosome. The ORFs are shown in red. The bottom screenshot shows the first 90 kb of synthetic chromosome XI as shown in the JBrowse website (b), which displays the mega-chunks(green), chunks (green), genes (yellow), restriction enzyme sites and PCR products. Also included in the screenshot are the positions of the loxPsym recombination sites found throughout the chromosome. This site (found at: http://syntheticyeastresource.com/main/JBrowse-1.11.5/index.php?loc=chr11%3A1..53700&tracks=tRNA%2CSite%20Specific%20Recombination%20Target%20Regions%2CMegachunks&highlight=) is only accessible to the collaborators and scientists part of the Sc2.0 project, as an online tool to access up to date information and progress related to the construction of the 16 synthetic chromosomes.

3.3 Results

3.3.1 The Design of Chromosome XI and the SynVI.L Starter Pack

As with all the chromosomes tackled by the Sc2.0 project, the design of the synthetic chromosome XI was done using Biostudio software (unpublished platform,[127]) and was overseen by Prof Jef Boeke and Dr Leslie Mitchell (then at Johns Hopkins University). Dr Mitchell oversaw the design procedure for dividing of mega-chunks into their smaller chunks, and for placing all auxotrophic selective markers into the far-right chunk of each mega-chunk. She also prepared a protocol "starter pack" which was provided to all new collaborators in order for them to train themselves on the construction approach. Dr Mitchell provided the core strains for the project to all involved groups, *i.e.* the original haploid strains BY4741 and BY4742, marker-swap plasmids, the Cre recombinase[128] enzyme on a common yeast replicative plasmid, etc. (all are listed in Chapter 2).

The encoding of all the unique restriction enzyme sites necessary for chunk assembly was performed computationally by Biostudio, using an algorithm to determine sites that were unique to specific mega-chunks and spaced appropriately according to chunk and mega-chunks sizes. The short synonymous codon changes regions that serve as a tool to distinguish between the genes of the wild type and synthetic genomes were designed to be 15 to 20 bases in length, and called PCRtags. These PCRtags were automatically designed by Biostudio and checked by Dr Mitchell that either recognise and amplify these or their wild type equivalent was automatically generated and provided for chromosome SynXI. The layout of the first 90 kb of the new designed synthetic yeast chromosome XI is shown in Figure 3.3b.

The Sc2.0 starter pack provided by Dr Mitchell held the components needed for the assembly and transformation of the first three mega-chunks of SynVI.L, allowing research teams to become familiar with the necessary protocols by following a previously tested case. To test mega-chunk assembly in out lab, I first used this starter pack to insert the first mega-chunk of SynVI.L into the wild type chromosome VI of BY4741 yeast. Following the protocols given in Chapter 2, I first mini-prepped all plasmid DNA containing "A" mini-chunks and chunks from bacterial stocks provided in the starter kit. I then used restriction digestions and agarose gel electrophoresis and gel purification to obtain cut linear fragments for each of these chunks or mini-chunks. Using standard ligation, I ligated mini-chunks A1.1 and A1.2 from the starter pack to make chunk A1 which was checked for correct size by agarose gel electrophoresis (Figure 3.4). I then further ligated A1, A2, A3 and A4 chunks together to form mega-chunk A, immediately transformed these into the yeast strain YFL054KC (BY4741 with KanMX integrated into the start of the left arm of

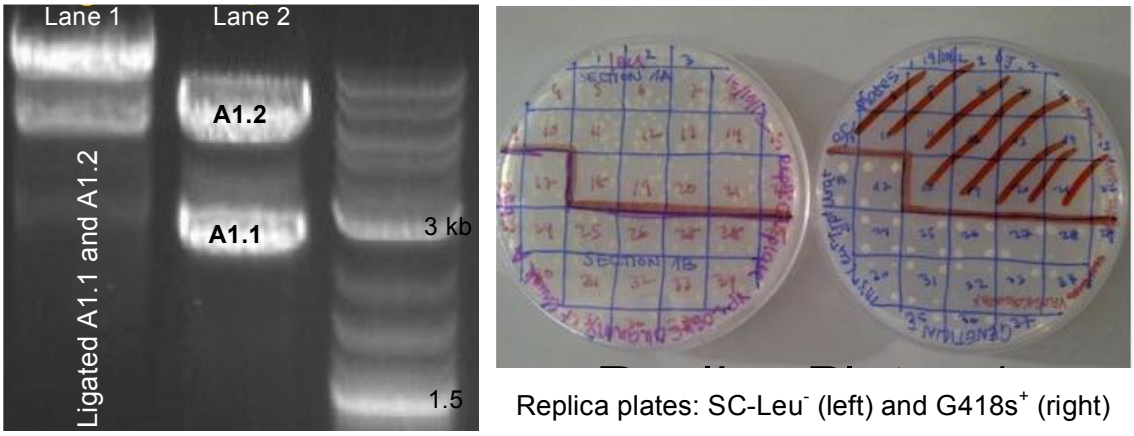


Figure 3.4 Assembly and Verification of SynVI.A. The gel electrophoresis image shows the analytical digestion of the two mini-chunks of chunk A1 digested (lane 2) and ligated (lane 1) run on a 1% agarose gel. NEB 2-log ladder is used for reference. The plates image depict the successful transformations following the integration of the mega-chunk A into the YFL054C strain. All colonies were subject to replica plating, in search of colonies that had lost the ability to grow on G418s media, whilst retaining the ability to grow on SC-Leu media. Phenotypically correct colonies were then PCRtag analysed (data not shown).

chromosome VI) using standard yeast transformation. Following transformation and selection of yeast on SC-Leu plates, colonies that grew were replica-plated on plates that select for the original marker (G418s plates) and for the new integrated marker (SC-Leu plates). This was done to distinguish colonies that had lost the KanMX marker (no longer grew on G418s) and have not gained the leucine marker (can grow on SC-Leu) ideally due to complete mega-chunk integration. During these experiments my work revealed an error in the starter pack, wherein the auxotrophic markers for mega-chunks were listed incorrectly, instructing for the selection of uracil selection following the integration of mega-chunk A. As such, during the initial transformation attempts there were zero viable transformants on any plates due to the use of the wrong auxotrophy plates. I streaked already integrated and provided strains out onto both SC-Leu and SC-Ura plates and they exhibited auxotrophic growth opposite to that described in the starter pack, as hypothesised. As we were the first international team to utilise the starter pack, the error was not carried forward to other groups. Figure 3.4 shows a summary of the results obtained during the test process described above, including the digesting, purifying and ligating of mini-chunks A1.1 and A1.2, the transformation plates and replica plates thereafter. Following protocols described in Chapter 2, I isolated and PCR-Tag validated two colonies that exhibited the correct selectable marker phenotype (not shown). This initial work showed that I could perform mega-chunk assembly by the standard protocol and successfully construct a partially-synthetic yeast chromosome.

3.3.2 Preparation and Assembly of Mega-chunks A and B for SynXI

To enable all members of the international consortium to understand and follow any correspondence between teams, a common nomenclature labelling system is followed for all chunks, mega-chunks, and strains [129]. For example, the full sequence, notes and additional information for chunk A2 (of chromosome XI) is kept in a file under the name "Chr11_34_3_A2". Figure 3.5 explains the meaning and significance of all the information provided in the name. Mega-chunks are written left to right along the chromosome and are labelled in alphabetical order (A, B, C…O, P, Q, etc."). Chunks are broken down according to numerical order within each mega-chunk; for instance mega-chunk B is comprised of 5 chunks (from left to right) B1, B2, B3, B4 and B5. Furthermore, synthetic yeast strains that have mega-chunks successfully integrated into them (and have been verified) are referred to as "Syn[chromosome number].[mega-chunks integrated]". So for example, synthetic chromosome III with mega-chunks A, B, C, D and E integrated is named as "SynIII.A-E". The last important note concerning nomenclature refers to a difference in the numbering systems. Prior to transformation of synthetic DNA into yeast, the plasmids and bacterial strains harbouring synthetic DNA are referred to with names containing Arabic numerals (Chr11_34_3_A1) whereas yeast strains transformed with synthetic DNA have names containing Roman numerals (SynXI.A).

The first mega-chunk (A) for our project – synthetic chromosome XI (SynXI) -

Figure 3.5 Nomenclature of the Sc2.0 chunks. Nomenclature is used to avoid confusion between teams and for logging of intermediate strains, changes, and other developments

is the smallest mega-chunk, comprised of only two chunks, A1 (5 kb) and A2 (11 kb). For assembly of this mega-chunk, the DNA for A1 and A2 chunks first needed to be made, digested and ligated together. The DNA chunks were synthesised by GenScript Inc. who were sent the chunk sequence by email, based on the sequenced designs produced by Biostudio. Notably chunk A1 contained a difficult-to-synthesise region that is repetitive and GC-rich. This region is known as the telomere seed sequence and when added to a chromosome inside yeast it promotes the growth of a stable yeast telomere at the chromosome end.

GenScript synthesised the A1 and A2 chunk DNA as inserts in common bacterial plasmid vectors, and delivered them as bacterial agar stabs rather than

purified plasmids as they struggled to produce them at high DNA yields. Due to the telomere seed sequence in A1 and other repetitive sequence regions in A2, both chunks were unstable as vector inserts and so were maintained by GenScript in specialised *E. coli* strains that are used for the maintenance of vectors at low copies. These usually help prevent any deleterious expression in *E. coli* or recombining-out of sequences. Upon receipt, I attempted to obtain plasmid DNA from these strains by growing them at 30°C (rather than the usual temperature of 37°C). Early drawbacks in obtaining the A1 and A2 chunks from GenScript included us needing to get chunk A2 resynthesised due to the insert being lost by recombination out from the host vector in the bacteria they sent us. Also a manufacturing error for chunk A1 led to the exclusion of the essential restriction site to remove it from the bacterial DNA plasmid. A1 chunk DNA was therefore reordered according to the designed sequence to ensure it contained the restriction site, and the assembly of SynXI.A then continued.

Following successful growth of plasmid containing bacteria from GenScript, a Qiagen Miniprep Kit was used for plasmid purification (see Chapter 2). The standard protocol was modified so that: i) after lysis buffer (P2) was added to the cells, a five minute incubation step (at room temperature) was included in order to fully lyse the large numbers of cells; (ii) elution water was preheated to 70 °C and 100 µl was used as the elution volume. These changes were made in order to maximise DNA purity and concentration, as many of the downstream reactions (such as the restriction digestions, gel extractions and DNA concentrations) all lead to loss of some DNA.

Following plasmid purification, chunk DNA was separated from the vector backbone DNA by restriction digestions with the appropriate enzymes. To ensure this had worked and to guarantee purity, all chunks were run on a 0.6 – 0.8% agarose gel, and the bands of interest were consequently gel extracted as described in Chapter 2. Table 3.1 shows chunks and the associated restriction enzymes that digest all A and B chunks from their plasmid backbones, as well as the selectable marker of each mega-chunk. A novel unpublished technique utilising PEG that combines DNA concentration of fragments with subsequent ligation of chunks was applied to assemble the mega-chunks, prior to their transformation into yeast (Chapter 2, section 2.3.10). Figure 3.6 shows the band sizes of pre-ligation chunks A1 and A2 using gel electrophoresis. Similar gels not shown here show the isolation of other chunks used to make mega-chunk B. For downstream work, getting good yields of chunks from digestion and gel purification is important, as is accurately quantifying their yield using a Nanodrop. During mega-chunk ligation, the last chunk needs to be added in the smallest quantity, as it contains the auxotrophic marker which is the limiting factor of transformations. By limiting the concentration of this

Table 3.1 List of chunks of mega-chunks A and B, and the associated restriction enzyme sites embedded within the genome. Also shown are any auxotrophic markers that are present in the final chunk.

Chunk	Chunk Sizes (bp)	Restriction Enzyme Site	Auxotrophic Marker
Chr11_34_3_A1	4,811	SfiI	
Chr11_34_3_A2	10,718	SfiI, BaeI	LEU2
Chr11_34_3_B1	7,768	NmeAIII, SfiI	
Chr11_34_3_B2	8,095	SfiI, BanI	
Chr11_34_3_B3	7,503	BanI, BsoBI	
Chr11_34_3_B4	7,572	BsoBI, Bsu361	
Chr11_34_3_B5	8,287	Bsu361, BsaXI	URA3

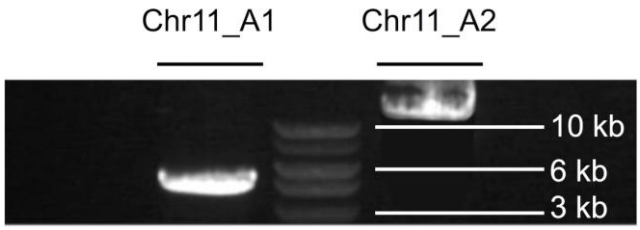


Figure 3.6 Gel Electrophoresis Analysis for Chr11_A1 and A2. Image of the extracted bands of Chr11_A1 (left band, 4,811 bp) and Chr11_A2 (right band, 10,718 bp). Bands were run on a gel once they were digested and gel extracted to verify sizes. The ladder for reference is NEB's 2-Log ladder.

chunk, we can increase the chances of colonies being correctly ligated and integrated transformants versus transformations where the only final chunk is integrated without ligating to others.

Digestion of chunk B1 from the plasmid provided by GenScript did not yield the expected band sizes, and there was a lack of the band representing the vector backbone (Figure 3.7a). Only a single band size was seen which suggested that plasmid DNA had only undergone a single cut rather than the double cut and had thus been linearised. Upon further inspection, it was determined that NmeAIII restriction enzyme had failed to cut. Further research revealed that NmeAIII is a rarely used Type IIG restriction endonuclease and requires a dimer of its recognition site (i.e. two copies of the same sequence) in the same reaction in order to cleave efficiently. As shown in Figure 3.7, I managed to resolve this interesting issue by including in the reaction a pair of annealed complementary primers that encode within them an NmeAIII recognition sequence. Addition of this short double-stranded DNA to the digestion reaction acted as the necessary second site required to allow NmeAIII to act on our plasmid DNA. Figure 3.7 shows the incorrect and correct

digestion of chunk B1 with and without the NmeAIII oligonucleotides (primers) provided. This resolved method of enabling NmeAIII cleavage was used for all future digestions with this enzyme in the construction of SynXI by the Imperial College London team.

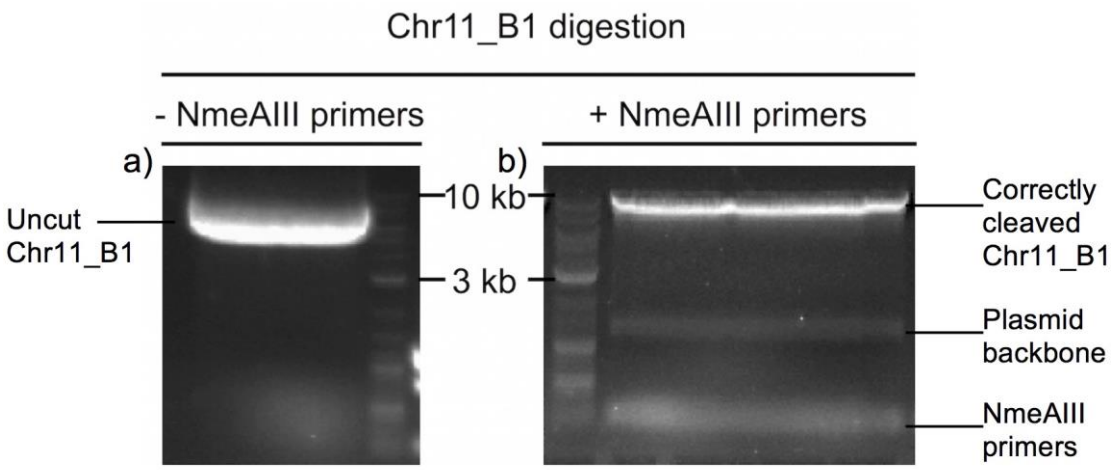


Figure 3.7 Gel Electrophoresis Analysis of NmeAIII Digestion. Image of the digestion of Chr11 chunk B1 with and without NmeAIII primers. Image (a) shows the incorrectly digested fragment, due to the lack of NmeAIII primers the enzyme, which requires two copies of its recognition site, can't cleave efficiently. Image (b) shows the same reaction supplemented with NmeAIII annealed primers, and the three bands (from top to bottom) show chunk B1 (7,768 bp), the plasmid backbone, and the annealed primers. The reference ladder is NEB's 2-Log ladder.

3.3.3 Re-iterative Homologous Recombination Using a Marker Swapping Technique

The marker swapping technique described above was used for the integration of mega-chunks A and B into BY4741 yeast. This requires transformation of yeast immediately after mega-chunk ligation followed by inspection of yeast colony growth first on the usual selectable marker plate and then by replica-plating. Conveniently, these latter steps also allow phenotypic fitness screening at the same time. In the event that any colony abnormalities are prevalent (e.g. small colonies) these signal that integration has likely led to a fitness defect and this is usually the first evidence that the integrated mega-chunk is problematic for the cell.

Because the wild type strain BY4741, which the Sc2.0 project is based around, lacks the presence of any selectable marker that the first mega-chunk (A) can replace, I began mega-chunk A integration by first ordering strain YKL220C from EuroScarf (European *Saccharomyces cerevisiae* Archives for Functional analysis; K. D. Entian, Frankfurt). This strain is genetically indistinguishable from BY4741, apart from the inclusion of a KanMX4 gene cassette in the sub-telomeric region of

Chromosome XI, which confers resistance to Geneticin (G418s). Correct integration of mega-chunk A into this strain can be screened by the marker-swapping method, screening for colonies that lose the ability to grow on Geneticin. I previously demonstrated this approach of looking for G418s-sensitive colonies for SynVI.A integration as shown in Figure 3.4

Following the same method as used for the starter-kit described above, fresh mega-chunk A ligation was transformed into YKL220C yeast, and cells were recovered on SC-Leu glucose agar plates. Successful colony growth was observed after three days of incubation at 30°C. One hundred colonies were selected at random from the transformation plate and replica-plated onto (in this order) onto: (i) G418s-containing yeast agar plates (at a selective working concentration of 200 µg/mL); and (ii) SC-Leu glucose agar plates. Replicating on the antibiotic plate first and then on the SC-Leu plate was done so that there was no chance of media carry-over. Two days later, from the replica-plated colonies, ten exhibited the desired phenotype of failing to grow on the G418s, while growing on SC-Leu. All of these ten colonies were subject to PCRtag analysis. Figure 3.8 illustrates the method of replica-plating that I devised for this work.

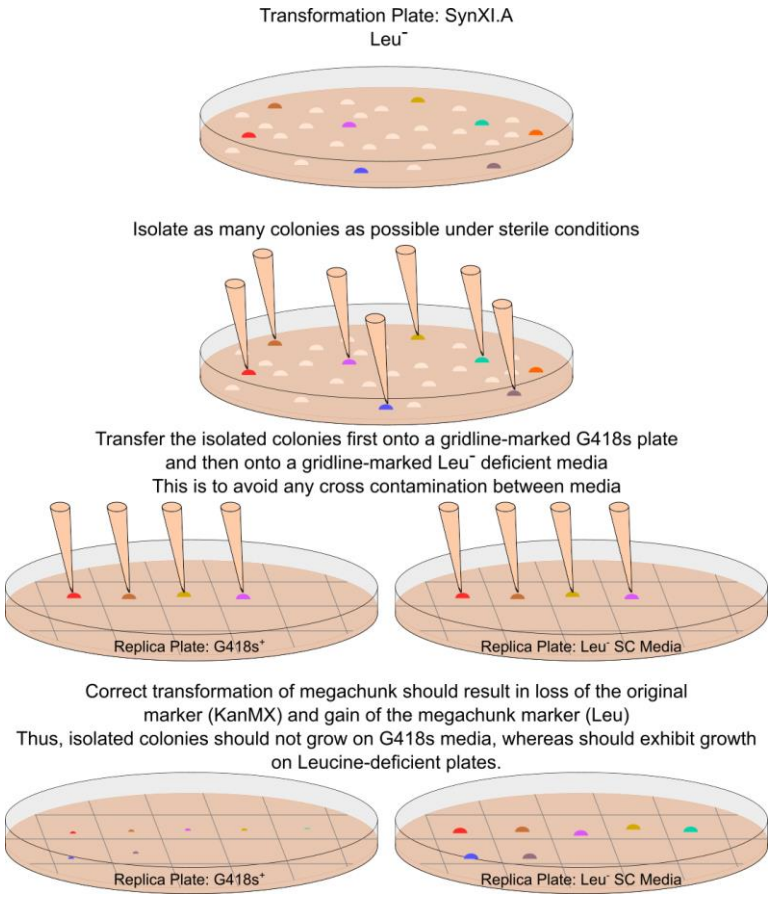


Figure 3.8 Replica Plate Method Schematic. Illustration showing the method used to replica plate transformants. It is important to note that all colonies were first plated onto G418s media, and then SC-Leu media to prevent from any carry over between plates. Plates were incubated for 1-2 days to allow for visible colonies to develop.

3.3.4 PCRtag Analysis to Verify SynXI.A

Replica plating established the correct phenotype, however it does not confirm the correct genotype. Therefore from the ten colonies exhibiting the correct phenotype for mega-chunk A integrations, I extracted genomic DNA following the protocol described in Chapter 2, and then moved on to using PCRtag analysis to confirm that the synthetic DNA had replaced the wild type DNA (see section 2.3.9). For each colony there are three possible outcomes following PCRtag analysis: (i) all wild type (WT) PCRtags are amplified and therefore none of the synthetic (SYN) DNA integrated (highly unlikely to occur); (ii) the amplification of both synthetic and wild type PCRtags occurs, suggesting that the artificial DNA has been incorporated into the yeast genome, but not into the designated position; and (iii) amplification of only SYN PCRtag primers and no WT equivalents – this is seen when all the synthetic DNA has integrated correctly and replaces the wild type DNA as intended. From the genomic DNA isolated from the ten colonies that replica plated correctly, one sample tested positive for the amplification of all mega-chunk A synthetic PCRtags and none of the wild types. The PCRtag products were separated and visualised using gel electrophoresis and UV Bio-image software (Figure 3.9). The assay used six primer pairs for both the SYN and WT PCRtags, split into three per chunk. The correct appearance of six SYN PCRtags and no WT PCRtags confirmed the successful construction of SynXI.A

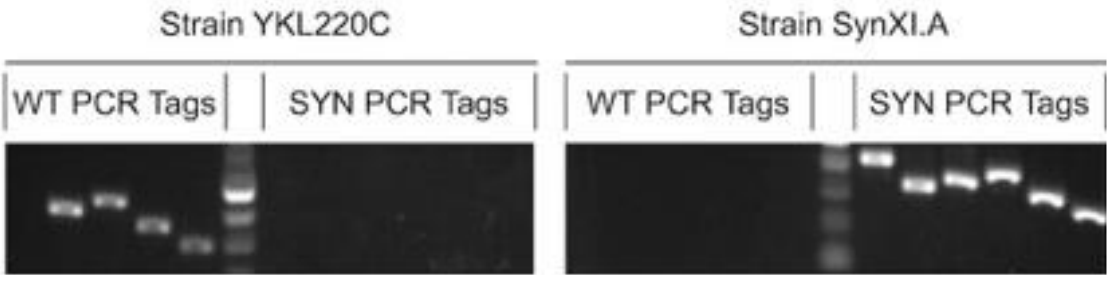


Figure 3.9 PCRtag Verification of SynXI.A. The WT and SYN primers were applied in a simple PCR amplification reaction to confirm the presence of the synthetic DNA (as shown in the right gel image, under Strain SynXI.A, SYN PCRtags) and the simultaneous loss of the wild type DNA (Strain SynXI.A, WT PCRtags). Strain YKL220C (left image) was also subject to PCR amplification to serve as a control strain, thus showing the amplification of the WT primers, but lack of for the SYN primers.

3.3.5 Integration of Mega-Chunk B and Lack of Transformants

With mega-chunk A having been integrated and confirmed by phenotype and genotype screening, efforts progressed to the assembly and integration of mega-chunk B. In contrast to the small mega-chunk A, this second mega-chunk consisted of five chunks and was nearly 40 kb in size. The gel-purified chunks B1, B2, B3, B4

and B5 were obtained at high yield as described above, combined together as before in ligation reaction and transformed into the SynXI.A yeast constructed in the last section. The transformed yeast was plated onto SC-Ura glucose agar plates, however, unlike the transformation of A which yielded hundreds of colonies, the inclusion of mega-chunk B yielded no viable colonies (data not shown). The ligation and transformation was repeated in triplicate to ensure reproducibility, but always produced no viable colonies, suggesting instead that mega-chunk integration gave a lethal fitness defect.

Upon closer scrutiny of the *in silico* DNA sequence, it was found that the design on synthetic chunk B5 (which also encodes the uracil auxotrophic marker), intentionally deletes a tRNA gene from the chromosome (illustrated in Figure 3.10). Although deletion of tRNAs by other teams working on synthetic chromosomes had not been shown to lead to any major fitness decreases, in all previous cases those tRNAs were present in multiple copies throughout the genome. In a first for the Sc2.0 project, the tRNA being deleted within chunk B5 (tRNA TRT2) was an essential tRNA gene, as it was the only copy found throughout the genome. Removal of this gene as per Sc2.0 criteria (all tRNA genes, which act as genome destabilising elements[130], are to be removed from chromosomes and translocated onto a neo-chromosome) was resulting in cell death as the tRNA becomes absent from the cell. This provided an explanation for the lack of viable transformants for mega-chunk B.

In order to continue with the assembly of the synthetic chromosome, Dr Ben Blount worked to reintroduce the unique TRT2 tRNA back into the genome. Due to the highly repetitive and unstable sequence structure of tRNAs, a special DNA

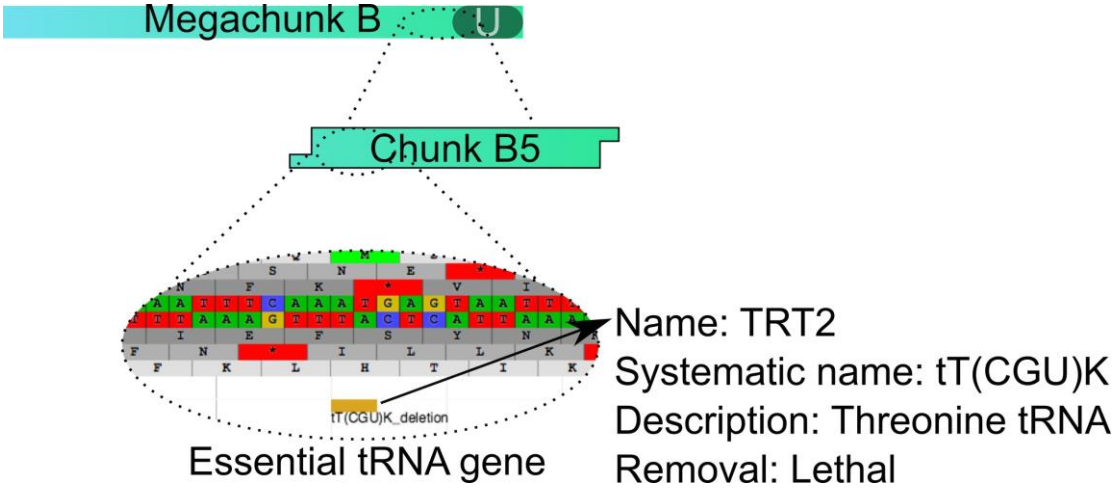


Figure 3.10 tRNA of Chunk B5 Schematic. Illustration showing the position of the tRNA TRT2 found at the beginning of chunk B5. In the synthetic design strategy, all tRNAs are removed from the synthetic chromosomes and translocated to a neo-chromosome. However, without the inclusion of the tRNA gene, the cell cannot proliferate and thus 100% morbidity is seen.

Polymerase (Kapa HiFi), specialising in amplifying structured regions, was used to amplify and clone the sequence. By cloning the TRT2 tRNA gene into a histidine-selectable integrating plasmid that inserted at the His locus, Dr Blount was able to prepare SynXI.A for transformation of mega-chunk B without subsequent lethality.

At this point, Dr Blount took over from me for SynXI construction, and successfully transformed mega-chunk B into the modified SynXI.A and verified it by PCRtag analysis. This time the removal of the TRT2 tRNA upon B5 chunk integration was not lethal, as the gene was provided elsewhere on the SynXI.A genome. Following the verification of SynXI.B, the reiterative recombination technique of subsequent mega-chunks was resumed.

3.3.5 Alternative Methods for DNA Assembly for Sc2.0

During the time the synthetic DNA for chromosome XI was being synthesised by GenScript, I tested alternative DNA assembly techniques using the then available synthetic DNA for chromosome VI.L, provided with the starter pack. I tested whether the ligation and DNA concentration step could be optimised in order to minimise steps involved, by using alternative DNA assembly approaches. The standard method uses classic restriction digestion and ligation of short sticky-end overhangs, but elsewhere in synthetic biology, homology-mediated DNA assembly methods such as Gibson Assembly[131, 132] are now widely used. I hypothesised that by lengthening the sequence at the ends of neighbouring chunks so that they have regions that overlap with enough homology, mega-chunk assembly could be carried out using homology-mediated DNA assembly. Chunk A, comprised of mini-chunks A1.1 and A1.2 were used to test this hypothesis using Gibson Assembly as the *in vitro* DNA assembly method. To generate overlap between chunks A1.1 and A1.2, I used two approaches that effectively provide a third double-stranded DNA piece that overlaps both A1.1 and A1.2 in any assembly reaction. The first (cheap) approach used two complementary 60mer oligonucleotides annealed together to make a 60 bp DNA fragment that consists of the last 30 bases of A1.1 and the first 30 bases of A1.2. The second approach (more expensive) used a purchased and 500 bp G-block (IDT, USA) DNA fragment that consists of the last 250 bases of A1.1 and the first 250 bases of A1.2. Both methods are in effect a three-piece DNA assembly reaction to bring A1.1 and A1.2 together, with one using 30 bases of overlap and the other having 250 bases of overlap.

Three reactions were set up, a standard ligation reaction and two Gibson reactions comprised of the mini-chunks with the G-block, and the mini-chunks with the oligos. Following the standard incubation times used for both the ligation and

Gibson reactions, any products formed were run on an agarose gel, for visualisation of band size and intensity. From Figure 3.11 the results showed that the ligation reaction remained as the most efficient, producing the correctly sized band at the highest intensity, when compared to the adjacent lanes representing the Gibson reactions with the G-block and oligos. In fact, the Gibson G-block and oligos assembly method did not yield the 10 kb band expected when the two mini-chunks are ligated together. Based on these results, the standard ligation method was continued to be utilised for the assembly of chunks.

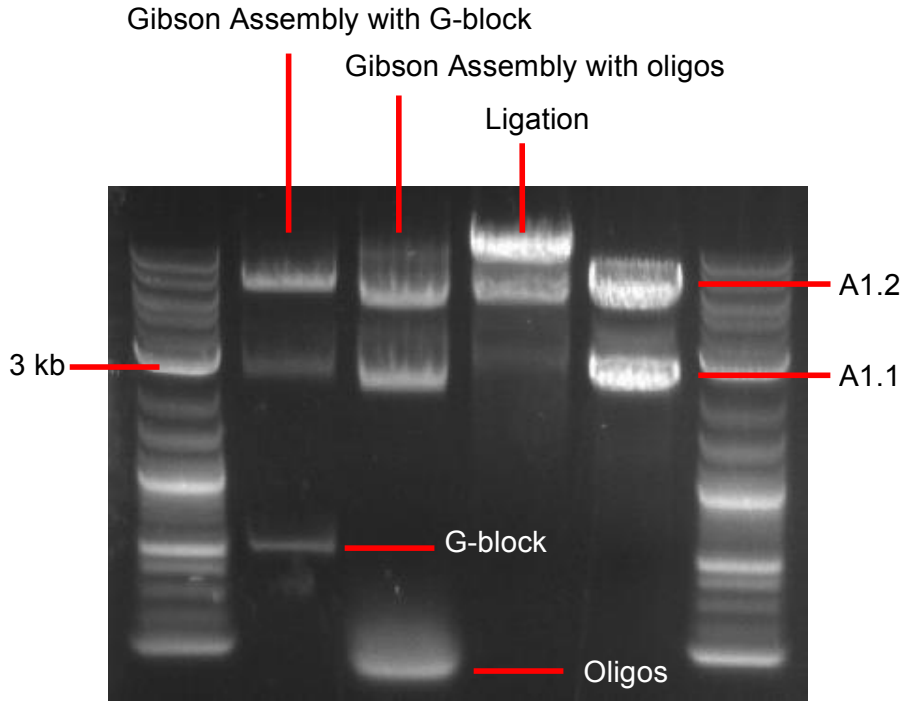


Figure 3.11 Using Gibson Assembly for SynVI.A1. Agarose gel analysis of alternative DNA assembly methods of SynVI mini-chunks A1.1 and A1.2 A Gibson assembly was set up with a 500 bp G-block (lane 1) and with 60 bp oligos (lane 2). Gibson assembly with the G-block failed to produce the 10 kb band of the expected chunk, and Gibson assembly with the oligos produced a very faint band. In parallel, a standard ligation reaction was also carried out, producing the brightest intensity of the desired 10 kb band (lane 3). Mini-chunks A1.1 and A1.2 are shown as size controls in Lane 4. A 2-log NEB ladder was used for reference.

By the time the synthetic DNA for chromosome XI arrived, and given the low efficiencies following transformation, replica-plating and PCRtag screening, as described above, I also tested whether the same method (using overlapping homology-based DNA assembly), could be repeated for SynXI.A construction, but rather *in vitro*, I hypothesised that it could be done *in yeasto*.

Mega-chunk A was used as the model system, as it comprised of only two chunks, and is the smallest synthetic fragment (~15 kb) of synthetic chromosome XI. I assessed whether I could improve assembly steps by adding overlapping regions of

homology between chunk A1 and A2 and using homology-mediated DNA assembly rather than the classic digestion and ligation. It is known that *S. cerevisiae's* can itself recombine regions of homology that are 38 bp or more[133], thus we presumed that the oligo-based approach would not work for DNA assembly by *in vivo* recombination in yeast but the G-block approach would. Both approaches may work for Gibson assembly. Whether either were better and more time- and cost-efficient compared to the classic method was the key question.

Firstly, to test the method for how well it enables DNA assembly by *in vivo* recombination in yeast, three equivalent transformations into strain YKL220C were set up, made up of equimolar amounts of the DNA components included. They were: a) the standard PEG precipitation and ligation reaction, comprised of digested, purified and ligated chunks A1 and A2; b) a non-ligated mix of purified 500 bp G-block, digested and purified chunks A1 and A2; and c) the non-ligated mix of annealed 60mer oligos, and digested and purified chunks A1 and A2 (Figure 3.12). Data was collected by counting the number of colonies that grew following transformation of the three reaction types and growth on SC-Leu glucose agar plates. Despite efforts to improve transformation efficiency using overlapping fragments for *in vivo* homologous recombination, the method yielding the highest number of colonies remained the original method of concentrating the digested DNA and ligating it prior to transformation in yeast, as shown in Figure 3.12d.

It is interesting to see the large difference in the number of colonies of the three types of reactions, despite the chunk with the auxotrophic marker having been added in equimolar amounts into each of the reactions. The pre-ligated chunks resulted in a total of 69 surviving colonies, but the transformations using overlapping DNA fragments totalled only nine colonies. At this stage, we would have expected to see similar numbers of surviving colonies, based on the presence of the marker encoded in chunk A2, rather than the large difference between the numbers of transformants from these methods. The colonies were replica plated onto counter-selective G418s media, to screen for integrants that had have lost the KanMX marker as a result of the transformation of a successfully ligated full mega-chunk. This confirmed that none of the overlap DNA based transformations showed correct mega-chunk assembly and integration, whereas two colonies constructed the standard way did not grow on the Geneticin media.

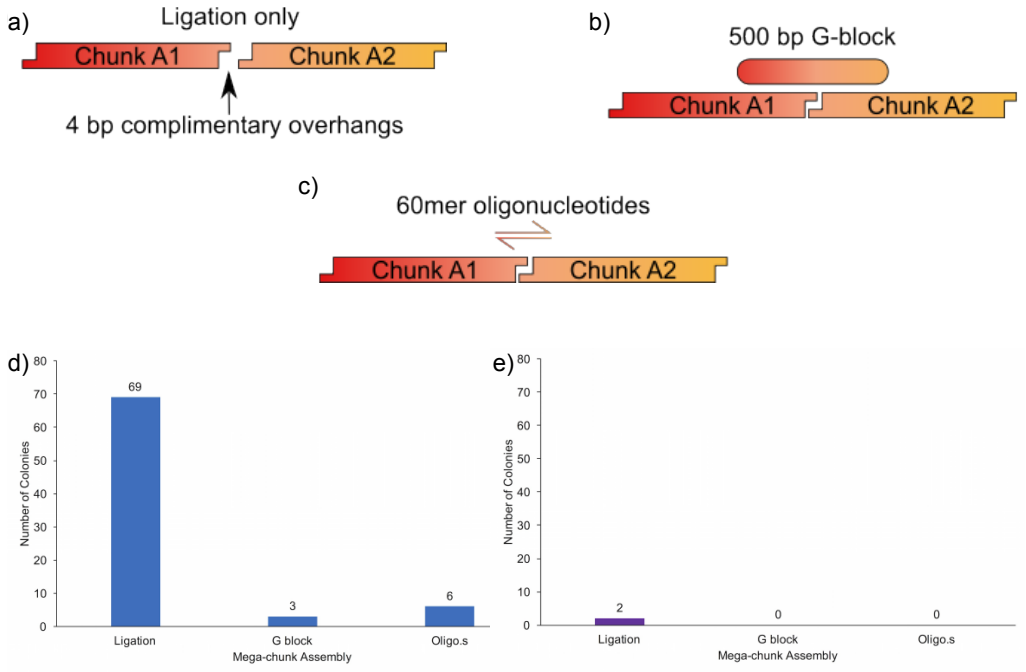


Figure 3.12 *In vivo* Overlapping DNA Methods Results. Synthetic chunks are provided in various modes into the yeast cell. Schematic representation of (a) the ligation of chunks prior to transformation as per Sc2.0 protocol; (b) Chunks A1 and A2 with no ligation, are instead provided with a 500 bp equally overlapping G-block; (c) The chunks are not ligated as in part (3.4b) but are provided with smaller overlapping DNA fragments in the form of 60mer annealed oligonucleotides. Methods (3.4b) and (3.4c) rely on exploiting yeasts' natural tendency to recombine overlapping fragments of DNA. (d) Graph showing results of a transformation following the differing modes of DNA preparation as outlined in the schematic. All colonies were replica plated onto G418s plates and the results are shown in (e). n=1

3.4 Discussion and Conclusions

Mega-chunk Integration Requires Screening High Numbers of Colonies

From the outset, mega-chunk A gave rise to a number of issues experienced at almost every stage from manufacture, to integration and to screening. The manufacturing and cloning of the chunks of A proved to be problematic for GenScript, requiring longer processing periods to meet the *in silico* design and DNA quality assurance testing compared to usual manufacturing time frames. The DNA was provided in low-copy recombination minus (recA1) strains, as is often done for the propagation of problematic DNA constructs. Chunk A1 was the less stable of the two chunks likely due to its GC-rich repetitive sequence.

Although only comprised of two chunks and being the smallest mega-chunk of SynXI mega-chunk A yielded low numbers of phenotypically and genotypically correct colonies, and required time-consuming rounds of screening. All transformants were replica-plated, and all colonies exhibiting lack of growth on counter selective media were screened through PCRtag amplification. Out of ten colonies subject to the PCR screening, a single sample was isolated as negative for WT primers, whilst retaining a positive result for the SYN primers. Although efficiencies are low for isolating a correct synthetic genotype, mega-chunk A required the least number of synthetic PCRtag reactions for its verification. We suspect that efficiencies were affected by the need to integrate A into the telomeric region of the left arm of chromosome XI. Telomeric regions tend to be highly repetitive, nucleoprotein complexes regulating chromosome length, gene silencing and genome stability[134, 135], as well as preventing chromosome fusion and exonucleolytic degradation[136]. They are therefore a repetitive complex and nucleoprotein-rich target that is likely harder for DNA to integrate accurately into compared to other chromosome regions.

The PCRtag analysis for the 15 kb synthetic fragment of A required a total of 24 forward and reverse primers (12 wild type and 12 synthetic) to make six pairs for the synthetic PCR reaction, and six for the wild type reaction. Thus screening the genomic DNA from ten colonies with each primer pair resulted in a total of 120 PCR reactions (not including the BY4741 as a control sample). While this was time-consuming, the other mega-chunks of SynXI are much larger, comprised on average, of 30 to 50 kb of artificial mega-chunks. Mega-chunk B for example is 37,945 bp in size, and has a total of 28 PCRtag products. Thus a single sample isolated for screening would require 112 primers, and a total of 56 PCR reactions (28 WT tags, and 28 SYN tags) to fully verify. Ten samples of potentially correct colonies would require 560 PCR reactions, being difficult to arrange in a typical thermocycler. The cost aspect must be considered as well, as the reagents for so many PCRs and gel

electrophoresis gels are a significant expense. Optimisation steps to minimise the number of reactions while simultaneously isolating false positives were devised early on, such as dividing screening into two stages, early and late stages. Working with Dr Blount, I devised that the early stage screening for mega-chunk B would use a single primer pair from each chunk (i.e. B1, B2…B5), thus reducing initial PCR reactions to a total of ten reactions per colony screened. Isolates showing amplification of only the synthetic primers were then subject to late stage screening, where all the PCRtags were included (56 reactions per sample).

What we observed during PCRtag screening was that a large portion of the amplicons resulted in presence of both the WT tags and the SYN tags, confirming that the mega-chunk A had been integrated, albeit not in the correct locus (and despite the replica-plating step). These findings may be explained by ectopic recombinations into other telomeric and sub-telomeric regions. For example, both chromosomes XII and IV share long TG repeat sequences similar to those in chromosome XI and are therefore recombination hotspots for mega-chunk A to elsewhere. It should be noted that following the integration of A (and the introduction of the TRT2 tRNA gene), that integration of the consequent mega-chunks proceeded at a much faster rate, although not necessarily at higher efficiencies, which may also reaffirm that the A mega-chunk integration was a particularly difficult case.

Lastly, it is important to note that certain re-designed features of mega-chunk A at the telomere region changes the behaviour of the chromosome, such as the elongation of the TG repeats, to make telomere seed sequences (TeSSs)[137]. By introducing these TESSs, the partially-synthetic chromosome theoretically gains a new streamlined telomere region, which has decreased chances of chromosome fusion. The risk of chromosome fusion will be particularly high during the final stages of completion of an entirely synthetic *Saccharomyces cerevisiae* genome, when 16 entirely synthetic chromosomes (including a neo-chromosome holding all the tRNA genes) need to be combined by mating into one cell to produce the *de novo* organism, Sc2.0. Although current engineered strains with partially- or fully-synthetic chromosomes (e.g. SynIII) have proved successful in their functioning and stability in their individual BY4741 backgrounds, it is currently unknown whether the finalised Sc2.0 strain (with 16 synthetic chromosomes) will be as genomically stable and functional as BY4741. There is some early unpublished evidence that shows synthetic chromosomes can exist in parallel with one another, such as a haploid strain containing two synthetic chromosomes SynV and SynX (thus named SynV/X), which have been produced by work of Tianjin University and New York University (unpublished data).

Integration of Mega-Chunk B led to the Exclusion of an Essential tRNA

Integration of mega-chunk B proved to be impossible until it was recognised that chunk B5 of the second mega-chunk recombined into the wild type genome at the locus of a unique tRNA, leading to the loss of that tRNA gene and no viable cell growth. This tRNA TRT2, encoding the only tRNA that brings threonine to the CGU codon (tRNA-Thr)[126], is essential and its deletion results in lethality[138]. This scenario of a unique tRNA gene needing to be removed for a synthetic chromosome, had not been encountered before in the Sc2.0 project as most tRNAs are present in multiple copies across the genome. Thus our efforts to resolve the issue and the workaround we achieved proved valuable to the rest of the Sc2.0 consortium and in particular for the plans for building the tRNA-encoding neo-chromosome.

Due to the RNA secondary structure genes, the isolation and PCR amplification of the TRT2 gene from the wild type chromosome was a challenge, and required a specialised polymerase (Kapa HIFI) and customised primers with high-specificity annealing. Dr Ben Blount, in the Ellis Lab, eventually managed to amplify the TRT2 tRNA gene and integrated it into the HIS3 locus of SynXI.A. Subsequent mega-chunk B transformations then became successful, and from thereon the continuation of constructing synthetic chromosome XI resumed unhindered with the work passing on to Dr Blount to lead.

One notable drawback from the reintegration of the tRNA-Thr into the HIS3 locus is the loss of a potentially useful auxotrophic marker. Histidine auxotrophic selection is usually used in the Sc2.0 project for the maintenance of the pCre_EBD plasmid which bears the recombinase that allows SCRaMbLE. The base BY4741 strain for SynXI construction is auxotrophic for histidine, leucine and uracil; uracil and leucine markers are used in the iterative homologous recombination methodology, while the pCre_EBD plasmid is normally selected for by histidine, as it is on a pRS413 backbone (YCp His). To deal with this situation for the following Chapters, the same Cre_EBD insert was also placed into two other vector backbones: cloning into pRS415 (YCp Leu) and pRS416 (YCp Ura). This means that the plasmid used for the expression of Cre_EBD can be chosen based on which auxotrophic marker is free. It is by chance that the partially-synthetic strains used in all subsequent work in this report all contain the same mega-chunk marker (Leu), and thus the pCre_EBD plasmid used in the rest of this study is always expressed on the pRS416 Uracil based vector.

The Standard Methodology for Building Sc2.0 was kept for SynXI

Mega-chunk construction from purchased chunks proved to be a cumbersome, repetitive process, although reasonably efficient. The aim of pursuing alternative DNA assembly methods was to establish whether construction could be accelerated or simplified and to increase efficiencies of mega-chunk integration to increase the rate of synthetic chromosome XI construction. Multiple research groups have reported the use of yeast recombination machinery to assemble fragments *in yeasto*, thus decreasing the number of steps between enzymatic restrictive digestion and transformation[60],[69],[139],[82]. Gibson *et al* (2008)[140] in particular described the *in yeasto* assembly of 590,011 bp of *M. genitalium* synthetic genome from 25 large DNA fragments (17 to 38 kb) using overlaps of 80 to 360 bp of double-stranded DNA.

The preliminary data shown in Figures 3.11 and 3.12 show that the standard ligation method is robust, and although it is also process-heavy and time-consuming, in all cases, it yielded the most efficient results for both *in vitro* and *in vivo* DNA assembly. When carried out in vivo it produced the highest number of growing colonies, as well as being the only method that produced colonies showing correct growth at replica-plating. The agarose gel image following attempts at ligation via Gibson Assembly clearly showed that the standard restriction/ligation method was far superior for producing the complete 15 kb mega-chunk A DNA. Although experiments shown here were not repeated multiple times, the magnitude by which the standard method was better than the alternatives confirmed to us that was not worth pursuing alternative DNA assembly methods.

Without pursuing further experiments to explore this, it is difficult to say why the homology-directed method used for SynXI.A construction was so inferior, especially when *in yeasto* DNA assembly is so widely used. One hypothesis is that the recombination was error prone, indeed other have reported that homologous recombination errors increase with number of fragments added[139]. Another explanation is that the method ends up producing less colonies after transformation. The low numbers of colonies associated by our recombination-based assemblies are also reflective of data seen in other published studies[141].

It is possible that the correlation between ligating the chunks correctly and the greater colony number is a result of the recombination that leads to mega-chunk integration. Chunk A1, chunk A2, the G-block and the 60mer oligonucleotides all share large homology with the wild type chromosome, and thus all have 5' and 3' regions that meet the prerequisites for *in vivo* recombination on their own. Perhaps the forming of the mega-chunk prior to transformation applies more pressure on the

yeast cell machinery to integrate a single large fragment, as opposed to each cell just integrating one of the fragments.

Another consideration could be that higher concentrations of DNA are needed for multiple recombination events[142]. The efficiencies of correct mega-chunk integration for Sc2.0 are already low; thus for multiple integrations more DNA is needed. Whilst it is important to abundantly provide exogenous DNA during yeast transformation, DNA overload in cells decreases transformation efficiencies; Kawai *et al* (2010) recognise that providing the suitable amount of DNA is critical for transformation efficiencies, and they showed that they experienced much higher efficiencies when providing 1 to 3 µg of DNA as opposed to 16 µg of DNA in their reactions.

Progress of Synthetic Chromosome XI

With the functional expression of the TRT2 tRNA restored by insertion into a different genomic locus, Dr Blount successfully integrated mega-chunks B and C into chromosome XI, completing the first 90 kb of SynXI. At the time of this report being

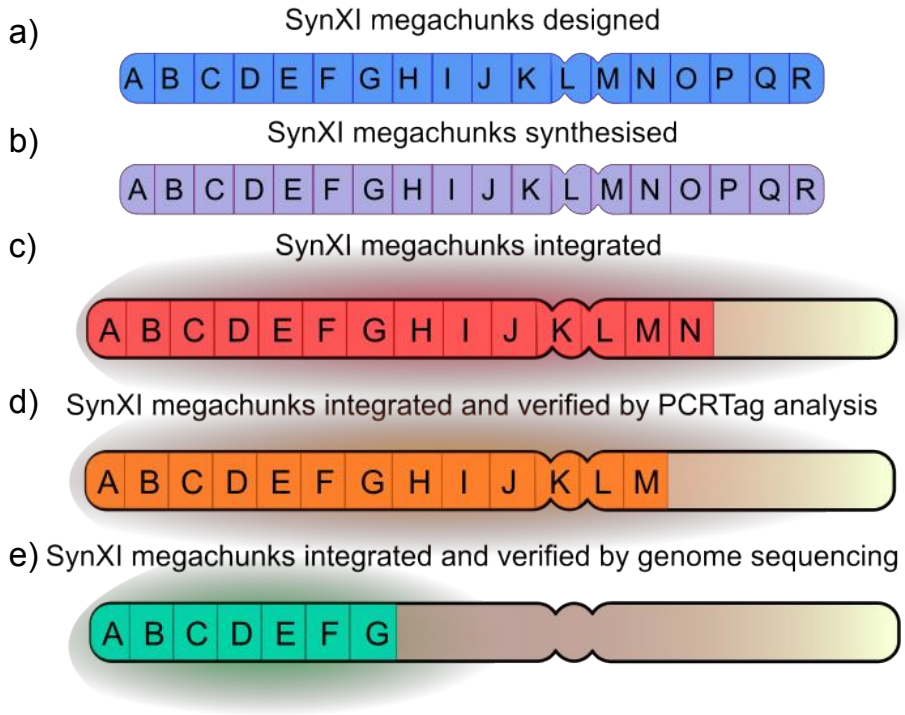


Figure 3.13 Progress of SynXI. Schematic diagram representing the progress of SynXI so far. At the time of publishing, the Ellis Lab at Imperial College London have received the completed design of the chromosome, supervised and received the full set of synthesised chunks and successfully integrated up to mega-chunk N. The PCRtag analysis includes all mega-chunks up to M, and several strains have been sent for genome sequencing to ensure no major errors are embedded within the genotype. So far construction shows a 99.9997% sequence accuracy rate. Work carried out by Dr Blount and Maureen Driessen.

written, rapid progress of SynXI had been made; the complete design was optimised and finalised (Figure 3.13a), all synthetic DNA chunks were ordered, manufactured and received (Figure 3.13b), 14 mega-chunks were transformed using the reiterative marker swapping technique up to mega-chunk N (Figure 3.13c), and then verified using the PCRtag assembly up to mega-chunk M (Figure 3.13d). Furthermore the full chromosome sequence of our yeast was verified using Illumina MiSeq genome sequencing[143] upon completion of mega-chunk G integration and creation of SynXI.A-G (figure 3.13e). This genome sequencing showed almost total accuracy in producing our chromosome to the specified DNA sequence, with only 1 bp error needing to be corrected following over 230,000 bp of synthetic DNA integration (unpublished data).

Currently Dr Blount has issues with mega-chunk M and mega-chunk O integration that are being resolved. As such, the most complete version of the chromosome that shows full fitness and is expected to be the correct sequence is SynXI.A-L, which is in the background strain BY4741 (plus TRT2 at HIS3 locus), and has an entirely synthetic left-arm of chromosome XI (~420,000 bp)

Genome sequencing the assembled synthetic genome will be a prerequisite for future SCRaMbLE experiments, acting as a parental comparison strain. It is also an obligatory validation step for determining the correct chromosome construction, and finalising endeavours to build the functional synthetic yeast. Mega-chunks A-G have been confirmed by sequencing, producing 99.9997% correct synthetic sequence alignment, of which an example sequencing alignment file is shown in Figure 3.14a. Fifteen synonymous nucleotide polymorphisms (SNPs) have been observed in over >230,000 bp of synthetic DNA. These have occurred exclusively at chunk ligation junction sites within mega-chunks, and were not detected in phenotypic screening or PCRtag analyses. These SNPs are (in order from left to right along the chromosome): i) between the border of chunk B4 and B5, a thymine instead of a cytosine, a cytosine instead of a guanine, and an adenine instead of a guanine; ii) between the C3 and C4 chunks' border, an adenine instead of a guanine, a guanine instead of a cytosine, and an adenine instead of a thymine iii) five mutations between the border of E4 and E5, a thymine replaces a guanine, a cytosine is replaced by a thymine twice, a guanine is replaced by a cytosine, and an adenine in substitute for a cytosine; iv) finally, between the F2 and F3 chunks' border, there are four nucleotide mutations, the first two and last being a thymine in place of a cytosine, and an adenine instead of a thymine. The listed SNPs are illustrated in Figure 3.14b.

A number of reasons could have given rise to these mutations, such as inconsistent and/or incomplete enzymatic digestion, insufficient ligation between the chunks, DNA degradation, or spontaneous mutations. However, based on the junction-exclusive occurrences, we can assume that the polymorphisms are due to T4 DNA Ligase infidelity[144]. In all cases the innocuous effects of these synonymous SNPs on the partially-synthetic genome means that no further action will be taken to rectify these, until the completion of the synthetic chromosome. Indeed, on close inspection all these mutations simply revert the sequence at these junctions to be back to the bases seen in the wild type sequence. The base changes introduced into the synthetic DNA sequence at these positions were only ever added to enable restriction digestion for mega-chunk assembly and so if they are absent from the synthetic chromosome then it shouldn't make a difference

Finally, out of 233,470 bp that have been assembled, integrated and sequenced, a single non-synonymous nucleotide polymorphism was observed (blue line in Figure 3.14a). This occurs between the junction of chromosome A1 and A2 (mega-chunk A). At this particular locus a cytosine replaces a thymine, causing a missense mutation from tyrosine to histidine. Despite this change, no unfavourable phenotypes were seen. The mutation lies in the 3' end of the MCH2 gene, a monocarboxylate transporter homologue, whose mutant is not deficient in monocarboxylate secretion and has no recorded fitness defects[145]. Despite no problems with this mutation, unlike the other SNPs, it will be repaired to the correct sequence at the time of completion of synthetic chromosome XI.

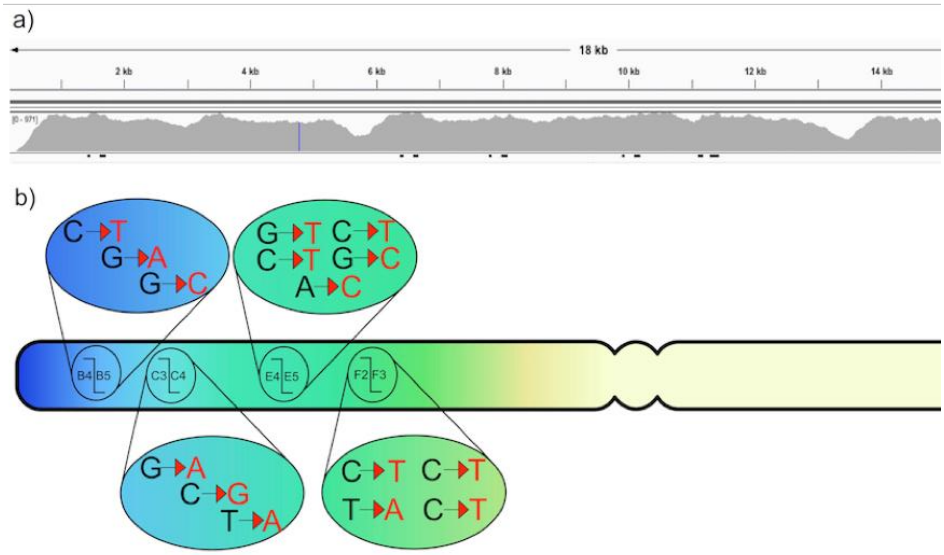


Figure 3.14 Genome Sequencing and Mutations Schematic. Screenshot of the genome sequencing alignment at SynXI mega-chunk A. The SynXI.A-G genomic DNA was sent off for Illumina sequencing to verify it had been correctly assembled (a). The grey histogram peaks show the strength of the sequencing reads (when compared to the reference sequence), thus confirming correct alignment. The vertical blue line seen close to 5 kb represents a misalignment, due to a nucleotide base error. A thymine nucleobase has been replaced by a cytosine, causing a shift in the codon frame from TAT to CAT. This recodes the amino acid sequence from a Tyrosine to a Histidine. (b) A schematic illustration of the sequenced SynXI chromosome depicting the chunk junctions B4/B5, C3/C4, E4/E5 and F2/F3 that have synonymous mutations. The black bases show the correct bases that have been mutated into the red bases.

Conclusions

The Sc2.0 project has expanded from students building a small chromosome to a global team of researchers working on the world's largest synthetic biology project. I was the first in the world to trial the Sc2.0 starter pack and then apply my knowledge to the assembly of an individual chromosome. Tasked with the assembly of synthetic chromosome XI, we successfully outsourced, cloned, assembled and achieved one of the aims of this project by integrating the first 90 kb of synthetic DNA, using the iterative marker recombination method according to suggested protocols. Strains were verified using PCRtag analysis, prior to ultimately being genome sequenced to validate the synthetic DNA sequence assembly. Since integration of the initial 90 kb, the SynXI chromosome has progressed to having more than 70% of its wild type DNA replaced by synthetic DNA, with minimal loss of fitness observed. The completion of synthetic chromosome XI is now well under way.

Design flaws and errors experienced during assembly, such as incorrect plasmid cloning of Chr11_A1 and the loss of an essential tRNA provided hurdles to my work, but were eventually debugged and aided in downstream improvements to the work in our lab as well as elsewhere within the Sc2.0 collaboration.

The first 90 kb of SynXI.A that I partly constructed (SynXI.A-C) contains a total of 23 loxPsym sites within the synthetic sequence. This forms part of the SCRaMbLE toolkit, the inducible recombination system that leads to the large-scale genomic rearrangements. In the following Chapters, I show the implementation of the SynXI.A-C strain constructed here as a suitable host for SCRaMbLE experiments.

4.0 CHAPTER 4: The Use of SCRaMbLE for the Introduction of Novel Genes into the Synthetic Yeast Genome

The design of synthetic chromosomes for the Sc2.0 project specifically introduces sequences that will allow large-scale genome “shuffling” via the system called SCRaMbLE. This has already been shown by others to be useful for rapidly-evolving new synthetic chromosome layouts by random Cre-lox mediated gene deletions, inversions, and duplications. Here, we hypothesised that with the correct formatting, heterologous genes could be automatically recombined into synthetic yeast at the same time as the Sc2.0 chromosomes are arranging due to SCRaMbLE. In this chapter, using the SynXI.A-C strain and strains containing other partially-synthetic chromosomes I develop a new approach called SCRaMbLE-in that allows foreign DNA to be incorporated into rearranged Sc2.0 chromosomes. This chapter focuses on characterising the effects of SCRaMbLE by predicting the gene outcomes of induced rearranged strains, and developing the methodology for SCRaMbLE-in by determining the optimal conditions for quick and efficient integration of heterologous DNA into synthetic yeast chromosomes.

4.1 Aims

- Develop and optimise a framework for the introduction of heterologous DNA fragments into a synthetic chromosome strain through SCRaMbLE-in.
- Use fluorescence cytometry to determine the effects of SCRaMbLE on three fluorescent genes
- Use the KanMX selectable marker as a model heterologous gene, determine the optimum parameters for highest yield and efficiency of genome integration, for SCRaMbLE-in induction.
- Examine the stability and viability of any successful transformants, investigating any genomic rearrangements and/or integrations due to SCRaMbLE.

4.2 Introduction

In the previous chapter, I reported the successful construction of synthetic chromosome XI strains as part of our lab's contribution to the Sc2.0 project. With the availability of partially-synthetic chromosome strains such as SynXI.A-C, the genomic recombination toolkit that uses Cre to recombine loxPsym sequences can theoretically be induced to bring on large-scale genomic rearrangements of the synthesised chromosome regions. The principles and methodology of the SCRaMbLE toolkit used in Sc2.0 were explained extensively in the introduction section. The Cre-lox recombination system has long been used as a site-specific recombination tool in molecular biology and is still the leading technology for inserting gene expression cassettes in transgenic mice[77, 146]. Cre-lox has been further adapted for use in the Sc2.0 project by modifications to both the Cre recombinase and to the lox sequence. For tight control of when recombination can occur an inducible Cre recombinase enzyme is used. The Cre recombinase is placed under the control of a daughter-specific promoter and thus theoretically only expressed in daughter cells[81] and therefore only does so as yeast is dividing and growing, rather than in stationary phase. For further control, the Cre enzyme is also fused to an oestradiol binding domain[128].

In the presence of β-oestradiol, the fused Cre recombinase undergoes conformational changes, allowing it to travel into the nucleus of the cell to attach to lox sites in the genome. Once in the nucleus, the Cre recombinase binds to the 13 bp palindromic sequences of the loxPsym sites, either side of the 8 bp core sequence. The core sequence of the loxP sites determines the directionality of recombination. The Cre recombinase contains an active site that allows it to recognise and bind to the DNA substrate (loxPsym sequences). Once the Cre recombinase is bound to the DNA it forms a dimer complex made up of two terminal domains, an amino terminal domain that interacts with the 13 bp palindromic DNA strand, and a carboxyl terminal domain which contains the active site of the Cre recombinase. The dimers created from one Cre-loxPsym complex will bind to a dimer of another loxPsym site, thus creating a tetramer. The Cre recombinase dimers first cleave one strand of the loxPsym DNA it is attached to, and recombine it to the adjacent sequence, creating a "Holliday Junction"[75-79]. The process is then repeated with the other loxPsym strand, and full recombination between the two loxPsym sites is completed.

Without the β-oestradiol the Cre protein accumulates in the cytosol away from the DNA. In the Sc2.0 project, rather than using standard lox sequences, a perfectly palindromic 'loxPsym' sequence (34 bp) is used, including the 8 bp core sequence. By being perfectly palindromic, it assures that the different rearrangements catalysed

by Cre (translocation, deletion, inversion) all occur with equal likelihood. On the downside, it means insertion of an inverted repeat sequence that can form a strong hairpin structure when single-stranded. As mentioned in the introduction, loxPsym recombination sites have been included throughout the synthetic genome design, placed at landmarks and in the 3'UTR of all non-essential genes (approximately 5000 genes), exactly three base pairs downstream of the gene's stop codon.

Previously, others have shown the use of the SCRaMbLE toolkit for rearranging and deleting genes from fully-synthetic and partially-synthetic chromosomes made by the Sc2.0 project. In this chapter, I investigate how SCRaMbLE can be used to simultaneously add heterologous genes from other sources while Sc2.0 chromosomes are rearranging. The aim is to develop a protocol for SCRaMbLE-in, a new method to bring foreign genes of interest automatically into rearranged synthetic yeast genomes.

The simplest heterologous gene to introduce and select-for is a dominant selectable marker. For the work in this chapter I used the dominant marker KanMX and selected for growth in the presence of G418s. The kanamycin gene is derived from *Escherichia coli* and is used in the well-characterised selectable KanMX cassettes. These are slightly altered versions of the bacterial kanamycin resistance transposon, where the expression marker gene is directed by a TEF promoter and terminator[147]. There exist six different versions of the KanMX cassettes, all which confer resistance to the geneticin antibiotic (G418s, part of the aminoglycoside family[148]) in eukaryotes, differing from one another by small modifications to the open reading frame, restriction sites and repeat sequences[149, 150].

Both kanamycin and G418s antibiotics operate by arresting the elongation stage of peptide synthesis, by binding to the 80S ribosomal complex, thus preventing the completion of eukaryotic translation. The advantage of using antibiotic markers, such as kanamycin over auxotrophic markers is that there is no prior need for gene deletion of disruption. These markers also can then remain available for alternative uses such as maintenance and expression of the pCre_EBD plasmid, and mega-chunk selection. The generally accepted working concentration of G418s in yeast is 200 mg/l. This is also known as the minimum inhibitory concentration (MIC)[151] and defined as the basal level of antibiotic to arrest non-resistant growth in yeast. An increased concentration of 350 to 500 mg/l (the 'selective' concentration) is used to kill non-resistant cells, which provides a useful further level of stringency to selection if needed. This particularly proved necessary for the work in this chapter.

4.3 Results:

4.3.1 Formatting KanMX4 to be Compatible for SCRaMbLE-in

For developing the SCRaMbLE-in method, I used the KanMX marker as a selectable heterologous gene. However, before beginning the work with the Sc2.0 chromosome strains, I first constructed six versions of the KanMX marker suitable for testing SCRaMbLE-in. Three of these were linear PCR fragments and three were the same fragments inserted into a plasmid that cannot replicate in yeast. In each case the three were KanMX with no loxPsym sites, KanMX with a 3' loxPsym site and KanMX with both 5' and 3' loxPsym sites. As with all other subsequent cloning work in this thesis, the design and cloning of the KanMX marker into appropriate vectors was initially carried out *in silico*, using the cloud-based vector software Benchling™ to allow for inspection of plasmid structures, orientations and sequences. A modular approach to cloning was employed, focusing on the core KanMX4 cassette sequence as a model insertion gene, and thus not modifying the TEF promoter and KanMX ORF in any of the constructs. Instead cloning focused on additions of restriction enzyme or loxPsym sites around these parts, and this cloning was done in a modular manner to facilitate future downstream cloning into all other possible plasmid vectors. The two DNA formats used for hosting the KanMX marker, were a bacterial (pUC19) plasmid with KanMX cloned into the MCS, and the other format simply being a PCR-amplified linear DNA fragment of KanMX. This project did not use YCp or YIp based yeast plasmids due to the presence of centromeric sequences (CEN) and autonomously replicating sequences (ARS) and auxotrophic markers, which provide stable expression of encoded genes in yeast.

The KanMX sequence was amplified from the genomic DNA of yeast strain YJR092W – a BY4741 based strain that encodes the KanMX4 gene on its chromosome X – using touchdown PCR (section 2.3.6) with primer DJ037 and DJ038. The endogenous (TEF) terminator for KanMX was not included in the design or cloning at this stage, due to the need to include a loxPsym site between the ORF and terminator in downstream constructs. Following gel electrophoresis and gel extraction, the amplified PCR product consisting of the TEF promoter, KanMX ORF, and a newly attached 3' loxPsym site, was cloned into destination vectors pRS413 and pDJ002 – a previously constructed plasmid already containing a 5' loxPsym site. This formed intermediate plasmids pDJ015 and pDJ016, respectively, as illustrated in Figure 4.1. Construction of these plasmids was confirmed by single-read sequencing analysis (data not shown). All cloning was carried out by digestion with NotI and XhoI restriction sites to create compatible sticky ends that are ligated

together using a short ligation protocol with T4 DNA ligase (section 2.3.4) followed by transformation into *E. coli* DH10B using electroporation method (section 2.3.14).

As a control the original KanMX4 cassette, lacking any loxPsym sequence and unmodified in any way, was also amplified from the same YJR092W strain, using a standard PCR reaction (Chapter 2, section 2.3.5) with primers DJ039 and DJ040, with encoded restriction enzyme sites. It too was cloned into pRS413 and the new plasmid named pDJ014. The common yeast pRS series of shuttle vectors[119], on which pDJ002-16 are based upon, all encode the common M13 primers binding sites, flanking the MCS of these vectors. This allowed amplification and extraction of linear versions of the cloned insert constructs by PCR with M13 primers. The linear versions of the cloned KanMX constructs were termed KanMX, KanMX_1L and KanMX_2L, for zero, one and two loxPsym sites, respectively. These were the first three constructs needed for SCRaMbLE-in experiments. For the other three constructs, the inserts from pDJ014-16 were cloned into bacterial vector pUC19, using restriction enzymes compatible with both pUC19 and pRS413. This cloning formed the non-yeast replicating bacterial plasmids for SCRaMbLE-in, pDJ017, pDJ018 and pDJ019 (Figure 4.1).

4.3.2 The Development of a SCRaMbLE-in Protocol

SCRaMbLE-in was tested experimentally in a variety of host strains, with varying a few parameters, in order to establish the optimal conditions for the introduction of heterologous DNA into a rearranging synthetic yeast chromosome. The process was divided into two stages, whereby the first stage centred on the transformation of the heterologous DNA into the yeast, and the second stage was the induction of SCRaMbLE with β-oestradiol, for recombination of heterologous DNA into the synthetic chromosomes. Once the KanMX DNA is in the yeast nucleus it is compatible with the Cre-lox recombination within the synthetic chromosome if it contains loxPsym sites. As no previously recognised procedure exists for SCRaMbLE-in using Sc2.0 strains, the method used and results obtained are detailed here.

A quickstep transformation (Section 2.3.16) was used to transform yeast using 500 ng of the purified DNA of interest. The modified KanMX cassette was provided in the form of a bacterial plasmid or simply as a linear fragment. As neither the bacterial nor the linear forms contain auxotrophic or an antibiotic selection markers, no selective pressure was exerted on the yeast to maintain the transformed DNA within the organism, post transformation. It was reasoned that the SCRaMbLE induction should occur immediately after the transformation of the foreign DNA, thus

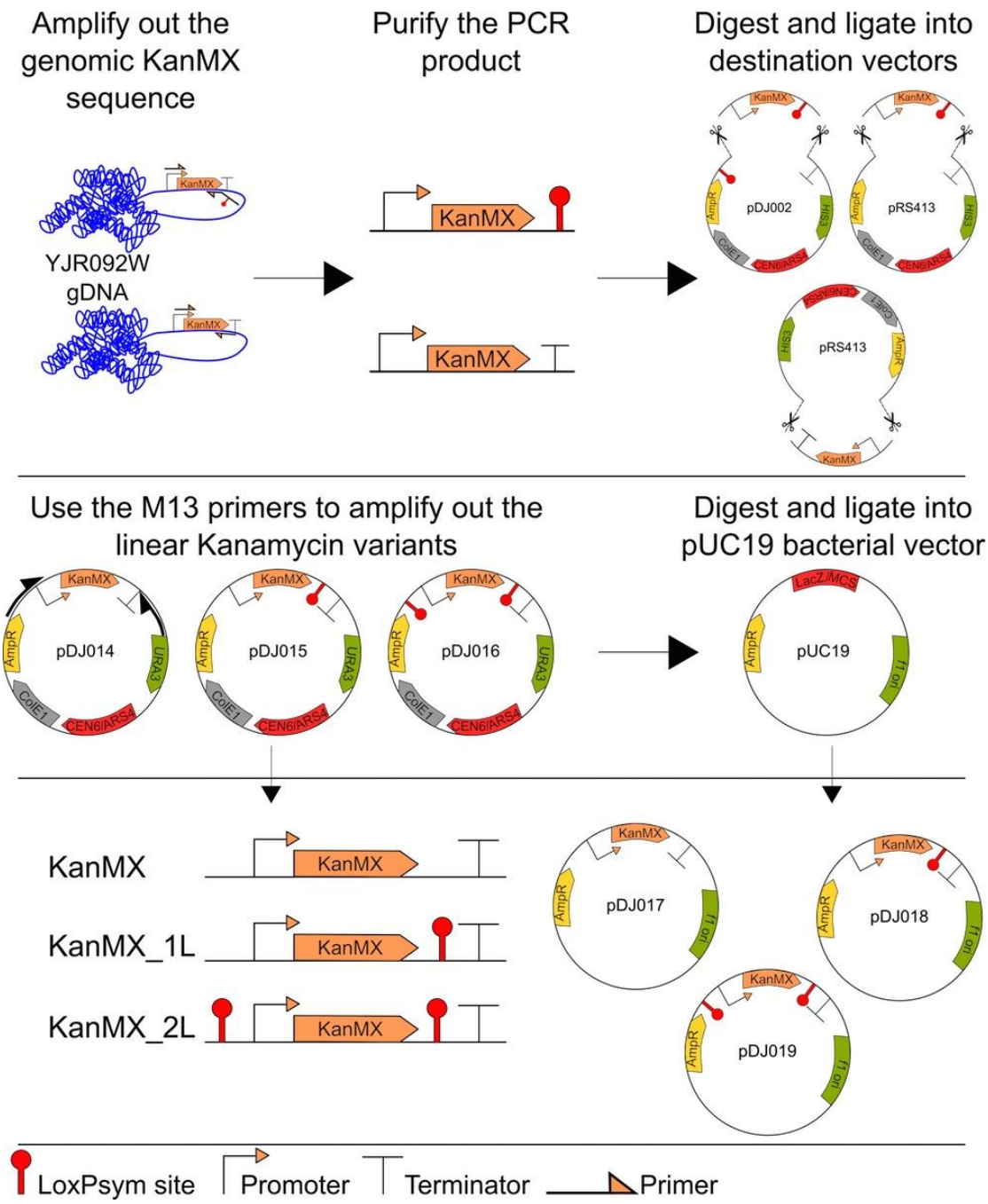


Figure 4.1 Schematic illustration of KanMX cloning technique. Genomic DNA was extracted and amplified using specially designed and modified primers to add on a loxPsym site where necessary. PCR products were then cloned into plasmids pDJ012 and pRS413, which were used as the reference vector for all constructs thereon. M13 primers amplified out the linear KanMX DNA (KanMX, KanMX_1L, and KanMX_2L) and restriction digestion and ligation was used to clone the modified KanMX sequences in pUC19, forming plasmids pDJ017 to pDJ019.

increasing the chances of genomic integration of the kanamycin cassette. Therefore, at the end of the transformation, all samples were resuspended in YPD media (1 ml). At this stage samples were allocated to two separate tube (5 ml FACS) for induced (SCR+) and non induced cells (SCR-) and 500 µl of the resuspended transformant

sample was transferred into each. SCRaMbLE is induced by the addition of β-oestradiol (final concentration 1 μM) to the SCR+ tube. Samples were incubated at 30 °C, shaking at 225 rpm, for one, two and three hours of induction. Due to the novel use of the SCRaMbLE system in yeast and lack of associated publications, our knowledge of this recombination system is somewhat limited and the optimal induction times for experiments were yet to be determined, thus why three different induction periods were used. After every hour, 150 μl of the YPD suspended cell (+/-β-oestradiol) mixture was transferred into 1.5 ml eppendorf tubes, and spun down for two minutes at 8000 rpm. This step was necessary in order to remove the YPD media and any β-oestradiol that may have been present. The pellet was resuspended in 200 μl of 5 mM $CaCl_2$ and the entire volume was plated onto a YPD plate. The cells were then allowed to recover for one hour on the plate, incubated at a temperature of 30 °C, before 200 μg/ml of geneticin antibiotic was spread, top-down, on to the plates surface. Plates were then incubated (at 30 °C) for two to three days, until colonies appeared. Figure 4.2 shows is an illustration of the protocol used for SCRaMbLE-in.

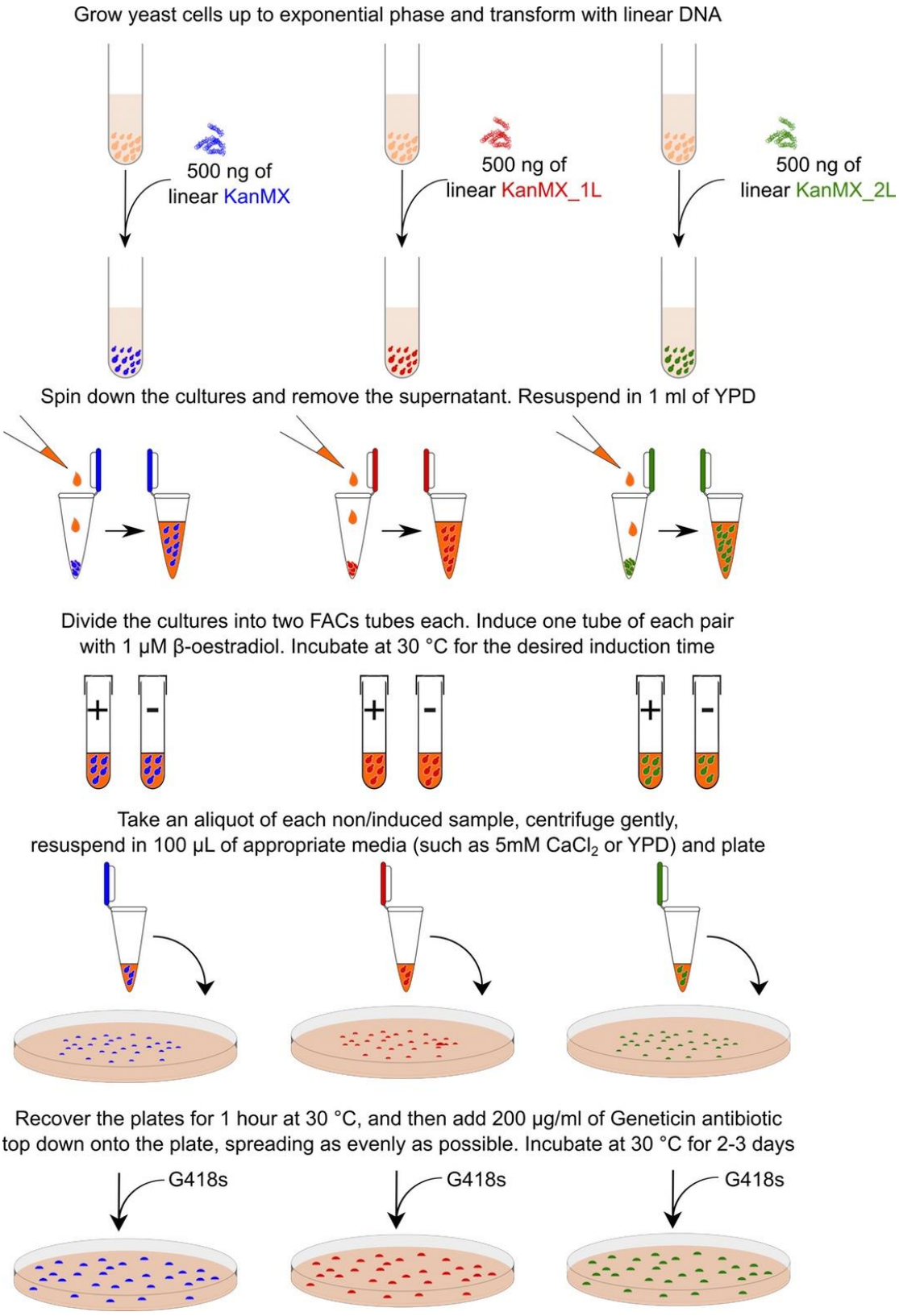


Figure 4.2 Process of SCRaMbLE-in Schematic. 500 ng of (KanMX) DNA is transformed into the cell, and cultures were resuspended in 1 ml of YPD. Cell solutions were divided into two FACs, representing the induced (+) sample, and the uninduced (-) sample. Induction time was varied between one, two and three hours, after which cells were spun down to remove the β-oestradiol, plated and incubated for one hour recovery period before spreading 200 µg/ml of G418s antibiotic top-down onto the cell-containing plates.

4.3.3 Using SynXI.A-C as a host for SCRaMbLE-in Reactions

We began the SCRaMbLE-in experiments, as outlined in Figure 4.2 in the haploid strain SynXI.A-C, constructed in Chapter 3, containing 90 kb of synthetic sequence, including >20 loxPsym sequences. The aim was to assess and optimise the SCRaMbLE-in methodology, using the cloned and constructed KanMX variants to produce G418s resistant strains that have incorporated the heterologous marker gene into the synthetic chromosome sequence during rearrangements. Figure 4.3a-c shows the average numbers of colonies following SCRaMbLE-in induction (one, two and three hours) using linear KanMX variants, which were recovered on 200 µg/ml G418s+ YPD agar. All colonies that showed growth were considered as viable, regardless of morphological differences that may have arisen as a consequence of essential gene deletion or silencing. There appear a higher number of surviving colonies from the uninduced (-) yeast in contrast to the induced (+) yeast that performed SCRaMbLE. Out of 2,530 colonies (across all induction times and states), 55% were from the non-induced cells. I evaluated whether the numbers of colonies of the induced states were of significance, and calculated a P value of 0.0456 (ANOVA). Based on the statistical analysis of the results presented in Figure 4.3a-c the numbers of colonies observed showed a lack of significance with respect to induced versus uninduced, and also in terms of the time of induction.

The presence of G418s antibiotic places a strong pressure upon yeast cells to express the KanMX gene and so it is possible that colonies are arising that have not integrated the KanMX gene as intended but somehow still can grow in the presence of G418s. In an attempt to negate false positive colonies from those that are truly resistant, selected colonies were restreaked onto the same media type (YPD), but containing a higher antibiotic concentration of 350 µg/ml G418s, and the results are shown in Figure 4.4. There was a large decrease of total colony count (decrease by 82%), from 2,530 to 457 colonies following restreaking, verifying that a large number of the colonies were likely false positives, surviving only on the MIC concentration plates (200 µg/ml). This was suspected to be the case as the majority of these colonies showed smaller sizes associated with phenotypic differences (not seen in the wild type strain) and decreased fitness.

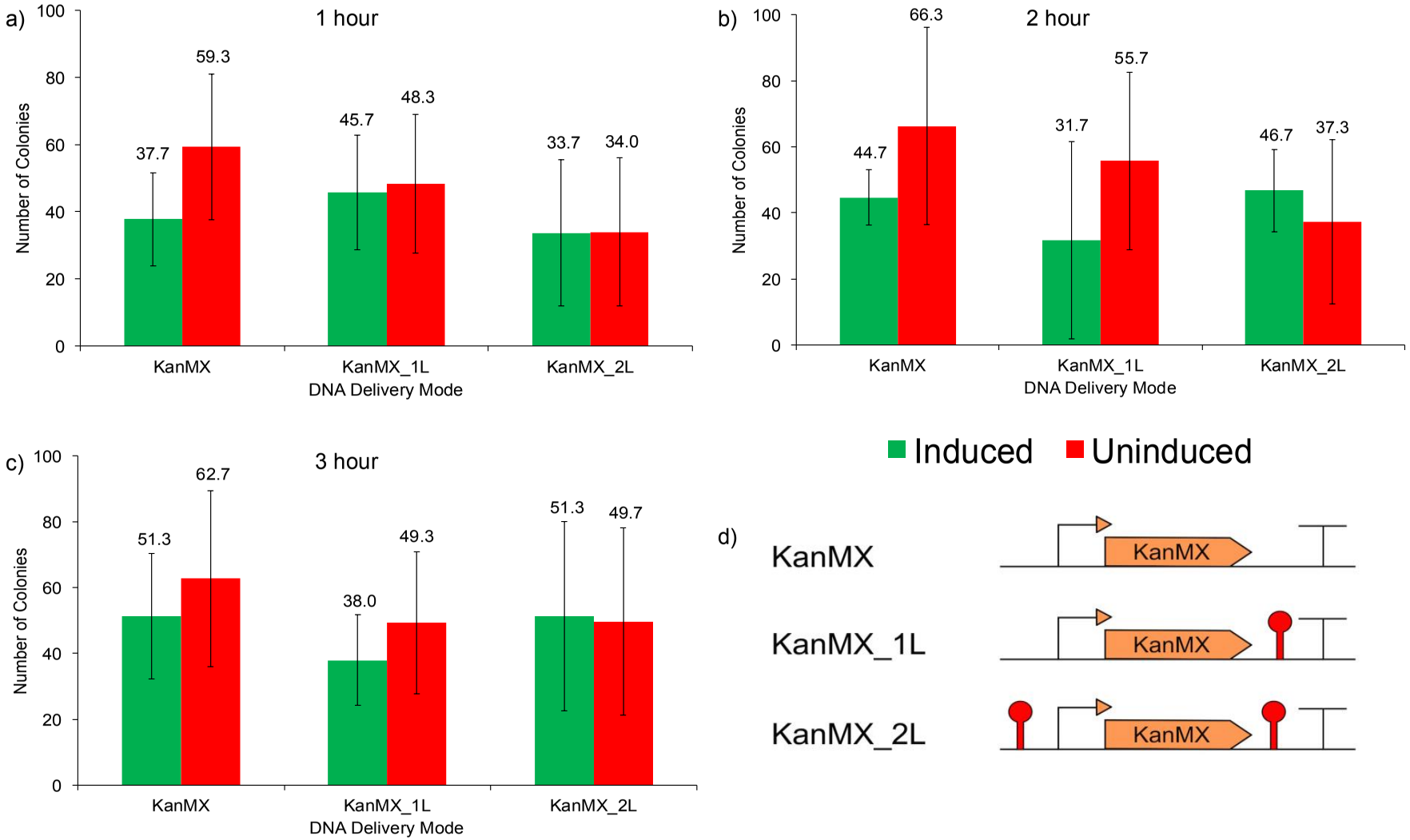


Figure 4.3 SCRaMbLE-in with Haploid SynXI.A-C Results 1. SynXI.A-C was transformed with linear KanMX, KanMX_1L and KanMX_2L (illustrated in d) and SCRaMbLE-in was induced for one (a), two (b), and three (c) hours, with β-oestradiol. The average colony numbers are represented in the three bar charts. Colonies were selected for in the presence of 200 μg/ml of G418s. The light green bars represent the induced colony numbers, and the red bars represent those from the uninduced controls. Induced colony numbers are not significantly different from uninduced (ANOVA; $P>0.04$). Error bars represent standard error (ERR=STDEV/√n, where n=3).

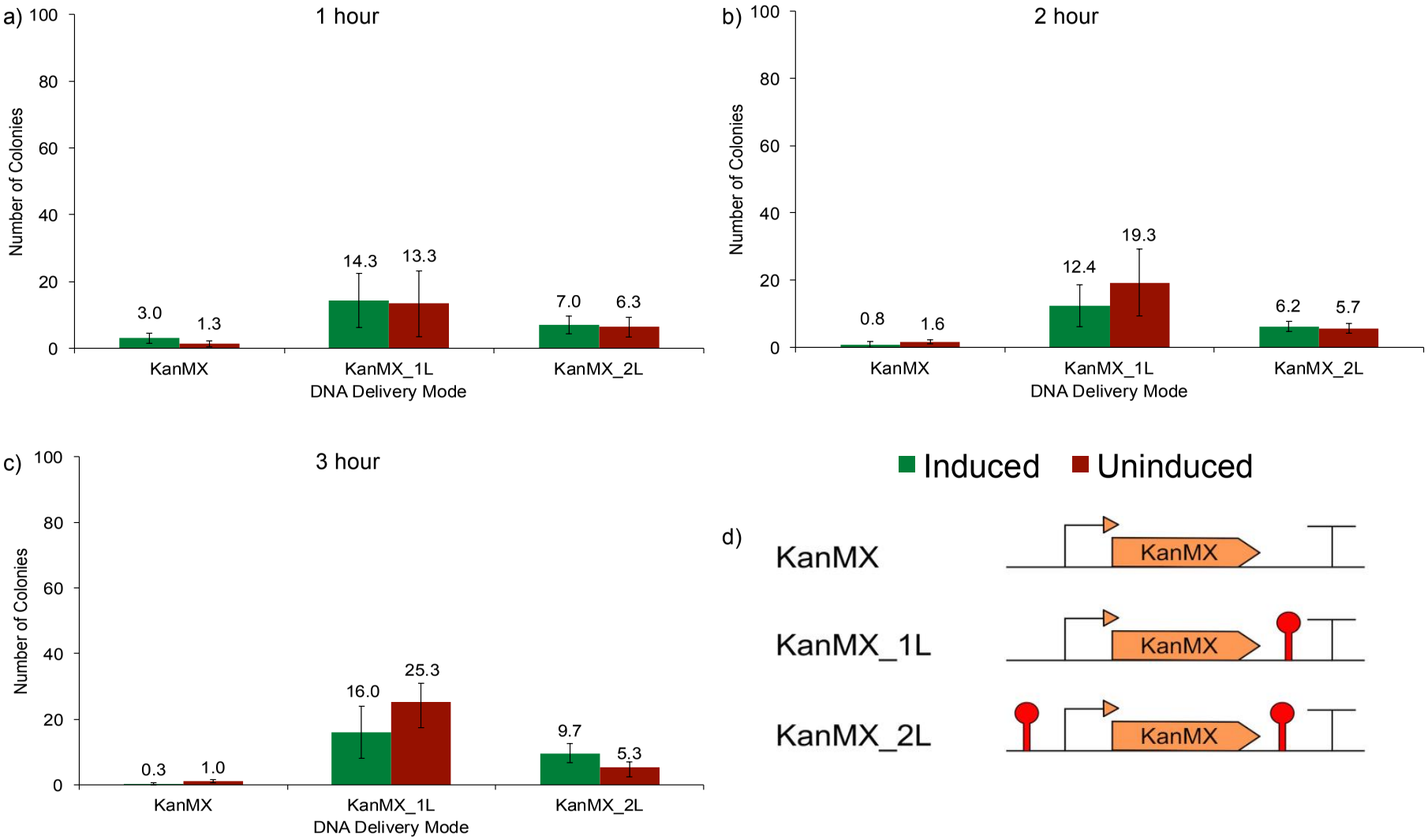


Figure 4.4 SCRaMbLE-in with Haploid SynXI.A-C Results 2. Restreaked average numbers of the haploid SynXI.A-C recovered colonies, in the presence of 350 µg/ml G418s. Dark green bars are induced colonies that showed growth post-restreaking, and the dark red bars show the number of uninduced surviving colonies. Error bars represent standard error (ERR=STDEV/√n, where n=3).

Following restreaking, the results show the greatest number of surviving colonies came from KanMX_1L transformants, with numbers of colonies from this variant exceeding the sum of KanMX and KanMX_2L. Secondly, the quantity of uninduced colonies exceeds that of the induced in two out of three cases (two hour and three hour induction times). β-oestradiol is not added to these samples, the SCRaMbLE system is not induced and therefore recombination between the loxPsym sites is not expected. Instead positive colonies here are likely to be the result of some other kind of integration or maintenance of the KanMX gene due to the high environmental pressure from the presence of G418s. An alternative explanation, that leaky Cre recombinase expression means that SCRaMbLE can occur without induction, is unlikely as there is tight control on Cre via the EBD system, and others have reported that the system is not leaky[60, 72, 82].

Statistical analysis on these initial results shows that induced colony numbers are not significantly different from uninduced numbers (ANOVA; $P>0.05$). The induced KanMX variants were compared to each other (regardless of induction time), and some statistical difference is observed ($P<0.05$). A *post-hoc* Tukey HSD test showed that all three mean colony numbers were significantly different from each other, suggesting that the presence of loxPsym sites affects numbers of surviving colonies, although it remains ambiguous as to which construct is the most favourable for correct KanMX integration.

In an attempt to decipher these unexplainable colonies, five colonies from the uninduced and five colonies from the induced set (or all if $n<5$) were screened further for the presence of the KanMX gene using colony PCR with KanMX amplifying primers. In both instances all colonies failed to show any band products, whereas positive control for KanMX DNA gave a clear band (data not shown). The data was thus inconclusive as G418s resistance was seen but the provided KanMX DNA couldn't be detected. Therefore using the haploid SynXI.A-C strain as the host organism for SCRaMbLE-in characterisation did not help to determine the optimal conditions for the introduction of heterologous DNA into the rearranging genomes, as no clear evidence for KanMX integration could be seen.

4.3.4 Production of a Diploid Synthetic Yeast Strain

It was hypothesised that the SCRaMbLE of the haploid SynXI.A-C strain could be too lethal for the transformed yeast population as the majority of the rearrangements within the 90 kb synthetic chromosome region would lead to cell death. Previously collaborators had shown that SCRaMbLE survival was significantly

higher when diploid yeast was used, where the synthetic chromosome is paired on the cell with its wild type equivalent which does not rearrange (*Y. Shen, personal communication*). Therefore a Syn/WT diploid version of SynXI.A-C was made by mating haploid SynXI.A-C with its wild type counterpart BY4742 (Chapter 2, Table 2.2). The diploid genotype provides an extra copy of the genome, that is immune to the negative effects of SCRaMbLE, thus increasing the tolerance of yeast to it. Each diploid strain was transformed with the pCre_EBD plasmid to provide the inducible Cre recombinase enzyme. Verification of the diploids was carried out using the PCRtag analysis (section 2.3.10) which was originally intended for Sc2.0 mega-chunks transformation screening. By using these available primers and doing PCR amplification on the diploid genomic DNA there was no need to carry out sporulation experiments, which are time-consuming and based on visual judgement. The use of PCRtag analyses provided confirmation of the presence of both the wild type (BY4742) and the synthetic (SynXI.A-C) chromosome in the newly formed diploid cell. Figure 4.5 shows the imaging of the gel electrophoresis results of the diploid PCRtag amplifications showing the amplicons of both the WT and SYN tags. The results for the other diploids that I generated are also shown.

4.3.5 SCRaMbLE-in using Diploid Synthetic Strain, dSynXI.A-C

Using the newly formed diploid synthetic strain, dSynXI.A-C, we repeated the SCRaMbLE-in experiments with the linear KanMX DNA, as well as the pUC19-based vectors pDJ017 – pDJ019. Again, 1, 2 and 3 hours inductions were used. With the average *S. cerevisiae* cell doubling time being approximately 90 minutes in YPD media[152] , and given that the pCre_EBD plasmid lies under the control of a SCW11 daughter specific promoter[81], an induction time of three hours allows for two doublings, maximum three per cell.

I first performed the SCRaMbLE-in experiments with the KanMX DNA provided on bacterial plasmids (pDJ017, pDJ018 and pDJ019). The protocol was done as before (Figure 4.2) and cultures were plated on YPD agar containing 200 µg/ml of G418s antibiotic. After 2 to 3 days of incubation at 30°C, plates were assessed for the number of surviving colonies (Figure 4.6). The sum of colonies across all plasmids and all induced hours was now limited to a total of only 28 colonies when SCRaMbLE-in was on, and zero colonies for when no induction occurs. Further analysis of colony numbers based on the KanMX formatting reveals that overall pDJ018 (one loxPsym site) results in the highest number of colonies (11), whereas pDJ017 and pDJ019 both result in a total of nine colonies (all hours combined), during induction. This was a vast difference from the 2,530 colonies seen

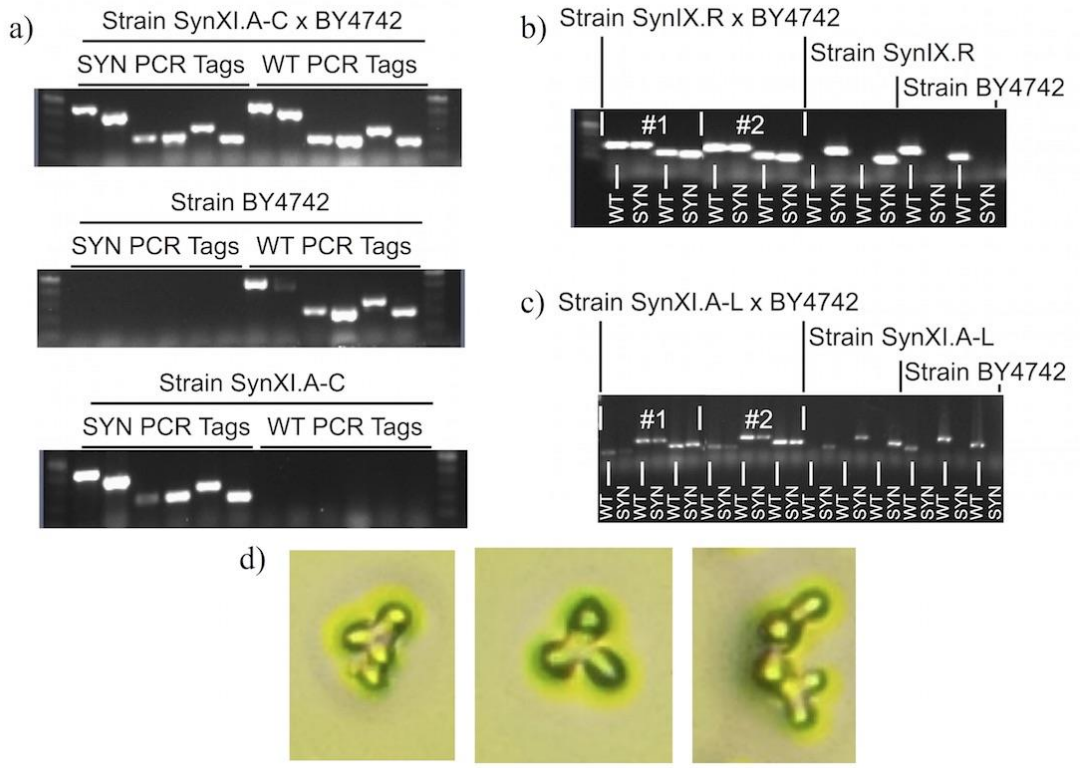


Figure 4.5 Diploid Strain Verification. Several diploid strains were verified using SYN and WT primers from the Sc2.0 kit, three examples are shown here (a) dSynXI.A-C, (b) dSynIX.R and (c) dSynXI.A-L. Image (d) shows microscopic images of zygotes that were isolated following mating, using the SporePlay Singer Instruments Microscope, screening for schmoo formation and multiplying cells.

in Figure 4.3. All colonies were isolated and restreaked onto higher strength antibiotic plates of 350 µg/ml G418s, to eliminate any false positives (Figure 4.7). Results showed a further decrease in colony numbers, but showed significantly more colonies (ANOVA; $P<0.05$) when KanMX_1L was introduced (pDJ018) compared to when KanMX (pDJ017) or KanMX_2L (pDJ019) were transformed pre-SCRaMbLE.

In contrast to the previous haploid SCRaMbLE data, these results showed low numbers of colonies, but gave significant data confirming a difference between induced and uninduced strains. Seemingly when KanMX is provided as circular DNA the highest number of resistant colonies post-SCRaMbLE come from the construct with a single loxPsym site (pDJ018). Different induction times gave no significant changes to the colony numbers.

Next, experiments were repeated but this time using linear KanMX fragments as the heterologous DNA, to test whether this DNA delivery mode showed a difference compared to circular DNA being provided. The linear KanMX constructs were transformed into the diploid yeast and SCRaMbLEd as previously, for one, two and three hours, and samples were then plated on YPD agar with 200 µg/ml G418s.

The results showed large increases in the number of viable colonies arising (Figure 4.8), with more colonies for induced and also colonies for uninduced samples. Although the increase in uninduced colonies is not sought after, the overall higher number of colonies is more favourable than low numbers seen for the pDJ017-19 experiments. Applying the same statistical analyses as before, we compare the total number of induced (+) colonies versus the total number of uninduced (-) colonies for all time points, and the data proves to be not significant ($P>0.05$). However, visually the KanMX_2L samples alone appear to be significantly more prevalent in induced samples than uninduced, especially at the later time-points.

As once again quite a large number of small colonies were observed, I next restreaked all colonies onto YPD agar plates containing the higher strength of 350 µg/ml G418s antibiotic. A total of 2,415 colonies were restreaked and screened for growth. Any observed growth was counted as a positive colony, whereas lack of growth was regarded as a previous false positive. Results are shown in Figure 4.9. Overall there was a 57% decline in total numbers of colonies observed (over all three times) dropping from 2,415 to 1,048 colonies. This is a much smaller decline than previously seen with the haploid strain. Of the 1,048 colonies that survived on the higher antibiotic concentration, 85% of these look to be as a direct result of SCRaMbLE-in as they are only present in the induced samples. This suggests that the SCRaMbLE-in method does lead to more G418s-resistant colonies. The number of loxPsym sites was found to be a significant parameter for the number of surviving colonies (ANOVA; $P=0.04$), and a *post-hoc* Tukey HSD test showed that the mean colony count for KanMX+ versus KanMX_1L+ were not significant, however the KanMX_2L+ mean colony count was significant from both its counterparts. Thus the highest numbers of surviving colonies occurred with SCRaMbLE induction when linear KanMX DNA was provided with two flanking loxPsym sites (KanMX_2L).

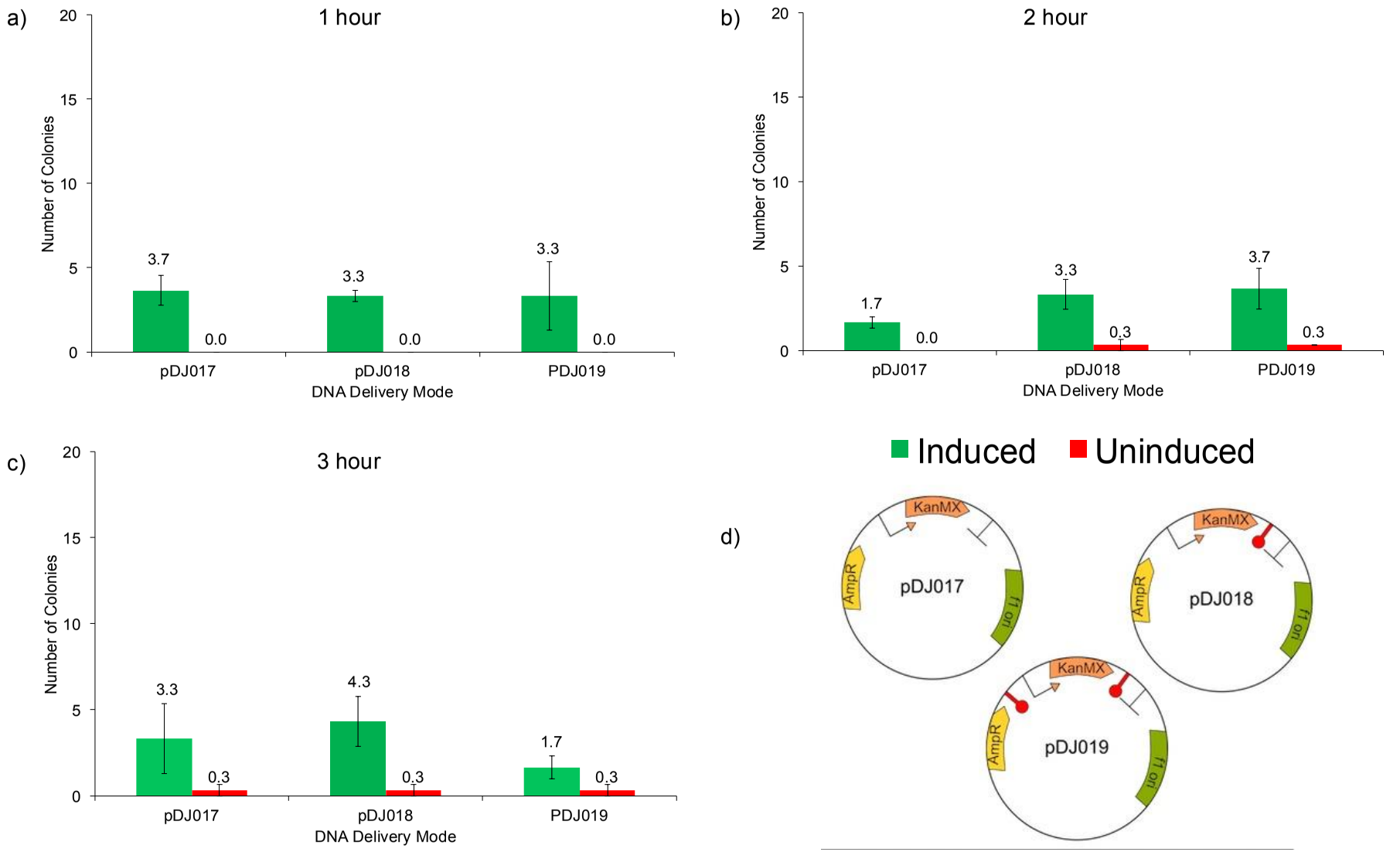


Figure 4.6 SCRaMbLE-in with a Bacterial Plasmid into dSynXI.A-C on 200 µg/ml of G418s. dSynXI.A-C was transformed with pDJ017, pDJ018 and pDJ019 constructs (d) and SCRaMbLE-in was induced for one (a), two (b), and three (c) hours, with β-oestradiol. The average colony numbers are represented in the three bar charts. (d) shows the bacterial plasmids used to deliver the KanMX constructs for SCRaMbLE-in. Induced colony numbers are significantly different from uninduced amounts (ANOVA; $P<0.05$). Error bars represent standard error (ERR=STDEV/√n, where n=3).

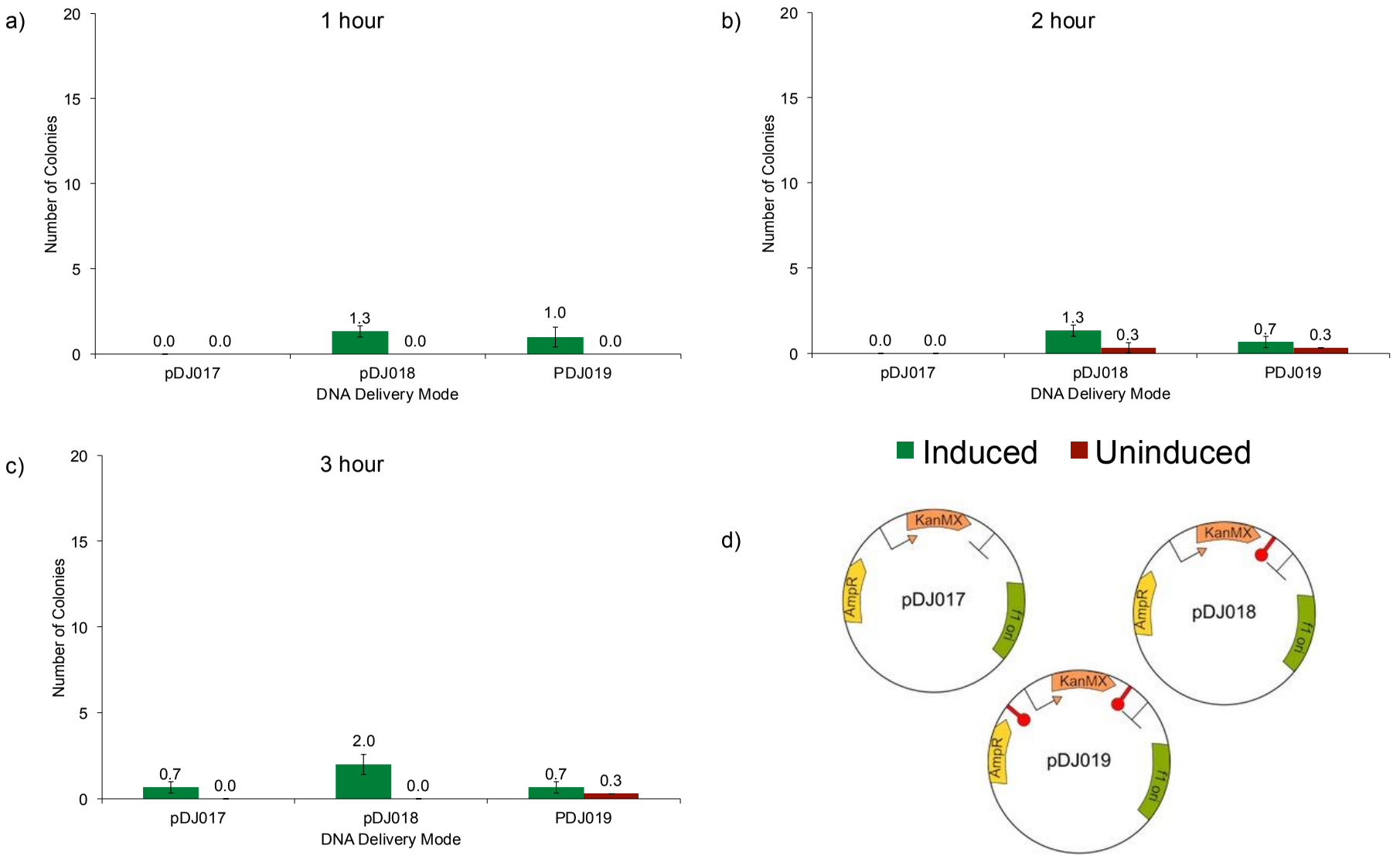


Figure 4.7 SCRaMbLE-in with Bacterial Plasmids into sSynXI.A-C on 350 µg/ml G418s. pDJ017, pDJ018 and pDJ019 (d) colonies from Figure 4.6 were restreaked onto higher antibiotic media containing 350 µg/ml G418s to select for colonies exhibiting strong G418s resistance. Results are shown for colonies that were induced for one hour (a), two hours (b), and three hours (c) of SCRaMbLE-in. Diagrams of the plasmids are shown in (d). Error bars represent standard error (ERR=STDEV/√n, where n=3).

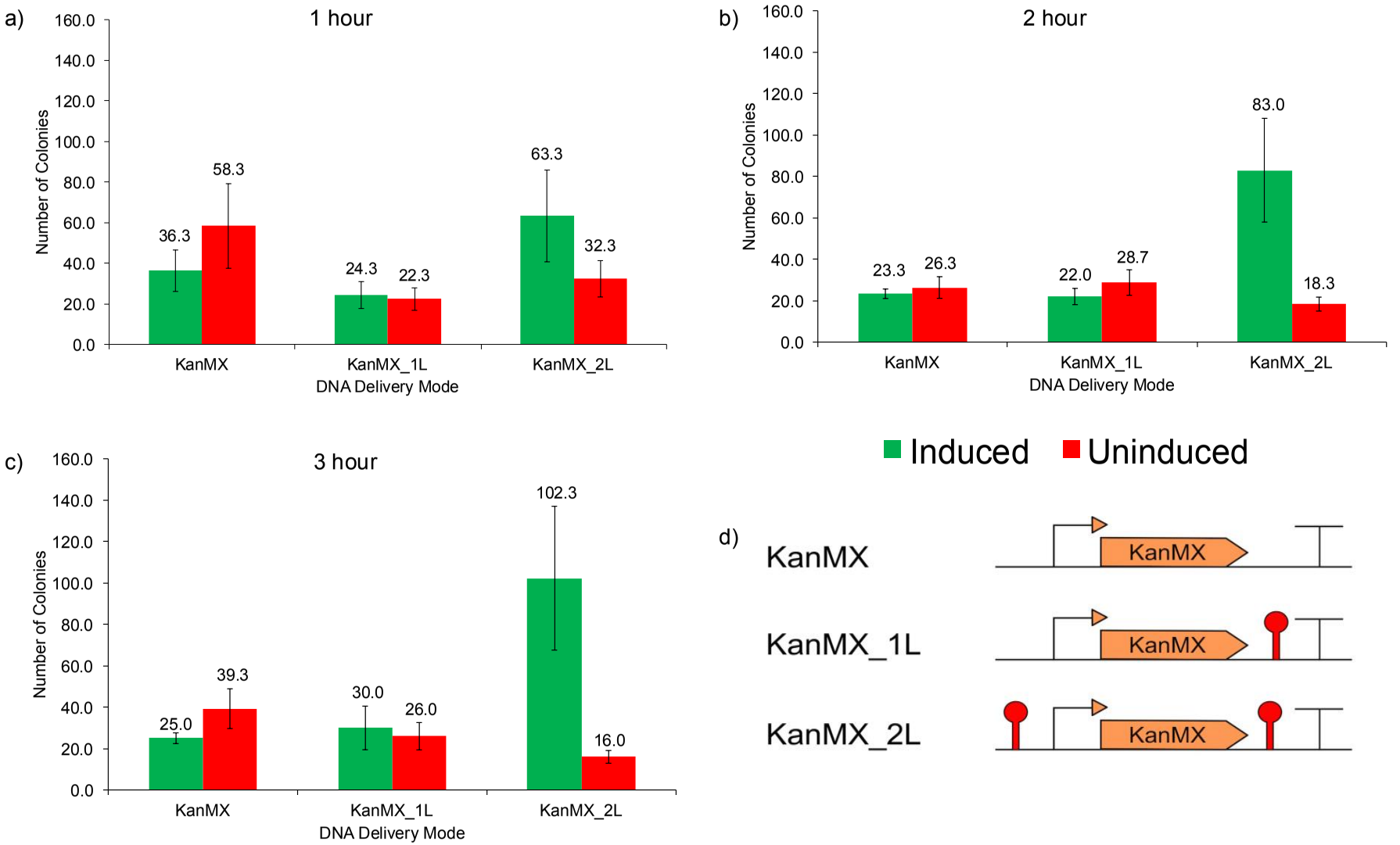


Figure 4.8 SCRaMbLE-in with Linear DNA into dSynXI.A-C on 200 µg/ml. Bar charts showing results obtained from SCRaMbLE-in of linear KanMX constructs for one hour (a), two hours (b), and three hours (c), in dSynXI.A-C. Light green bars represent the average number of surviving colonies following induction of SCRaMbLE-in on 200 µg/ml G418s media, whereas light red bars represent the uninduced controls. Varying the number of loxPsym sites also appears to lead to significantly higher numbers of colonies, and the mean of KanMX_2L+ is significant with the mean of KanMX_2L+ significantly greater that that for KanMX_1L+/KanMX+. Part (d) is a schematic illustration of the linear KanMX constructs used in this study. Error bars represent standard error (ERR=STDEV/√n, where n=3)

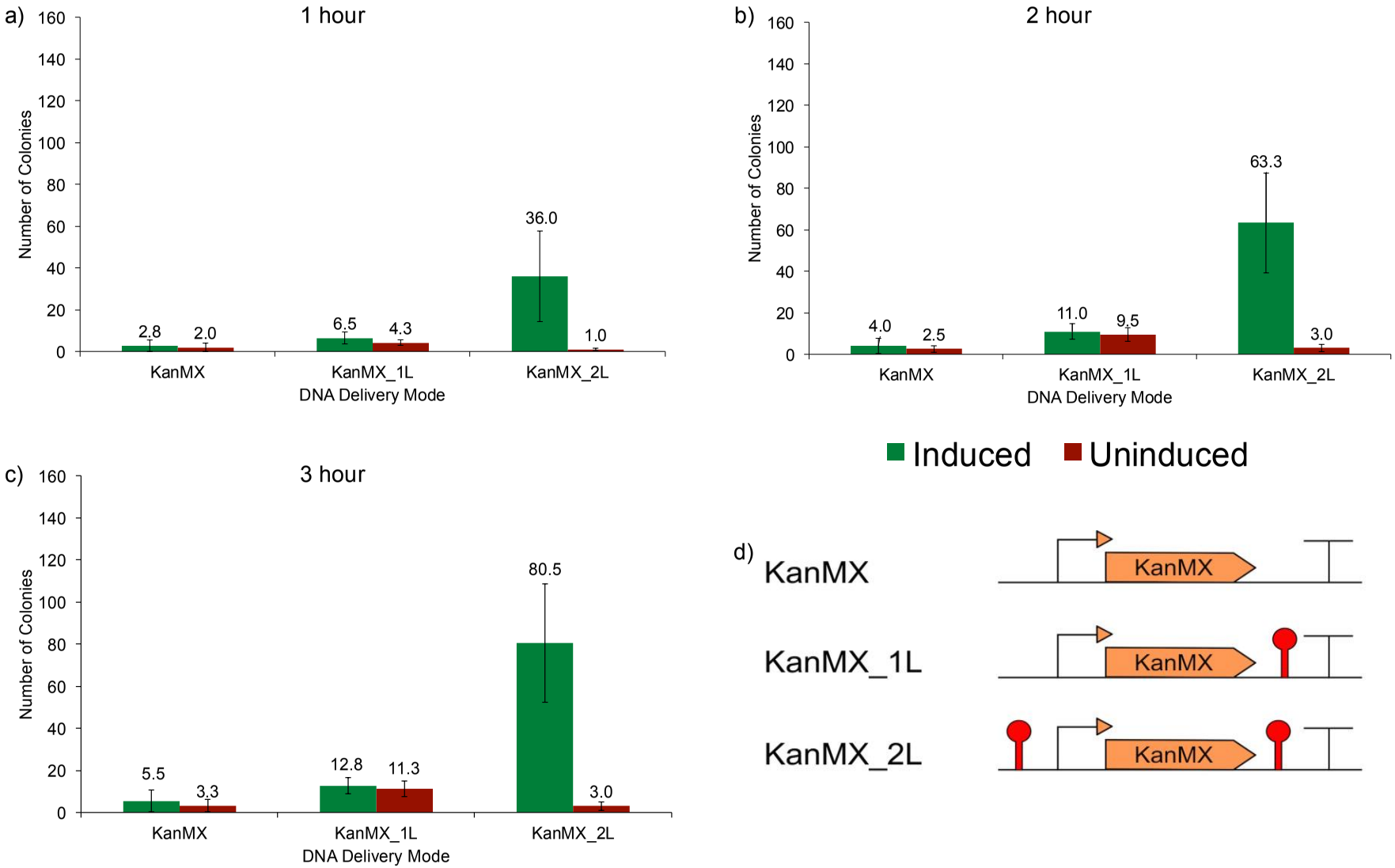


Figure 4.9 SCRaMbLE-in with Linear DNA into dSynXI.A-C on 350 µg/ml. Average restreaked colony count for SCRaMbLE-in of linear KanMX constructs on plates containing a higher antibiotic concentration of 350 µg/ml for one (a), two (b) and three (c) hours of induction. No significant difference is observed as a factor of time, however KanMX_2L linear constructs show the highest average number of surviving colonies, especially when compared to the uninduced state. ANOVA and Tukey HSD testing show that the KanMX_2L+ mean colony number is significant over all others ($P<0.05$). A schematic illustration of the KanMX constructs is shown in (d). (n=3 and error bars represent standard error).

4.3.6 Determining the Genomic Sites of SCRaMbLE-in

With diploid yeast, the G418s resistance results strongly suggested that the SCRaMbLE-in protocol developed here worked to integrate heterologous DNA from circular plasmids with one loxPsym site and linear DNA fragments with two loxPsym sites. However, for complete confidence in this assertion we need to provide evidence for the integration at the genomic level. Therefore, selected colonies from the 350 µg/ml G418s plates from the above experiments were grown and their genomic DNA extracted and screened for the presence of KanMX. To simultaneously determine if the KanMX DNA is in the genome and where it is, I used an inverse PCR (iPCR) protocol (Chapter 2) with primers designed to anneal within KanMX but amplify away from each other (divergent). When genomic DNA is digested with a common-cutting restriction enzyme (that doesn't cut KanMX) and then re-ligated, it forms circular DNA. The circles of genomic DNA containing the KanMX gene will be amplified by the divergent primers and the resulting PCR product will contain surrounding DNA sequence, pin-pointing the genomic locus where the KanMX DNA was integrated.

A number of restreaked surviving colonies were used in the inverse PCR (iPCR) experiments. The genomic DNA of samples was extracted as described in Chapter 2 and digested using enzyme XbaI, which cleaved the DNA 35 times to produce bands of 2,496 bp on average. The digested mixture was slowly re-ligated with T4 ligase to form circular products. The ligation reactions were used as DNA template for PCR amplification with divergent primers pDJ047 and pDJ048. Any successfully amplified bands were gel extracted and sent for single-read sequencing.

Figures 4.10 to 4.14 show the sequencing traces and iPCR schematics of five isolated and successfully amplified samples. The samples were randomly isolated from 350 µg/ml G418s-agar, and represent KanMX_2L samples that have been induced for two hours (Figure 4.11), and three hours (Figure 4.10, 4.12 and 4.13), as well as the PCR amplification of a dSynXI.A-C sample SCRaMbLEd with pDJ018 (Figure 4.14). The sequencing results show that three of the five examples show evidence of genomic integration (Figures 4.10, 4.11 and 4.13), in all cases providing sequencing of the flanking DNA of KanMX_2L. The high quality of the sequencing data facilitated the matching of the flanking sequences to loci within dSynXI.A-C, and provided evidence for both the integration and orientation of genes following SCRaMbLE-in.

iPCR example 1 from Figure 4.10, amplified the genomic region upstream of the 18th loxPsym of dSynXI.A-C, and the downstream sequence of the 19th loxPsym

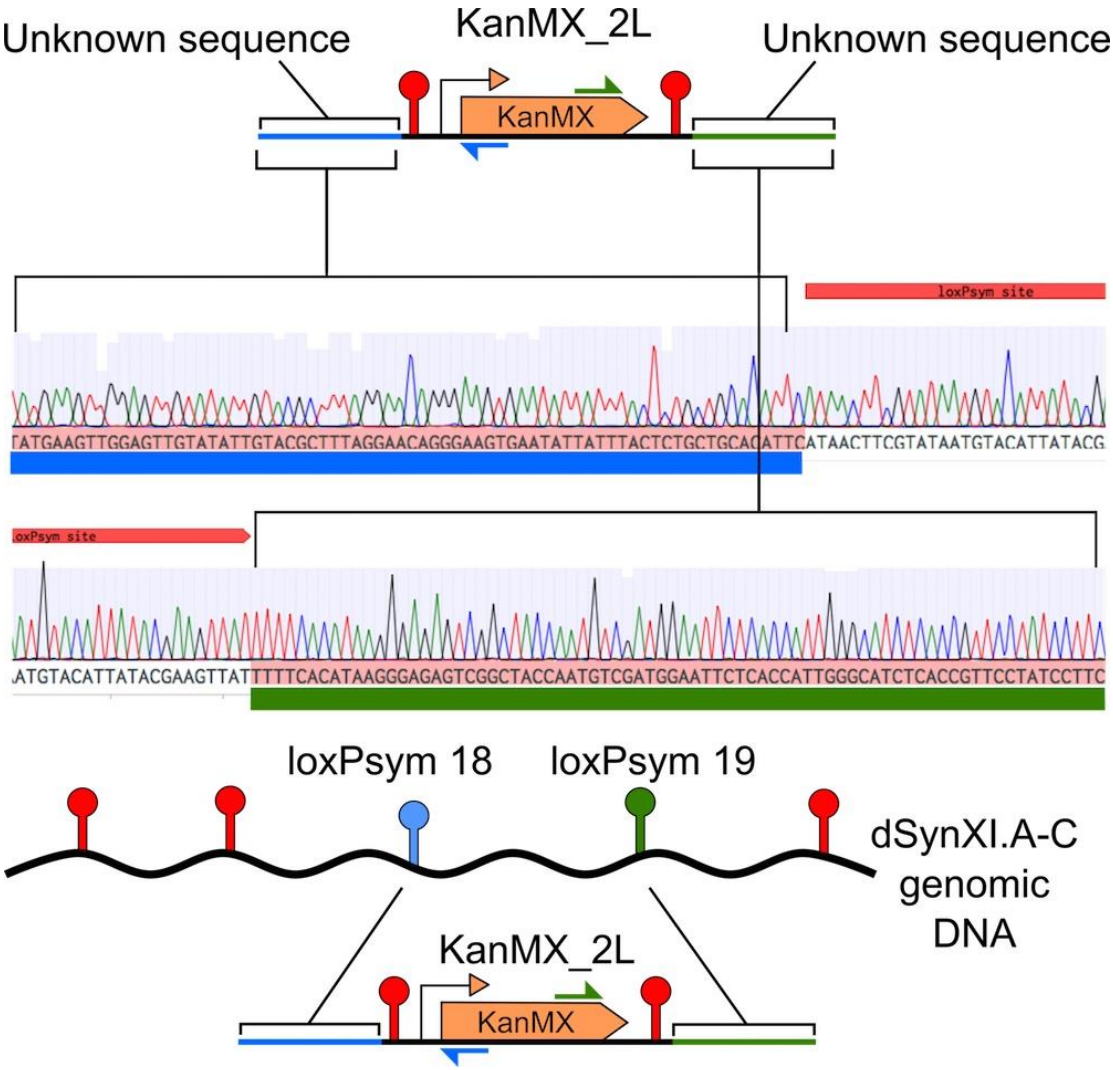


Figure 4.10 iPCR Analysis 1. Inverse PCR and sequencing data showing genomic integration for KanMX_2L following SCRaMbLE-in for three hours, example 1. Genomic DNA of a SCRaMbLEd-in colony was amplified using divergent primers (shown as green and blue arrows). PCR products were gel extracted (not shown), and sequenced. The sequence ABI traces shown above (underlined green and blue) show the surrounding sequence of the KanMX_2L fragment, found either side of the loxPsym sites. This sequence is matching to the dSynXI.A-C genomic sequence found before and after loxPsyms 18 and 19, respectively.

site, supposedly replacing 2,502 bp of synthetic chromosome sequence between those

two sites with the KanMX fragment by homologous recombination. The second example (Figure 4.11) also shows the same type of integration into the dSynXI.A-C genome, although presumable between two loxPsyms positioned further away from each other than example 1. The amplified DNA of dSynXI.A-C from the inverse PCR for both the forward and reverse primers was traced to the 2nd and 7th loxPsym sites, respectively. If no previous recombination occurred between the two sites prior to the

insertion of the heterologous DNA, a total of 20.9 kb of the synthetic chromosome will have been deleted, representing 25% of the synthetic sequence.

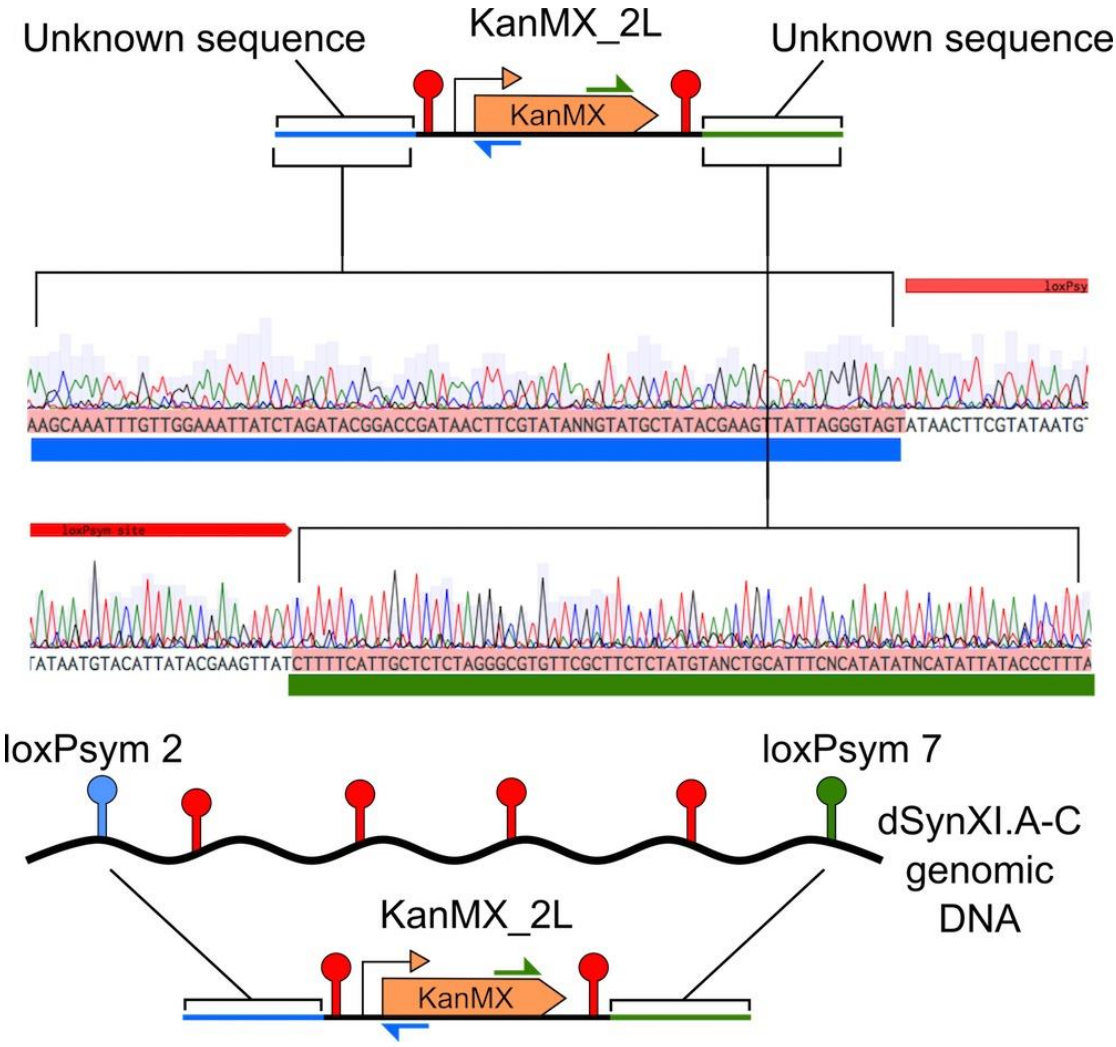


Figure 4.11 iPCR Analysis 2. Inverse PCR and sequencing data showing genomic integration for KanMX_2L following SCRaMbLE-in for two hours, example 2. As shown in Figure 4.10, the flanking sequence of the sequencing traces show that integration has occurred between loxPsyms 2 and 7. Given no other recombination has occurred between these two loxPsym sites, this would suggest a 20.9 kb of synthetic chromosome has been replaced by the KanMX_2L DNA fragment.

Another sample that showed integration of the KanMX_2L into the genome between two loxPsym sites was example four (Figure 4.13), amplifying the upstream and downstream genomic regions of loxPsyms 12 and 18, respectively. However, in contrast to the previous genome-integrating examples, the divergent primers DJ047 and DJ048, based on their directionality, show that the KanMX_2L gene has recombined in an inverted orientation. Furthermore, the synthetic chromosome length found between the 12th and 18th loxPsym sites, is a total of 31.9 kb. Any previous rearrangements that occurred in the genome prior to the integration of KanMX_2L is not known from iPCR amplification. This is an example of greatest spatial-distance between two loxPsym sites undergoing recombination.

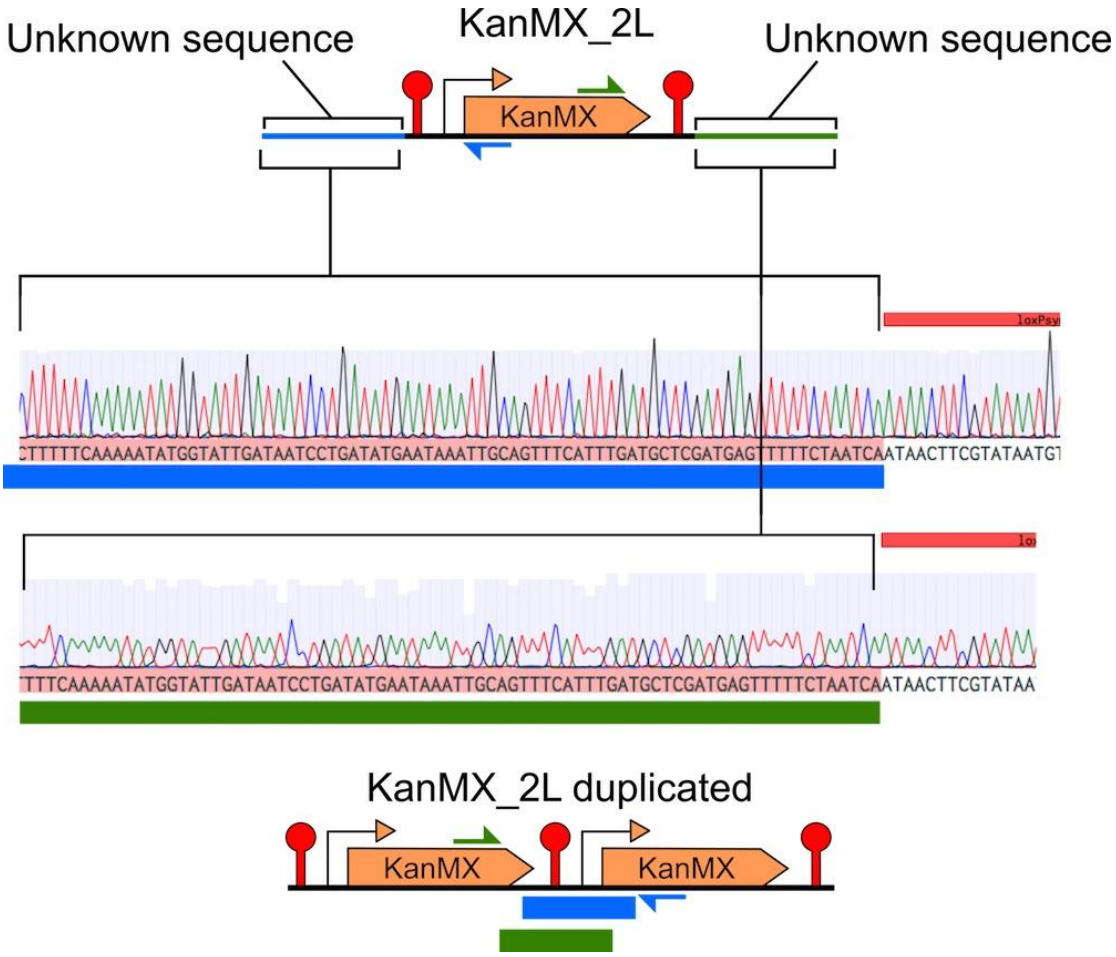


Figure 4.12 iPCR Analysis 3. Inverse PCR and sequencing of a three hour SCRaMbLEd-in KanMX_2L, example 3. Unlike examples 1 and 2, this sequencing trace did not match any genomic DNA of dSynXI.A-C, but rather amplifies two Kanamycin gene sequentially showing recombination has occurred between two KanMX fragments. Although it is not known where in the genome integration has occurred, recombination can be confirmed as the presence of the loxPsym site is not detected.

The iPCR and coupled sequencing data also provided evidence of a different type of recombination event during SCRaMbLE-in. The third example in Figure 4.12 provided a smaller band of approximately 1.8 kb which was consequently gel extracted and sent for sequencing. The resulting trace showed the amplification of the 3'-end of the KanMX_2L gene (close to the stop codon), the loxPsym, and then the KanMX promoter sequence, and no evidence of the genomic integration of this construct. The two possible explanations that can be suggested for this scenario is that either the KanMX_2L have recombined its flanking loxPsym sites to form a circularised plasmid, or two KanMX_2L fragments has recombined in tandem in the dSynXI.A-C chromosome. The scenario more likely to occur is the integration of sequential KanMX_2L fragments into the synthetic chromosome.

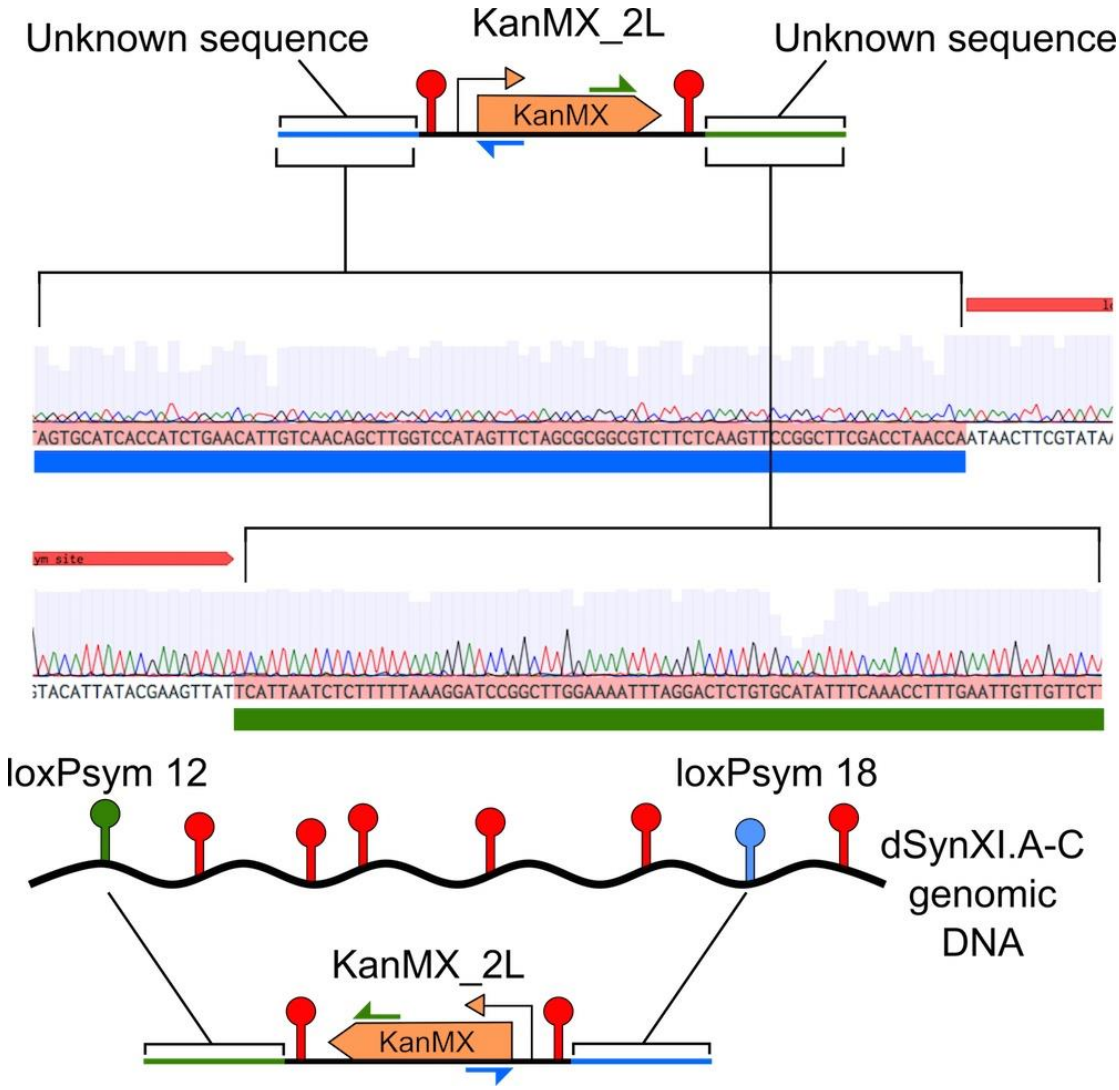


Figure 4.13 iPCR Analysis 4. Inverse PCR and sequencing data for a three hour SCRaMbLEd-in KanMX_2L colony, example 4. The flanking sequences of this sample showed that this gene has been integrated into the genome in an inverted orientation, and amplifies the genomic DNA upstream and downstream of loxPsyms 12 and 18, respectively. This resulted in the largest deletion (assuming no recombination prior to the KanMX_2L fragment recombining in) of 31.9 kb.

Lastly, one of the colonies that was recovered from 350 µg/ml G418s-agar following three hours of SCRaMbLE-in with pDJ018, was assayed using iPCR and produced a band that was successfully gel extracted. I sequenced this sample, and the data (Figure 4.14) shows both primers amplify the unaltered plasmid vector containing the KanMX gene with a single loxPsym site. It can be concluded in confidence that this plasmid had not recombined due to the fact the terminator had not been separated from the KanMX gene's downstream loxPsym site. Given the strains ability to survive and proliferate on the high antibiotic media, it is likely that expression of the KanMX is occurring from elsewhere, possibly from another integrated plasmid.

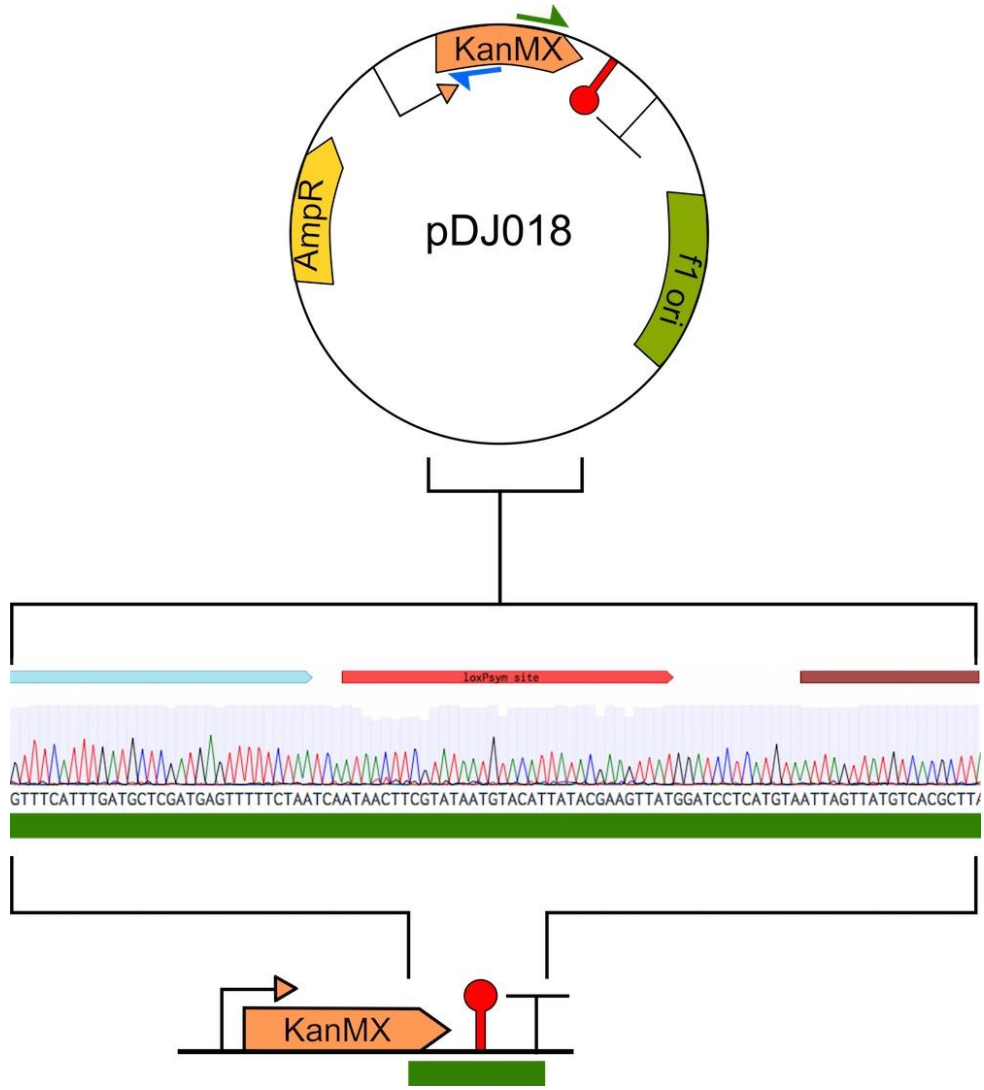


Figure 4.14 iPCR Analysis 5. Inverse PCR and sequencing traces for SCRaMbLEd-in pDJ018, following three hours of induction. A recovered genomic DNA sample analysed for dSynXI.A-C following SCRaMbLE-in with pDJ018, yielded a sequencing trace that showed the presence of the intact plasmid. The forward trace of primer DJ047 is not shown as it amplified the plasmid backbone. The reverse trace, illustrated above as green, shows the presence of the kanamycin gene, loxPsym site, and terminator in the same order as provided. Had SCRaMbLE-in occurred, the tCYC1 provided terminator would not have been downstream of the loxPsym site.

4.3.7 Constructing Fluorescent Strains to Study Effects of SCRaMbLE

Despite developing a SCRaMbLE-in protocol that appears to work for introducing heterologous genes into diploid yeast form circular and linear DNA, it was challenging to determine what rearrangements were occurring on the genome as the iPCR method proved difficult and gave results that were at times complex to interpret. We reasoned, that it would be easier and quicker to gain insight into what happens on the genome during SCRaMbLE and SCRaMbLE-in, if we had marker genes on our chromosome that were easy to detect in terms of whether they had been deleted, inverted or duplicated. Thus, to investigate the effects of SCRaMbLE on gene rearrangements at the genome level, I next used three genes where constitutive promoters drive the expression of different fluorescent proteins with different terminator sequences. The genes encode sfGFP (green fluorescence), mTagBFP (blue fluorescence) and mRuby (red fluorescence). These genes were each designed to contain a loxPsym site three base pairs downstream of their stop codon, upstream of their terminator (i.e. in the standard Sc2.0 placement for loxPsyms). By integrating these genes into a chromosome, we placed loxPsyms in known genomic positions and could determine if these loxPsyms had been recombined by Cre based on changes to the fluorescence of yeast cells. The three colours of fluorescent protein chosen could all be distinguished individually by flow cytometry.

Four YIp-based vectors were assembled under my direction by visiting undergraduate student Robert Chen. These were designed to contain all three fluorescent genes, but in different orders. The DNA assembly produced plasmids pRC1841, pRC1842, pRC1844 and pRC1845, which are shown in Figure 4.15. Each of these plasmids were transformed into the non-synthetic BY4741 yeast strains, and the plasmid DNA was integrated into the URA3 locus to make the four recombinant *S. cerevisiae* strain, yRC1841, yRC1842, yRC1844 and yRC1845. While these strains are not part of the Sc2.0 construction, they contain the three genes in SCRaMbLE format inserted into the genome at a well-characterised site.

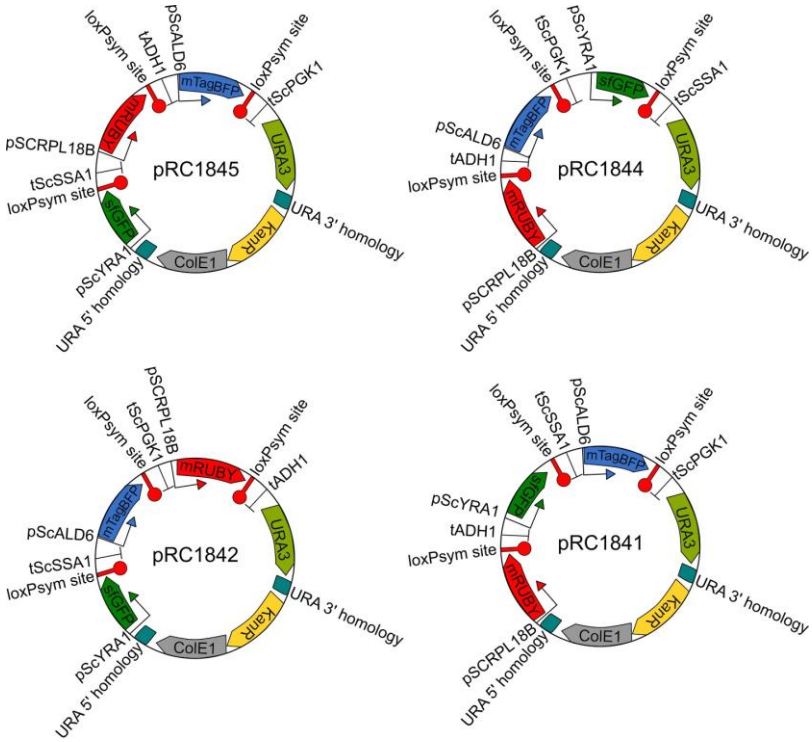


Figure 4.15 Fluorescent Plasmids Schematic. The four integrating plasmids used for fluorescence based studies of SCRaMbLE and SCRaMbLE-in are shown in the four diagrams above. The four plasmids are identical apart from the order of the three fluorescent genes they encased. Promoters and terminators are paired with the fluorescent genes, and are not altered between plasmids. The orders of the sfGFP (G), mRuby (R) and mTagBFP (B) per plasmid are: pRC1845 – GRB, pRC1844 – RBG, pRC1842 – GBR, and pRC1841 is RGB.

Without SCRaMbLE being induced, each fluorescent gene retains its promoter and terminator when on the plasmid and then subsequently when integrated into the genome. Thus the levels of fluorescent expression remain constant and constitutive. Shen *et al* (2015)[82] previously reported that there was no evidence of recombination events between loxPsym sites and the intrinsic BY4741 genome (which lacks loxPsym sequences), thus it is assumed the induced Cre recombinase enzyme will only catalyse recombination events between the provided loxPsym sites, which in these cases are just the three sites confined to the inserted region containing the fluorescent protein genes.

Before assessing SCRaMbLE with these strains, I first considered the likely fluorescence profiles that may arise as outcomes of SCRaMbLE. To be able to predict the genetic rearrangements, and what the subsequent flow cytometry data of the genomically-rearranged yeast will look like, several assumptions were made: (i) in normal SCRaMbLE, the three events that can occur are deletions, inversions and duplications as a result of recombination between two loxPsym sites; and (ii) a gene can only be rearranged once, *i.e.* it can be inverted **or** duplicated, but it cannot be inverted **and** duplicated.

In reality, the SCRaMbLE rearrangements possible are far more complicated, and not restricted to these assumptions, however without complete genome sequencing there is no way of tracking the extensive rearrangements that may have occurred. The assumptions made here are based more on what rearrangements are likely to be detected as fluorescence changes by flow cytometry. Figure 4.16 lists the possible gene rearrangements that can occur as a result of SCRaMbLE for strain yRC1841. Notably, the first gene of every plasmid is left unaltered, as it is not flanked by loxPsym sites like the second and third genes and so cannot be deleted or inverted. However, it could still possibly lose its 3'UTR due to the other genes being inverted. While this logically would result in loss of expression as the 3'UTR encodes the polyA tail for the gene mRNA, the most recent SCRaMbLE research paper[82], reported that the fitness of cells was not affected when the 3'UTR was disrupted, so

		Gene Order		
	No.	1st Gene	2nd Gene	3rd Gene
1841	Control	R	G	B
	1	R	g	B
	2	R	G	b
	3	R	g	b
	4	R	-	B
	5	R	G	-
	6	R	-	-
	7	R	GG	B
	8	R	G	BB
	9	R	GG	BB
	10	R	g	-
	11	R	-	b
	12	R	g	BB
	13	R	GG	b
	14	R	-	BB
	15	R	GG	-

Figure 4.16 Prediction Matrix for SCRaMbLE of yRC1841. In this figure we have outlined possible rearrangements between the loxPsym sites and listed how the genes are affected, using yRC1841 as an example. We assume that there are only three types of events that can occur, deletions (G = -), inversions (G = g) and duplications (G = GG). A total of 15 phenotypes arise as a result, each with an equal likelihood of 6.7% occurring. There are also an equal number of inversions, deletions and duplications occurring of the second and third gene. The same principles of this figure can be applied to the other constructs, yRC1842, yRC1844 and yRC1845, whereby the same probabilities are acquired albeit with different order of genes.

in the prediction table here, the first genes always maintains constant expression.

The simplified list of predicted rearrangements resulted in 15 different phenotypes, each with a 6.7% equal chance of occurrence. Further analysis of the recombination events shows that the likelihood of an inversion, deletion or duplication of the second or third gene is equal at 26.7% each (calculated by four phenotypes whereby a deletion, duplication or inversion has occurred in the second or third gene, out of a total of 15). In all cases these is a 46.7% chance that a phenotype has loss of one or two fluorescent genes (seven possibilities out of 15). The results from SCRaMbLE experiments can therefore be compared to this theoretical prediction table to determine whether the real-life data is more skewed towards a particular type of rearrangement, or whether certain recombination events at the different loxPsym sites happen at higher or lower tendencies than expected.

4.3.8 Fluorescence Analysis of Genomic SCRaMbLE

Before investigating SCRaMbLE-in, I first assessed how SCRaMbLE changed the fluorescence phenotypes of the constructed yeast strains and compared the results to the predictions above. The pCre_EBD plasmid was added to the four constructed strains yRC1841, yRC1842, yRC1844 and yRC1845 as before. These were grown and then SCRaMbLE was induced for three hours by the addition of 1 μM β-oestradiol to the growth media. After three hours, cells were centrifuged for two minutes at 8,000 rpm and washed with YPD (to remove the β-oestradiol and prevent further recombination events). These cells were then plated on YPD plates, and incubated 2 to 3 days and 30°C. An uninduced control was also performed in parallel. Seven colonies were isolated from each SCRaMbLE-induced plate for each construct. These colonies were grown overnight in YPD, and then analysed by flow cytometry using the BD Fortessa X20 Flow Analyser, measuring single cells from each sample for green, red and blue fluorescence (section 2.3.20).

I screened for fluorescence output from all SCRaMbLE induced strains, and compared histogram count peaks to the uninduced control strains. Figure 4.17 shows the mRuby, sfGFP and mTagBFP histogram traces for each un-SCRaMbLEd construct (control), alongside a randomly-selected SCRaMbLEd samples of that construct. All fluorescent histogram traces are shown in the order of which they are found within the constructs (yRC1841: R-G-B; yRC1842: G-B-R; yRC1844: R-B-G; yRC1845: G-R-B), and the colours represent the fluorescing colour (i.e. red, green and blue for mRuby, sfGFP and mTagBFP, respectively). These histogram traces illustrate what it looks like when a change in the fluorescence is detected, which typically means that the histogram moves either left (decrease in fluorescence per

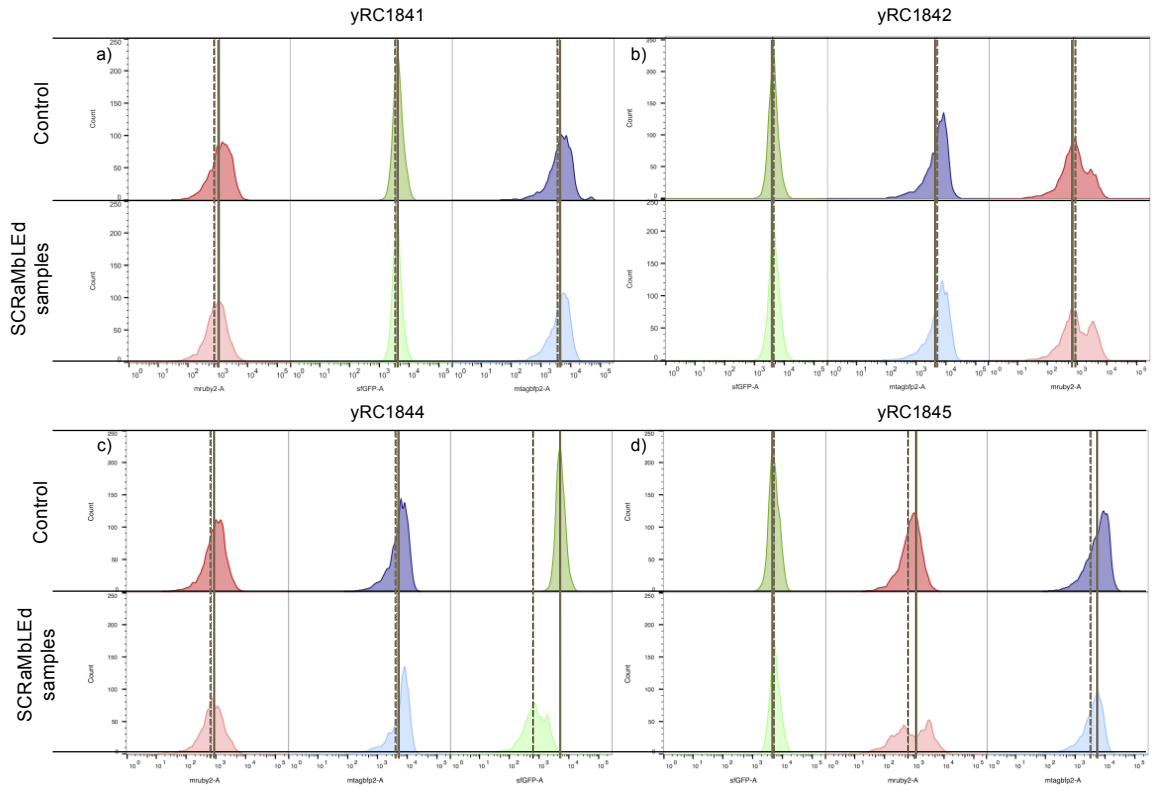


Figure 4.17 Flow Cytometry Analysis for SCRaMbLE. Histogram traces for mRuby, sfGFP and mTagBFP fluorescence. Random examples of each SCRaMbLEd construct yRC1841 (a), yRC1842 (b), yRC1844 (c) and yRC1845 (d) were isolated for peak fluorescence comparison. Each construct control (the uninduced counterpart) is shown in dark red (mRuby), dark blue (mTagBFP), and dark green (sfGFP). The coloured order of histograms' fluorescence profiles of each construct are shown according to the order of fluorescence-producing genes in each of the yRC184N strains (i.e. yRC1841: R-G-B; yRC1842: G-B-R; yRC1844: R-B-G; yRC1845: G-R-B).

cell) or right (increase in fluorescence per cell) compared to its equivalent control histogram.

To assess all the screened colonies, I took the flow cytometry data from all samples run and used FlowJo software to calculate statistical measures for the fluorescence of each sample. As the main measure, I looked at the geometric mean (GM), which is indicated as the grey lines in Figure 4.17. In this sample set the largest change of florescence was observed when SCRaMbLE was induced in yRC1844 where sfGFP fluorescence declines greatly and is effectively completely lost. This is shown by a 92% decrease in the GM of green fluorescence, whereas the first and second genes only exhibit GM decreases of 34% and 22% (mRuby and mTagBFP, respectively).

The further statistical values of coefficient of variance (CV) and mode (peak) were also determined to be able to identify possible bimodal distributions which would be seen as samples with a high CV and a mode that is quite different to the

GM. To visualise the data set as a whole, and to show which SCRaMbLE induced strains exhibited what changes, the statistical data from each sample were compared to one another by using a heat map format (Figure 4.18). Given that all genes contained the same promoters and terminators throughout all constructs, the strains were treated as a single data set. Thus, heat maps of GM and mode were formatted to colour code according to lowest value (white) and highest value (darkest red, green or blue) across the datasets of all constructs for each statistic. The CV has the same format applied, but in grey scale. Figure 4.18a shows the four heat maps generated for strains yRC1841 to rRC1845. The uninduced control for each strain is included in each case, as in the BY4741 background strain. The geometric mean of sample 6 from strain yRC1845 is shaded white as it is the lowest recorded value of mRuby (221 AU) across all the constructs, unlike the highest, which is found in yRC1841 sample 5. Some samples show a high coefficient of variation (CV) suggesting that some bimodality exists within the recorded yeast samples. Ideal results would show a correlated GM and mode, indicating a narrow distribution, further confirmed by a low CV. We see this pattern in the majority of the samples in Figure 4.18a, however there are examples of a low GM but a high mode (second gene, sample 2 of yRC1842, GM is 3925 AU whereas mode is 9533 AU), and this is reflected in the high CV for that sample (77.6%).

Figure 4.18b shows histogram traces of the samples containing the highest and lowest CV values of each gene as an example. As well as being an indication of bimodality, the CV can also point out noisy data, or long-tailed distributions. The process of SCRaMbLE may likely be producing noisy and varied data based on the rearranging of the genes due to the disruption of the 3'UTR or other means. The CV acts as a statistical representation of this.

It must be noted at this stage that the detection of mTagBFP2 (blue fluorescence) by our flow cytometer was not as sensitive as detection of sfGFP or mRuby, so values of GM seen for the positive cells and negative cells were not as separated as I would have liked. Furthermore, mTagBFP2 fluorescence also typically had a high CV due to a long-tail distribution measurements.

Using the geometric mean (GM) colouring intensity of the SCRaMbLE samples as visual comparison to the control GM, the rearrangements of each gene (based on the fluorescence production or lack of production) can be roughly determined without needing to inspect the DNA sequence, allowing me to estimate the likelihood of deletions, inversions and duplications.

Deletions and inversions were differentiated by the presence or absence of fluorescence, in comparison to the low fluorescence seen in each channel when a

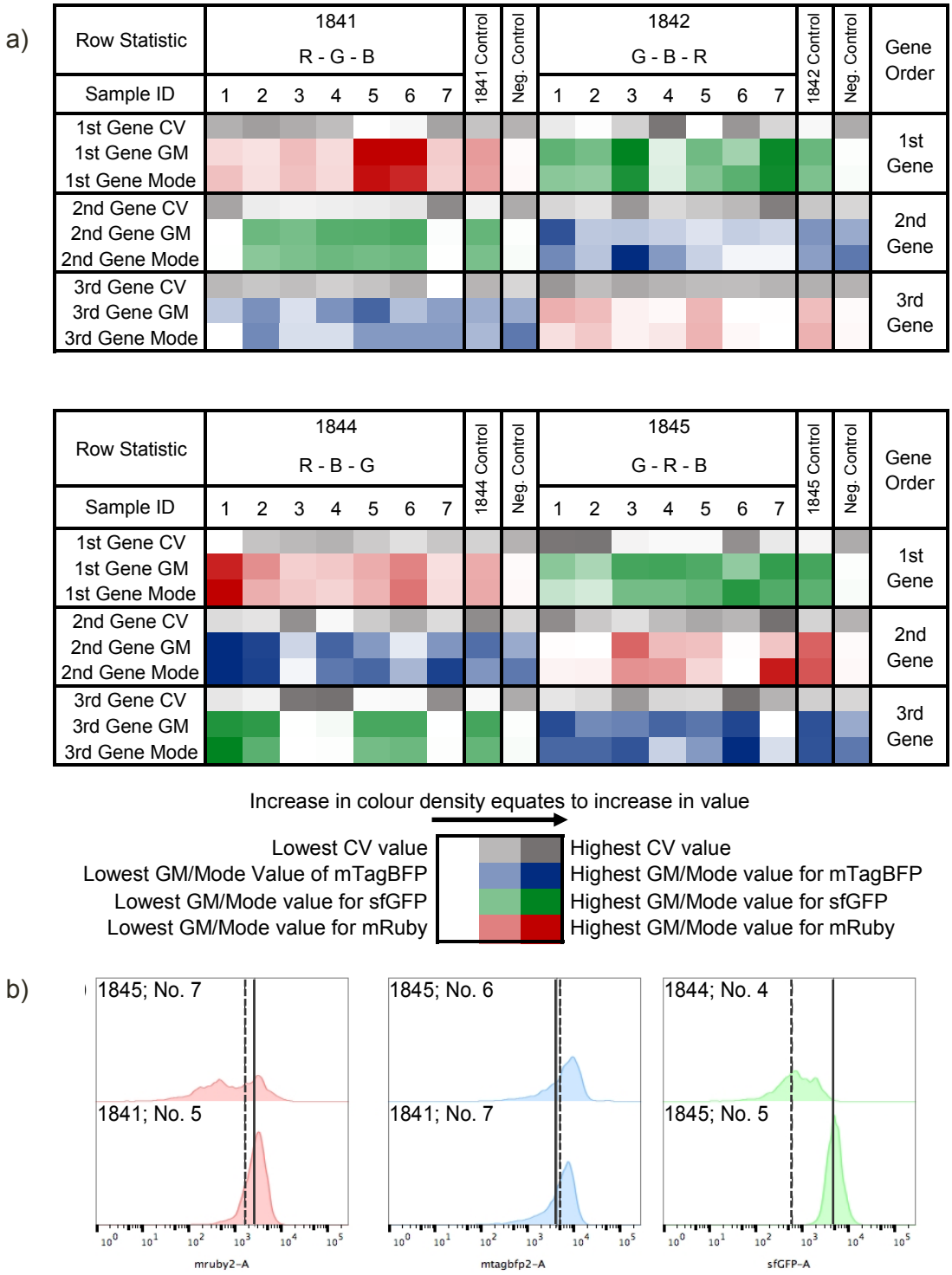


Figure 4.18 Heat map of SCRaMbLE. yRC1841, yRC1842, yRC1844 and yRC1845 fluorescence based on the coefficient of variance (CV), geometric mean (GM) and mode calculated for each fluorescence channel for each sample. In (a) we have applied colour coded formatting based on the values of the CV, GM or mode ranging from the lowest value (shaded in white) to the highest (shaded in the darkest grey, red, green or blue). The formatting was applied using the four constructs as a single data set, with each respective control listed at the end of column for use as a reference. Gene order is reflective of how it is found within the construct. The CV values are based on a grey scale, whereas green, red and blue is used for sfGFP, mRuby and mTagBFP, respectively. Histogram traces (b) were samples used to compare distributions with the highest CVs (top row), versus those with the lowest CVs (bottom row) – the No.s represent the sample ID from (a). Annotated on the traces are the GMs of the top (dashed line) and bottom (solid line) histograms.

negative (BY4741) strain was measured. This low value was considered to be no fluorescent protein expression (i.e. autofluorescence) and values close to this were deletions (loss of protein expression). The heat map data in Figure 4.18a suggest that deletions occurred in the highest frequency, as these are observed in 20 of the 28 cases for the second or third gene in the construct. Of the 20 deletions, 55% occur at the second gene locus, and 45% result in the deletion of the third gene. The highest deletion rate occurred with yRC1842, which also had the greatest number of double deletions (where both the second and third genes are lost). To simplify analysis, I next attributed inversions as the case for samples where fluorescence was detected (GMs significantly greater than the BY4741 control) but it was seemingly below the level seen for the uninduced control. This was further supported by a high CV for these samples and a difference between mode and GM indicating bimodality (and thus potential 3'UTR disruption). The strain yRC1845 holds the highest number of inversion events in samples 4, 5 and 7, isolated exclusively to the second gene. We only see two likely duplication events, in samples 5 and 1 of yRC1841 and yRC1844, respectively. This possibly may be due to a limited SCRaMbLE induction time of three hours. In cases where outcomes are not obvious, such as yRC1842 sample 1 (Figure 4.18a), I referred back to the histogram traces in order to compare these in more detail.

One interesting observation from these data was that the output of expression from the first gene in each construct was not constant in all samples as expected. The first gene itself cannot be inverted, deleted or duplicated due to only having a 3' loxPsym site and no 5' site, so why do some samples show less expression and others considerably more? This finding was a surprise but may be explained by downstream rearrangements to the other two genes that affect the 3'UTR of the first gene and thus alter its expression. More consideration of this is given in the chapter discussion.

Based on the inspection of the heat maps from this experiment and from further investigation of the histogram traces, I condensed the data from Figure 4.18 into a table of likely SCRaMbLE outcomes which is given in Figure 4.19 in the same format as the prediction table in Figure 4.16. From what is seen in this table, deletions appear to happen almost three times more than inversions, despite the use of symmetrical loxPsym sites. Duplications are also occurring and the first gene can have altered gene expression despite only having a downstream loxPsym.

1841			
Sample ID	1st Gene	2nd Gene	3rd Gene
1	r	-	b
2	r	G	B
3	r	G	-
4	r	G	B
5	R(R)	G	BB
6	R(R)	G	b
7	r	-	B

1842			
Sample ID	1st Gene	2nd Gene	3rd Gene
1	G	B	R
2	G	-	R
3	G(G)	b	-
4	g	b	-
5	G	-	R
6	g	-	-
7	G	-	-

1844			
Sample ID	1st Gene	2nd Gene	3rd Gene
1	R(R)	BB	G
2	R	B	G
3	r	-	-
4	r	B	-
5	r	B	G
6	r	-	G
7	r	B	-

1845			
Sample ID	1st Gene	2nd Gene	3rd Gene
1	g	-	B
2	g	-	B
3	G	R	B
4	G	r	B
5	G	r	B
6	g	-	B
7	G	r	-

Figure 4.19 Matrix of Outcomes for SCRaMbLEd Fluorecent Strains. Phenotypes of the SCRaMbLEd yeast (y) strains yRC1841 (R-G-B), yRC1842 (G-B-R), yRC1844 (R-B-G) and yRC1845 (G-R-B). Here I have listed whether genes have undergone an inversion, duplication or deletion, according to the GM value heat maps from Figure 4.18a. In instances where the first gene of every construct has a particularly high GM indicative of increased fluorescence, this has been noted and annotated by encasing an extra copy of that gene in brackets, e.g. “R(R)”. Inversion of genes are shown in small letters (e.g. “b” or “g” or “r”). Deletions are represented by dashes (-), and genes that have not experienced change have been left in their original states. From these four tables yRC1842 undergoes the highest number of individual deletions (of the second or third gene, or both combined), whereas yRC1845 has the highest number of inversions. yRC1841 and yRC1844 both contain a single duplication.

4.3.9 Fluorescence Analysis of the Effects of SCRaMbLE-in

Having established a non-sequencing method for investigating genomic SCRaMbLE events by following the fluorescence profiles of constructed yeast strains after the induction of SCRaMbLE, I next looked to see what this could tell us about SCRaMbLE-in. Previously I had determined that the most efficient method of SCRaMbLE-in was to use a linear DNA fragment with two flanking loxPsym sites, and induced SCRaMbLE-in strains transformed with this for one, two or three hours. Therefore here I once again performed SCRaMbLE-in with the KanMX genes with TEF promoter and with two loxPsym sites on a linear DNA piece (KanMX_2L). But this time I used strains yRC1841, yRC1842, yRC1844 and yRC1845 as the host. In contrast to the SCRaMbLE experiments with these strains, in these experiments I recovered the cells on YPD agar with 200 µg/ml G418s to specifically select only for the colonies where KanMX has been inserted.

The SCRaMbLE-in protocol as outlined in Figure 4.2 was followed but for the yRC strains. Following this, randomly selected colonies were analysed using the same methods as above for the previous SCRaMbLE experiments, and the same gates were applied to the flow cytometry data (section 2.3.20).

The same methods and assumptions for analysing the flow cytometry data were applied as before. However, it is important to mention that integration of a linear DNA fragment flanked by two loxPsym sites requires rearrangements at two pairs of loxPsym sites unlike the usual single rearrangement at one pair of sites (e.g. for a deletion).

Example fluorescence histograms traces from four randomly selected G418s-resistant colonies analysed on the flow cytometer are shown in Figure 4.20, as well as the corresponding histograms for the non-SCRaMbLE control for each strain (these controls are the same that were used in Figure 4.17). The histogram trace results showed that the first gene (mRuby) of SCRaMbLEd yRC1841 does not experience a notable change of its fluorescence peak (Figure 4.20a), whereas the second gene shows reduction in fluorescence, seen in the leftward shift of its fluorescence distribution, decreasing from a GM of 4294 AU to 1765 AU. This could be a deletion, however as there is still significant fluorescence it is probably another rearrangement: either the gene has been inverted or it has experienced an unfavourable disruption of its 3'UTR that has led to the decrease of green fluorescence detected. The final mTagBFP gene does not show much change, and the GMs are overlapping. Looking at the next example in Figure 4.20, yRC1842 shows a marginal increase of first gene (sfGFP) mode and GM post-SCRaMbLE-in, relative to the control. A possible explanation could be that the terminator of the first

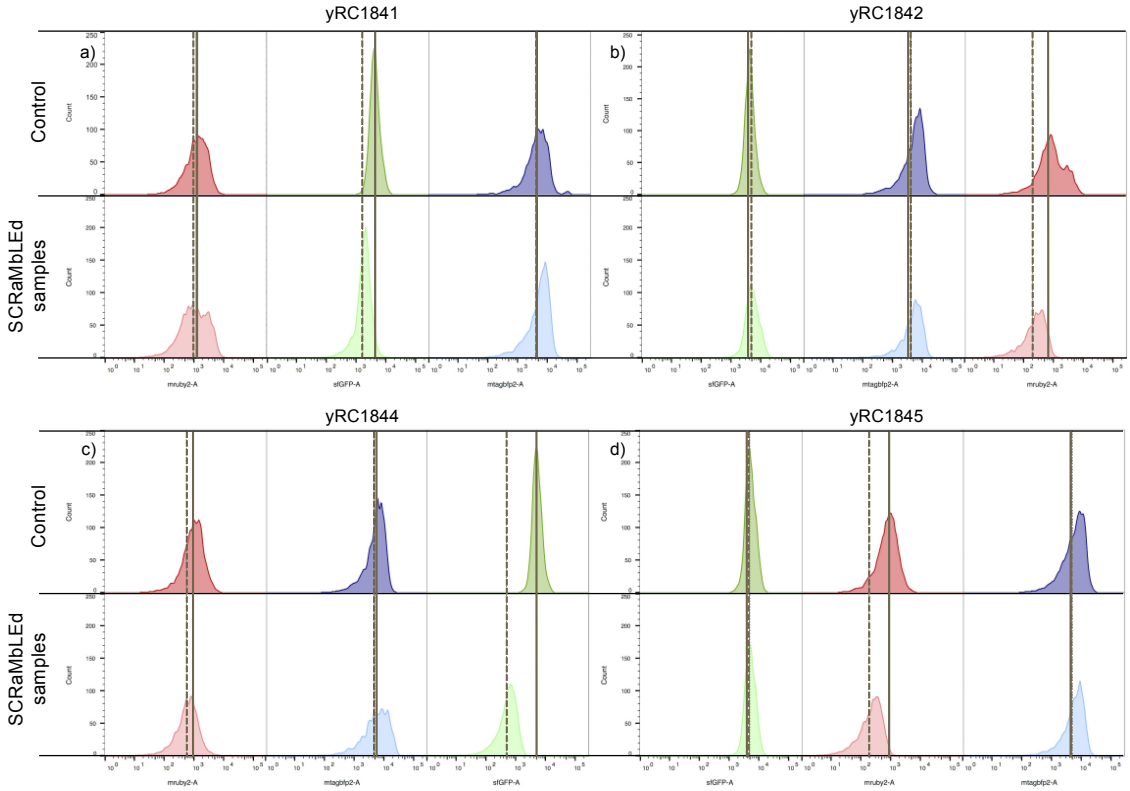


Figure 4.20 Flow cytometry analysis of SCRaMbLE-in. A random selection representing each strain, yRC1841 (a), yRC1842 (b), yRC1844 (c) and yRC1845 (d) was isolated and histogram traces for each fluorescent gene shown. The top histograms of a-d represent the fluorescence distributions for the uninduced (control) strains, of which each respective GM is shown as a solid grey line. Bottom histograms are traces for samples that have been induced for SCRaMbLE-in, and their GMs are shown as dashed lines. Shifts in the histograms that indicate loss of fluorescence can be seen in the last two genes of (b) and (c), and the second gene of (d). In all three cases the GMs drop to basal levels that are reflective of the negative control values. In strain yRC1841 (a), the sfGFP gene in the second position also shows a decrease in GM, as well as a marginal shift to the left of the histogram trace, however it is still a fluorescing population, therefore we assume that in such instances the gene has undergone an inversion which could offer an explanation for the decrease in fluorescence. The coloured order of histograms' fluorescence profiles of each construct are shown according to the order of fluorescence-producing genes in each of the yRC184N strains (i.e. yRC1841: R-G-B; yRC1842: G-B-R; yRC1844: R-B-G; yRC1845: G-R-B).

gene may have been rearranged by SCRaMbLE leading to the increase in green fluorescence. The mTagBFP histogram is unchanged once more, whereas mRuby fluorescence has dropped down to autofluorescence levels, suggesting it has been deleted.

Based on the histogram shift and decreased GM values for yRC1842, yRC1844 and yRC1845 in Figure 4.20, it appears that the third, third and second genes, respectively, have been deleted and in exchange the KanMX marker has SCRaMbLEd-in. In the case of yRC1841, the GM value is only decreased for the second gene (sfGFP) but not to the level associated with deletion. This suggests that

KanMX insertion has occurred in a manner that disrupts (rather than deletes) the sfGFP gene (e.g. by removing its 3'UTR).

As in the previous section, I next produced heat maps to plot the intensity of fluorescence of mRuby, sfGFP and mTagBFP based on the GM, mode and CV values of all colonies that arose on the post-SCRaMbLE-in G418s plates (Figure 4.21 to 4.24). These figures firstly show that the number of surviving colonies is directly correlated to the induction times for SCRaMbLE, as in almost all instances

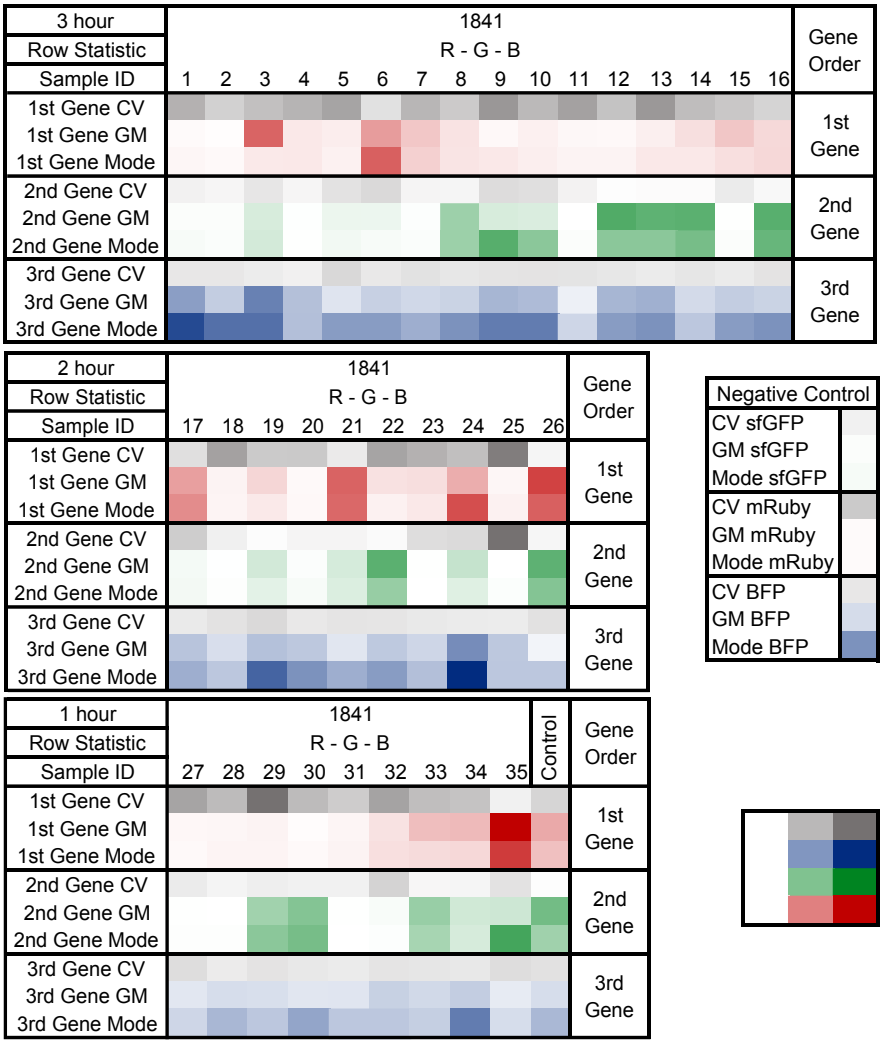


Figure 4.21 Heat Map of SCRaMbLE-in for yRC1841 yeast strain following 3, 2 and 1 hour of SCRaMbLE-in. A total of 35 colonies exhibited growth on the G418s agar, and the CV, GM and mode of each gene is shown in the diagram above. The control is referenced as the final square of the last set of results. The highest occurring event is deletions, a total of 19 are seen in (15) sfGFP and (4) mTagBFP. Where we see decreased fluorescence, but still with evidence of a fluorescing GM or mode, we consider these as inversions, which we see in 15 instances, of which 67% are of the second, sfGFP, gene. There is a single duplication of the second and third genes in sample ID's 12 and 24, respectively. Changes of the fluorescence observed in the first gene are discussed further on in the chapter; in short lack of first gene fluorescence is likely to arise from unwanted recombination events or 3'UTR disruption.

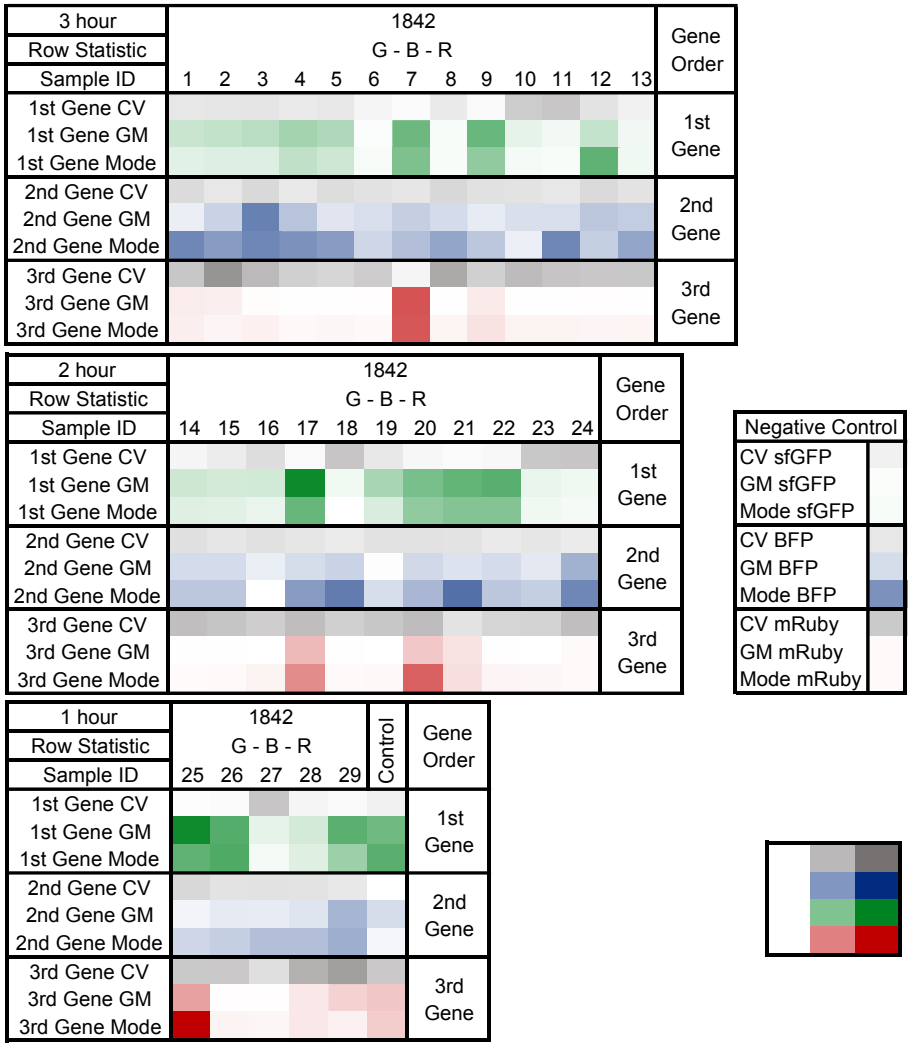


Figure 4.22 Heat Map of SCRaMbLE-in for yRC1842 following 3, 2 and 1 hour of SCRaMbLE-in. A total of 29 colonies exhibited growth on G418s agar, and the CV, GM and mode of each gene is shown in the diagram above. The control is shown after the final sample. This strain contains the highest number of gene deletions of its second or third gene, with a total of 43% showing complete loss of fluorescence. However unlike the previous strain where the majority of rearrangements occurred in the second gene position, yRC1842 has 96% of deletions in the third gene locus. The second gene has a high percentage of inversions, with 82% of all inversions seen in mTagBFP. There are two notable duplications, in sample 3 of the mTagBFP gene and in sample 7 of the mRuby gene. Both GMs increase by 0.5 and 2.5 times greater than the control.

(apart from yRC1844) as the induction time is increased, so are the number of surviving colonies. This suggests that the likelihood of the KanMX_2L being integrated into the yRC strain genome (via rearrangements at two loxPsym sites) increases with longer induction times. Another general observation is that increased induction times also increase the number of deletions, where significant fluorescence is lost from the strains. It is possible that the two observations are directly linked to each other, so that as KanMX_2L integrations increases, Cre-mediated recombination results in the deletion of a fluorescent gene. Alongside the expected

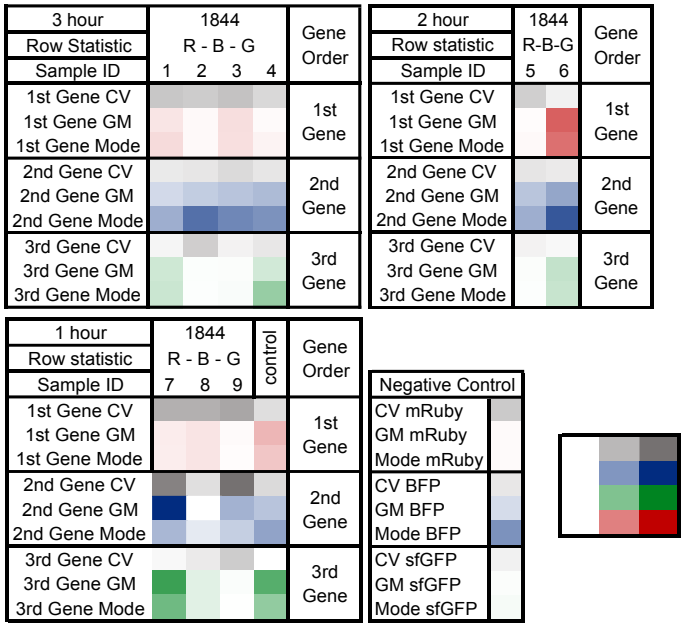


Figure 4.23 Heat Map of SCRaMbLE-in for yRC1844 following 3, 2 and 1 hour of SCRaMbLE-in. A total of 9 colonies exhibited growth on G418s agar, and the CV, GM and mode of each gene is shown in the diagram above. Valid conclusions based on this strain are difficult to make as there are a significantly lower number of surviving colonies when compared to the other three yeast host strains. There are a total of five deletions, four inversions and a single duplication observed. In eight instances of the nine samples, the third gene has undergone a rearrangement, whereas the mTagBFP gene shows one duplication (sample 7), and one deletion (sample 8).

changes in the expression of genes two and three in each strain (which are flanked by loxPsym sites), it is notable that in most samples we also see a significant decrease in expression from the first gene. While this was not initially expected, several explanations for this are possible. In most cases the fluorescence of the first gene decreases but is not completely reduced to autofluorescence levels. This may be due to SCRaMbLE-in adding the KanMX marker into the 3' loxPsym of the first gene in a way that removes its 3'UTR, preventing correct mRNA expression (as appears to be the case for pRC1841 in Figure 4.21). In cases where the fluorescence of the first gene disappears all together (where the mode and the GM are both as low as the BY4741 negative control), one obvious explanation is complete loss of the entire construct from the yeast cell, for example by non-SCRaMbLE recombination. This appears to be the case for sample 11 of yRC1841 which shows autofluorescence levels of mRuby, sfGFP and mTagBFP. In this case the KanMX_2L gene may not have recombined in between two loxPsym sites as predicted, but instead produced sufficient expression from integration and maintenance elsewhere in the cell's genome. With no subsequent pressure for the yeast to maintain the red-green-blue construct in its genome at the URA3 locus (uracil selection is not applied post-SCRaMbLE), the construct could theoretically

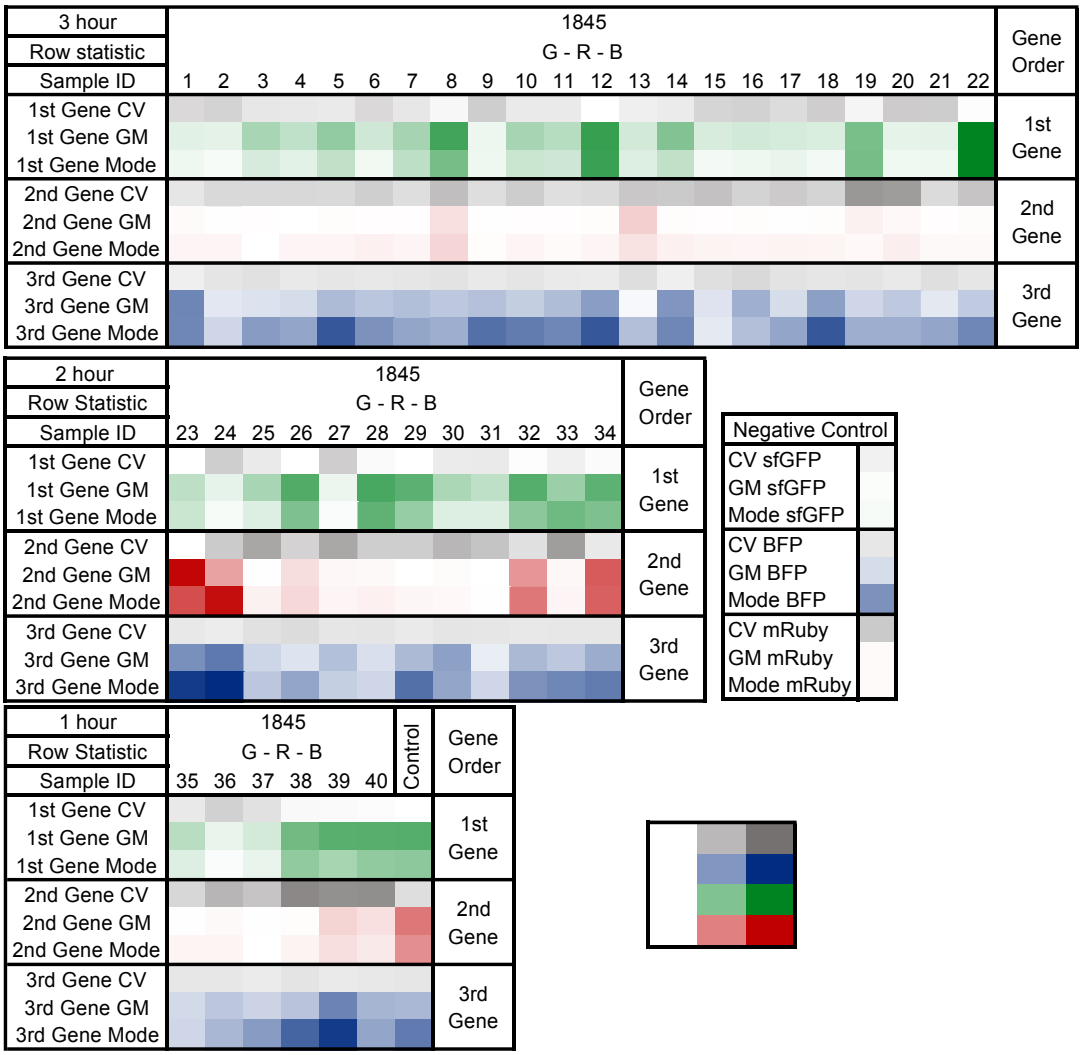


Figure 4.24 Heat Map of SCRaMbLE-in for yRC1845 following 3, 2 and 1 hour of SCRaMbLE-in. A total of 40 colonies were isolated, 22 from three hours of induction, 12 following two hour induction, and six colonies from one hour induction period. This strain results in the biggest data set of all four constructs. There are 34 deletions, of which the majority (91%) occur in loss of mRuby. There are also a high number of inversion events, of which the majority recorded are observed in the mTagBFP gene (81%). A single duplication is seen in sample ID 23 (2 hours), where the GM increases from 1653 AU (control, mRuby GM), to 2894 AU (mRuby GM, mode is valued at 3108).

have been lost from the genome entirely while the KanMX had somehow integrated elsewhere.

The interpretation of the phenotypic outcomes based on the heat map data for all four strains and all time-points are summarised in the table of Figure 4.25. A plethora of different phenotypes arise from SCRaMbLE-in in the four strains. Even with just three loxPsyms in each of the strains, there is a substantial amount of diversity in the subsequent fluorescence output, demonstrating how SCRaMbLE can dramatically alter and diversify genomic gene expression just with a few hours of induction. Interpreting the phenotypes is challenging, especially when there is only a small difference between positive and negative fluorescence for mTagBFP

measurements, and also when there is the unanticipated complication of the first gene from each construct changing expression levels despite only having a single 3' loxPsym.

To further understand the outcomes shown in Figure 4.25, results were tallied-up to see whether a particular combination of rearrangements occurred with the most frequency. Across all strains, there are 88 rearrangements that occur in the

yRC1841

Time	Sample ID	1st Gene	2nd Gene	3rd Gene
3 HOURS	1	-	-	BB
	2	-	-	B
	3	R	g	BB
	4	r	-	-
	5	r	g	b
	6	R	g	b
	7	R	-	-
	8	r	G	b
	9	-	G	B
	10	r	g	B
	11	-	-	-
	12	-	G	B
	13	r	G	B
	14	r	G	-
	15	R	-	b
	16	R	G	b
2 HOURS	17	R	-	b
	18	-	-	-
	19	R	g	B
	20	-	-	B
	21	RR	g	-
	22	r	G	b
	23	r	-	-
	24	R	g	BB
	25	-	-	-
	26	RR	G	-
1 HOUR	27	-	-	-
	28	-	-	-
	29	r	G	-
	30	-	G	b
	31	r	-	-
	32	r	-	-
	33	R	G	-
	34	R	g	B
	35	RR	G	-

yRC1842

Time	Sample ID	1st Gene	2nd Gene	3rd Gene
3 HOURS	1	g	b	r
	2	g	b	r
	3	g	B	-
	4	g	B	-
	5	g	b	-
	6	-	-	-
	7	G	b	RR
	8	-	-	-
	9	G	-	r
	10	g	-	-
	11	g	B	-
	12	G	-	-
	13	g	b	-
2 HOURS	14	g	-	-
	15	g	-	-
	16	g	-	-
	17	GG	b	R
	18	g	B	-
	19	g	-	-
	20	G	b	R
	21	G	B	R
	22	G	-	-
	23	g	-	-
	24	g	B	-
1 HOUR	25	GG	-	RR
	26	g	-	-
	27	g	-	-
	28	g	-	r
	29	G	B	r

yRC1844

Time	Sample ID	1st Gene	2nd Gene	3rd Gene
3 HOURS	1	r	-	g
	2	-	B	-
	3	r	b	-
	4	-	b	g
2 HR	5	-	b	-
	6	RR	B	g
1 HR	7	r	BB	G
	8	r	-	g
	9	-	-	-

yRC1845

Time	Sample ID	1st Gene	2nd Gene	3rd Gene
3 HOURS	1	g	-	B
	2	g	-	-
	3	g	-	-
	4	g	-	-
	5	G	-	B
	6	g	-	b
	7	g	-	b
	8	G	r	-
	9	g	-	B
	10	g	-	b
	11	g	-	b
	12	G	-	B
	13	g	r	-
	14	G	-	B
	15	g	-	-
	16	g	-	-
	17	g	-	-
	18	g	-	B
	19	G	r	-
	20	g	-	-
	21	g	-	-
	22	GG	-	B
2 HOURS	23	g	RR	BB
	24	g	R	BB
	25	g	-	-
	26	G	r	-
	27	g	-	b
	28	G	-	-
	29	G	-	B
	30	g	-	b
	31	g	-	-
	32	G	R	b
	33	G	-	b
	34	G	R	B
1 HOUR	35	g	-	-
	36	g	-	-
	37	g	-	-
	38	G	-	B
	39	G	r	B
	40	G	r	b

Figure 4.25 Matrix of Outcomes for SCRaMbLE-in. A summary of the recorded phenotypic changes following SCRaMbLE-in of KanMX_2L into yRC1841, yRC1842, yRC1844 and yRC1845 fluorescent strains. Here I have listed whether genes have undergone an inversion, duplication or deletion, according to the GM value heat maps from Figure 4.21 to 4.24. In instances where the first gene of every construct has a particularly high GM indicative of increased fluorescence, this has been noted and annotated by encasing an extra copy of that gene in brackets, e.g. R(R). Inversion of genes are shaded the lightest blue, green or red and duplications are coloured in darker shades. Deletions are represented by dashes (-), and genes that have not experienced change have been left in their original states.

second gene, and 90 that occur in the third gene out of a total of 176 rearrangements that result in a changed fluorescence of the second or third gene. The second gene experiences 70.5% deletions, 27.2% decreased fluorescence, and 2.3% of the recorded samples show increases in fluorescence. Similarly, in the third gene we observe 63.3% deletions, 28.9% decreased fluorescence and 7.8% increased fluorescence. Given the limited sample size analysed here, conclusions are not able to be made from these numbers about the events at the second versus third genes, despite the deletion of fluorescence being prevalent over the decrease or increase of fluorescence. Increases in fluorescence are classified as duplications, which occur in the least numbers as expected as these are events that require the cell to be replicating their chromosomes at the same time (so that two copies are present in the nucleus).

From the interpretation of the SCRaMbLE outcomes shown in Figure 4.25, I next tried to determine where KanMX_2L had likely been inserted in to the genome of each sample. For this I assumed the KanMX required insertion by loxPsym recombination at both ends of its linear DNA, meaning that two Cre-mediated recombination events needed to occur in all cases (i.e. it was assumed that no events with three or more recombinations occurred). It was not attempted to interpret events with duplications leading to much greater fluorescence as these are likely to be more complex recombinations due to more than one copy of the construct present in the nucleus. Attributions based on the decreases in fluorescence (rather than loss) were either due to inversions or insertion of KanMX in a manner that disrupts or removes the 3'UTR from the upstream gene, preventing its proper expression. Events occurring at the loxPsym sites are summarised in Figure 4.26, which shows the predictions of where the KanMX gene has integrated into the genome based on the phenotypes of Figure 4.25. Tables 4.1 and 4.2 show the number of rearrangements that have occurred with respect to loxPsym sites, as well as the total number of rearrangements based on type that I have interpreted from this data. There does not seem to be a significant difference between the total number of events occurring at the three loxPsym sites (Table 4.1), however there is a difference in the type of rearrangements occurring, with the most prevalent being replacements. Replacements were classified as the KanMX gene recombining in between two loxPsym sites and replacing the gene originally found there. Insertions are where the KanMX has inserted into a single loxPsym site. Given the presence of two, flanking loxPsyms, it is not surprising that replacements happen three-fold more than insertions.

Deletion of the second gene of fluorescent constructs indicates that recombination has occurred between the first and the second loxPsyms, thus replacing the second gene with the KanMX gene, observed in samples such as yRC1841 numbers 1, 2 and 20 (Figure 4.25). The 3'UTR of the first gene is also disrupted which is why fluorescence from the first gene is also seemingly lost. This phenotype is classified as a replacement between the first and the second loxPsym sites. The other two replacements that can occur are between the second and the third loxPsyms (thus resulting in the loss of fluorescence produced by the third gene) and the first and the third loxPsyms (resulting in no fluorescence detected from either the second or third gene). Examples of these can be seen from samples yRC1841 sample 33 and yRC1842 sample 14, for second and third gene replacements and first and third gene replacements, respectively.

Phenotypes that show decreases in fluorescence but not the loss of a gene are considered to be insertions of the KanMX gene at loxPsym sites one, two or three and are seen in a number of samples such as yRC1841 numbers 8, 16, and 19. A few samples show complete loss of the integrant, suggesting that i) KanMX has integrated elsewhere in the genome, off-target of the loxPsym sites provided and ii) that the fluorescent integrant has been lost from the sampled strain. Overall, seven phenotypes arose whose phenotypes could not be determined, thus were categorised as undetermined (such as sample 2 of yRC1844).

While the information in Figures 4.25 and 4.26 provide valuable insights into what is occurring during SCRaMbLE-in, they do not for certain indicate where and how KanMX has been integrated into the genome. Further development of the use and analysis of these fluorescent strains will likely be needed to shed light upon how SCRaMbLE and SCRaMbLE-in works at the genome level. While SCRaMbLE is likely to be a very powerful tool for strain evolution and SCRaMbLE-in further adds heterologous genes into this, their outcomes are a challenge to interpret due to the complexity of multiple recombination. The fluorescence approach I developed here attempts to provide a non-genome sequencing method to obtain information on the genotype outcomes (via phenotype analysis). However, without parallel verification of the outcomes by KanMX inverse PCR or better yet full genome sequencing it is difficult to conclude if this approach works.

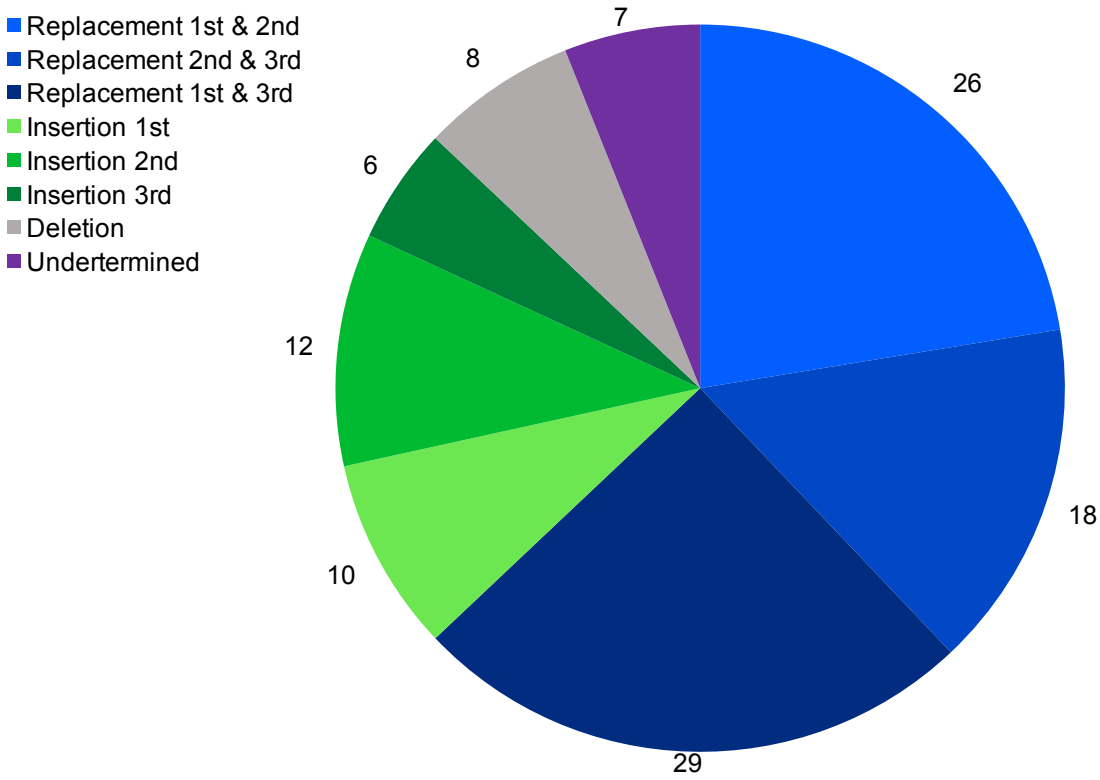


Figure 4.26 Pie Chart Data of Rearrangement Types showing rearrangement types and frequencies at specific loxPsym sites of all constructs. From Figure 4.25 interpretations have been made to enable predictions about where KanMX_2L has inserted. Replacements (shown in blue) represent the insertion of KanMX_2L between two loxPsyms that has resulted in the deletion of a gene, and these can occur between the first and second loxPsym sites (Replacements 1st and 2nd), between the second and third loxPsyms (Replacements 2nd and 3rd) or between the first and the third loxPsyms (Replacements 1st and 3rd). It is difficult to differentiate between inversions and 3'UTR disruptions, therefore these have been classified as inversions, wherein the KanMX_2L shows evidence of inserting at the first (Insertion 1st), second (Insertion 2nd) or third (Insertion 3rd) loxPsym site. Deletions (grey) represent examples where no fluorescence is detected at all, suggesting the integrant has been lost. Undetermined samples could not be interpreted (purple)

Table 4.1 Total Number of Rearrangements at LoxPsym sites of yRC constructs

Construct	loxPsym 1	loxPsym 2	loxPsym 3	Total
yRC1841	14	16	12	42
yRC1842	13	12	19	44
yRC1844	3	3	2	8
yRC1845	32	22	20	74
Total	62	53	53	**336**

Table 4.2 Total Number of Rearrangements

Type	yRC1841	yRC1842	yRC1844	yRC1845	Total
Insertion	10	4	2	6	22
Replacement	16	20	3	34	73
Deletion	5	2	1	0	8
Undetermined	4	3	3	0	10
Total	35	29	9	40	**226**

4.4 Discussion

4.4.1 The Utility of SCRaMbLE-in

In this Chapter, a new method called 'SCRaMbLE-in' for automatically integrating heterologous genes into *S. cerevisiae* during synthetic yeast chromosome rearrangements was developed. It was hypothesised that with the correct formatting of heterologous genes, the induction of the Cre-lox recombination system in the synthetic yeast could lead to large genomic rearrangements such as deletions, inversion and duplications, and also lead to the insertion of genes, especially those offering an environmental advantage. From the data presented in this chapter, the hypothesis appears to be confirmed, as heterologous genes formatted with loxPsym sites are incorporated into the genome upon induction with SCRaMbLE. The engineering of this new SCRaMbLE-in method led to the laboratory-based evolutionary adaptation of the partially-synthetic diploid yeast strain to incorporate the beneficial KanMX gene into the synthetic chromosome, and thus ensure survival of the cell on G418s YPD agar.

It was demonstrated that the SCRaMbLE-in method could be applied to a single gene where the flanking loxPsym sites are optimally placed upstream and downstream of the promoter and gene, respectively. It was also shown that longer induction times resulted in greater numbers of G418s-resistant colonies. The process of SCRaMbLE and SCRaMbLE-in are still in the premature stages of characterisation and optimisation, however a proof of concept was demonstrated here, which will be invaluable to the future SCRaMbLE and SCRaMbLE-in development for Sc2.0 teams and others using the Sc2.0 strains. SCRaMbLE was built into the synthetic chromosome design in order to allow for "subsequent evolution of a minimal genome…that can be controlled by the experimentalist at will" (synthetic yeast website; syntheticyeast.org/designs/principles). Following completion of all the 16 synthetic chromosomes and the neo-chromosome, researchers will combine the 16 strains into one, to create the first eukaryotic synthetic genome. With 16 synthetic chromosomes in a single strain the induction of SCRaMbLE will have unprecedented effects on the gene rearrangements of Sc2.0. Furthermore, in the endeavour to evolve a minimal genome with particular functions, researchers may be able to also implement the SCRaMbLE-in system developed here to adapt the synthetic strain under environmental pressures to express heterologous genes whilst retaining or co-generating a minimal genome.

4.4.2 Developing and Optimising the SCRaMbLE-in Method

The results obtained from the experiments showed that the average highest colony yield was obtained from formatting the KanMX resistance gene with flanking loxPsym sites on a linear piece of DNA, and transforming it into the diploid strain containing partially-synthetic chromosome XI strain. Although data was not significantly different for the three different induction times of SCRaMbLE-in, there was a direct correlation between the number of colonies generated and the induction period of SCRaMbLE-in, therefore it was concluded that the optimal length of induction was three hours, based on the data from Figure 4.9. In theory, longer induction times could lead to higher numbers of colonies.

However, initial SCRaMbLE-in experiments were carried out in the haploid partially-synthetic yeast strain, SynXI.A-C and were unsuccessful. This is likely due to essential gene loss occurring in combination with the environmental pressure to grow on G418s YPD agar. As a result, very few healthy colonies arose and the overall results were inconclusive. In particular SCRaMbLE-in of the KanMX linear fragments (KanMX, KanMX_1L and KanMX_2L) resulted in large numbers of sick colonies that did not exhibit the same fitness or phenotype as the SynXI.A-C strain (or BY4741), showing small colonies size, and unexplained phenotypes (not shown).

A reoccurring observation of SCRaMbLE-in using the SynXI.A-C haploid strain was that a large percentage of colonies that were false positives, surviving on 200 µg/ml G418s agar, but failing to grow on the higher concentration of 350 µg/ml G418s agar (Figure 4.4). I believe there was a link between the strange phenotypes of the recovered colonies from the first 200 µg/ml G418s-agar and the ability of colonies to survive on this media. Firstly, a large number of colonies were very small in size – this suggested that the genome had undergone deleterious rearrangements, most likely of essential genes, resulting in stunted and sickly growth. Secondly, although the negative control (SynXI.A-C transformed and SCRaMbLE uninduced) did not show growth on the 200 µg/ml G418s agar, the KanMX SCRaMbLE-in samples seem to reflect cell growth arrest, but not death, as is common when using the Geneticin antibiotic. Lastly, a number of the colonies that arose were column shaped and upward-growing suggesting that the surviving cells (farthest away from the plate) were surviving and proliferating off the nutrients provided by the dead/arrested cells, closest to the plate (and G418s). This theory was further confirmed by the restreaking of these particular colonies on the higher 350 µg/ml G418s-agar, and the lack of growth observed on this media type.

A side experiment was carried out to test the death rate of the haploid strain which showed that there were large differences in the surviving numbers of colonies

when SCRaMbLE was switched on (for one, two and three hours; data not shown). Overall, SCRaMbLE induction on haploid SynXI.A-C caused a 2.5-times, 3-times and 10.5-times decrease in surviving colony numbers when induced for one, two and three hours, respectively. If we factor in the transformation process (whereby only a small fraction of cells will transfect the DNA into the nucleus[153]) combined with the pressure to integrate the KanMX marker (required for survival on G418s) as well as the high death rates from SCRaMbLE (presumably due to essential gene loss), the likelihood of generating SCRaMbLE-in G418s-positive cells is greatly reduced. The high levels of cell lethality and sick phenotypes observed in the haploid strain were assumed to be a consequence of essential gene loss. To compensate for this, the diploid strain dSynXI.A-C was generated, in order to reduce cell death by providing a copy of the genome immune to the effects of SCRaMbLE (consequently, death rate experiments on the diploid strain showed no differences in cell numbers between the induced and uninduced states). At the same time, recent results from Shen *et al* (2015)[82] also showed that there was no evidence of the Cre recombinase catalysing ectopic or off-target recombination in the BY4741 strain, and that this strain was not affected by the presence of the Cre enzyme. Therefore, we can be sure that the non-synthetic chromosomes in the diploid are not affected by SCRaMbLE.

One particular observation made during the development of SCRaMbLE-in was the importance of including a recovery incubation step following induction of the cell mix with β-oestradiol prior to applying an antibiotic selection pressure. Preliminary experiments during method development that lacked a one-hour recovery step resulted in zero viable colonies (in both the haploid and diploid SynXI.A-C versions: data not shown). It wasn't ascertained exactly why this recovery step was so vital for the survival of the cells (especially the diploid strain), but it is presumed to be mainly due to cell shock from the Lithium-Acetate (LiOAc) transformation process to transport the heterologous DNA intracellularly. The heat-shock stage of the LiOAc-based transformation protocol is used to increase cell wall permeability, and encourage endocytosic uptake of the provided DNA[142]. However, it is also well known that the heat-shock stage (usually carried out at 42°C) is shock-inducing to the cells, and some high-efficiency transformation protocols have been adapted to include a recovery period following the heat shock stage. For the purposes of this study, and optimising the sequential events and experiments, I settled on using the standard transformation protocol as described by Gietz *et al* (2007)[153].

SCRaMbLE recombines the DNA between two loxPsym sites, but the cells seemingly need time to establish enough expression from these newly provided

genes. Therefore, G418s antibiotic was spread top-down onto the agar plates, one hour after the cells were plated on the YPD agar. Striking the balance between recovery times was important, and so 60 minutes was selected based on average yeast replicating times (90 minutes on YPD, although probably even slower following a transformation and SCRaMbLE), with the thinking that after this time the surviving cells would be entering the S phase of replication and a bud will be starting to form. Elongated recovery periods (more than 90 minutes) could lead to the increase of the KanMX gene(s) being lost by DNA degradation or dilution if no selective pressure is applied on time. It would be interesting to test in the future whether longer recovery periods such as 90, 120 and 180 minutes result in greater or lower number of G418s resistant colonies, or indeed if a recovery period could be integrated into the transformation process, enabling plating of cells directly onto G418s-containing YPD plates.

The diploid strain dSynXI.A_C was tested using two types of DNA delivery modes, the bacterial plasmid and linear constructs, in order to provide the heterologous genes into cells. Ultimately, based on the number of colonies that passed both the 200 µg/ml and 350 µg/ml of G418s growth, the linear constructs of KanMX_2L were significantly better at incorporation into yeast than any other DNA tested. Not only did this method result in greater numbers of KanMX integrated colonies, but it is also the cheaper and easier method of DNA preparation, as a simple PCR can provide the DNA (as opposed to cloning being needed to provide a plasmid). Furthermore, using linear fragments of DNA has greater potential to be scaled into larger experiments, possibly where a variety and multitude of genes are used for SCRaMbLE-in to produce libraries of synthetic strains. The most significant difference between the pUC19-based constructs (pDJ017-19) and the linear KanMX DNA fragments is the conformation of the provided DNA, *i.e.* a circular plasmid versus double stranded small linear fragments. There is a lack of publications that tests the transformation efficiencies of bacterial vectors versus linear DNA, primarily due to the fact that the bacterial pUC19-based plasmids are not designed to be incorporated and replicated in the yeast cell. The SCRaMbLE-in experiments are not good indicators of the differences in transformation efficiencies but they offer some insight into what expected results might yield, as the differences in numbers of colonies between the two DNA delivery modes are so large. However, further investigation of these would be necessary, as well as a more subsequent iPCR data, to determine what portion of the pDJ017, pDJ018 and pDJ019 SCRaMbLE-in strains have in fact integrated the KanMX gene into the synthetic genome region, or whether

they are surviving some other way – e.g. as a ectopic circular plasmid, maintained by the yeast cell for expression of the KanMX gene.

Possible ways to improve the methodology would be to apply an evolutionary pressure on the yeast cells during SCRaMbLE, rather than selecting for viable cells after. Before the inclusion of the one-hour incubation stage on YPD prior to the addition of G418 antibiotic, no viable cells were recovered when plated immediately on G418s-YPD agar. Thus it was established that the effects of SCRaMbLE in combination with the transformation process are lethal and resulted in no viable cells. Nonetheless, a small dosage of G418s antibiotic could be added to the SCRaMbLE media (for example experiments could be set up to test cell yields with injections of 10 µg/ml, 20 µg/ml, 50 µg/ml or 100 µg/ml), to influence and encourage the yeast to include genes that would confer resistance to G418s. Inducing in the presence of an environmental pressure, even a small environmental pressure, could lead to better survival rates. Cells that have SCRaMbLEd-in the KanMX gene, even in the weak environmental pressures, will gain an advantage over those that haven't, with early expression of the KanMX gene. Plating these cells on selective media will isolate the SCRaMbLE-in cells faster and easier than those that have not done so. Thus, the "survival of the fittest" type of evolution is implemented in laboratory conditions.

4.4.3 Determining Genomic Loci of the KanMX Gene of SCRaMbLEd Strains

Following the recovery of KanMX resistant SCRaMbLEd strains, it was attempted to determine the exact integrating locus of the heterologous DNA. Genome sequencing was not an option for this study due to the expenses associated and the large number of samples we hoped to test. Therefore, instead we adapted an inverse PCR protocol that has often been used before to elucidate flanking genomic sequences of expressed heterologous genes. Using divergent primers (those amplifying outwards, away from each other), genomic DNA of several samples was digested to cleave the genomic DNA regularly and frequently (but not within the KanMX cassette), ligating the cleaved fragments to circularise the DNA and amplify the surrounding sequence of the KanMX marker.

There were 15 colonies screened in total, which included colonies recovered from both the haploid SCRaMbLE-in strains and the diploid SCRaMbLE-in strains with both the circular and linear DNA delivery modes. Of those 15 colonies five resulted in bands generated by iPCR (Figures 4.10 to 4.14), and all the samples represented either two hours or three hour induction periods. Three of the samples gave sequencing results that informed as to where the KanMX DNA had been inserted. In the first case (Figure 4.10) it has inserted between the 18th and 19th

loxPsym sites of the synthetic region, and presumably this happened by simultaneous deletion of the DNA between these sites – a 5kb region in mega chunk C that contains 4 genes (including two mitochondrial genes). In the second case (Figure 4.11) the insertion is between two more distant loxPsym sites – the 2nd and 7th. These are 21 kb apart on the chromosome and if the recombination was done in a manner that led to deletion of the region between these sites then 5 genes (all with metabolic functions) would be lost. The final case (Figure 4.13) shows a reverse direction insertion between the 12th and 18th loxPsym sites, potentially deleting out 31.9 kb of DNA. This would be a large deletion if we presume that no other recombination occurred prior to the KanMX integration. If this region was indeed deleted then 11 annotated genes would be lost including at least one essential gene. However, as this work was done with a diploid strain, the loss of any genes, whether essential or not, should not be lethal.

While these results show that this approach can determine the SCRaMbLE-in location, the method exhibited low efficiency and was time-consuming, requiring high quality and purity of extracted genomic DNA, long periods of digestion, an overnight gradient ligation, and extended touchdown PCR amplifications that require long elongation periods to compensate for the amplification of fragments up to 10 kb. Furthermore, there is uncertainty about i) the efficiency of the gradient ligation and how the circularisation of the genomic DNA has occurred and ii) whether the inverse PCR is successful. A higher degree of confidence could have been achieved should a positive control, such as a plasmid of a known size, been used in the inverse PCRs. Having said that, the bands amplified from the five successful inverse PCRs shown in Figures 4.10 to 4.14, showed that the primers annealed as intended to the KanMX gene.

Further analysis using the inverse PCR method could have been attempted with the PCRtag primers that anneal to the synthetic sequence. However, the number of possible reactions and the combinations of primers would be very large for this, and the interpretation of any potential results that might be obtained would be challenging. To reiterate this point further, there are approximately 288 primers designed for the screening of the recoded open reading frames that are found up to (and including) mega-chunk C. The set up and combination of primers to attempt to screen for the genomic rearrangements would require in excess of 1,000 reactions. The input required in balance with the output gained did not make this analysis method worth pursuing.

The ideal analysis I would have chosen for this project to ascertain the SCRaMbLE-in strain genomic rearrangements would be long-read genome

sequencing, and in particular the new type of nanopore sequencing (Oxford Nanopore Technologies). This uses an electrical current applied to a nanopore protein to sequence DNA molecules with no major read length limitations (in contrast to iPCR) in real-time. Fragments of the synthetic chromosome could be extracted and run through the nanopore sequencers, which would not only reveal the exact chromosomal integration loci of the KanMX gene but also the equally important other rearrangement brought on during SCRaMbLE. This method would also be able to track multiple integrations, which is an inherent drawback of the iPCR method.

4.4.4 SCRaMbLE and Determining SCRaMbLE-in using Fluorescent Proteins

Fluorescent protein constructs in a non-synthetic strain (BY4741) were used to try to predict and map the effects of SCRaMbLE and SCRaMbLE-in on the expression of three genes formatted with loxPsym sites. The rationale behind the experiments was to follow the gene rearrangements caused from the induction of the Cre-lox system to understand how SCRaMbLE works and effects expression. This could then be used to phenotypically predict where the KanMX gene integrates and whether it is more likely to integrate by a certain type of rearrangement. The use of fluorescent proteins in the BY4741 strain for the study of SCRaMbLE is more advantageous than using (partially-)synthetic chromosome strains as i) the mRuby, sfGFP and mTagBFP genes are not essential genes and ii) cell viability is not affected, thus their rearrangement, whether it be deletions, inversions or duplications, does not negate the survival of the cell. Shen *et al* (2015)[82] provides the most-recent and only published extensive study of SCRaMbLE, reporting an equal number of deletions and inversions from their sequenced strains. I assumed here that the more likely scenario is that there were a higher number of deletions, but due to detrimental essential gene loss in the Shen *et al.* study, these were not recorded as those cells do not survive. Initial inspection of the fluorescence data in this chapter suggests a far higher number of deletions than inversions in this study, shown by the lack of the detected fluorescence from the mRuby, mTagBFP or sfGFP genes both in the SCRaMbLE and SCRaMbLE-in experiments. However, on closer inspection this cannot be claimed, as we do not know for sure if loss of fluorescence is always due to deletions. Inversions and insertions could also both appear as loss of fluorescence caused by disruption of the 3'UTR of the upstream mRNAs.

The comparison between the predicted phenotypes (prior to SCRaMbLE, and listed in Figure 4.16) and the actual outcomes (in Figure 4.18 and 4.19) obtained and based on my interpretations of the flow cytometry GM values and heat maps, shows that that the predictions made for the second and third gene matched the data

collected post-SCRaMbLE. Surprisingly, changes in fluorescence observed in the first gene were common and this was not predicted. This provided the first indication that the design of the loxPsym sites with respect to their location between genes and terminators was an issue for SCRaMbLE-in. A new design is proposed in Chapter 6, but in short, all heterologous genes introduced into the synthetic genome should be designed to contain a terminator that will not be recombined out, thus not silencing or affecting the expression or 3'UTR of the upstream gene of the locus it recombines into.

The design that was employed for the fluorescent constructs, wherein a loxPsym site was situated three base pairs downstream of the stop codon of a gene was done so in order to adhere to the design principles and methods used overall in Sc2.0[72]. Reflecting back on the aims, data and the BY4741 strain used as the host for the fluorescent constructs, it would have been more optimal had the loxPsym site been placed between the terminator and promoter of the first and second, and second and third genes, respectively, and downstream of the terminator of the third gene. One drawback of this approach would have been an issue with the compatibility of the cloning steps with the YTK and Golden Gate method that was used to construct the fluorescent constructs, but this is not a significant hurdle. More importantly, although it could provide insight as to where the heterologous DNA is being inserted into, the changes of expression caused by SCRaMbLE or SCRaMbLE-in of the chromosomal genes would not be applicable or comparable to the designed synthetic chromosomes of the Sc2.0 project.

The complexity of fluorescence matrices was even more pronounced with the SCRaMbLE-in of the KanMX_2L gene. Decreased fluorescence from the first gene was seen in some case with SCRaMbLE in Figure 4.19, but with SCRaMbLE-in of KanMX_2L, particularly in yRC1841 and yRC1845, decreased expression and even loss of expression of the first gene was a regular occurrence (as summarised in Figure 4.25). Consider sample 12 of yRC1841 (in Figure 4.25) – expression of mRuby (the first gene) is not detected, whereas sfGFP and mTagBFP showed similar levels of fluorescence detected when compared to the control. The deletion of a gene is interpreted when the GM of the fluorescence produced by that gene is close to the GM obtained for the wild type BY4741 negative control, at the same excitation wavelength for respective lasers.

The interpretation of this data was difficult, but one explanation that I can suggest is that KanMX_2L fragment inserted at the first loxPsym site, possibly by circularising itself first, then by recombining the single loxPsym site into the first loxPsym site (three base pairs downstream of the stop codon of the mRuby ORF).

The orientation of the KanMX gene inserting into this first loxPsym site determines whether the first gene will have a terminator sequence to complete transcription or not. If KanMX inserted in with its promoter immediately after the mRuby ORF (i.e. pointing in the same direction), then this means that the mRuby ORF no longer has a working 3'UTR. This would prevent the mRuby mRNA from correct expression and cause a decrease in red fluorescence. Expression of the genes downstream of the newly-inserted KanMX_2L would not be affected. The loss of fluorescence at the first gene was not seen as often in SCRaMbLE-only experiments, but common with SCRaMbLE-in. From Figures 4.18 and 4.19, the first gene of some samples does undergo a decrease in fluorescence, indicative of transcriptional interference, possibly by the inversion of the second gene to create convergent ORFs sharing a terminator. However, in none of the samples was there a clear deletion. This is further confirmed when we refer back to the heat map of the GM, of which all give evidence of some fluorescence. Therefore, we suspect that the orientation of the integrating KanMX DNA determines whether the first gene is able to translate an effective fluorescent producing protein from the disrupted 3'UTR region: if it inserts in with the same direction of the first gene it removes its terminator and so fluorescence is lost, whereas if it inserts in in the opposite direction, it shares a terminator so fluorescence is probably just reduced.

Despite helping with understanding the SCRaMbLE outcomes, there were downsides to the heat mapping approach presented in section 4.3. Firstly, the extent of the transcriptional interference of the 3'UTR brought on by the inversion of genes or terminator loss is not known, and at this stage is speculation. Secondly, there were several phenotypes whose outcomes were difficult to interpret and were thus classed as undetermined, such as sample 2 of yRC1842 (Figure 4.25). Lastly, differences in the detected blue fluorescence produced by mTagBFP between the samples, BY4741 (negative control) and the unscrambled positive control, were very small, as indicated by the large amount of blue in the heat maps of Figures 4.18 and 4.21 to 4.24. Elucidating concise conclusions from these data was thus difficult, and required the production of a secondary heat map to form clearer results, abstracting the data even further. The reads were repeated, and the negative control (BY4741) was regrown several times to ensure no contamination was present, which was confirmed not to be the case. The mTagBFP gene of the fluorescent constructs is placed under a constitutive promoter (ALD6) which causes medium expression of the gene, therefore it was expected to produce a level of blue fluorescence that was significant over any autofluorescence. However, this was not seen. Instead the integrated constructs produced only a weak blue fluorescence making it harder to

determine cells with mTagBFP from those without. In order to check that the flow cytometer could detect the protein correctly, a strain holding a plasmid that encoded only the mTagBFP gene under the control of a strong constitutive promoter was also measured (data not shown), and this showed much clearer differentiation between the strong blue fluorescence produced from mTagBFP and autofluorescence. For future work, I recommend that the mTagBFP is placed under a stronger constitutive promoter to generate clearer blue values and avoid a data set that required such high levels of thought in order to interpret.

Time constraints prevented further analysis of these colonies, but another simple screening method to verify the integration of the KanMX gene(s) within the fluorescence-encoding region constructed here would not have required an inverse PCR reaction but rather a standard PCR with convergent primers from KanMX annealing with primers annealing either side of the genomic region. These PCRs would determine i) genomic location of integration and ii) the presence of other genes. Generated PCR products could have been sent off for sequencing to shed further insight to the orientation of those genes, and would help verify and understand any conclusions from the SCRaMbLE and SCRaMbLE-in experiments done with these fluorescence-expressing strains.

4.4.5 Conclusions

SCRaMbLE is a novel technology implemented as part of the design of all synthetic yeast chromosomes part of the Sc2.0 project. The work presented in this chapter aimed to implement and manipulate the SCRaMbLE toolkit of synthetic chromosomes and further develop it to also be able to introduce heterologous genes into synthetic chromosomes, and in doing so provide the yeast with new functions. This new method, called SCRaMbLE-in, will allow simultaneous one-pot genome rearrangement and integration of heterologous genes, providing a new way that synthetic yeast genome strains can be evolved and adapted under laboratory conditions. This will aid in the engineering of yeast genomes for novel functions.

The KanMX marker was used as the model heterologous gene to implement in SCRaMbLE-in studies, and was cloned under a range of formats to investigate which gave rise to the highest number of G418s-resistant colonies. SCRaMbLE-in was primarily induced in the haploid SynXI.A-C strain, however this generated high strain lethality, and was not suitable for the development of this new methodology. To counter this, a diploid version, dSynXI.A-C, was created and SCRaMbLE-in was shown to function best in this strain by using linear PCR-generated DNA fragments

encoding the KanMX gene flanked by an upstream and downstream loxPsym site, deemed KanMX_2L. Furthermore I showed that induction for three hours with KanMX_2L resulted in the highest numbers of colonies that were truly resistant to G418s antibiotic.

In an attempt to characterise SCRaMbLE, and map whether there was a likelihood of KanMX_2L integrating into a particular locus or in a particular way, fluorescent constructs were produced to contain a loxPsym site three base pairs of after the ORF of every gene and were integrated into the genome of the non-synthetic BY4741 strain to create 'test' strains for assessing SCRaMbLE. Both SCRaMbLE and SCRaMbLE-in were induced in the generated fluorescent strains. The complexity of the phenotypes that arose made the following and analysis of the data difficult without genome sequencing, however it exemplified the diversity of phenotypes produced in the presence of only three loxPsym sites, which was further incremented by the inclusion and SCRaMbLE-in of KanMX_2L into the fluorescent constructs. It reiterated the large genomic diversity that arises when there are multiple loxPsym sites part of synthetic chromosomes.

5.0 CHAPTER 5: SCRaMbLE-in of Heterologous Pathways: Expression of Xylose-Utilisation Genes in a Synthetic Yeast Strain

Having demonstrated that the SCRaMbLE-in system developed in this study can be used for the automatic integration of a heterologous gene (KanMX) into a partially-synthetic yeast genome, I next expanded the SCRaMbLE-in system to integrating multiple heterologous genes. In this chapter I used heterologous genes from fungal and bacterial sources that encode xylose utilisation pathways known to allow yeast to grow on the lignocellulosic sugar xylose. I cloned and tested four xylose utilisation genes, and demonstrated that they enable growth on xylose for our engineered yeast. I then further showed that SCRaMbLE-in can add these genes in one go into the rearranging genome of the diploid synthetic yeast strain, dSynXI.A-C and automatically produce a strain that can grow with xylose as the carbon source. This demonstrates how SCRaMbLE-in can take multiple genes provided as DNA fragments and rearranges them into functional pathways in synthetic yeast.

5.1 Aims

- To isolate and express relevant xylose-utilisation genes in a synthetic yeast strain with the purpose of enabling growth on xylose media.
- Characterise the aerobic growth of the synthetic yeast in xylose media and in glucose/xylose media and compare to appropriate controls.
- Format the xylose-utilisation genes for integration into synthetic yeast via the SCRaMbLE-in method developed in Chapter 4.
- Perform SCRaMbLE-in with multiple heterologous genes and screen, isolate and investigate any colonies that now grow on xylose media.

5.2. Introduction

Lignocellulosic biomass is a complex and environmentally abundant plant matter, found in wood and agricultural waste, and is a rich source of pentose and hexose sugars. Lignocellulosic hydrolysate is composed of three complex macromolecules that are cellulose, hemicellulose and lignin. Lignin is an organic and hydrophobic polymer that varies in size and complexity, depending on where it is found. The other two polysaccharide components, cellulose and hemicellulose, are the two sugar-rich compounds; cellulose is made of long chains of glucose, and is easily fermented and metabolised by a large number of microorganisms, as well as being the structural foundation of bacterial biofilm[154] production. Hemicellulose (also known as polyose due to its structure) is formed of a matrix of hexose and D-pentose sugars such as glucose (60 – 70%), xylose (30 – 40%)[114, 155], arabinose, galactose and mannose. After glucose, the most prolific monomer is xylose and it is particularly abundant in certain grasses and hardwoods.

The saccharification of lignocellulose biomass is currently a significant research effort due to its potential as a renewable source material for the production and fermentation of biofuels and bioethanol. The bioethanol market in the U.S. alone has been estimated as a multi-billion dollar market and at present bioethanol manufacture is typically achieved using sugars from six-carbon sugar substances such as sugar cane and corn. However these sources are already used extensively for human and animal feed, therefore their consumption in bioethanol production is undesirable as this leads to a competition between the foods and energy markets. Thus lignocellulosic biomass, due to its sugar high content and low use in food production, has been recognised as the alternative option for biofuel or bioethanol production, and global research efforts are being pursued in order to optimise the metabolic engineering of bioethanol and biofuel-producing cells to be able to efficiently convert lignocellulosic sugars such as xylose into fuel replacement molecules[112, 156].

The most studied organism for engineering in catabolism of lignocellulosic sugars is Baker's Yeast (*Saccharomyces cerevisiae*). It is the most industrially-exploited yeast in biofuel production[86] to date, due to its fermentation abilities, high ethanol tolerances and productivity[157], and because it is generally regarded as safe (GRAS; a definition of GRAS organisms can be accessed on: http://www.fda.gov/Food/IngredientsPackagingLabeling/GRAS/). Despite the fact that we can engineer yeast to such a degree that we can build entirely synthetic chromosomes, a major drawback for *S. cerevisiae* is that this yeast does not possess the ability to naturally utilise xylose as a carbon source. While there are a variety of

natural bacterial and fungal metabolisers of xylose (for example, *E. coli*[158], *Pyromyces*[104] and *Clostridium*[159] species, *Bacillus subtilis*[160] and the yeast *Scheffersomyces (Pichia) stipitis*)[161] fermentation of xylose by *S. cerevisiae* would be much more desirable due yeast's suitability as an industrial host and due to its well-studied genome[93, 94, 162-164].

The catabolism of xylose by cells requires conversion of the pentose sugar into (smaller) useful compounds for central carbon metabolism through two possible pathways: (i) the isomerase pathway, or (ii) the oxidoreductase pathway. The isomerase pathway is predominantly found in fungi and bacteria[165, 166], and uses a single gene (xylA), encoding for a xylose isomerase (XI), an enzyme that converts D-xylose into D-xylulose. The oxidoreductase pathway is found in organisms such as *Scheffersomyces stipitis*[167] yeast and requires the genes XYL1 (encoding the xylose reductase (XR) enzyme driving catalysis of xylose into xylitol), and XYL2 (encoding for xylitol dehydrogenase (XDH) which continues the catabolism of xylitol into D-xylulose). *S. cerevisiae* posses versions of both of these genes encoded in its genome, but due to their low basal expression, they do not enable yeast to grow on xylose. An analog of xylose reductase is encoded on chromosome VIII (32.3 to 32.4 kb) as a GRE3 gene – an aldose reductase[168], and the XYL2 gene is found on the 12th chromosome (27.4 to 27.5 kb), and has a high sequence similarity to the *S. stipitis* gene[169]. A third gene, XYL3, is a xylulokinase and is the completing gene of both the isomerase and oxidoreductase pathways, phosphorylating D-xylulose into D-xylulose-5-phosphate which then enters central carbon metabolism via the Pentose Phosphate Pathway (PPP) (Figure 5.1). It too is found intrinsically in *S. cerevisiae* (chromosome VII 88.6 to 88.7 kb)[110], and has been shown to optimise growth on xylose media through moderate overexpression[115].

In recap from Chapter 1, the oxidoreductase pathway consists of three genes, XYL1, XYL2 and XYL3, which encode for xylose reductase (XR), xylitol dehydrogenase (XDH), and xylulokinase (XKS). The xylose reductase enzyme catalyses the reduction of xylose into xylitol, which is then converted into xylulose by xylitol dehydrogenase. The last oxidreductase gene, XKS, breaks down the xylulose further in a reaction requiring ATP to produce xylulo-5-phosphate. This is the compound that enters the pentose phosphate pathway (PPP) to be converted into ethanol, glycerol, and other substances. In contrast, the isomerase pathway is made up of only two genes, the bacterial xylose isomerase (xylA) which converts xylose straight into xylulose, which is then catalysed by XKS once more, to make xylulo-5-phosphate (Figure 5.1). Both the isomerase (IS) and oxidoreductase (OR) pathways have advantages and disadvantages associated with their recombinant expression in

yeast, as summarised in Chapter 1. Overall, current research suggests that the OR pathway is better suited for xylose utilisation in yeast, in comparison to the IS pathway[108]. It does however require more heterologous genes to be expressed and suffers from an inherent redox imbalance due to the two enzymes, XR and XDH, having different preferences for cofactor utilisation[170].

In Chapter 4, I demonstrated that the SCRaMbLE-in system developed in this study could be utilised for the automatic integration of the a heterologous gene (KanMX) into the synthetic genome of dSynXI.A-C. In this chapter I aim to expand the SCRaMbLE-in system from incorporating single genes, to integrating multiple genes, specifically heterologous genes that encode xylose utilisation pathways into yeast. Upon isolation and testing of four xylose utilisation genes (xylA*3, XYL1, XYL2 and XYL3), I first established that their recombinant expression in the diploid synthetic yeast strain, dSynXI.A-C was successful, allowing growth on xylose. I then proceeded to format the heterologous genes according to the best layout determined in Chapter 4 for SCRaMbLE-in, and induced the recombination process in dSynXI.A-C cells after providing the xylose utilisation genes as linear DNA flanked with loxPsym sites. In this chapter I report my procedures and my findings for the SCRaMbLE-in of multiple genes that create a xylose-utilisation pathway in yeast. Successful expression of a SCRaMbLEd-in metabolic pathway shows an exciting

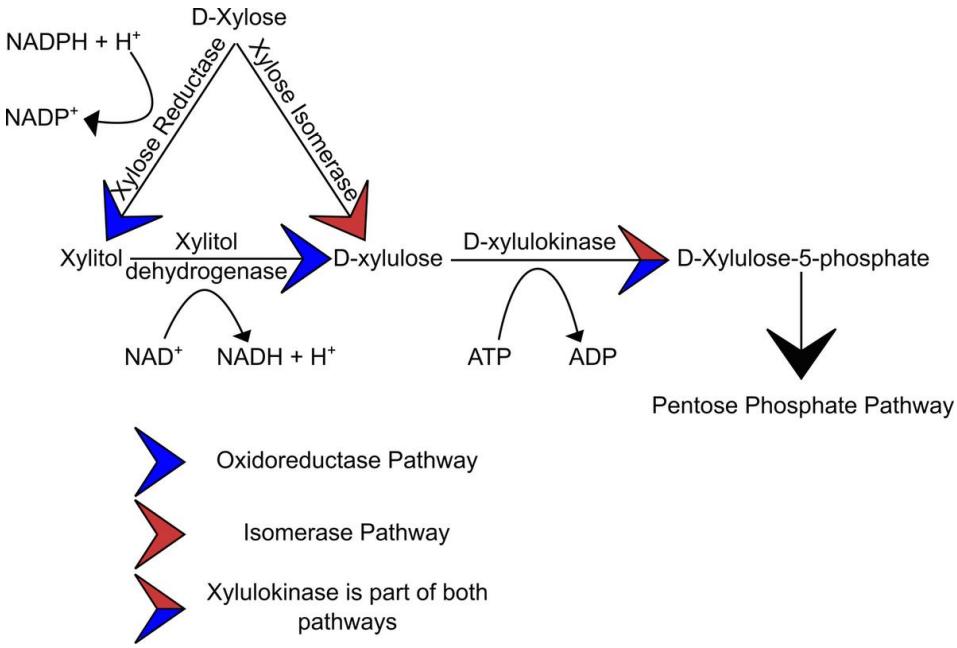


Figure 5.1 Xylose Pathways Schematic. Illustration showing the Isomerase and Oxidoreductase pathways different routes to the break down of xylose, to produce Xylulose-5-phosphate that enters the Pentose Phosphate Pathway. The Oxidoreductase pathway suffers from a redox imbalance due to different cofactor preferences by the xylose reductase and xylitol dehydrogenase enzymes.

preliminary use of Sc2.0 synthetic yeast, and this method could be used for further development of more complex pathways in future work.

5.3 Results

5.3.1 Utilising the Yeast Tool Kit for Efficient Cloning of Xylose-Utilisation Genes

A modified version of the *E. coli* isomerase gene, xylA, was acquired from Prof H Alper (Texas, USA), and is listed as xylA*3 in their research[111]. The Alper group modified this gene through three rounds of PCR mutagenesis, developing a final version of the gene that exhibited a nine-fold greater growth rate on xylose than the original when expressed in an isogenic BY4741 strain (also developed by them).

The XYL1 gene was the first to be commercially synthesised, outsourced (Life Technologies, Cal. USA), as a "GeneArt Gene String". Its native sequence was derived from the *S. stipitis* genome[171] and the ORF of the gene was codon optimised for *S. cerevisiae* expression, as well as programmed to exclude any major endonuclease restriction sites which may potentially be used in downstream cloning. The final designed ORF sequence had no DNA homology to any genes in natural genomes or DNA present in GenBank[172]. Using primers pDJ043 and pDJ044, XYL1 was PCR amplified to add the type 3 prefix and suffix sequences necessary for compatibility with the Yeast Tool Kit (YTK)[121], a modular Golden Gate system for the design and cloning of genes to be expressed in yeast that was previously described by Lee *et al* (2015)[111]. PCR purification was carried out to purify the PCR product according to Section 2.3.2. Following this, YTK cloning was performed to place the amplified synthetic ORF sequences into plasmids and transformed into *E. coli*, and then transformed into yeast for expression. The intended YTK plasmids for this study were designed *in silico* using Benchling software[173], and their designs are shown in Figure 5.2. The XYL2 was outsourced in the same way, and type 3 prefixes and suffixes were added to its sequence using primers pDJ045 and pDJ046. The last OR gene, XYL3, was outsourced to include the prefix and suffix sequences, and did not require any subsequent PCR amplification.

All the xylose utilisation genes (xylA*3, XYL1, XYL2 and XYL3) were formatted with the appropriate YTK prefix and suffix sequences to make them compatible with the YTK cloning system. They were then cloned to be downstream of constitutive promoters (e.g. TEF1p, TEF2p, TDH3p) and upstream of strong terminators (e.g. ADH1t), and hosted on shuttle plasmids that are a modification of the YCp yeast vectors. All the YTK-generated plasmids are listed in Table 5.1 describing their component parts. The OR genes XYL1, XYL2 and XYL3 are cloned into plasmids pDJG1, pDJG2 and pDJG3, respectively. pDJG1 and pDJG3 have the inclusion of loxPsym sequences upstream of XYL1's promoter and downstream of XYL3's stop codon so that when they are further combined into a single multi-gene

cassette the three OR genes are flanked by two loxPsym sites (Figure 5.3). The single IS gene xylA*3 in cloned into plasmid pDJG4 with a loxPsym added upstream of its promoter and downstream of its stop codon.

The three OR genes were also cloned as single genes each flanked by an loxPsym site upstream of the promoter and downstream of the stop codon in plasmids pDJG5 (XYL1_2L), pDJG6 (XYL2_2L) and pDJG7 (XYL3_2L). This allowed for their use in SCRaMbLE-in, described later on in this section. Their design is included in Figure 5.2. The Golden Gate protocol for cloning using YTK is outlined in Chapter 2, section 2.3.8. Plasmids were isolated from white bacterial colonies following ligation and transformation, and were analysed using analytical restriction digestion with BsmBI as the cleaving enzyme, using gel electrophoresis to visualise correct construction by identifying the appropriate band sizes. An illustrative example of a gel digest confirming correct cloning is shown in Figure 5.4a.

Table 5.1 The Cassette and Multi-Gene Cassettes Structure and Parts of the Xylose-Utilising Plasmids Used in this Chapter

CASSETTES								
Construct	**Connector**	**Promoter**	**Gene**	**Terminator**	**Connector**	**Yeast Replication Origin**	**Marker**	**Bacterial Replication Origin (BAC RO)**
Cassette	Part 1	Part 2	Part 3	Part 4	Part 5	Part 6	Part 7	Part 8
pDJG1	ConS-loxPsym	TEF2	XYL1	ADH1	Con1	CEN6/ARS4	URA3	AmpRColE1
pDJG2		TEF1	XYL2	TDH1	Con2	CEN6/ARS4	URA3	AmpRColE1
pDJG3		TDH3	XYL3	loxPsym-ENO2	ConE	CEN6/ARS4	URA3	AmpRColE1
pDJG4	ConS-loxPsym	TEF1	xylA*3	ADH1	Con2	CEN6/ARS4	URA3	AmpRColE1
pDJG5	ConS-loxPsym	TEF2	XYL1	loxPsym-ADH1	ConE	CEN6/ARS4	URA3	AmpRColE1
pDJG6	ConS-loxPsym	TEF1	XYL2	loxPsym-TDH1	ConE	CEN6/ARS4	URA3	AmpRColE1
pDJG7	ConS-loxPsym	TDH3	XYL3	loxPsym-ENO2	ConE	CEN6/ARS4	URA3	AmpRColE1

MULTIGENE CASSETTES				
Multigene Cassette	**Cassette 1**	**Cassette 2**	**Cassette 3**	**Backbone**
pDJM1	pDJG1	pDJG2	pDJG3	KanRColE1-URA3 CEN6/ARS
pDJM2	pDJG4	pDJG3	N/A	KanRColE1-URA3 CEN6/ARS
pDJM3	pDJG1	pDJG2	pDJG3	KanRColE1-URA3 hom

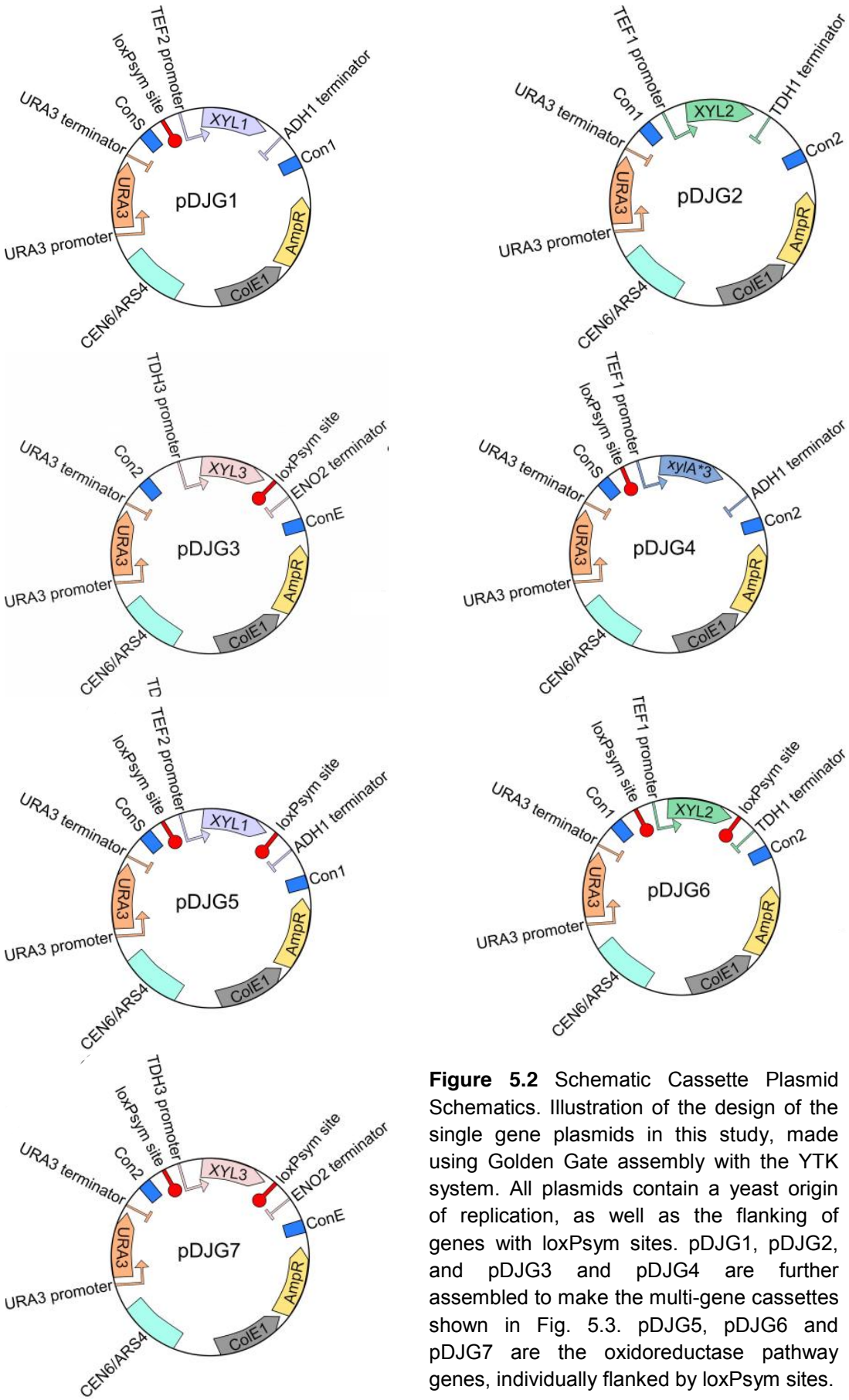


Figure 5.2 Schematic Cassette Plasmid Schematics. Illustration of the design of the single gene plasmids in this study, made using Golden Gate assembly with the YTK system. All plasmids contain a yeast origin of replication, as well as the flanking of genes with loxPsym sites. pDJG1, pDJG2, and pDJG3 and pDJG4 are further assembled to make the multi-gene cassettes shown in Fig. 5.3. pDJG5, pDJG6 and pDJG7 are the oxidoreductase pathway genes, individually flanked by loxPsym sites.

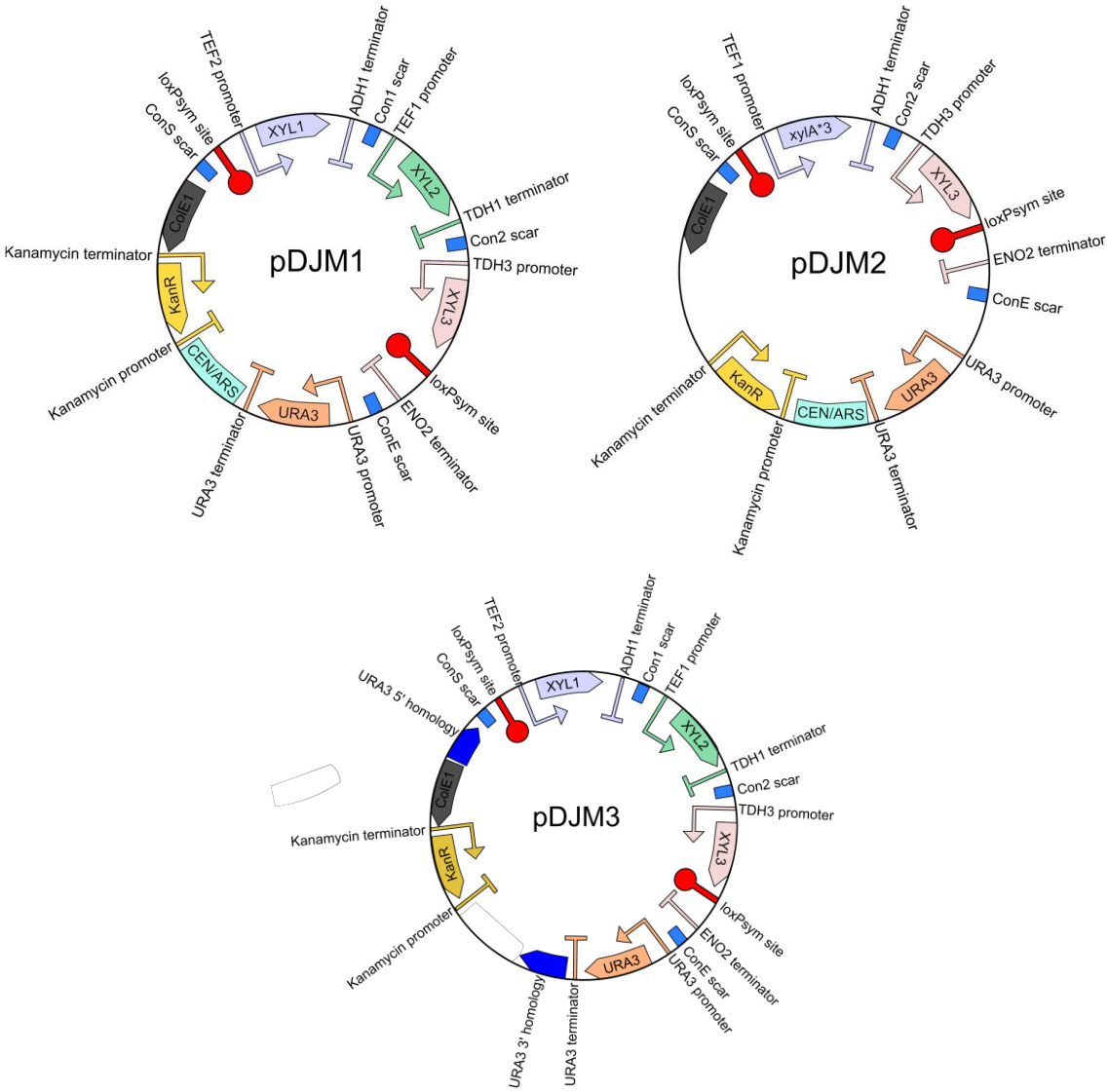


Figure 5.3 Xylose Multi-gene Cassette Plasmid Schematic. Illustration of the multi-gene cassettes, pDJM1, pDJM2 and pDJM3, used for expression of the oxidoreductase and isomerase xylose genes. pDJM1 and pDJM3 are almost entirely identical, differing only in yeast replication origin: pDJM1 is expressed as a centromeric plasmid, encoded by the CEN/ARS sequence, whereas pDJM3 contains URA3 5' and 3' areas of homology for chromosomal integration into the URA3 locus. The multi-gene cassettes are transformed into the dSynXI.A-C strain to make yDJM1, yDJM2 and yDJM3 strains.

Having constructed the xylose utilisation genes into plasmids with appropriate promoters, terminators and loxPsym sites, I next used the YTK system to take these gene constructs and combine them into larger multi-gene cassettes. I constructed three multi-gene plasmids, two containing the oxidoreductase pathways genes (pDJM1 and pDJM3) and one containing two genes for the isomerase pathway genes (yDJM2). Diagrams for these plasmids are shown in Figure 5.3. Following construction, analysis for correct plasmids was carried out by isolating colonies and extracting and digesting plasmid DNA. Test digests were done using BsmBI/NotI as

a cleaving enzyme, and an example of the correct band sizes confirming construction is shown in Figure 5.4b for three of the plasmids. It should be noted that pDJM1 and pDJM3 are effectively the same plasmid, containing the same three OR genes, and differing only in their yeast replication origin part. pDJM1 is a YCp based vector (with a centromeric sequence), which is a low copy plasmid in yeast with one to two copies

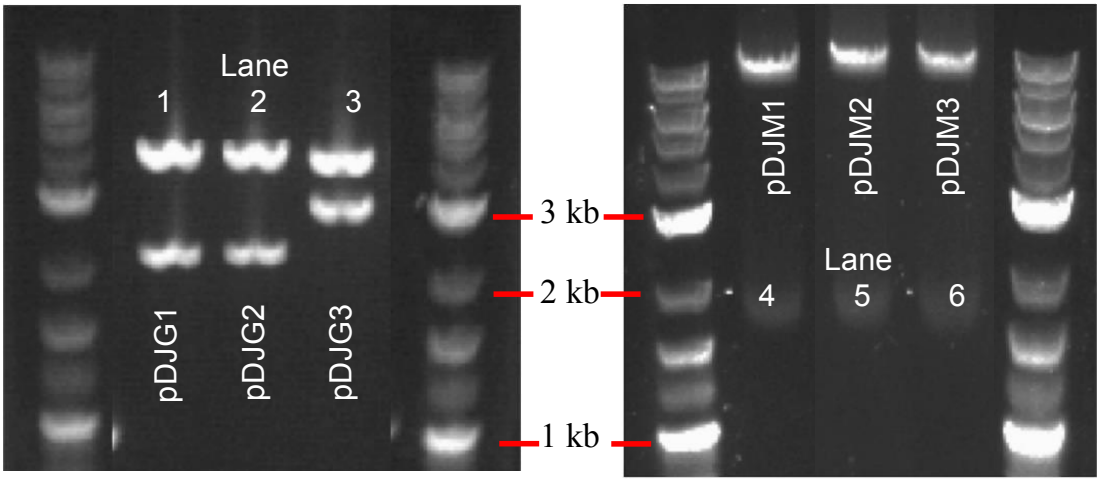


Figure 5.4 Gel Electrophoresis Analysis of Xylose Plasmids. Image showing analytical digestions. Here I show some examples of correctly cut Golden Gate assembled plasmids (a), pDJG1, pDJG2 and pDJG3, cut with BsmBI produces bands at 2 kb and 4 kb (lanes 1 and 2), and 3 kb and 4 kb (lane 3). The gel image to the right is the analysis of multi-gene cassette plasmids, pDJM1-3, using NotI to produce bands of 8.5 kb and 2 kb (b). A 2-log ladder is shown as reference for band sizes.

present per cell[174, 175]. pDJM3 is a YIp based plasmid, which integrates its DNA into the URA3 locus of the host yeast strain, and thus is expressed stably at single copy. These plasmids were made to ensure no growth discrepancies arose as a result of different yeast replication origins. Reasons for the same not being applied to the pDJM2 plasmid are made clear in the next section.

5.3.2 Growth Assays of Xylose Metabolising Synthetic Yeast Strains

The three multi-gene plasmids pDJM1, 2 and 3 were transformed into yeast using the Gietz protocol[153] to make synthetic yeast strains yDJM1, yDJM2 and yDJM3, respectively. Transformed colonies that exhibited correct auxotrophic growth were resuspended in liquid SC-Ura Glu media and grown-up overnight. Prior to growth assays in xylose media, all overnight saturates required centrifugation and a minimum of two washing rounds, to get rid of glucose traces from cultured cells. The washed samples were resuspended in xylose media at a starting optical density (OD_{600nm}) of 0.03 A.U., in 250 ml baffled Erlenmeyer flasks containing 50 ml of SC-X or YP-X media (minus URA for plasmid selection where applicable). SC-X is minimal

yeast media comprised of synthetic yeast drop out mix and nitrogen base source without amino acids, and YP-X media is made of yeast extract and peptone; both media types include 4% of xylose as the main carbon source (see section 2.26 to 2.2.9). Every growth experiment also included the untransformed strain, dSynXI.A-C as a control. Assays and media were set up to ensure that xylose was the only limiting parameter; thus in SC-X growth was done in the presence of all amino acids, ensuring cell death was not as a result of essential amino acid starvation. Flasks were incubated shaking at 225 rpm with aeration, and 1 ml specimens were taken at regular time intervals for measurement of optical density. A comprehensive description of the protocol can be found in Chapter 2, Section 2.3.18. From the optical density data, growth curves for the four tested yeast strains were plotted and are shown Figures 5.5 for growth in SC-X and Figure 5.6 for growth in YP-X.

The untransformed strain fails to show successful catabolism of xylose in SC-X media as there is no visible growth. The strain containing the isomerase pathway genes (yDJM2) also lacks growth in SC-X media, and even shows a decrease in OD_{600} that significantly different from the lack of growth of dSynXI.A-C ($P<0.05$;

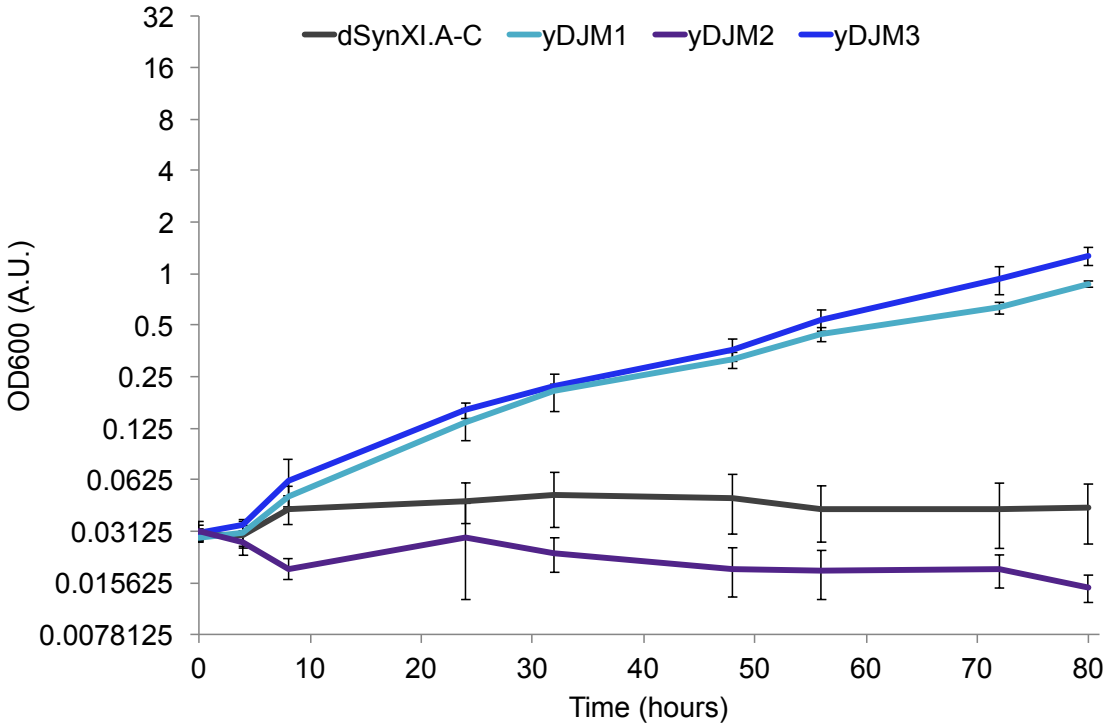


Figure 5.5 Growth Curve of Strains in SC-X Media. Here I have shown the growth data from the differing strains, untransformed (dSynXI.A-C, grey), pDJM1 (light blue), pDJM2 (purple), and yDJM3 (dark blue). Samples were cultured for a total of 80 hours, and cell density was recorded for every sample (OD_{600nm}). The y-axis is shown as a log scale, and n=3. Error bars represent standard error (ST ERR = SD/√n).

ANOVA). Both the untransformed strain and yDJM2 do show growth in YP-X media and this growth is comparable for these two strains. However, this growth is unlikely due to xylose catabolism as it is seen for dSynXI.A-C, instead this growth to OD 0.25 is likely from the other carbon sources available in YP-X media.

When comparing the growth rates of dSynXI.A-C versus yDJM1, yDJM2 and yDJM3 in SC-X media, I obtain data that shows a significant difference (P=0.001; ANOVA), visibly showing that yDJM1 and yDJM3 grow over the 80 hours of incubation in SC-X, unlike the other strains. A Tukey's HSD test shows that the mean ODs of dSynXI.A-C and yDJM2 are significantly different to those from yDJM1 and yDJM3, but not from each other (dSynXI.A-C vs. yDJM2: non-significant; yDJM1 vs. yDJM3: non-significant), confirming that the two pairs of strains behave differently. This is also seen in YP-X Media (Figure 5.6), where all strains grow but after 36 hours, yDJM1 and yDJM3 grow visibly better reaching a final OD_{600} that is nearly 4 times higher than that of the other two strains. The growth curves of all four yeast strains share similarity in the initial increase of cellular density as expected due to the richer, more nutritious composition of YP media compared to SC (Figure 5.5). Equally, yDJM1 and yDJM3 proliferate even more in YP-rich media, with very similar doubling times of 14.11 and 12.06 hours, respectively. Although the OD_{600} plots appear to be of closer proximity to each other, both the doubling times and the growth rates of yDJM1 and yDJM3 are significantly different from dSynXI.A-C and yDJM2 ($P<0.001$; ANOVA) over the experiment. The doubling times and growth rates for each media type are listed in Table 5.3.

These growth curve results were important to establish that the xylose-utilisation pathways when transformed into dSynXI.A-C could metabolise xylose and give growth, and to also assess the correct growth media for further selection of xylose-utilizing strains. While SC-X showed the best distinction between cells growing on xylose and those not growing (better than YP-X where distinction is only after 36 hours), the growth rates in SC-X were seen to be very low, meaning that at least 2 to 3 days of growth was needed to get up to workable ODs. As this was not ideal for experiments, different media types were next tested in order to verify which media gives reasonable growth rates but still allows distinction between xylose-utilizing strains and those that are untransformed.

Due to the lack of growth on xylose exhibited by the isomerase pathway strain yDJM2 in the initial experiments, it was not cloned into a YIp based plasmid for chromosomal integration (to make an isomerase equivalent of yDJM3), nor was it carried forward for further growth analyses. Lack of growth also indicated that

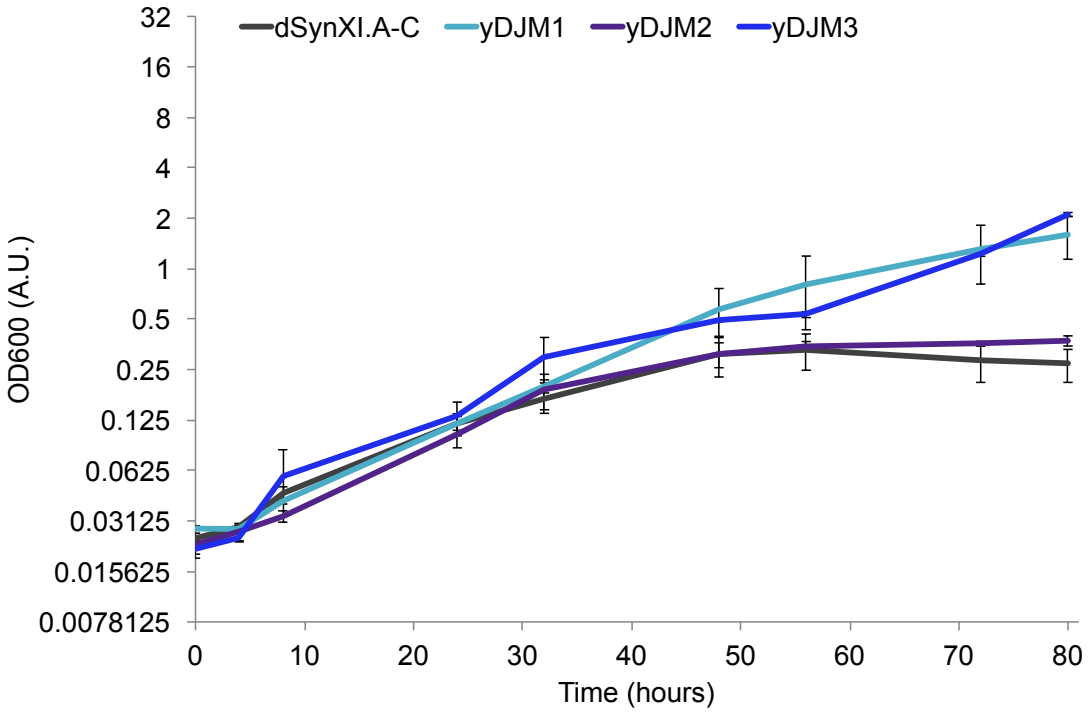


Figure 5.6 Growth Curve of Strains in YP-X Media. Here we have shown the growth data for the oxidoreductase and isomerase strains alongside the dSynXI.A-C untransformed strain in YP-X 4% xylose media. The oxidoreductase pathway strains (yDJM1 and yDJM3) exhibit similar growth curves, whereas the isomerase pathway strain (yDJM2) shows growth reflective of the untransformed strain. n=3 and error bars represent standard error.

downstream SCRaMbLE-in of the isomerase genes would likely not yield viable growth or colonies during screening steps, thus this strain was not pursued further.

In both SC-X and YP-X, the growth on xylose has been linear with large doubling times when compared to glucose growth. With the aim of boosting growth to shorter doubling times, to get an exponential growth curve, diauxic growth experiments were next set up, supplementing the xylose media with small amounts (0.05% or 0.01%) of glucose. Table 5.2 shows the composition and names of the mixed media (called M1 to M4) used for this study, and the growth results using these media are plotted in Figures 5.7 (SC-X mixed media) and 5.8 (YP-X mixed media). The reintroduction of the logarithmic phase was achieved from adding small amounts of glucose to the media. This means that all samples including the untransformed strain now show improved growth. The new exponential phase spans from 0 to 24 and 0 to 32 hours in SC-X and YP-X mixed media, respectively. After this a diauxic shift is seen and the cells transgress to the post-diauxic phase, presumably shifting to xylose catabolism. At this point the untransformed strain ceases to proliferate in the media in both cases, and exhibits a stationary growth

profile while the strains with the oxidoreductase pathway genes continue growth, albeit with a linear profile.

The quantity of the glucose added is directly related to the optical density, thus higher quantities of the initial glucose inevitably lead to greater ODs of all samples. Growth of dSynXI.A-C in M4 media experiences a significant drop at 56 hours following 24 hours of continual decline, and then appears to rise to again. We do acknowledge that the data of this sample shows relatively large error bars from the triplicate recordings obtained, causing a skewed average reading of 0.436 A.U at 72 hours. Despite the variance, statistical tests showed these data to be significant from each other (all M media type growth rates and doubling time $P<0.05$, ANOVA). The skewed readings at 56 hours for this sample could be as a result of a contaminant in the media after the long growth periods. This is not seen in any of the other samples or media.

The two oxidoreductase pathway plasmids, pDJM1 and pDJM3, showed similar growth in all the mixed media types, but in the majority of cases were close enough in values to not be significantly different from each other (under all conditions) despite one being present on a yeast plasmid and the other integrated single copy into the yeast genome. The boosting of ODs and growth rates seen with this mixed-media is advantageous in that it increases cells optical densities within the first 24 hours of growth. However, the growth of the untransformed strain to relatively

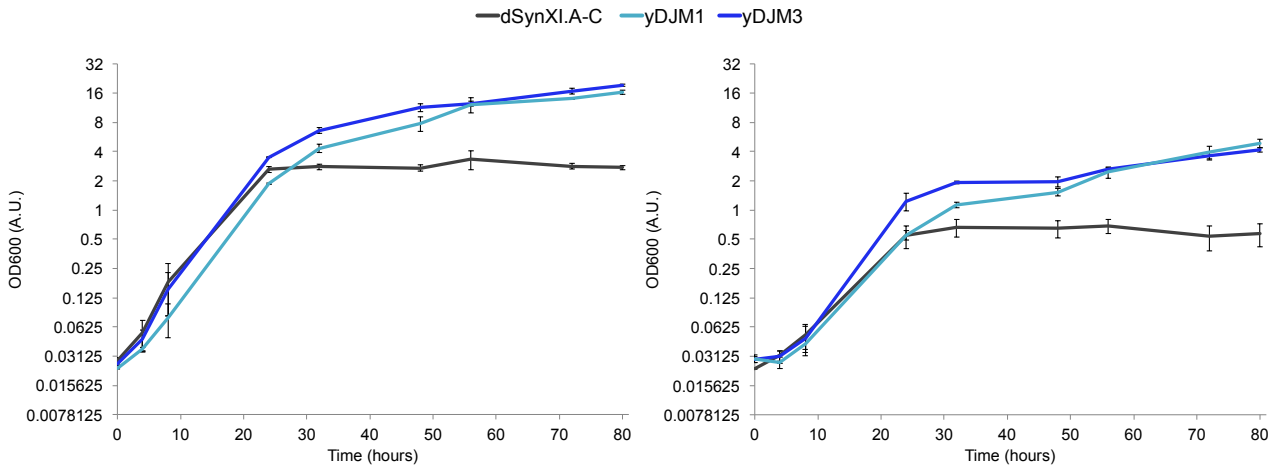


Figure 5.7 Growth Curve of Strains in SC-X Mixed Media. The growth assays were supplemented with 0.05% glucose (left) and 0.01% glucose (right) in SC-X media containing 4% xylose. The higher glucose percentage, the higher OD_{600} was reached, and in both media the untransformed strain also exhibited growth in the presence of glucose. n=3, and error bars represent standard error.

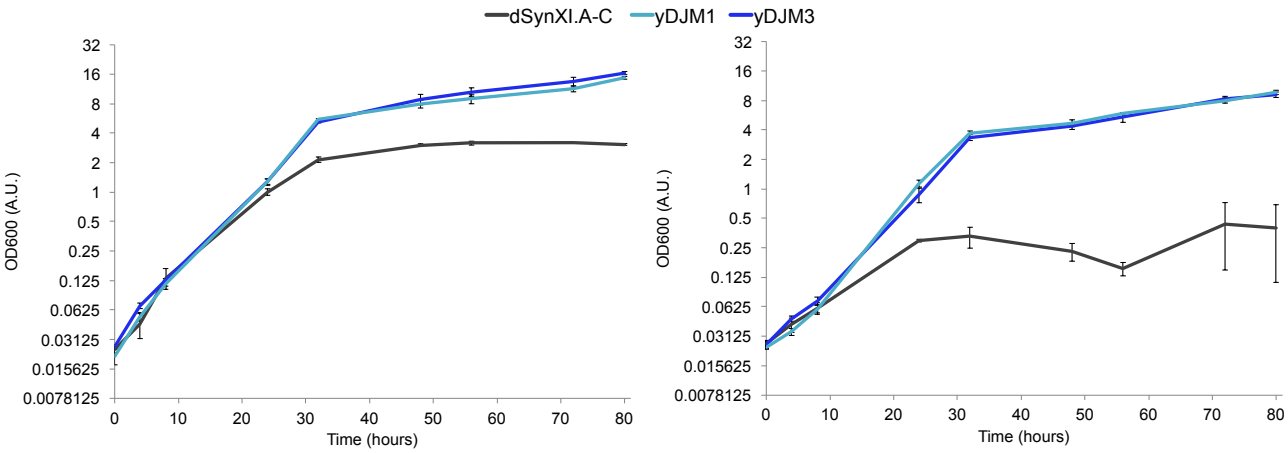


Figure 5.8 Growth Strains in YP-X Mixed Media. The glucose additions of 0.05% (left) and 0.01% (right) were repeated in YP-X 4% xylose media. The untransformed strain once again grows in the presence of glucose but fails to provide evidence for catabolism of the xylose carbon source at later time points. n=3 and error bars represent standard error.

high optical densities makes it potentially challenging to identify xylose catabolising cells from those simply growing-up on glucose. For example, following a SCRaMbLE-in experiment that adds xylose utilisation and then grows cells in one of these media, what would be the ratio of untransformed dSynXI.A-C cells compared to those now catabolising xylose as desired? The ideal media allows for good initial growth of cells to aid in post-SCRaMbLE recovery, but has a clear separation in growth of those using xylose and those not. SC-X shows the best separation but slow initial growth. Mixed media M4 shows reasonable initial growth and then clear separation beyond 24 hours, where >4 times more xylose-utilising cells are seen then untransformed cells.

Table 5.2 Summary of Growth Assay Results of Synthetic Yeast Strains harbouring Xylose Metabolising Genes

Name	Media Type	Glucose (%)	Xylose (%)
M1	Synthetic Complete (SC)	0.05	4
M2	Synthetic Complete (SC)	0.01	
M3	Yeast Peptone (YP)	0.05	
M4	Yeast Peptone (YP)	0.01	

Table 5.3 Summary of Growth Assay Results of Synthetic Yeast Strains harbouring Xylose Metabolising Genes

Strain	Genes	Media	Doubling Time (h)	μ Growth Rate (h^{-1})
dSynXI.A-C (WT)	None	4% SC-X	74.56	0.000
		4% YP-X	24.17	0.003
		M1	12.10	0.034
		M2	18.20	0.007
		M3	11.49	0.038
		M4	30.88	0.005
yDJM1	XYL1, XYL2, XYL3	4% SC-X	16.26*	0.011*
		4% YP-X	14.11*	0.020*
		M1	8.47*	0.204*
		M2	10.91*	0.060*
		M3	8.31*	0.184*
		M4	9.25*	0.121*
yDJM2	xylA*3, XYL3	4% SC-X	N/A	0.000
		4% YP-X	19.92	0.004
yDJM3	XYL1, XYL2, XYL3 int	4% SC-X	15.00*	0.015*
		4% YP-X	12.06*	0.026*
		M1	8.39*	0.241*
		M2	11.16*	0.052*
		M3	8.61*	0.206*
		M4	9.43*	0.112*

* Significant values when compared to the equivalent wild type dSynXI.A-C strain

5.3.2 Competitive Growth of Xylose Metabolising Synthetic Yeast Strains

Thus far all the conditions assayed showed that yDJM1 and yDJM3 had significantly different growth curves from the untransformed strain. It was hypothesised that under competition in xylose-rich media, the yDJM1 or yDJM3 strains would outcompete non-xylose metabolising untransformed cells. A competition assay was set up to investigate at what different ratios can a xylose catabolic strain outcompete a non-metabolising one. This experiment was essential to determine the likelihood of identifying xylose-utilising strains following SCRaMbLE-in.

To be able to track competitive growth, I first needed to label one of the strains using fluorescent protein expression so that they can be tracked using flow cytometry. In order to facilitate simple screening, the untransformed dSynXI.A-C strain was transformed with a construct where sfGFP is expressed from a strong constitutive promoter on a URA3 YCp based plasmid (see section 2.1, Table 2.3). This was constructed using the same Golden Gate method of the YTK system and gave the strain yDJGFP. By adding the sfGFP as a URA3 plasmid, this strain became equivalent in its auxotrophic selection to strain yDJM1 (which does not express the GFP protein). Thus the two strains yDJM1 and yDJGFP, both possessing YCp vectors, and selection via uracil, can both be subject to the same parameters and conditions together. We used 4% SC-X Ura as the growth media, thus selecting for the presence of both plasmids and only allowing growth (slow growth) of xylose catabolizing cells. Strains were inoculated in this solution and incubated shaking in flasks for five consecutive days and measured using the Attune NxT Acoustic Focusing Cytometer Tee and Autosampler (Life Technologies).

Experiments were set up such that the xylose metabolising strain (yDJM1) typically began as the minority in the media, and yDJGFP was the outnumbering strain using five different ratios. The seven separate samples were yDJGFP (only; control one), yDJM1 (only; control two), a 1:1 ratio of yDJGFP : yDJM1, a 5:1 ratio (yDJGFP : yDJM1), a 10:1 ratio (yDJGFP : yDJM1) a 50:1 ratio (yDJGFP : yDJM1), and a 100:1 ratio (yDJGFP : yDJM1). All solutions were standardised to have a starting OD 0.03 (A.U) as was carried out for previous growth assays, readings were made every 24 hours, and samples were grown for a total of four days. We used the Attune to analyse the number of cells per sample over time, as well as the change in the fluorescence of the population in order to study whether yDJM1 successfully outcompeted yDJGFP over time growing in SC-X. The results of the competition experiments are shown in Figures 5.9 and 5.10.

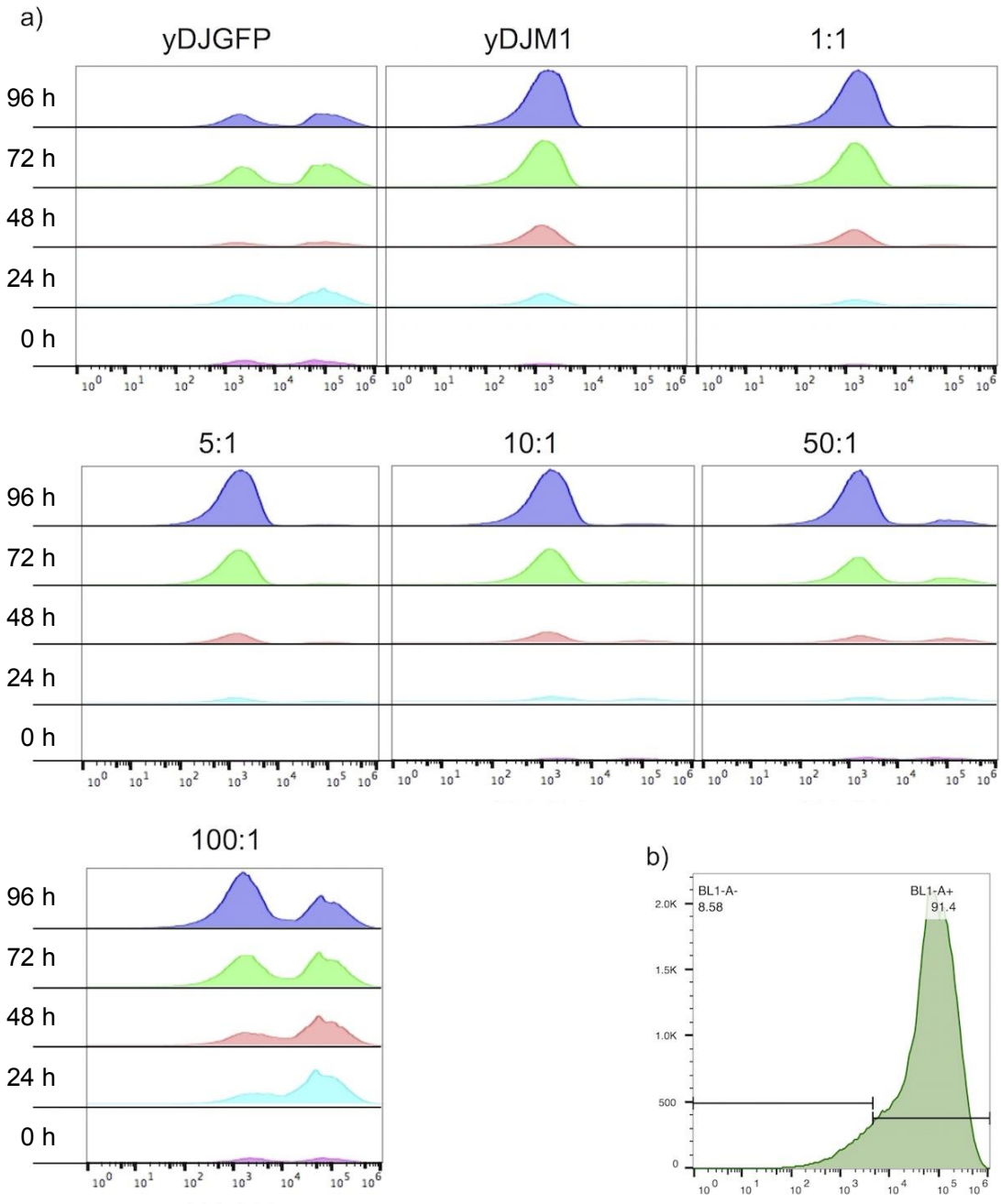


Figure 5.9 Flow Cytometry Data for Competition Assay. Histograms of cells following competitive growth in minimal xylose media. (a) yDJM1 is always shown as the minority sample to make it as difficult to recover as possible. yDJGFP is the GFP positive sample although it does not possess the ability to metabolise xylose. (b) A histogram trace showing GFP fluorescence following overnight growth in glucose, whereby the population is predominantly GFP positive, and lacks the GFP negative population seen in the competition readings.

The fluorescence histogram profiles of the seven samples are illustrated in Figure 5.9, and overall they show that the number of non-fluorescing cells increases over the assay period, as the numbers of fluorescing cells decline. However, noticeably from the yDJGFP control sample there seem to be an equal number of non-fluorescing and fluorescing cells at zero hours. This bimodal trait only seems to arise when there are low numbers of cells being measured (starting OD_{600} is 0.03 A.U, which is approximately 5,000 cells in a 150 µl measuring volume) and may be related to plasmid loss or lack of gene expression in these cells after they have been placed in media in which they cannot grow well. A fully saturated, steadily grown yDJGFP sample grown in glucose under optimal conditions was also assayed, and the histogram trace is shown in Figure 5.9b where over 90% of the cells are GFP positive (GFP+).

The presence of this non-fluorescing population in the yDJGFP control makes it slightly more complicated to analyse the results of the assay. The small peak of non-fluorescing yDJGFP control cells will of course overlap the peak of the true non-fluorescing, xylose metabolising cells. However, despite this in all cases it can be visually seen from Figure 5.9 that over time growing in SC-X media, the non-fluorescent peak always rises to become dominant and thus the majority of the cells in the media are yDJM1. It takes more time for this to occur if the initial dilution of yDJM1 compared to yDJGFP is larger.

These data were further analysed to show the total number of cells measured at each time-point for each sample, along with whether they are GFP+ or GFP-. To quantify the GFP positive and GFP negative cell numbers, I set a gate in FlowJo analysis software according to sample yDJM1 0h wherein 97% of the cells were classified as GFP-, and I applied this gate to all samples. The gate was set to allow 3% of the entire population to fall into the GFP+ group, to prevent any skewing of data towards yDJM1 cells. This gating facilitated the calculation of the percentage of cells that were GFP+ versus GFP- and the results are shown in the five-stacked bar charts of Figure 5.10. Analysing the data from this perspective reiterates quantitatively how the non-fluorescing cell percentage increases over time, as a result of growth in SC-X media by yDJM1. In particular, it also provides a means of assessing cells numbers variation following SC-X growth. As expected, yDJGFP reaches a maximum cell number of approximately 70,000 units, in contrast to yDJM1 recorded a maximum of over 450,000 events. Overall, we see a decrease in the GFP+ cell percentages, and an increase in GFP- cells. Equally, in all samples cell numbers increase, although one sample that is noticeably slower to increase in numbers is 50:1. However this is explained by the smaller starting OD_{600} that was

measured at the start of the experiment of 0.02 A.U. (data not shown). The other samples ranged in starting optical densities of 0.028 to 0.035 A.U.

The Figure 5.10 data also shows that at zero hours for yDJGFP instead of 100% GFP+ cells, we only have 60.9% of cells due to the bimodal population seen above. This means that when I set up the 1:1 ratio at 0h, instead of having an equal percentage of GFP+ versus GFP- cells, there is actually 65.7% of GFP- cells and thus 34.3% of GFP- cells, despite having 50:50 inoculation. The same issue is seen in the 100:1 sample, wherein there should only be 1% GFP- population, but instead the percentages are 60% and 40% of GFP+ and GFP-, respectively. I recognise that the GFP- cells of yDJGFP will skew the data and intended ratios, however this does not affect the overall outcome, and I did not gate or apply bottlenecks to exclude this GFP- populations, and it was included in all statistical calculations.

The results of Figure 5.10 are explicit enough to allow for the following observations: (i) the yDJGFP strain does not successfully grow in SC-X media. This is further confirmed by the OD_{600} readings of the specimens (data not shown), wherein the OD_{600} only reached a maximum of 0.089 A.U., which is comparative to zero growth seen before (Figure 5.5); (ii) the yDJM1 strain outgrows yDJGFP in all of the experiments that were set up, becoming the dominant cells within each competitive population. In all the competition traces the outcome is one and the same: the yDJM1 results in producing the greatest peak of non-fluorescing cells, albeit at different rates depending on the initial ratio. For example, if we compare the peaks of 1:1 we can see that by time 48 hours yDJM1 growth has exceeded yDJGFP by almost double; however 48h for the 100:1 sample shows that the yDJGFP strain is still the dominant yeast strain in that mixture, and the fluorescing population is still greater than the non-fluorescing until time 96 hours.

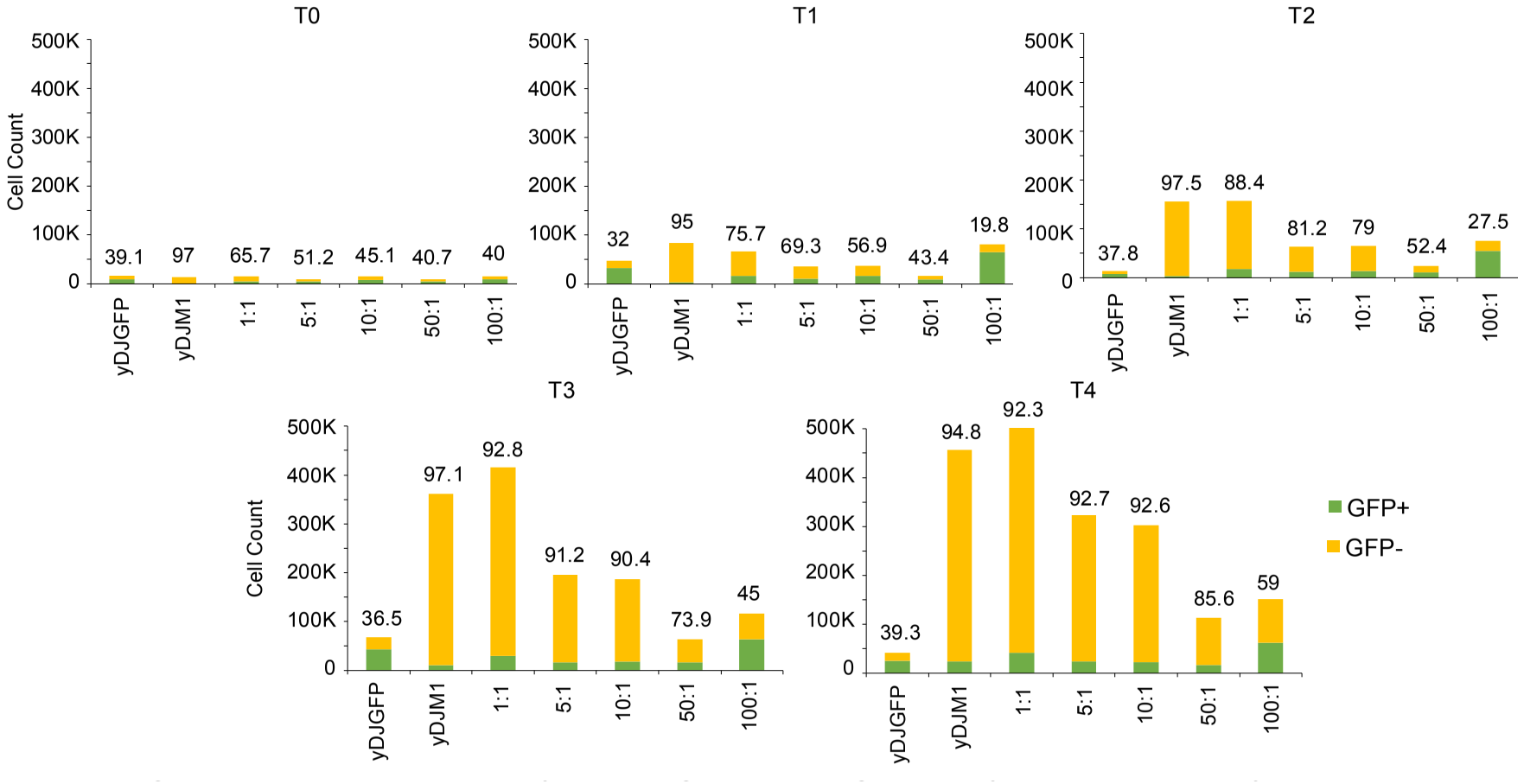


Figure 5.10 Summary of Competition Assay. Stacked bar charts showing the number of cells that are GFP+ (green) and GFP- (yellow) for each time interval. Apart from yDJM1 and 1:1, the yellow coloured stacks are the minority of cells. However, by T1 they are starting to dominate three out of five competition experiments (1:1, 5:1 and 10:1) and by T3 they have also dominated the 50:1 sample. In T4 all the mixed assays are mainly composed of non-fluorescing cells, in varying concentrations. The percentage of non-fluorescing cells of each sample are labelled above each bar. (n=1).

5.3.3 SCRaMbLE-in using Xylose Utilisation Genes XYL1, XYL2 and XYL3

Following the development of the SCRaMbLE-in method from Chapter 4, I next intended to expand SCRaMbLE-in from a single gene to the combinatorial recombining of multiple genes, potentially showing a link between metabolic engineering using loxPsym guided recombination in synthetic strains. Introduction of metabolic pathways and simultaneous evolution of host strain could lead to advantageous phenotypes in the synthetic strains not seen before, as well as providing evidence for the suitability and malleability of Sc2.0 strains as metabolic pathway hosts. In section 5.3.2 I showed that xylose metabolism via the oxidoreductase pathway was functional in our new dSynXI.A-C strain. I also concluded from the subsequent competition assays that yDJM1 always outcompeted non-xylose metabolising cells in SC-X media, albeit at different rates depending on initial ratios. Based on these findings, I hypothesised that the xylose-utilisation genes could be individually SCRaMbLEd-in, and then selected-for as a one-step method to develop a novel synthetic yeast strain capable of xylose catabolism.

Based on the results from Chapter 4, the best method to provide heterologous DNA for SCRaMbLE-in was as loxPsym-flanked linear DNA. The plasmids yDJG5, yDJG6 and yDJG7 were used as the template DNA for SCRaMbLE-in experiments (Figure 5.2). However, to prevent PCR-based mutations, the linear DNA encoding the genes was prepared by cleaving the expression cassettes from the plasmid DNA with BsmBI and a single cutter enzyme that cleaves the plasmid backbone. In each case, the desired resultant linear DNA band containing the expression cassette was gel extracted to separate the product from the plasmid backbone and ensure purity of DNA. The linear DNA cassettes recovered were named XYL1_2L, XYL2_2L and XYL3_2L and had the loxPsym format seen for KanMX_2L in Chapter 4.

These three DNA fragments were co-transformed into yeast in equimolar amounts and SCRaMbLEd-in as described in Section 4.3.4. Following a short recovery, transformants were plated on SC-X agar minimal media (with all amino acids supplemented) or on YP-X agar media. Plates were incubated at 30°C until colonies appeared, which took a long time due to the very slow growth rates with xylose as the carbon source. Following two weeks of incubation, all plates showed over ~1000 small satellite colonies, including the negative plates. Experiments were repeated wherein dilutions of SCRaMbLEd samples were made, but the same results were observed. Therefore, I sought alternative methods that would facilitate screening of SCRaMbLEd-in cells that could potentially metabolise xylose, but didn't require on having to wait many days for very slow growth on SC-X agar plates.

In order to place a quicker, stronger screening technique into the experiment that can select for cells that are transformants and are capable of SCRaMbLE-in, I added the KanMX_2L DNA along with the three xylose-utilisation gene DNAs. This time following SCRaMbLE-in, samples were first plated on YPD agar with G418s, allowing growth only of cells that had taken up the DNA and begun expressing KanMX. I reasoned that instead of screening immediately for slow xylose growth, I could first use KanMX to isolate the sub-population of cells that had taken up any DNA and had undergone SCRaMbLE-in. Within this sub-population, my hypothesis was that there would be a higher percentage of colonies that could have integrated the oxidoreductase pathways genes. This should therefore reduce the number of colonies needed to be screened, opening up the possibility of colonies screened in liquid SC-X media which shows faster growth than on solid SC-X agar plates (data not shown). The protocol I developed is illustrated in Figure 5.11 and the transformation efficiencies associated with this experiment are shown in Table 5.4.

In total, 51 colonies grew after SCRaMbLE-in on the G418s plate (6 on the control) when only KanMX_2L DNA was provided. This was an equivalent number of colonies to those seen for the same experiments in Chapter 4. When the KanMX_2L was not the only DNA provided (i.e. when the 3 xylose-utilisation genes were also added) there were only 12 viable colonies that grew on the G418s plate (1 on the negative control). The lower number of colonies could be related to KanMX_2L now only representing ¼ of the available DNA for SCRaMbLE-in. Indeed, 12 colonies is close to ¼ the number of colonies seen when only KanMX_2L is provided.

The 12 colonies were next screened for their ability to grow on xylose in two stages. Firstly, small scale growth assays were carried out in 10 ml of SC-X media, followed by larger scale assays, in 50 ml SC-X media (as performed for the growth assays in 5.3.1). Following both rounds of screening for growth, a single colony (named KX7) was identified that grew successfully in SC-X media with 4% xylose.

KX7 was assayed at regular time intervals for growth in 4% SC-X media and 4% YP-X media as had been done previously. The results for this are shown in Figure 5.12. From the graphs, we can see that the KX7 strain (red) grows visibly better than the untransformed dSynXI.A-C strain (grey) when in SC-X media. It also appeared to grow better in YP-X media, especially after 48 hours. A significant difference is seen when the growth rates and doubling time are compared in each media type ($P<0.05$; ANOVA), thus verifying this as a xylose catabolic strain.

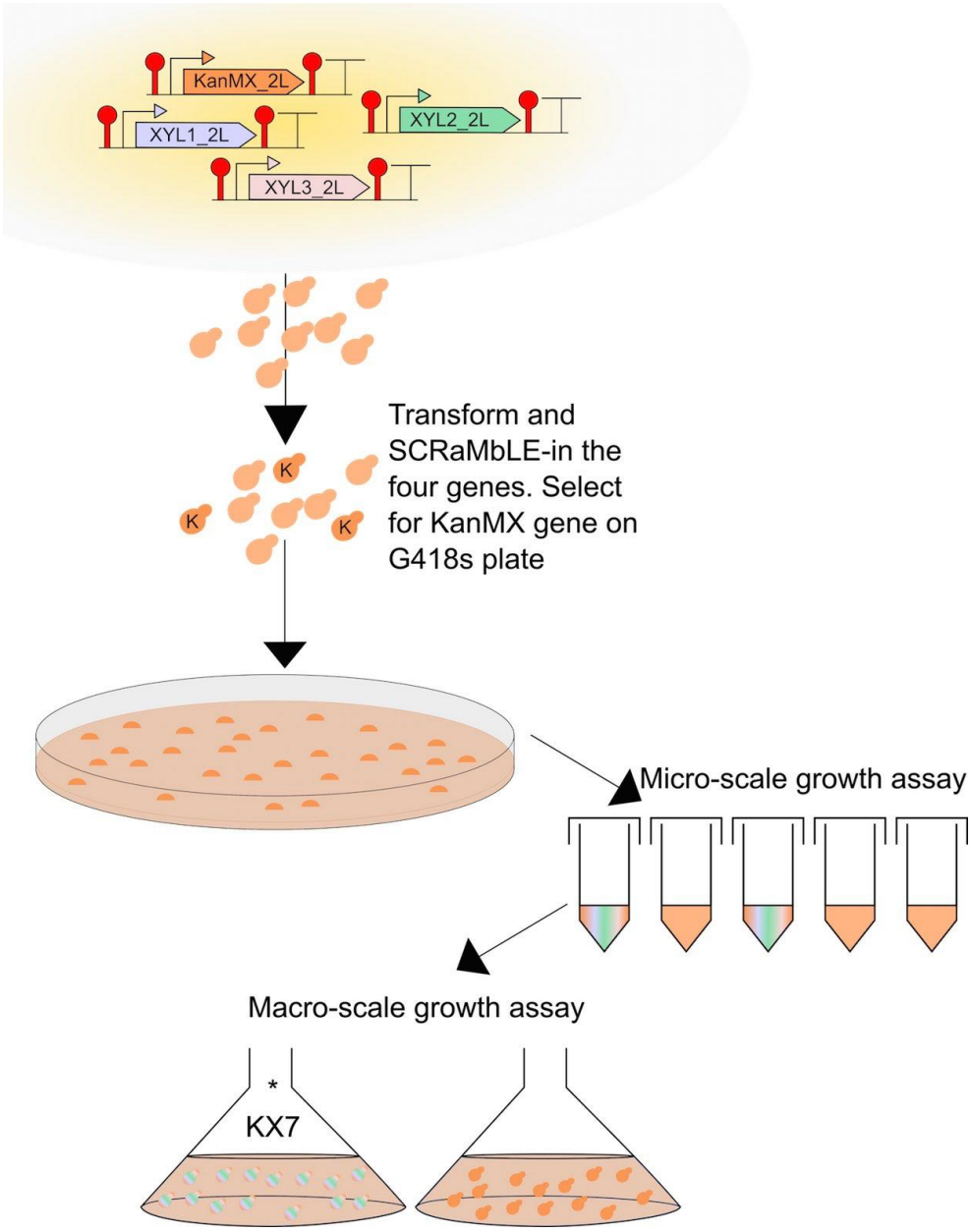


Table 5.4: Transformation Efficiencies followoing SCRaMbLE

	SCR+	SCR-
KanMX_2L	51	6
KanMX_2L + XYL1_2L + XYL2_2L + XYL3_2L	12	1

Figure 5.11 Schematic of SCRaMbLE-in of Multiple Xylose-Utilising Genes. dSynXI.A-C cells were transformed with 250 ng of KanMX_2L, XYL1_2L, XYL2_2L and XYL3_2L (each) and induced for SCRaMbLE. Cells were then plated on G418s media to first screen for those that had undergone SCRaMbLE. The transformation efficiencies are shown in Table 5.4. All colonies were inoculated in small volumes of SC-X media to assay for potential xylose consuming strains. Following increases in OD_{600} cells were then inoculated in large 250 ml Erlenmeyer flasks and screened for growth. The single strain KX7 was verified as xylose metabolising.

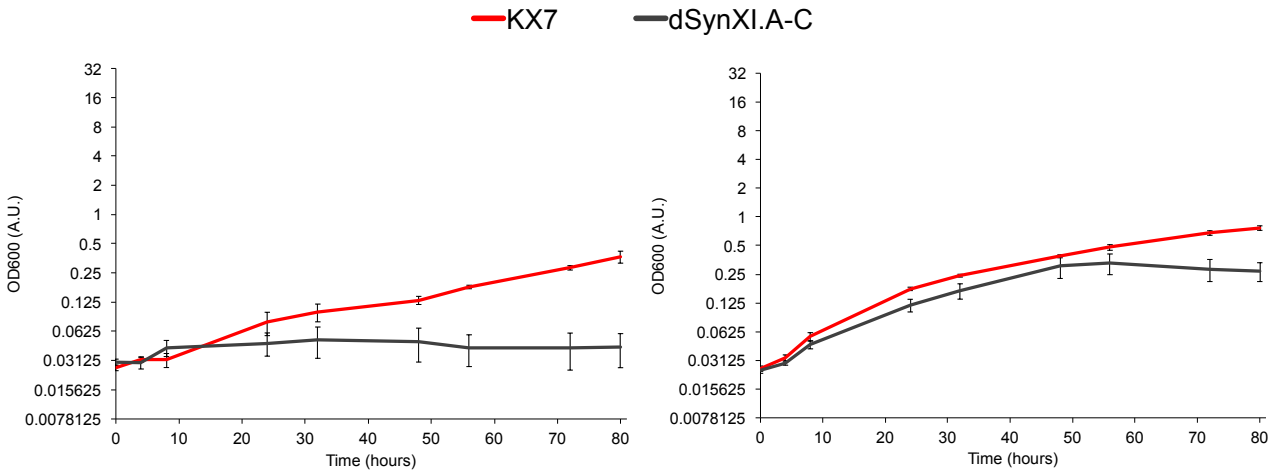


Figure 5.12 Growth curves of KX7 in SC-X media (left) and YP-X media (right). The average growth rate of KX7 in SC-X is 0.0043 and in YP-X is 0.0092, which are both significantly different from the untransformed growth rates. The average doubling times are 21.25 and 16.35 in SC-X and YP-X media, respectively, and significantly different when compared to dSynXI.A-C growth. ($P<0.05$; ANOVA). n=3 and error bars represent standard error.

These results thus show for the first time that SCRaMbLE-in can be used to added a combination of genes, and that these can encode a heterologous metabolic pathway.

The growth rate measured in SC-X for the KX7 strain was however notably less than that for yDJM3, which may indicate that the SCRaMbLE-in process had not introduced all three genes as expected or has lowered growth rate by causing unfavourable arrangements of its own chromosomal genes. Despite producing a xylose-catabolising colony, the loci of the three genes within the genome of this strain are not known, nor is what the new chromosome arrangement is post-SCRaMbLE. To verify the integration of the XYL genes and identify the surrounding sequences of each of these genes, KX7 genomic DNA was extracted and inverse PCR was carried out using divergent primers as in Chapter 4. PCR products from the inverse PCR were gel purified as before and single-read sequenced. Figure 5.13 shows the results of the PCR and schematic illustrations of the returned sequence ABI traces.

The inverse PCR successfully amplified all three XYL genes suggesting that these were all present in the KX7 genome as was hoped. However, the sequencing of the inverse PCR reads gave difficult to interpret results. The forward primer (pDJ052) of XYL1_2L was seen to amplify the 3' end of the XYL1 gene, a loxPsym site (which has presumably been recombined with another loxPsym site), and then the TEF2 promoter of XYL1. Its reverse counterpart primer, pDJ051, amplified the 5' end of the XYL1 gene and TEF2 promoter but did not produce a sequence read long

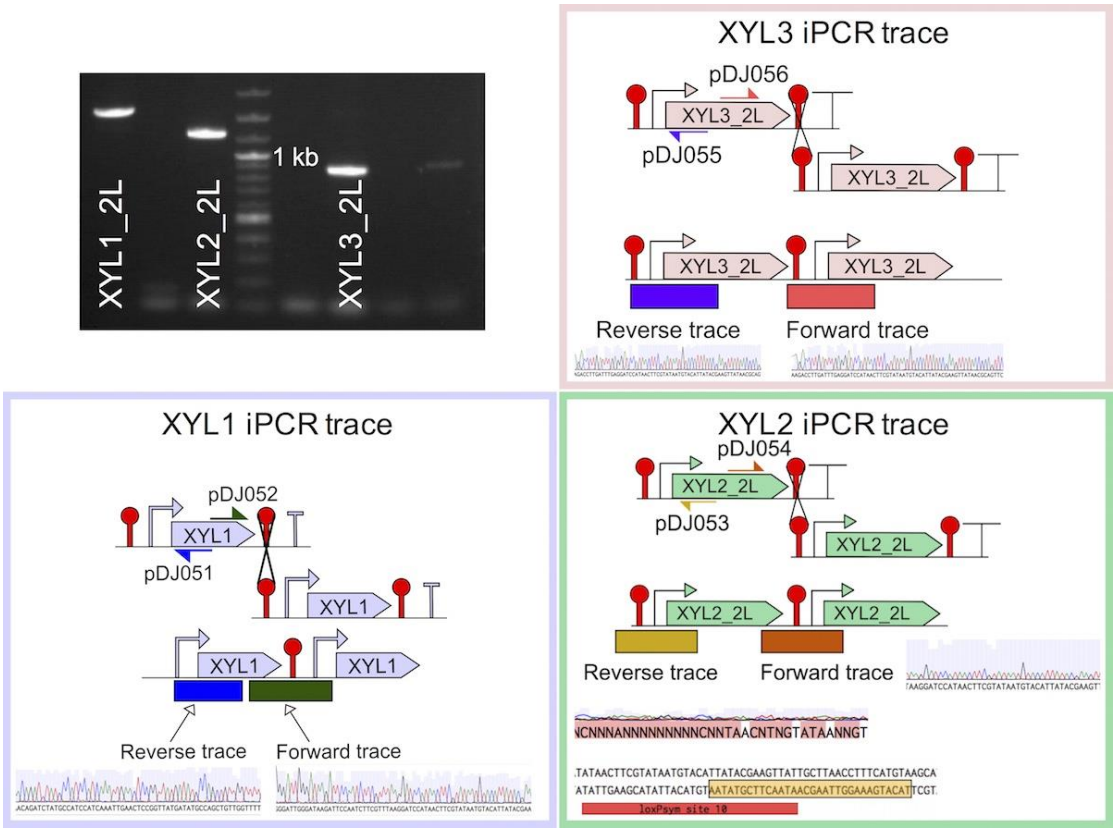


Figure 5.13 iPCR Analysis for KX7. Gel electrophoresis image of the iPCR results of XYL1, XYL2 and XYL3 amplification from KX7. Each band was purified and sequence verified. The three schematic boxes represent the results from the sequencing ABI traces. In almost all sequencing traces the results show that the linear fragments of DNA are duplicated. The primer of XYL2, pDJ053, has an inconclusive end trace, which suggests it could have been integrated downstream of the tenth loxPsym site of dSynXI.A-C, however evidence for this is not strong. A 2-log ladder is used as reference.

enough to show what was upstream. The trace from pDJ052 suggests that recombination at the downstream loxPsym sequence has occurred, deduced by the loss of the terminator, which was part of the linear DNA fragment at the time of transformation. Sequencing data at this loxPsym site also does not provide evidence for genomic integration, thus recombination has either occurred between two separate XYL1_2L fragments which then integrated into the chromosome at a location that this sequencing cannot determine, alternatively the XYL1_2L DNA could have simply circularised on itself. The traces for pDJ054 (XYL2_2L) and pDJ056 (XYL3_2L) show the exact same scenarios as described for XYL1_2L, although in both cases the forward primer traces are longer and provide better evidence of a second XYL2 or XYL3 gene, respectively. Interestingly, the reverse primer of XYL2 (pDJ053) loses trace integrity towards the end of the sequencing data, however using a Benchling algorithm, there is weak evidence suggesting that the 5' loxPsym site of the first XYL2_2L fragment has recombined with the 10th loxPsym site of

dSynXI.A-C. However, without further evidence, conclusions from this data cannot be made in confidence. Given more time, further sequencing of the KX7 strain would be pursued.

5.3.3 SCRaMbLE-in of Oxidoreductase Pathway Genes into dSynXI.A-L

Following the seemingly positive SCRaMbLE-in of the oxidoreductase pathway using dSynXI.A-C, I next investigated the effects of repeating the same process in diploid strain dSynXI.A-L. I hypothesised that given the longer region of synthetic chromosome and the availability of more loxPsym sites, the three linear XYL genes could potentially integrate into multiple loci and at multiple copies, and so possibly produce a more-optimised version of a synthetic xylose metabolising strain. Experiments were repeated as for dSynXI.A-C, and colonies were plated on G418s YPD agar as before.

A good colony yield was obtained (shown in Table 5.5), and numbers were similar to the previously obtained transformation efficiencies of dSynXI.A-C. Due to quite large numbers of colonies obtained, 11 randomly isolated colonies were picked at random, and were screened for the presence of the xylose-utilising genes that were SCRaMbLEd-in. In an attempt to reduce the rounds of screening, growth and false positives, a PCR amplification was set up using convergent primers pDJ057 and pDJ058 (XYL1), pDJ059 and pDJ060 (XYL2), and pDJ061 and pDJ062 (XYL3) to amplify the presence of any of the genes from extracted genomic DNA. The PCR failed to amplify the presence of XYL1 and XYL3, but did produce bands for four of the 11 colonies for XYL2 (data not shown). Prior to setting up growth curves, an inverse PCR was set up for all the colonies, however no bands were produced by the end of this process (except for KX7 which was used as a positive control). Further growth analyses were not pursued, however future experiments are suggested for the continuation of SCRaMbLE-in screening in section 5.4 of this Chapter.

Table 5.5 Transformation Efficiencies of SynXI.A-L using KanMX_2L only or KanMX_2L with the oxidoreductase pathway genes.

Heterologous DNA	SCR+	SCR-
KanMX_2L	78	13
KanMX_2L, XYL1_2L, XYL2_2L, XYL3_2L	49	6

5.4 Discussion

5.4.1 The Oxidoreductase Pathway in Yeast showed Growth on Xylose Media

In Chapter 4, SCRaMbLE-in was shown to work with a single gene, and in this Chapter the aim was to show SCRaMbLE-in could be applied to multiple genes part of a metabolic pathway. The isomerase and oxidoreductase xylose-utilising pathways were perfect models to test SCRaMbLE-in with, as the isomerase pathways was made up of two genes, and the oxidoreductase pathway was made up of three genes, showing a gradual increase in the number of incorporated genes. However, the isomerase pathway genes, xylA*3 and XYL3, did not enable growth when assayed following transformation into dSynXI.A-C. The previously reported growth from Lee *et al* (2012)[111] was based in an isogenic BY4741 strain that overexpressed a TAL1 gene (a transaldolase enzyme part of the non-oxidative PPP) and the XKS1 gene (XYL3), as well as having a GRE3 knockout (an aldose reductase enzyme). This obtained a growth rate of 0.16 h^{-1}.

For the purposes of this study, no knockouts or changes were made to dSynXI.A-C genotype extending past the synthetic chromosome, as the host strain needed to be reflective of an unmodified Sc2.0 strain. Furthermore, any changes made to the host strain for the isomerase pathway genes would need to be included to the host for the oxidoreductase pathway genes as a fair comparison. The deletion of the GRE3 gene and the overexpression of the TAL1 gene has also been reported to improve growth of *S. cerevisiae* yeast on xylose when harbouring the oxidoreductase pathway[176], and the likelihood is that the modified genetic background used by others would have facilitated better growth for both the xylose-utilising pathways in this study. A decision was made not to alter the dSynXI.A-C strain for two reasons: i) the dSynXI.A-C strain should remain as similar to the Sc2.0 design as possible – reflective of the end strain that will be generated with the combination of all synthetic chromosomes and ii) the overexpression of TAL1 would require the presence of auxotrophic markers that were not available in the dSynXI.A-C (without further knockouts, which again, would lead to the dSynXI.A-C strain deviating away from the Sc2.0 based strains). In an ideal scenario the SCRaMbLE process could theoretically generate a strain that results in the deletion and overexpression of favourable genes, based on the environmental selective pressures. Deletions and duplications are now acknowledged gene rearrangements that arise as a result of SCRaMbLE[82], therefore the deletion of the GRE3 gene (found chromosome VIII) and the duplication (and hence overexpression) of the TAL1 gene (found chromosome XII) are possible SCRaMbLE outcomes. This case underlies the value of being able to rapidly diversify the yeast genome.

Without these changes, sustainable aerobic growth of the yDJM2 strain in pure xylose media (SC-X or YP-X) was not seen, and the optimisation of this strain was not pursued, nor was it used for consequent experiments. The oxidoreductase pathway based strains, yDJM1 and yDJM3, did show significant growth in SC-X and YP-X media when expressed in dSynXI.A-C. Furthermore, their growth rates were similar to those reported previously[105, 176], in unaltered *S. cerevisiae* strains, which exemplified that the presence of the partially-synthetic chromosome XI in dSynXI.A-C did not seemingly affect the functional expression of the xylose metabolising pathway genes. The growth rates for yDJM1 and yDJM3 were 0.011 h^{-1} and 0.015 h^{-1}, respectively in SC-X, comparable to the industrial strain TMB3001 which grew at 0.01 h^{-1} as reported by Johansson *et al* (2002)[105].

Despite driving transcription using the same constitutive promoters and terminators, the expression of the chromosomal integrated plasmid is likely to be more stable than expression of the YCp plasmid. The plasmid is theoretically present in 1 to 2 copies per cells matching the copy numbers of chromosomes, however centromeric plasmids are on average lost every 10^2 divisions, and bad mitotic fidelity can result in the transmission of none or double the plasmids into daughter cells, leading to either a growth advantage or decreased growth[175]. The results I saw are more inclined to show decreased growth, mirroring the probability that plasmids are being lost from cells every few generations. Nonetheless, growth rates of both the yDJM1 and yDJM3 strains were not significantly different from one another, and showed a high degree of similarity.

5.4.2 Establishing the Conditions for Selecting Post-SCRaMbLE-in

Having established growth of the yDJM1 and yDJM3 strains in xylose media, several attempts to optimise the growth rates were made, and to reintroduce a logarithmic growth phase during the assays. The hypothesis was that by implementing a greater difference in optical density and growth rates obtained between the background (dSynXI.A-C), yDJM1 and yDJM3 strains, the screening and isolation of post-SCRaMbLE colonies that grow on xylose would be faster, easier and more efficient. Initially, mixed media was used, injecting growth assays with small doses of glucose (Figures 5.7 and 5.8), which resulted in greater growth rates and greater final ODs. Although this was the desired effect, the small glucose injections of 0.05% and 0.01% also resulted in much greater growth of the background strain, which would have proved problematic in the differentiation of post-SCRaMbLE colonies that were truly xylose-utilising, had it been used.

Furthermore, the background strain does not experience death in pure xylose media, probably as a result of intracellular carbohydrate storage and residual glucose from the overnight cultivation[176]. This theory is further supported by the fraction of growth seen in the untransformed strain, increasing from a starting OD_{600} of 0.03 A.U to 0.051 A.U. at 32 hours in SC-X media, after which the optical density begins to decline (Figure 5.5). Despite the dSynXI.A-C untransformed strain possessing orthologs of the oxideoreductase genes within its genome[114], the expression of these does not produce functional enzymes that enable metabolism of the D-xylose. This strain is grown with all essential amino acids provided, but given that xylose is not a lethal environmental factor (for example like G418s) and will not cause cell death, yeast cells have the capacity to survive approximately 24 hours of carbon starvation[177]. However, continuous depletion of carbon leads to cell morbidity due to the shutting down of intracellular machinery as a result of loss of protein, carbohydrate, lipid and/or nucleic acid function[178]. This decay effect is seen in Figure 5.5.

Given that there is evidence of basal growth from the non-xylose catabolising strain, I aimed to determine if the xylose-metabolising strains were able of outcompeting background strains. To facilitate quick and easy screening by flow cytometry, the sfGFP gene producing green fluorescent protein was transformed into dSynXI.A-C to make yDJGFP and used as the background strain. yDJM3 and yDJGFP were produced to be otherwise exactly the same in genotype, and were cultured in SC-X media at different ratios. The results (shown in Figure 5.10) indicated that i) the starting ratio of yDJM1 to yDJGFP greatly affected the rate that yDJM1 outcompeted the non-xylose metabolising strain, and ii) yDJM1 outcompeted yDJGFP in all scenarios. Useful insights were gained from this experiment, as it showed that as the quantity of yDJGFP increased relative to yDJM1, it required yDJM1 much greater periods of time to outcompete the non-xylose metabolising strain.

The highest ratio assay carried out was 100:1 and showed that the yDJGFP strain possessed the ability to increase in both cell numbers and GFP production for a longer period of time prior to dying out. Whilst these experiments showed a positive result in that yDJGFP was always outcompeted, the likelihood is that following SCRaMbLE-in of three separate xylose-utilising genes, each provided at 250 ng, few yeast cells will recombine all three genes into a single cell to generate a xylose-metabolising strain. The more realistic ratios of non-xylose-metabolising strains to those that were able to grow are closer to 10,000:1, possibly even 100,000:1 or

more, and much longer cultivation times would be required for yDJM1 to outcompete vast numbers of non-metabolising strains.

One minor flaw in the design of competition assay is that SCRaMbLE-in will in theory cause the chromosomal integration of xylose-utilisation genes. Therefore, the yDJM3 strain, which has the three genes integrated into the URA3 locus in its genome, is likely to be a better representative of growth following SCRaMbLE. However, it can be argued that difference in growth between the yDJM1 and yDJM3 strains is negligible, and thus results would have reached the same conclusions. Furthermore, the weaker strain of the two was used, and yet it still showed successful competitive growth over the non-xylose metabolising strain.

Selection on SC-X or YP-X agar plates was not useful for isolating xylose SCRaMbLE-in integrants, as there was noisy background growth as a result of the survival of the dSynXI.A-C strain and very long incubations were required. Following the competition assay that showed yDJM1's capability of outcompeting a non-xylose catabolising strain, a potential method that could have been used to isolate a SCRaMbLEd xylose-metabolising strain was to resuspend the SCRaMbLEd cells directly into liquid SC-X media, rather than plating. If a xylose-metabolising SCRaMbLE-in strain was successfully generated, it could be isolated following serial dilutions and incubations. However, this method was not used due to the uncertainty of whether a positive strain would arise from the experiments as well as the consideration of the invested time that would have been necessary for this method. As mentioned before, it took a long time for yDJM1 to outcompete yDJGFP at the 100:1 ratio, taking 96 hours of incubation for the GFP- percentage to go above 50% (Figure 5.10). In reality in the post-SCRaMbLE-in culture the number of background cells that haven't integrated xylose-utilising genes successfully is likely to be far greater than those that have, and so starting ratios will greatly exceed 100:1. Thus on average, it is estimated that for a xylose-metabolising strain to start outcompeting its competitors is likely to require a minimum of seven days, followed by subsequent serial dilutions and cultivation to dilute the cells and repeat the process, until a culture comprised of only xylose-metabolising cells is generated. I estimated that at least two dilutions would be necessary to increase numbers of xylose metabolising SCRaMbLE-in cells, and on average seven days per cultivation round. This would result in a total of three weeks to obtain a single, optimised, xylose-metabolising colony. The efficiency of this method was too low to pursue, thus I tried different methods to see if I could isolate a SCRaMbLEd xylose-utilising strain in a more time-effective manner.

5.4.3 SCRaMbLE-in can be used for the integration of metabolic pathways

The work in this Chapter showed that SCRaMbLE-in can be applied to simultaneously induce metabolic and genome engineering. I successfully provided, induced, integrated, isolated, confirmed and assayed a xylose-metabolising yeast strain following SCRaMbLE-in. From the previous experiments and sections in this discussion, screening for successfully integrated colonies proved problematic, therefore a more rigid and severe screening method was implemented – using KanMX from Chapter 4 as one of the genes during part of the SCRaMbLE-in procedure. Selection on G418s media did not allow for background colony growth, and I hypothesised that this preliminary step could be included to screen for cells that had been transformed and had undergone SCRaMbLE. Within resistant colonies, it was possible that a sample had SCRaMbLEd-in all four genes at one.

It was a big step to go from SCRaMbLE-in of a single gene to SCRaMbLE-in of four separate genes with the expectation of all four genes being integrated, in order to input a new function in the yeast. But given the issues that were experienced from the background growth when plating on xylose-containing agar, this was attempted nonetheless. Following G418s screening and then growth cultivation in xylose-media, I isolated the resultant strain KX7, which was eventually confirmed as a xylose catabolic strain by both growth assay and by preliminary iPCR results. The first observation was that the number of cells that arose from the SCRaMbLE-in of four different linear DNA fragments, was a quarter of the number of recovered colonies as opposed to when only one DNA fragment was provided. The KanMX_2L gene was decreased in quantity from 500 ng (used in Chapter 4) to 250 ng (used in Chapter 5) to compensate for the inclusion of the other three linear genes. The 52 colonies that arose from 250 ng of KanMX_2L only (Table 5.4) is halved when compared to the numbers of colonies obtained in Chapter 4 (Figure 4.8) where the smallest number of colonies from 500 ng of DNA was 102 for three hours of induction. This is expected, as the amount of DNA added to the mixture is also halved. However, when the xylose-utilising genes are included in SCRaMbLE-in as well, the number of G418s-resistant colonies is a quarter than when KanMX is the sole heterologous DNA source. Possibly, SCRaMbLE-in is more being affected by the concentration of DNA added to the cell, affecting the actual transformation efficiency. The bottleneck could therefore be at transformation rather than during SCRaMbLE.

This is very interesting data, as this was not an aspect of the development of SCRaMbLE-in that was investigated. Limitations of SCRaMbLE-in, such as the concentrations of DNA and number of different DNAs that can be provided are

currently not established fully yet, and this Chapter reiterated this and also provided scope for investigating these factors further. The findings in this chapter suggest that the provided DNA concentration of linear genes could be a major factor in the driving of successful SCRaMbLE-in of genes. It is also speculated that the reduced number of (12) colonies that arose from the SCRaMbLE-in of KanMX_2L, XYL1_2L, XYL2_2L and XYL3_2L was directly affected by the size of the region of available synthetic chromosome in dSynXI.A-C. There are 23 loxPsym sites in 87,356 bp of synthetic DNA. The complexity of the gene rearrangements occurring between these loxPsym sites is nearly infinite, and remains to be ascertained. This theory is further backed up by the repeat of SCRaMbLE-in of KanMX_2L, XYL1_2L, XYL2_2L and XYL3_2L in SynXI.A-L, which has a much greater region of synthetic chromosome XI sequence available (425,015 bp) and approximately 112 loxPsym sites encoded in that sequence. Despite not confirming a colony as positive for the presence of the xylose-utilising genes, the number of G418s-resistant colonies that arose from SynXI.A-L was four times greater than those from SynXI.A-C, which is directly proportional to the greater number of loxPsym sites available for site-specific recombination of heterologous genes into the synthetic chromosome, i.e. the length of synthetic chromosome of SynXI.A-L is four-fold longer than SynXI.A-C.

Thus, the concentration of heterologous DNA is seemingly an issue with partially-synthetic chromosomes, but might not be an issue when there will be 5,000 loxPsym sites available, as in the final Sc2.0 strain containing all 16 synthetic chromosomes. I predict that the SCRaMbLE-in system that I developed here will be able to be scaled-up and implemented to include in one reaction a multitude of different loxPsym-flanked genes when larger regions of synthetic chromosomes are available. This would in a single step create a large number of diverse libraries of Sc2.0-based strains, each with minimised, extensively rearranged genomes, and potentially each providing new functions encoded by incorporated heterologous genes.

KX7 was assayed for growth in SC-X and YP-X media, and the results are shown in Figure 5.12. This was the first step that I used to show that a xylose-catabolising strain was generated from the SCRaMbLE-in of the three xylose utilisation genes into the dSynXI.A-C host strain. The aim of SCRaMbLE-in of the xylose metabolic pathway genes was to generate a laboratory-evolved, optimised, xylose-metabolising, partially-synthetic strain. The above was achieved, although the proliferation of this strain in SC-X and YP-X media was not optimal, and the new growth rates (0.0043 h^{-1} and 0.0092 h^{-1}, respectively) were reduced when compared to either the yDJM1 or yDJM3 strains. Despite being more similar to the yDJM3

strain, as in theory the xylose-utilising genes were integrated into the genome, the growth rate of KX7 was more reflective of yDJM1, from xylose-utilising genes expressed from a centromeric-based plasmid. At this point the reasons for this reduced growth were not clear: gene rearrangements that were detrimental to the KX7 could have occurred, however these should be compensated for by the diploid strain, and the presence of an extra non-synthetic genotype, therefore this was not considered to be the case.

A possible explanation for the decreased growth rate seen from Figure 5.12 could be as a result of gene copy number of XYL1, XYL2 and/or XYL3, as well as the expression of the three genes. With respect to the copy number, iPCR sequencing (Figure 5.13) of KX7 seemed to reveal that all three genes were duplicated (at least once), and that there was no evidence provided for integration into the genome between two loxPsym sites. These are preliminary data that ideally should be confirmed. However, they could indicate that the genes are present in more copies than planned. Overexpression of XYL1 and XYL2 could potentially lead to a greater cofactor imbalance – something that is an inherent drawback of the oxidoreductase pathway. The cofactor imbalance of the oxidoreductase pathway leads to the formation and build up of xylitol, which is toxic to cells at high concentrations. It is possible that the overexpression of XYL1, which reduces xylose to xylitol, is causing this build up and resulting in slower growth.

Lastly another explanation, and what I consider the most likely, is that SCRaMbLE-in of the xylose-utilising genes causes disruption in the 3'UTRs of both the XYL genes and native genes, which in turn results in transcriptional interference and therefore reduced expression of the enzymes encoded by one, two or all three genes. The limitations of using iPCR have been previously discussed in Chapter 4, however it did provide enough evidence to draw two conclusions from the data obtained in this Chapter, which were: i) iPCR sequencing showed that recombination of the XYL1, XYL2 and XYL3 genes had in fact occurred – confirmed by the lack of a terminator sequence (which was provided on the linear DNA fragments prior to SCRaMbLE-in) in the ABI traces; and ii) there was a lack of evidence to conclude that integration of the xylose-utilising genes into the genome had occurred, as the ABI traces revealed the consecutive amplification of each gene. If the iPCR data is to be trusted this means that there are two possibilities. Either the XYL genes have integrated into the synthetic region of the genome but always have inserted with at least two copies at each site (i.e. integration of doubles or more), or a more radical outcome is that the Cre recombinase has instead circularised the LoxPsym-flanked

DNA fragments and that these are somehow present and expressed in the cell without being chromosomally integrated.

In terms of the first possibility, which seems more likely, the integration of multiple copies could reduce gene expression by causing disruption of the 3'UTRs of all of the XYL genes. While the SCRaMbLE study conducted by Shen *et al* (2015)[82] claimed that the disruption of the 3'UTR regions of native genes did not result in loss of gene expression, my work in this thesis appears to be this finding for non-native genes, especially when we examine the flow cytometry data from Chapter 4 which shows that SCRaMbLE clearly alters gene expression of genes that are not deleted. The work in this Chapter thus shows a possible new phenomena for SCRaMbLE, the formation of un-natural 3'UTR regions around heterologous genes that in the least reduces their expression in an undesired way.

The second possibility seems less likely. How can the DNA be maintained and expressed if they are simply just small DNA circles without integration anywhere in the genome? The initial iPCR data strongly suggests this. For a start, all three genes show recombination only with themselves (instead of for example, XYL1 recombination with XYL2 or XYL3 recombination with XYL1), and in none of the ABI traces are there evidence for two loxPsym sites. Furthermore, the sizes of the bands generated from the iPCR provided some evidence that this is the case, based on the annealing sites of each primer provided per gene. For example, the small band generated from the amplification of XYL3_2L shows amplification of only the end of the XYL3 gene, and the promoter sequence, forming a band of approximately 900 bp. This is directly representative of the amplification of primers pDJ055 and DJ056 that were designed to be divergent, but have become convergent with the circularisation of the XYL3 gene. The counter-argument against this though is that the same result would be generated (and probably has been) from the recombination of sequential genes, one next to the other, wherein the DJ056 primer of the upstream gene convergently amplifies any sequence between itself and the DJ055 primer of the downstream gene (Figure 4.12 provides a schematic for this, using KanMX_2L as an example). The identical conclusions can be made for XYL1 and XYL2, based on the annealing sites of their respective primers and the bands generated as a result of these.

Interestingly, a very recently published study by Demeke *et al* (2015)[179] showed that the successful expression of a xylA gene was achieved in yeast through the formation of an extrachromosomal circular DNA element (eccDNA), showing for the first time that yeast has the ability to form *in vivo* plasmids from genomic DNA that contains heterologous genes. However, this study also showed that all the

observed eccDNA contained a replicating site in the form of an ARS – and thus can replicate and maintained in the yeast cell. As the XYL genes SCRaMbLEd-in are done so as linear fragments of DNA, it is highly unlikely that they are replicating as eccDNA, concluding that either they exist as integrated and doubled genes in the synthetic chromosome, or the iPCR assay has not worked properly, and has generated inconclusive results.

Ideally, to confirm that this is the case in my KX7 strain, further genome sequencing would be done to provide the full insight as to whether this was in fact the case or whether the multiple integration theory is true. Both Shen *et al*[82] and Demeke *et al* (both 2015)[179] were able to report extensively about the effects of SCRaMbLE and the production of extrachromosomal circular DNA plasmids (respectively) following genome sequencing analysis, providing them with information about gene copy number, the full extent of rearrangements occurring, and the underlying mechanisms associated with each publication. Adding genome sequencing to this study would have determined whether the iPCR products generated were truly representative of the results presented in this thesis, or whether it was more the case that the apparent circularisation of the genes was a result of favoured amplification of now-convergent primers, thus outcompeting the amplification of any potentially integrated genes that may or may not have occurred. While expensive and time-consuming, future work on this strain should ideally be accompanied by a determined genome sequence. This would also help determine the gene copy number in each case as well (from the depth of sequencing reads). The iPCR amplification of the circularised xylose-utilising genes only confirms that the linear fragments are there (and that they may have circularised), but it does not confirm the number of these fragments in the cell. Any future studies that continue the development of metabolically-engineering SCRaMbLE-in of strains will find it invaluable to genome sequence any generated synthetic strains in order to understand the mechanisms and consequences of SCRaMbLE-in better. Despite trying to use both flow cytometry with a novel fluorescent strain and use iPCR in this thesis, neither method proved to be as conclusive as genome sequencing would be.

5.4.4 Conclusions

Previous publications on introducing xylose catabolism into yeast tended to focus on engineering and manipulating a more suitable *S. cerevisiae* host cell for xylose utilisation pathway expression[166, 180-184]. However, in this study I instead aimed to enable dSynXI.A-C to metabolise and grow on xylose in one step by using SCRaMbLE-in. I focused on using this pathway as a test case for SCRaMbLE-in and as such there were no industrial motives in terms of producing a commercially-competitive fermenting, ethanol secreting yeast strain. The work here instead exemplified that SCRaMbLE-in could be used to metabolically engineer a synthetic yeast strain, providing it with a novel function.

Three xylose-utilising genes that are part of the oxidoreductase pathway and a KanMX_2L gene formed in Chapter 4 were SCRaMbLEd-in to dSynXI.A-C, and produced a xylose catabolising strain, capable of growth in both SC-X and YP-X media with xylose as the main carbon source. Further analysis of the xylose-metabolising generated strain KX7 confirmed the presence of the xylose-utilising genes through iPCR. Subsequent sequencing of the amplified bands suggested either that all genes had integrated into the genome in multiple copies or that we were seeing formation of self-maintaining, extrachromosomal circular DNA by the DNA recombining back with itself.

While the iPCR data is only preliminary and is somewhat inconclusive, either scenario would offer interesting insights into recent breakthroughs, especially if extrachromosomal circular DNA is present, which is an interesting new area of research. It could also offer new insights if disruption of the 3'UTRs and the generation of transcriptional interference causes decreased growth on xylose-media when the XYL genes are integrated in multiple copies. The fact that in either scenario there is still sufficient expression of the XYL genes to at least produce a xylose-metabolising partially-synthetic strain (KX7) but it's performance is below what would be expect, means that the reasons for the reduced growth would likely be a measure of the effects of the way the genes are arranged or where they are maintained.

Ultimately the work here shows that SCRaMbLE-in can be done for multiple genes and can create heterologous pathways in a single step. SCRaMbLE-in of multiple pathways paves the way for the use of further more fully-synthetic strains in metabolic engineering projects. Theoretically these may uptake and integrated large numbers of heterologous genes provided on linear DNA, and can evolve and adapt into new rearranged, minimal genomes with the ability to carry out functions not seen naturally in yeast.

6.0 CHAPTER 6: Discussion and Conclusions

6.1 Construction of a Synthetic Chromosome XI

6.1.1 Summary

The Sc2.0 project aims to produce a fully synthetic, redesigned eukaryotic genome completed by the end of this decade. The Ellis Lab at Imperial College London took on the synthesis, assembly and integration of synthetic chromosome XI, and in Chapter 3 of this report, I reported on my work and the results for doing the first few mega-chunks of this project. The findings from this chapter did not deviate from our expectations of this project based on the methods previously demonstrated by the other teams. Essentially, the mega-chunk assembly method worked and the chromosome design meant that we could replace wild-type sequence with synthetic sequence without loss of fitness. Apart from some initial synthesis and cloning problems associated with the repetitive sequence of the first mega-chunk A, the process of assembly worked efficiently. The discovery that the synthetic design excluded an essential tRNA from mega-chunk B helped us understand why this step was not working. Progress on the building of synthetic chromosome XI has remained unhindered, and the Ellis Lab is on track to conclude the construction of synthetic chromosome XI, likely this year.

6.1.2 Essential tRNA Genes Require Re-integration for Cell Viability

Following the receipt of the synthetic chromosome XI design from JHU, the synthetic DNA sequence for the first 90 kb was outsourced for synthesis by GenScript. The aim of Chapter 3 was to extract, assemble and integrate the synthetic DNA in the form of assembled mega-chunks and transform into the BY4741 strain. Transformation of the synthetic mega-chunk DNA needed to integrate into the target chromosomal locus and replace the wild type chromosome sequence using homologous recombination. Mega-chunk A showed that this process of recombining synthetic DNA to recombine wild type DNA out was functional in yeast strain BY4741, and initiated the sequential building of the synthetic chromosome.

There was a design flaw encountered during the integration of mega-chunk B that arose as a result of the planned exclusion of all tRNA genes from the synthetic sequence of chromosome XI and the relocation of these genes onto a neo-chromosome under-construction. Usually this would not have posed an issue, as many copies of the same tRNA genes exist scattered around the yeast genome. However, the TRT2 tRNA gene was an essential tRNA and so needed to be encoded

elsewhere into the genome, prior to the integration of the synthetic mega-chunk B. During my work on this project, we were the first group to experience total cell lethality due to the exclusion of a tRNA gene, as no other teams had yet got to a point of deleting a tRNA gene that was unique in the genome. We also became the first team to reintroduce a tRNA gene back into the synthetic genome in order to solve this problem.

The native chromosome XI contains a total of 16 tRNA genes, of which the first 13 are deleted by the synthetic mega-chunks that have so far already been PCR verified (chunks up to mega-chunk M as of March 2016). Thus far only the TRT2 (threonine tRNA) gene has shown cell lethality following integration of a synthetic mega-chunk and so it is confirmed that this is the only one that is essential. In addition, no fitness defects have been seen by our lab so far following removal of the other 15 tRNAs.

When the other Sc2.0 collaborators experience similar problems (TRT2 is not only essential single copy tRNA gene in the *S. cerevisiae* genome) they can follow the methods we took here to relocate the gene to another chromosome to allow chromosome progress to continue. However, in the long term it will be more desirable to ensure that all the tRNA genes have functional expression from the neo-chromosome that is under construction at Edinburgh University. For completion of the full Sc2.0 project, pairs of chromosomes will be brought together by mating and ideally, the starting strain will be the one containing the neochromosome. Likely the synthetic strains will first need to be mated with the strain containing the neochromosome. Any strains that have been altered to reintroduce tRNA genes during their construction of their respective synthetic chromosomes will also need to be restored to their original designs following successful mating with the neochromosomal strain.

6.1.3 Verification of Integrated Mega-chunks Requires Time-consuming Screening

There is some scope for improvements and optimisations to be made to the processes of constructing Sc2.0, from the integration of mega-chunks to the time-consuming verification steps, such as replica-plating. In my work, I replica plated simply by re-spotting picked colonies, however, I also tested doing replica plating in a single step with the use of a velveteen cloth to simultaneously transfer all the colonies from one plate type onto another. This method is now in use in our lab. The transfer of colonies from one plate to another with this method proved fast and simple, however the incubation of the two freshly replica-plated samples required at

least 18 hours to allow for sufficient growth for any positive colonies. Following this, any phenotypically correct samples are grown overnight to full saturation in glucose rich media, and then their genomic DNA is extracted using a fast, simple and effective chelex prep[123]. The ensuing PCR verification step is the most-time consuming and labour-heavy process carried out throughout the sequence of processes for Sc2.0 construction. On average, assuming that each step associated with the preparation, construction, transformation and verification of synthetic DNA is successful (or rather yields a single successful colony), the total time necessary per mega-chunk is approximately two weeks. In reality, the low yields and low efficiencies of the various Sc2.0 steps mean that verification on a correctly integrated mega-chunk is closer to four weeks, even when no issues arise. The long time per mega-chunk, further reiterates the need for multiple collaborating labs to work in parallel to complete construction of all 16 synthetic chromosomes.

I briefly carried out experiments to test whether assembly of synthetic chunks could be recombined *in vivo*, thus eliminating the DNA ligation and concentration stages. The results I obtained resulted in inconclusive data, but generally they indicated that the Sc2.0 outlined methodology for standard mega-chunk preparation was the best, yielding the greatest number of colonies. Furthermore, the approach of using overlapping DNA fragments between chunks for *in yeasto* recombination would not have been a cost-effective method as extra overlapping DNA would always need to be synthesised at the start and end of each chunk, probably adding a 5% increase to the amount of synthetic DNA required. Further attempts at optimising or improving upon mega-chunk assembly were not pursued in the rest of my thesis.

The leading research group of Sc2.0, now situated at NYU and led by Prof. Jef Boeke, has since expanded efforts to accelerate mega-chunk assembly protocols, and notably has focused on the time-consuming verification steps, rather than on DNA assembly. They have optimised the PCRtag analyses by utilising real-time PCR, as well as using a bulk liquid dispenser to automate distribution of reagents into 1,536 wells of a quantitative PCR machine. The gel electrophoresis step is wholly eliminated from the process, as results are displayed in real-time by using the computational interpretation of the quantitative PCR data[185]. Although equipment is expensive for this set-up, the authors claim that the original PCRtag analyses that we do actually cost over time are more than the low-cost, high-throughput, automated method of qPCR that they run. This is a prime example of upgrading technology to faster methods and in this case is used to suit the screening demands of Sc2.0, allowing the analysis using a large numbers of primer pairs simultaneously for multiple numbers of potentially correct colonies. While the

increased scale of analysis obtainable with this equipment is impressive, it should be noted that it theoretically doesn't reduce the time by much. This is because the automated sample prep stages and the qPCR combined are only a couple of hours faster than the PCR plus agarose gel methods done in Chapter 3. Considering the time-scale for a round of mega-chunk assembly is measured in weeks, a saving of a few hours is not that significant. Ultimately the days required to recover yeast on plates after transformation and then to grow it again further on replica plates, takes up the bulk of the time. Without finding ways to make yeast grow faster, it is difficult to consider how time can be saved further.

6.1.4 The Applications and Utility of Generated Synthetic Chromosome Strains

The applications of a completely synthetic, redesigned, functional eukaryotic cell are the most exciting aspect of the Sc2.0 project, which has already received considerable media and academic attention. Despite the well-documented and good progress of all synthetic chromosomes (a summary of these can be found in Figure 1.5), the Sc2.0 project is still in its premature phases of completion. In reality, the applications of creating an entirely synthetic eukaryotic strain are not yet fully established or understood, and remain to be elucidated as and when people begin to use the strains in their projects, rather than just construct them. The inclusion of the inducible synthetic chromosome rearrangement and modification by loxPsym evolution (SCRaMbLE) system that has been included as part of the synthetic chromosomes' design is perhaps the most powerful aspect of the Sc2.0 design. As demonstrated in my thesis SCRaMbLE can cause an unprecedented variety of strain rearrangements and modifications, and could be implemented in many different ways e.g. to better understand genomic stability and dynamics, as well as genome-wide interactions. A fully synthetic eukaryotic strain will have approximately 5,000 loxPsym sites included three base pairs downstream of each non-essential gene's stop codon i.e. even if we assume (incorrectly) that each loxPsym site recombines only once, that will bring about 5,000 new inter- and intra-chromosomal rearrangements in each cell. These will include deletions, inversions, duplications and translocations (where regions of synthetic chromosomes are translocated from one synthetic chromosome to another) which all can be minor changes or major changes depending on what genes are involved and how many. In reality, the number of rearrangements per cell is unknown and will likely be influenced by the concentration of Cre in the nucleus and the time it is induced for. Far more than 5,000 recombinations will be possible, but also less may occur if there are not more than 5,000 Cre recombinases in the

nucleus. One issue I predict that will arise as people SCRaMbLE when more and more of the genome becomes synthetic is that the magnitude of the possible rearrangements will increase to a point where it becomes very difficult to accurately follow i) the screening of all the "new" genomes generated and ii) the data tracking, analysis, and storage of the rearrangements. The bioinformatics workload that will be associated with analysing SCRaMbLE Sc2.0 strains should not be underestimated, and Sc2.0 may eventually move from being a mainly wet-lab based project, to a primarily dry-lab based project, requiring a large global team of experts to deconvolute bioinformatics data such as next generation genome sequencing.

One obvious long-term aim of applying SCRaMbLE to the finished Sc2.0 strain will be to create a minimal yeast genome and ideally to delete any unnecessary metabolic processes so that the cell survives off the most basal reactions to maintain itself in standard lab media. This strain could be used to host engineered metabolic pathways, such as for the production of therapeutic drugs. It would be desirable as not only would side-reactions from native metabolism be minimised, but that it would be easier to predict how a new metabolic pathway would behave in this cell. This would especially be aided by the minimal yeast genome and any subsequent characterisation and modelling that could be done with this strain. A reduced-genome yeast would likely be a good candidate for the first whole-cell model and simulation of a eukaryote.

However, in reality there will likely be a large number of diverse "minimal" genomes generated by long-term SCRaMbLE experiments, rather than just one single 'winning' minimal genome. Determining the sequences of these minimal genomes will aid researchers in understanding what makes a eukaryote, what processes are essential and what degree of genome plasticity, flexibility and genome engineering can be tolerated. The power of SCRaMbLE lies in the ability to bring about simultaneous and large genomic changes that would have previously required extensive laborious cloning to achieve duplications, rearrangements or knockouts of many genes.

6.2 Developing SCRaMbLE-in using Synthetic Chromosome Strains

6.2.1 Summary

With the construction of the partially-synthetic chromosome strain XI, the SCRaMbLE system was implemented into the design of the synthetic sequence and became available to us for further use and development. The aim of Chapter 4 was to show that SCRaMbLE could be induced to not only cause extensive synthetic chromosome rearrangements but also the simultaneous insertion of heterologous

genes given an appropriate format. The aim was to develop a new method of 'SCRaMbLE-in'. The heterologous gene used as a model system was the Kanamycin resistance gene cassette KanMX, which was formatted in a number of different modes with zero or one loxPsym site or two flanking loxPsym sites to create KanMX, KanMX_1L and KanMX_2L constructs. These were made as linear fragments of DNA and as bacterial circular plasmids unable to replicate in yeast. Ultimately, it was shown that SCRaMbLE-in was most successful in the diploid dSynXI.A-C strain, and yielded the highest number of colonies when using linear two loxPsym constructs as the heterologous DNA format (i.e. KanMX_2L).

I also aimed to characterise the SCRaMbLE and SCRaMbLE-in processes in more depth, by using fluorescent constructs to map and track the gene rearrangements and expression changes that can occur as a result of the two processes. I used three fluorescence-producing genes mRuby (red), mTagBFP (blue) and sfGFP (green) each designed to have a loxPsym site three base pairs downstream of its stop codon, and with these I induced both SCRaMbLE and SCRaMbLE-in with KanMX_2L. My results showed the generation of large genomic diversity and subsequently a multitude of phenotypes were detected by flow cytometry. Although the phenotypes generated were difficult to decipher, they did provide information to the affects and complexity that SCRaMbLE and SCRaMbLE-in can have on gene expression, even in the presence of only three loxPsym sites.

6.2.2 SCRaMbLE-in of Heterologous Genes by Flanking LoxPsym Sites

The synthetic chromosome rearrangement and modification by loxPsym evolution toolkit is made up of symmetrical loxPsym sites always situated three base pairs downstream of every non-essential gene's stop codon. The recombination between these sites is catalysed by the inducible CreEBD version of the Cre recombinase enzyme. The Sc2.0 publications that reported the construction of synthetic chromosomes III[60], IX.R and VI.L[72] all briefly mentioned the implications the SCRaMbLE system could have on the finished Sc2.0 genome, and both papers showed induction of the Cre-lox recombination led to deletion of genes situated between two loxPsym sites. Dymond *et a*l (2012)[85] provided a brief review of SCRaMbLE and reasons for including it as part of the Sc2.0 design, along with its potential applications and possible future uses, e.g. implementing as an alternative method to genome minimisation and chromosome modification. However they do not provide any scientific data on SCRaMbLE applications, and the publication is purely theoretical. To date, the only two extensive studies that have been published about SCRaMbLE are the recent paper by Shen *et al* (2015)[82] and this report. One of the

major hypothesis the authors of the SCRaMbLE research paper tested was whether ectopic recombination or off target activity of the Cre recombinase enzyme could be detected. For SCRaMbLE experiments to be applicable, in this report as well as all others, there needs to be confidence that the combinatorial phenotypes that arise are solely limited as a result of SCRaMbLE. Of the 64 strains they genome sequenced, Shen *et al* found no evidence of ectopic or off-target recombination, proving that all phenotypes and genotypes that are generated in their study and this one, are due to the homologous recombination between the recombinase sites.

If we further compare the findings between this report and Shen's *et al*, we can conclude that the genome sequencing data they provide is incredibly valuable, as it has led to the discovery of new types of 3'UTR junctions that were formed as a result of SCRaMbLE. Although the results between their report and the results in this study are not in agreement, it could be possible that the disruption of 3'UTRs does not switch gene expression off completely but rather renders it decreased. From the results presented here, the expression of the fluorescent proteins in Chapter 4 or the metabolising rates of Chapter 5, are both decreased suggesting that the disruption of the 3'UTR has affected the ability to produce fully functional proteins. Shen *et al* not only exhibited novel junctions formed by SCRaMbLE, such as two coding regions facing each other separated by a loxPsym site, but showed that the SCRaMbLE of auxotrophic markers did not affect cell fitness or prototrophic growth. However, one critical observation is that apart from the sequencing of the SCRaMbLE generated genomes, there was no further expressional work carried out, such as mRNA quantification (qPCR) or comparative analysis between the quantity of the SCRaMbLEd and quality of proteins produced before and after SCRaMbLE. These experiments could further provide insight as to the exact effects of SCRaMbLE on gene expression.

In this study, we used partially-synthetic chromosome XI of strain SynXI.A-C, and exploited SCRaMbLE to do insertion of genes. SCRaMbLE-in was first attempted in the haploid SynXI.A-C strain, but this proved to not be fruitful due to large cell mortality rates. This underlined the previous reports for other synthetic chromosomes that showed that cell death was common after SCRaMbLE. Taking the whole picture into consideration, the process of SCRaMbLE-in is likely to be very challenging for the health of most cells. Over just a few hours the cells are transformed, then SCRaMbLEd and then subjected to G418s selection. How can we reduce this stress? One consideration is to SCRaMbLE for less time, or to SCRaMbLE with less regions of the genome in synthetic format. This would reduce the mortality due to lethal rearrangements. However, as SCRaMbLE is induced for

longer and especially as the number of loxPsym sites goes up as more chromosome regions become synthetic we would also expect that there would be a higher likelihood of insertion of the KanMX gene allowing survival on G418s plates. So SCRaMbLE-in probably has to consider a trade-off: more loxPsym sites and longer SCRaMbLE induction times means more cell death, but also a greater likelihood of the heterologous genes being incorporated. Striking the balance between these will be crucial.

To get around the issue of SCRaMbLE leading to cell death, a diploid version of SynXI.A-C was created in this study, dSynXI.A-C. This proved to be much more hardy against the effects of the Cre-lox recombination system, and was used as the host organism for all consequent SCRaMbLE and SCRaMbLE-in experiments. Initially it was predicted that the circular bacterial plasmid containing a single loxPsym site (pDJ018) alongside the KanMX gene (and promoter and terminator) would result in the highest number of colonies when induced in YPD media for one, two and three hours. This was because only a single recombination event would be needed to insert this DNA into a single loxPsym site. The DNA would effectively linearise so that the whole plasmid sequence would end up inserted in the genome.

The single recombination event between the loxPsym site on pDJ018 could insert into any of the 23 loxPsym sites provided in the sequence of synthetic chromosome XI of dSynXI.A-C, whereas with no loxPsym sites recombination shouldn't occur. With two loxPsym sites in this plasmid, recombination into the genome could also occur, but equally likely (or perhaps even more so) would be the deletion of the KanMX gene by recombination simply between the two sites of the plasmid. This meant that the single loxPsym site was expected to yield the most colonies. The data from Figure 4.7 shows indication of this, however the surviving, G418s resistant colonies that were recovered by the second screening step on 350 µg/ml of G418s-agar was very low, yielding an average of 2 colonies after three hours of SCRaMbLE-in induction for pDJ018. Although it does seem from the data that having just 1 loxPsym site is the favoured layout when the heterologous DNA is on the plasmid, ultimately the plasmid-based DNA performed much worse than the linear DNA and always gave a disappointing number of colonies in these experiments. With the same process repeated using linear KanMX constructs (KanMX, KanMX_1L and KanMX_2L), the surviving colony efficiencies that arose were much greater on 350 µg/ml G418s-agar, and this is shown in Figure 4.9. The total average colony number for all three constructs over all time-points increased 33-fold. This time the linear DNA delivery mode did not produce many colonies in single loxPsym format, and instead gave by far the greatest number of G418s-

resistant colonies when 2 loxPsyms were present in the KanMX_2L construct. An investigation into why the linear DNA format outperformed the circular DNA format was not done, and possibly is suitable work for further studies.

The existence of a few colonies that were a result of transformation of KanMX or pDJ017, the two constructs lacking any loxPsym sites, was initially not anticipated, but can be presumed to be a consequence of some other rare form of recombination of the DNA into the genome that likely occurs during DNA replication and repair. Yeast has the ability to maintain, integrate and express essential genes to ensure its survival under strong environmental pressures, so it is perhaps unsurprising that a few cells from the millions transformed managed to find a way to insert the selectable marker. It was interesting to also see a few colonies arising when the linear KanMX_1L construct with only 1 loxPsym was added to the SCRaMbLE-in reaction. As this is a linear fragment, the recombination of the DNA into the genome at its loxPsym site would in theory then cause a double-stranded break in the DNA at the other end of the fragment. Full integration and expression would therefore not be straight-forward. The number of surviving colonies on 350 µg/ml G418s-agar (Figure 4.9) was greater than the number seen for KanMX with no loxPsym sites, however there was no difference between the induced or uninduced cells, so these colonies can not be as a result of SCRaMbLE-in. The SCRaMbLE-independent generation of G418s colonies in all cases was likely a consequence of yeast-directed NHEJ recombination, under the strong selective pressure of the G418s antibiotic.

Having determined that linear DNA with two flanking loxPsyms sites was the best format for the provided DNA, we next looked at the types of rearrangements in SCRaMbLE that would cause the heterologous DNA to be incorporated into the genome. Theoretically there are three types of possible integrations that can happen when flanking loxPsym sites are provided with heterologous genes. The first, and seemingly the most likely, is that recombination happens between the two flanking loxPsym sites and two genomic loxPsym sites to recombine-in the provided gene whilst, simultaneously recombining-out the DNA between the two genomic loxPsym sites. This scenario supported by Figures 4.10, 4.11 and 4.13, where the inverse PCR results show that the amplified flanking genomic regions that were sequenced reveal KanMX integration between the 18th and 19th loxPsyms of SynXI.A-C (deleting a 5 kb region), as well between the 12th and 18th loxPsyms of SynXI.A-C (deleting a 31 kb region) and between the 2nd and 7th loxPsyms of SynXI.A-C (deleting a 21 kb region). Another possibility is that KanMX_2L could also recombine its two loxPsym sites to circularise itself, and then recombine and insert a second time into a single loxPsym site. This would be shown to be the case if the upstream and downstream

genomic region of one loxPsym site was amplified and subsequently sequenced by iPCR. In this study we did not see any evidence of this scenario in sequencing results.

The final mechanism that could potentially yield expression from the KanMX_2L gene is circularisation of the linear fragment by recombination of the flanking loxPsym sites, but then instead of integration into a genomic loxPsym locus, this circular plasmid is maintained in the yeast cell as an extrachromosomal circular plasmid (eccDNA). This plasmid would consist of the KanMX promoter, KanMX gene and the single remaining loxPsym site and importantly it would lack any replication initiation sequences (i.e. ARS sites). In the study of Demeke *et al* (2015)[179] they showed that tandem repetitive regions of the *S. cerevisiae* genome are potential substrates for homologous recombination either side of genomic ARS sites, and that this recombination can give rise to circularisation of DNA to create self-replicating eccDNAs. These eccDNA are capable of functional expression of any genes they may contain, and are maintained in the yeast cell when there is a selective advantage. In the study by Demeke *et al* that advantage is growth on xylose, by maintenance of xylose-utilisation genes, showing parallels with work done in this thesis.

The existence of eccDNAs is relatively new knowledge, and several groups are now studying these in yeast, with Demeke *et al* being the first to demonstrate their involvement with expression of heterologous genes. How they arise is not fully defined but Demeke *et al* postulated a variety of possible mechanisms in their study, with the strongest candidate being the duplication of regions during improper sister-chromatid exchange. Could the SCRaMbLE-in system, which catalyses recombination between loxPsym repeats, be forming synthetic equivalents of the eccDNA plasmids that support the expression of heterologous genes required for selection?

There were two examples of iPCR samples that gave evidence that the provided KanMX gene was indeed circularised in the cell. Firstly, in Figures 4.14 and 4.12 we show the amplification and presence of the KanMX gene from the DNA provided as a bacterial plasmid (pDJ018) and we also show the presence of a “duplicated” KanMX gene, respectively. However, both iPCR samples fail to show evidence of any flanking genomic sequence, thus it cannot be established whether the sequences have been integrated or not. Further inspection of the iPCR data for dSynXI.A-C SCRaMbLE-in with pDJ018 is that this bacterial plasmid has not been integrated, as the terminator sequence placed on the plasmid downstream of the KanMX ORF appears to be still present. This suggests that no recombination has

occurred at the plasmid's loxPsym. The same is seen for pDJ048. The second iPCR example was of dSynXI.A-C SCRaMbLE with KanMX_2L. This time the amplified band size and the primer sequence traces for both the forward and reverse primers showed that i) SCRaMbLE of the amplified loxPsym site had occurred – the terminator provided with the KanMX linear DNA fragments had been lost, and ii) based on the band size there was only a single gene present, rather than a duplication. In both these cases, the iPCR evidence is pointing to circular DNA. In the first case the bacterial plasmid either remains as provided or is duplicated but stays unintegrated, and somehow continues to be maintained in the cell. In the second case, the linear DNA is circularised and also is maintained in the cell.

Given that the DNA in all these cases contains the KanMX gene and the promoter, it is probable that expression of this selectable marker continues. However, in the second case the lack of the 3'UTR sequence will likely effect its expression. This could potentially be countered by having higher copy numbers of these DNA circles in the cell. However, the crucial difference between the possible circular DNAs we are seeing in our experiments and the previously reported eccDNA plasmids is that all previous reported eccDNAs were shown to contain ARS sites. Indeed, they always form around ARS sites. These sites provide the critical sequences that allow the DNA to replicate and be maintained in the cells. In contrast, the circular DNAs we are seeing have no ARS sites and there is no obvious way that they could replicate and be maintained over many generations (except if a genomic ARS region recombined out of the synthetic chromosome and into the circular DNAs somehow). Given the lack of any ARS or alternative replication mechanism for the circular DNAs that may have arisen here, it is difficult to justify their existence. The alternative explanations are that (i) the iPCR approach is somehow inadvertently giving the appearance of circle DNA from the genomic DNA, even though it does not exist, or (ii) the added DNA has duplicated (or more) before then being inserted in.

6.2.3 High Genome Diversity is Generated from SCRaMbLE Strains

In an attempt to understand SCRaMbLE and SCRaMbLE-in in more depth, fluorescent construct strains were generated in the BY4741 non-synthetic strain and induced for three hours (with KanMX_2L for SCRaMbLE-in). Four constructs included the three fluorescent producing genes mRuby, mTagBFP and sfGFP, which differed only in the order of the genes. The interpreted phenotypes following SCRaMbLE of the constructs (summarised in Figure 4.19) showed that twenty samples (out of the 28 phenotypes recorded) underwent a decrease in fluorescence produced by the second or third gene, to the extent that the geometric mean (GM)

values were effectively the same as wild type BY4741 recordings. In 7 cases the second or third gene produced less fluorescence following SCRaMbLE than the unSCRaMbLEd control. At this stage we can speculate that complete loss of fluorescence is reflective of that gene being recombined out (especially as this is seen only in the third and second gene, as the first gene should not be able to be recombined out), however without further sequencing or PCR analysis, this cannot be confirmed as the genes in question could have been inverted or experienced a detrimental 3'UTR disruption that means that a complete mRNA of the fluorescent protein could not be produced and expressed from. Without the further sequencing or PCR data, full analysis of the characterised colonies is not possible, but the SCRaMbLE results for yRC1841, yRC1842, yRC1844 and yRC1845 at least did show that interesting and unique gene expression phenotypes arise. This underlines how the presence of just three loxPsym sites can have a wide variety of effects on gene expression from a partially-synthetic yeast chromosome.

Interestingly in the majority of the SCRaMbLE samples that appear to have deleted the second or third gene (or both) within their constructs, the upstream gene also appears to be affected, undergoing a decrease in fluorescence. This is an example where the likelihood is that the upstream genes (upstream of gene 2 is gene 1, upstream of gene 3 is gene 2; the genes differ with respect to the strain type) are probably not being inverted (especially in the case of the first gene who's orientation cannot be altered), but rather are producing less fluorescence as a result of the altering of the 3'UTR sequence, which has either lost a terminator, but more likely has been replaced with a different terminator that does not produce a mature mRNA that gives expression with the same fluorescence intensities. This observation was seen for colonies from yRC1845, yRC1844 and yRC18421. Furthermore, the four constructs which differed in gene order, did not provide any evidence that the order of the genes affected the phenotypes that occurred. This same conclusion also came from the SCRaMbLE-in experiments, which are summarised in Figure 4.25.

In order to counteract the decreased expression of the genes in the synthetic chromosome from SCRaMbLE-in, a possible solution to this issue could be to redesign the format of the linear heterologous DNA that is provided so that it now contains a terminator sequence between the upstream loxPsym site and the promoter of the gene being inserted (Figure 6.1). By adding a terminator sequence after each loxPsym site, it ensures that when the foreign DNA integrates into the genome by SCRaMbLE-in, the 3'UTR of the upstream gene that is disrupted will be replaced by another working terminator sequence. This will ensure that the upstream gene still generates mature mRNAs whose expression is not affected by the events

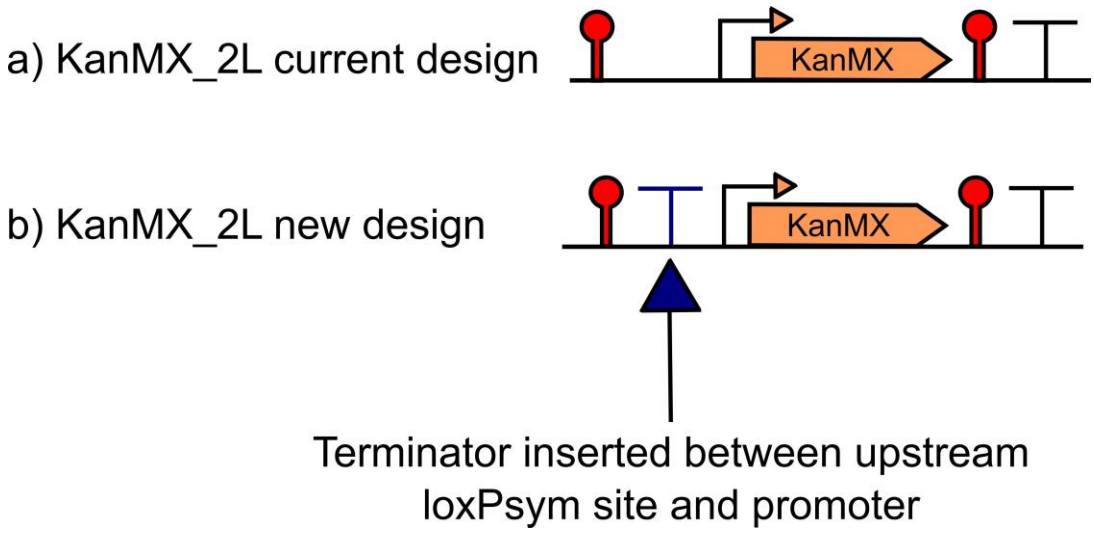


Figure 6.1 Proposed new design of heterologous DNA for SCRaMbLE-in. The current design of heterologous DNA (a) causes decrease in gene disruption due to disruption of the 3'UTR region. The new design (a) proposes to include a terminator sequence between the upstream loxPsym site and promoter, to provide a terminator sequence for the upstream gene following SCRaMbLE-in

of SCRaMbLE-in. So long as the rearrangements caused by SCRaMbLE don't invert the genes, the swapping of 3'UTRs by recombination shouldn't have a great affect on gene expression. Indeed, the MRes project of a student that I supervised in 2015 specifically investigated this effect[186] by using SCRaMbLE to swap 3'UTR sequences between fluorescent reporter genes. No significant changes in expression were seen from the reporter genes when only their 3'UTRs were changed.

For SCRaMbLE-in with the KanMX_2L fragment (data summarised in Figure 4.25 and 4.26) the majority of the phenotypes generated when grown on 200 µg/ml G418s-agar were as a result of either the third gene or second gene or both being swapped out. I classed this rearrangement type as a "replacement", where the KanMX_2L gene recombines in between two loxPsym sites to replace the DNA between the genomic loxPsym sites (total rearrangements and insertions are summarised in Tables 4.1 and 4.2). In fact, it is important to note that when SCRaMbLE-in was done with the KanMX heterologous DNA, 72% of the phenotypes were best interpreted as a replacement; replacing between the first and second, second and third, or first and third loxPsym sites. Thus, any researchers using the SCRaMbLE-in method developed in Chapter 4, should expect that the majority of SCRaMbLE-in products will also result in the loss of the chromosomal synthetic DNA between any two loxPsym sites that recombine the KanMX_2L between them. Table 4.1 summarised whether more activity was seen at particular loxPsym sites, depending on their order in the construct. Triplicates of the data sets weren't made,

however the numbers show that 62 events (whether replacements or insertions) occurred at the first loxPsym, and 53 events were seen at the second and also at the third loxPsyms. In short, there didn't seem to be a discrepancy between which loxPsym sites underwent the rearrangement, suggesting that they were each as likely to be used as each other. It would be interesting to see in future SCRaMbLE experiments done on the Sc2.0 strains, whether some loxPsym sites are used more often than others. For example ones near ARS sites would be presumably present in more copies per cell as these regions are duplicated first. Also loxPsym sites associated with regions that are protected with chromatin may be accessed less.

SCRaMbLE-in with KanMX_2L yielded a total of 35, 29, 9 and 40 viable colonies for the strains yRC1841, yRC1842, yRC1844 and yRC1845, respectively (Figure 4.25). Even though the data presented in Chapter 4 is based on my interpretations of statistical analyses, no two samples yielded the same GM values for either of the three genes showing that there is a large diversity of the phenotypes produced with each construct. The generation of genome diversity from SCRaMbLE was established early on in the Sc2.0 project and has been reiterated in all the consequent reports[60, 72, 82, 85], and this report provides a small insight of that possible diversity from three loxPsym sites.

As a final note, a design flaw was noticed when fluorescent constructs were integrated into the URA3 locus of BY4741, which was that all the promoters and terminators used for the expression of mRuby, mTagBFP, and sfGFP were found to also occur naturally in the BY4741 genome. On average a promoter or terminator sequence is about 300 bp, and given yeast's natural tendency to recombine areas of homology, there is a chance that rearrangements could have arisen as a consequence of the recombination of the promoters or terminators out of yeast.

6.3 SCRaMbLE-in of a Xylose Metabolic Pathway

6.3.1 Summary

Having demonstrated SCRaMbLE-in being able to integrate a single gene into the dSynXI.A-C strain (KanMX), we then demonstrated that SCRaMbLE-in could be used to integrate entire metabolic pathways. The oxidoreductase pathway genes were formatted under to the same design as KanMX_2L, and SCRaMbLE-in into dSynXI.A-C was performed. Following subsequent screening of colonies, KX7 was isolated and confirmed as xylose metabolising in pure xylose media, with growth rates that were similar to the previously established growth of the dSynXI.A-C expressing the pathway from a standard integration. iPCR analysis of KX7 generated bands that were sequenced, but there was no evidence that the genes were in fact

integrated into the genome. Nonetheless, in Chapter 5 of this report I showed that SCRaMbLE-in of a metabolic pathway is possible, and theoretically this permits researchers to simultaneously carry out metabolic engineering and genome evolution to create a diverse libraries of synthetic yeast cells with new functions.

6.3.2 Growth on Xylose Media of a Synthetic Strain using the Oxidoreductase Pathway

The two types of xylose metabolising pathways that allowed for the growth of *S. cerevisiae* on pure xylose media were the isomerase pathway and the oxidoreductase pathway, which are explained extensively in Chapter 5. The isomerase pathway was not functional in dSynXI.A-C. Presumably this was because the previously described genome knockouts associated with this pathway functioning efficiently were not done in our strains[111]. Instead I showed that the codon-optimised genes of the oxidoreductase pathway that I generated for this study allowed the host strain dSynXI.A-C to grow on xylose media, producing growth rates that were significantly different from the background strain, comparable to those previously reported and without the need for any specific gene knockouts from the host genome.

Despite the functionality of the oxidreductase pathway gene in dSynXI.A-C, growth rates were still low, although reflective of those published thus far. They were still too slow to facilitate efficient screening between the background strain and the xylose-metabolising strains (yDJM1 and yDJM3) on SC-X agar or YP-X agar. This was an issue that directly impacted the downstream processes, such as the screening of colonies following SCRaMbLE-in of these genes.

6.3.3 SCRaMbLE-in of the Oxidoreductase Pathway Genes

The knowledge gained from Chapter 4, about cell survival on G418s-agar following integration of the linear KanMX_2L gene yielding the highest number of colonies from three hours of induced media with β-oestradiol was carried forward and applied to the SCRaMbLE-in experiments with XYL1_2L, XYL2_2L and XYL3_2L. It was hypothesised that if the KanMX_2L gene was included in the transformation and SCRaMbLE-in mix, we could select for colonies having generally undergone SCRaMbLE-in, having previously failed to isolate a positive xylose-metabolising colony with SCRaMbLE-in using only the xylose-utilising genes. Our theory was proved correct, and a single sample was isolated as positive for growth on xylose.

To characterise the growth further, LC-RID experiments were attempted to quantify the amount of xylose being utilised in the media of the xylose-metabolising

synthetic yeast strains in comparison to the background strain. Xylose and related compounds were detected and measured using an Agilent 1290 Infinity Binary LC and 1260 Infinity II Refractive Index Detector (RID). However, the data obtained did not yield results that showed useful information, and after three biological replicated the assay was not pursued. Reasons for the failure of the quantification of the xylose in media is possibly due to the interference of peaks or detection generated by other compounds, such as the elution of xylitol, or compounds in the media. Perhaps better sample preparation is needed or an alternative LC buffer (H_2SO_4 was used).

6.3.4 Applications of SCRaMbLE-in using Multiple Gene Fragments

The data generated from the KX7 strain lacks genome sequencing to be able to draw solid conclusions from, however it does provide some insight as to the possible applications that SCRaMbLE-in of multiple genes, especially those of a pathway, could lead to. This work demonstrated that a xylose-metabolising strain, shown to otherwise not be able to utilise this carbon source with its wild type genes, was generated through the process of SCRaMbLE-in with heterologous genes. Although the genomic locations of the three genes remain elusive, this work does demonstrate that multiple genes (in this case four separate genes, including the KanMX gene) could be SCRaMbLEd-in. This could be of particular use to researchers generating new yeast strain (to possess novel functions), or for metabolic engineers like DeLoache[91] and Wei[100], mentioned in Chapter 1 for their pioneering work with yeast metabolic engineering, to provide several pathway genes simultaneously, and adaptively evolve strains possessing desired traits. Furthermore, large combinatorial yeast libraries could be generated easily, through single recombination events. The current limiting step is the transportation of the DNA into the cell, and upon the optimisation of this step, the SCRaMbLE-in of yeast can proceed much faster and possibly more effectively, than present methods.

6.4 Future Work

6.4.1 Completion of SynXI.A-C and Sc2.0

Future experiments for the finalisation of SynXI will be to complete the modular assembly and construction of the remaining right chromosome arm, as well as to sequence verify the remainder of the chromosome. Upon sequence verification, any missense (or major mutations, should they arise) should be corrected to match the *in silico* designed genotype.

With the availability of the SCRaMbLE toolkit, the synthetic chromosome versions can be induced to undergo mass genome recombinations, and the phenotypes that may arise as a result of these could be of interest. Tracking the genetic changes that have occurred as a result of SCRaMbLE could potentially reveal new gene interactions and networks, as well as novel phenotypic links, previously unknown[85]. Inductions could also be made under certain conditions for the evolution of new strains, such as thermo-tolerant strains (for example, induction of strains at 37 – 42 °C) or pH tolerant strains (such as in acidic media).

Furthermore, with the availability of other synthetic chromosomes from Sc2.0 collaborators, synthetic chromosome strains will be mated and sporulated to produce haploid versions that contain multiple synthetic chromosomes within a single nucleus. For example, the mating of SynXI.A-F with SynIII to produce SynIII/XI.A-F, thus creates a strain where induction of the Cre-lox recombination system can explore the inter-specific chromosome rearrangements and interactions, and compare these to the previously intra-specific SCRaMbLE-d chromosome.

The ultimate goal of Sc2.0 is to produce a fully functional, synthetic eukaryotic cell with 16 synthetic chromosomes and a neo-chromosome. Currently teams are on track to finish this aim by 2017/2018, at which point the real investigations of what can be done with synthetic genome engineering will begin.

6.4.2 Further Development and Verification of SCRaMbLE-in

The SCRaMbLE-in methodology is a new method developed in this study, and ideally requires further experimentation in order to gain better understanding and knowledge of this procedure and how to optimise it. There are a range of further experiments that could be done to aid researchers in developing a more optimised protocol for the introduction of heterologous genes into the synthetic chromosome strains.

Firstly, the pDJ017, pDJ018 and pDJ019 plasmids were not tested in the haploid SynXI.A-C strain. If in theory the single loxPsym based bacterial vector inserts into loxPsym sites without resulting in the recombining out of subsequent

synthetic chromosome regions, I predict that the pDJ018 plasmid could potentially be used to generate haploid strains that recombine the heterologous KanMX gene into its genome, without causing as much lethality. In this case it is not the number of loxPsyms that are the primary limiting factor, but rather the circularised DNA may be more suitable for haploid synthetic strain inclusion in order to minimise genome deletions associated with the recombining of two loxPsym events (in combination with the already extensive essential gene deletions/silencing that are deleterious to the strain).

In continuation of this point, I also suggest experimenting with different concentrations of heterologous fragments to test whether number of colonies or SCRaMbLE-in efficiencies could be increased. A standard quantity of 500 ng DNA concentration was used in this study for the integration of KanMX_2L and 250 ng for the integration of KanMX_2L, XYL1_2L, XYL2_2L and XYL3_2L (to produce a total of 2 µg of linear DNA). There were differences observed between the two concentrations of heterologous DNA used, especially for the generation of G418s-resistant colonies in dSynXI.A-C, which experienced a 2-fold drop when KanMX_2L was added in half the amount. Interestingly, there were more colonies observed in the dSynXI.A-L strain when all four heterologous genes were SCRaMbLEd-in at the same time. I suspect this is due to the greater region of synthetic DNA available in dSynXI.A-L (and thus the greater numbers of loxPsym sites), allowing for greater flexibility and scope for integrating heterologous DNA. It would be insightful if a small, simple and quick study was carried out to test whether there was correlation between the quantity of heterologous DNA added to SCRaMbLE-in reactions and the number of positive colonies output.

One drawback that was noted for the SCRaMbLE-in process was the total time needed (in a single day) to transform and SCRaMbLE-in heterologous DNA. A standard Gietz transformation from overnight cultures requires approximately four hours to grow the yeast cells to exponential phase, followed by approximately three hours (optimal) period needed to perform the transformation. Following this, SCRaMbLE-in can be induced, for one to three hours, cells are plated, recovered for an additional hour, and then antibiotic is added prior to their incubation. In total, a single day of SCRaMbLE-in requires an average of 13 hours from start to completion. The 13-hour day also impacts other future experiments that could prove insightful to characterising and understanding SCRaMbLE-in better, such as longer induction times, like Shen *et al* who induced their synthetic strain for four hours. Based on their observations that SCRaMbLE reduced colony numbers following six hours of induction, for SCRaMbLE-in work here, two extra hours (totalling to four

hours and five hours) of induction could be investigated to see whether the number of colonies expressing provided heterologous genes can increase or decrease. Using the protocol developed in Chapter 4, that would bring the total number of hours required in a single day to 15.

Therefore, in order decrease the time-investment aspect, the transformation protocol could be altered. There are a number of other yeast transformation methods that i) do not use chemically-competent cells and ii) do not require a four hour exponential growth phase[142]. I suggest the SCRaMbLE-in be coupled and tried with an electroporation transformation, that requires electrocompetent yeast cells to be prepared in advance, and doesn't require several incubation stages[187].

Diverging away from the DNA preparation and handling, I suggest that synthetic chromosome strains are mated to produce diploid strains with multiple synthetic chromosomes. Following this, induction of SCRaMbLE-in should be repeated using the new strains as the hosts, particularly concentrating on the use of DNA formatted to match either KanMX_2L or pDJ018.

Lastly, and most importantly, all the strains generated and used for SCRaMbLE-in, including the KX7 strain generated from the SCRaMbLE-in of the oxidoreductase metabolic pathway genes, should ideally be genome sequenced, in order to fully understand the rearrangements and new chromosome topologies that are generated. Furthermore, important information such as the integration loci, the copy number, the 3'UTR regions and the orientation of genes can be ascertained with genome sequencing. This SCRaMbLE-in study will be invaluable to the Sc2.0 project, but our data lacked detailed information about the process and mechanism, therefore this genome sequencing will be an essential aspect to any future work.

6.4.2 Growth of a Xylose-Utilising Synthetic Strain

The growth of the isomerase strain failed in dSynXI.A-C. Given the aims of the experiments carried out in this report, and the issues that were associated with screening of a xylose-utilising strain, it would be useful to i) enable growth of the isomerase pathway in yeast and ii) increase the difference of growth between the isomerase or oxidoreductase strain and the background strain. By doing so, it is possible that the screening and SCRaMbLE-in with the KanMX_2L gene also provided would not be necessary, and the pathways could be selected for based on their ability to grow on xylose agar. Therefore, although previously not done in order to maintain genome similarity as much as possible to the Sc2.0-based genomes, I suggest that the host strain – dSynXI.A-C – be altered to include the overexpression and knockouts that have been reported as beneficial to xylose metabolism. Karhuma

et al (2005)[176] showed that by overexpressing the XYL3 and pentose phosphate pathway genes and deletion of the GRE3 gene, they increased the growth rate of their engineering strain to 0.16 h^{-1} in SC-X media, which is a 10-fold greater growth rate than the best growth rate obtained in this report, which was 0.015 h^{-1} for yDJM3 (Table 5.3). Furthermore, their aerobic growth rates showed a logarithmic phase that was prominently lacking in our pure-xylose media growth curves.

Implementation of mutated and altered cofactor preference was shown by Bengtsson *et al* (2009)[180], and Jeppsson *et al* (2006)[188] to slightly improve growth of *S. cerevisiae*-based strains on xylose media. Therefore, with the same aims, these gene versions could be utilised in the SCRaMbLE-in experiments and for the growth assays of Chapter 5, in order to increase aerobic growth and facilitate easier isolation and screening of SCRaMbLE generated strains.

Enzymatic profiling of the non-SCRaMbLE xylose-metabolising strain that I constructed will provide a control for the same experiment of SCRaMbLE-in strains. This will particularly provide information about the activity of enzymes following SCRaMbLE, and to what extent does 3'UTR disruption lead to reduced enzyme activity – is this the reason that a decreased growth rate of the KX7 strain is seen?

Further on from the above point, an obvious experiment to do for generated SCRaMbLE-in strains will be to quantitatively assess mRNA production by qPCR, which will show whether any transcriptional interference can be detected in these assays post-SCRaMbLE.

6.4.3 Implementations and SCRaMbLE-in of Metabolic Pathway Genes

At the beginning of SCRaMbLE-in of the oxidoreductase pathway, the three genes were placed together on one single linear fragment of DNA (~ 10 kb in size) flanked with two loxPsym sites. As no viable colonies were isolated, the format that performed best in Chapter 4 (flanking of individual genes with loxPsym sites) was applied to the SCRaMbLE-in of the xylose pathway genes. However, given that a method using the KanMX_2L gene to screen for colonies post-SCRaMbLE was used successfully, it would be interesting to employ this method once more, with the xylose-utilisation genes provided as a single linear fragment, rather than as individuals. Furthermore, if the host strain was altered to include the gene knockouts and overexpressions that enable faster catabolism on xylose media (mentioned in Section 6.5.2), then the SCRaMbLE-in of the linear single gene fragment could potentially show efficient genomic integration.

One aspect of this project that was brief was the repeating of SCRaMbLE-in in the dSynXI.A-L strain. With more time, further screening of the recovered G418s-

resistant colonies that SCRaMbLEd-in KanMX_2L, XYL2_2L, XYL2_2L, and XYL3_2L is likely to yield a xylose-metabolising strain. Only a fraction of the total colonies (22%) were screened for growth on xylose. Furthermore, prior to initiating any screening of colonies via PCR or amplification or digestions, the recovery of SCRaMbLEd-in colonies could be adjusted to a single-pot assay, competitively assaying for xylose-catabolising strains through continual rounds of SC-X media growth to isolate the fittest SCRaMbLE-in strain, that have integrated the desired cells into the genome. The competition assays summarised in Figure 5.10 showed that in all scenarios, the xylose metabolising strain outcompeted the background strain.

6.4.4 Other Future Work

The applicability of SCRaMbLE-in for the insertion of metabolic pathways was demonstrated in this work for the oxidoreductase pathway. The inherent drawback of this pathway is that the environmental pressure, the presence of xylose, does not cause immediate cell death or arrest, as observed in the presence of G418s. Therefore, the background strain exhibits a small amount of basal growth, and this severely hindered the isolation and screening of any xylose catabolising strains generated by SCRaMbLE-in on media that facilitates fast growth. Nonetheless the data demonstrated that SCRaMbLE-in could be applied for the inclusion of metabolic pathways. Therefore going forward, SCRaMbLE-in should also be tried on pathways with more obvious selectable phenotypes. For example, use of the lycopene biosynthesis pathway gives rise to red colonies and so would facilitate easy screening. The limitation of the lycopene pathway however is that it does not endow the yeast cell with a useful new function, beyond making a molecule that is already widely-available. Another possibility is a pathway producing a medical molecule. The penicillin biosynthesis pathway would perhaps be more suitable, as it is made up of four genes and can be selected for in the presence of bacteria that will outcompete yeast unless it can secrete the antibiotic. Furthermore, SCRaMbLE of this pathway could theoretically lead to an optimised SCRaMbLE genome, secreting large amounts of penicillin, which would be a very useful biotechnological application. Work towards this is underway in our group.

Lastly, the extent of genome rearrangements will be huge in the final Sc2.0 strain when induction of the Cre-lox recombination system is implemented. I suggest that for the preliminary stages of the SCRaMbLE-ing experiments, researchers might want to apply an added level of control to where the Cre-recombinase induces SCRaMbLE. One idea that I developed during my work with the SCRaMbLE system,

and the rise of CRISPR-Cas9 technology, was how interesting it would be if the Cre recombinase could be designed to be coupled with guide RNA sequences, which would direct the Cre enzyme to be more specific in its choice of loxPsym. This would enable researchers to cause specific genome rearrangements whilst conserving the presence of the Cre-lox system.

6.4.5 Final Recommendations for Successful SCRaMbLE-in

The above described future works are likely to lead to an optimised SCRaMbLE-in method for the (combinatorial) insertion of heterologous gene(s) into synthetic chromosome strains. Based on the observations I have gathered thus far during the development of the SCRaMbLE-in system, I can suggest the following guidelines that should be used to increase efficiency and yield of desired colonies by researchers wishing to implement this system:

1. In all cases of SCRaMbLE-in throughout this chapter, an induction time of three hours yielded the highest number of colonies, even though induction times proved not to be significantly different to each other. I suggest the use inducing SCRaMbLE between three and five hours, to increase the chances of gene insertion.
2. SCRaMbLE-in is still in the developing stages of research and understanding. High cell death and fitness-decrease was seen when SCRaMbLE-in was carried out in the haploid strain. Therefore, for now, I advise SCRaMbLE-in continue to be assayed in diploid partially (or wholly) synthetic strains.
3. The concentration of the DNA was not assayed in this study, however as yeast handles up to 3 µg of foreign DNA well, I suggest using between 0.5 – 1 µg of DNA. Considerations as to the number of different genes being included should be made.
4. Depending on the genes being inserted, a 60 to 90 minute recovery should be included, in order to allow cells to establish growth on the media, prior to selection.
5. Based on the findings of this thesis, I suggest that linear fragments encoding heterologous DNA, lacking replication origins and flanked by two loxPsym sites with downstream terminators (Figure 6.1) be used as the introducing DNA delivery mode.
6. Lastly, when it is a challenge to stringently select for a new function post-SCRaMbLE-in, I recommend co-transforming with KanMX_2L DNA and initially screening survivors for resistance to G418s. This narrows the pool of

survivors to only the yeast cells that have been successfully transformed with DNA and have undergone SCRaMbLE.

7. I have not mentioned the type of yeast transformation to be used, but I believe that this step can be significantly improved, and it is up to the researcher to decide what method they wish to pursue.

6.5 Overall Conclusions

The scope of the work presented here aimed to initiate synthesis, assembly and integration of the designed synthetic DNA for chromosome XI of the *S. cerevisiae* yeast genome, as part of the international project to create the first synthetic eukaryotic genome, Sc2.0. In this project we investigated if we could repeat the assembly and integration of synthetic mega-chunk DNA, as reported in Dymond *et al* (2011)[72]. We showed that the integration of the synthetic DNA was successful, and thus far we have generated more than half of synthetic chromosome XI, which is functional in the BY4741 strain.

With the construction of partially-synthetic strains during SynXI construction, the SCRaMbLE system was used to bring about rearrangements to the synthetic chromosome XI DNA. Furthermore, SCRaMbLE was modified to also have an extra functionality not seen before, and so in this study we report the first uses of SCRaMbLE to also introduce heterologous DNA into the synthetic chromosome. Using the KanMX gene flanked by two loxPsym sites we showed that the Cre-lox system could be induced to cause recombination between the loxPsyms flanking the provided gene and pairs of genomic loxPsyms. This allows automatic recombination of heterologous DNA into the synthetic chromosome. We have called this new method SCRaMbLE-in.

Lastly, in an endeavour to scale-up the inclusion of heterologous genes from a single gene type to multiple genes, we hypothesised that SCRaMbLE-in could also be applied for the inclusion of metabolic pathways. In this study we confirmed that using the same format we developed for the inclusion of the KanMX gene, we additionally introduced three more heterologous gene into the synthetic chromosome strain, that enabled dSynXI.A-C to metabolise and proliferate on xylose media, using the oxidoreductase pathway. This study showed that SCRaMbLE-in can be implemented for simultaneous metabolic engineering and genome evolution, and going forward could be a powerful tool for synthetic biology and industrial biotechnology.

7.0 BIBLIOGRAPHY

1. Watson, J.D. and F.C. Crick, *Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid*. JAMA, 1993. **269**(15): p. 1966-1967.
2. Reese, C.B., *The chemical synthesis of oligo- and poly-nucleotides by the phosphotriester approach*. Tetrahedron, 1978. **34**(21): p. 3143-3179.
3. Letsinger, R.L. and K.K. Ogilvie, *Nucleotide chemistry. XIII. Synthesis of oligothymidylates via phosphotriester intermediates*. Journal of the American Chemical Society, 1969. **91**(12): p. 3350-3355.
4. Khorana, H., *Total synthesis of a gene*. Science, 1979. **203**(4381): p. 614-625.
5. Edwards, R.G., *MATURATION IN VITRO OF HUMAN OVARIAN OOCYTES*. The Lancet. **286**(7419): p. 926-929.
6. Wilmut, I., et al., *Viable offspring derived from fetal and adult mammalian cells*. Nature, 1997. **385**(6619): p. 810-3.
7. Arber, W. and S. Linn, *DNA modification and restriction*. Annu Rev Biochem, 1969. **38**: p. 467-500.
8. Smith, H.O. and K.W. Welcox, *A Restriction enzyme from Hemophilus influenzae: I. Purification and general properties*. Journal of Molecular Biology, 1970. **51**(2): p. 379-391.
9. Johnson, I., *Human insulin from recombinant DNA technology*. Science, 1983. **219**(4585): p. 632-637.
10. Biology, S. *Synthetic Biology Community*. 2007 [cited 25th July 2013; Available from: <http://syntheticbiology.org/%3E.
11. Andrianantoandro, E., et al., *Synthetic biology: new engineering rules for an emerging discipline*. Mol Syst Biol, 2006. **2**: p. 2006 0028.
12. Carlson, R., *The changing economics of DNA synthesis*. Nat Biotech, 2009. **27**(12): p. 1091-1094.
13. Perry, N. and A.J. Ninfa, *Synthetic networks: oscillators and toggle switches for Escherichia coli*. Methods Mol Biol, 2012. **813**: p. 287-300.
14. Hillenbrand, P., G. Fritz, and U. Gerland, *Biological Signal Processing with a Genetic Toggle Switch*. PLoS ONE, 2013. **8**(7): p. e68345.
15. Basu, S., et al., *A synthetic multicellular system for programmed pattern formation*. Nature, 2005. **434**(7037): p. 1130-1134.
16. Saeidi, N., et al., *Engineering microbes to sense and eradicate Pseudomonas aeruginosa, a human pathogen*. Molecular Systems Biology, 2011. **7**: p. 521-521.
17. Joshi, N., et al., *Novel approaches to biosensors for detection of arsenic in drinking water*. Desalination, 2009. **248**(1–3): p. 517-523.
18. Paddon, C.J., et al., *High-level semi-synthetic production of the potent antimalarial artemisinin*. Nature, 2013. **496**(7446): p. 528-532.
19. Kwok, R., *Five hard truths for synthetic biology*. Nature, 2010. **463**(7279): p. 288-90.
20. Canton, B., A. Labno, and D. Endy, *Refinement and standardization of synthetic biological parts and devices*. Nat Biotech, 2008. **26**(7): p. 787-793.
21. Müller, K.M. and K.M. Arndt, *Standardization in Synthetic Biology*, in *Synthetic Gene Networks: Methods and Protocols*, W. Weber and M. Fussenegger, Editors. 2012, Humana Press: Totowa, NJ. p. 23-43.
22. Purnick, P.E. and R. Weiss, *The second wave of synthetic biology: from modules to systems*. Nat Rev Mol Cell Biol, 2009. **10**(6): p. 410-22.

23. Hodgman, C.E. and M.C. Jewett, *Cell-free synthetic biology: Thinking outside the cell*. Metabolic Engineering, 2012. **14**(3): p. 261-269.
24. Richmond, K.E., et al., *Amplification and assembly of chip-eluted DNA (AACED): a method for high-throughput gene synthesis*. Nucleic Acids Research, 2004. **32**(17): p. 5011-5018.
25. Li, J., et al., *Improved Cell-Free RNA and Protein Synthesis System*. PLoS ONE, 2014. **9**(9): p. e106232.
26. Heyman, Y., et al., *Cell-free protein synthesis and assembly on a biochip*. Nat Nano, 2012. **7**(6): p. 374-378.
27. Gartner, Z.J., et al., *DNA-Templated Organic Synthesis and Selection of a Library of Macrocycles*. Science, 2004. **305**(5690): p. 1601-1605.
28. Forster, A.C. and G.M. Church, *Synthetic biology projects in vitro*. Genome Research, 2007. **17**(1): p. 1-6.
29. Ceroni, F., et al., *Quantifying cellular capacity identifies gene expression designs with reduced burden*. Nat Meth, 2015. **12**(5): p. 415-418.
30. Urnov, F.D., et al., *Genome editing with engineered zinc finger nucleases*. Nature Reviews Genetics, 2010. **11**(9).
31. Boch, J. and U. Bonas, *Xanthomonas AvrBs3 Family-Type III Effectors: Discovery and Function*. Annual Review of Phytopathology, 2010. **48**(1): p. 419-436.
32. Joung, J.K. and J.D. Sander, *TALENs: a widely applicable technology for targeted genome editing*. Nat Rev Mol Cell Biol, 2013. **14**(1): p. 49-55.
33. Jiang, W., et al., *RNA-guided editing of bacterial genomes using CRISPR-Cas systems*. Nat Biotech, 2013. **31**(3): p. 233-239.
34. Hwang, W.Y., et al., *Efficient genome editing in zebrafish using a CRISPR-Cas system*. Nat Biotech, 2013. **31**(3): p. 227-229.
35. Wang, Y., et al., *The CRISPR/Cas System mediates efficient genome engineering in Bombyx mori*. Cell Res, 2013. **23**(12): p. 1414-1416.
36. DiCarlo, J.E., et al., *Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems*. Nucleic Acids Research, 2013. **41**(7): p. 4336-4343.
37. Ran, F.A., et al., *Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity*. Cell, 2013. **154**(6): p. 1380-1389.
38. Cong, L., et al., *Multiplex Genome Engineering Using CRISPR/Cas Systems*. Science, 2013. **339**(6121): p. 819-823.
39. Sander, J.D. and J.K. Joung, *CRISPR-Cas systems for editing, regulating and targeting genomes*. Nat Biotech, 2014. **32**(4): p. 347-355.
40. Shen, B., et al., *Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects*. Nat Meth, 2014. **11**(4): p. 399-402.
41. Gibson, D.G., *Programming biological operating systems: genome design, assembly and activation*. Nat Meth, 2014. **11**(5): p. 521-526.
42. Gibson, D.G., et al., *Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome*. Science, 2008. **319**.
43. Gibson, D.G., et al., *Creation of a bacterial cell controlled by a chemically synthesized genome*. Science, 2010. **329**(5987): p. 52-6.
44. Lajoie, M.J., et al., *Genomically Recoded Organisms Expand Biological Functions*. Science, 2013. **342**(6156): p. 357-360.
45. Isaacs, F.J., et al., *Precise manipulation of chromosomes in vivo enables genome-wide codon replacement*. Science, 2011. **333**(6040): p. 348-53.
46. Wang, H.H., et al., *Programming cells by multiplex genome engineering and accelerated evolution*. Nature, 2009. **460**(7257): p. 894-898.

47. Porro, D., et al., *Recombinant protein production in yeasts*. Molecular Biotechnology. **31**(3): p. 245-259.
48. Cherry, J.M., et al., *Genetic and physical maps of Saccharomyces cerevisiae*. Nature, 1997. **387**(6632 Suppl): p. 67-73.
49. Stanley, D., et al., *The ethanol stress response and ethanol tolerance of Saccharomyces cerevisiae*. Journal of Applied Microbiology, 2010. **109**(1): p. 13-24.
50. Mattanovich, D., M. Sauer, and B. Gasser, *Yeast biotechnology: teaching the old dog new tricks*. Microbial Cell Factories, 2014. **13**: p. 34-34.
51. Prinz, B., et al., *Establishing a versatile fermentation and purification procedure for human proteins expressed in the yeasts Saccharomyces cerevisiae and Pichia pastoris for structural genomics*. Journal of Structural and Functional Genomics. **5**(1): p. 29-44.
52. Oliver, S.G., et al., *The complete DNA sequence of yeast chromosome III*. Nature, 1992. **357**(6373): p. 38-46.
53. Goffeau, A., et al., *Life with 6000 Genes*. Science, 1996. **274**(5287): p. 546-567.
54. Lemoine, F.J., et al., *Chromosomal translocations in yeast induced by low levels of DNA polymerase a model for chromosome fragile sites*. Cell, 2005. **120**(5): p. 587-98.
55. Louis, E.J., *The chromosome ends of Saccharomyces cerevisiae*. Yeast, 1995. **11**(16).
56. Parenteau, J., et al., *Deletion of Many Yeast Introns Reveals a Minority of Genes that Require Splicing for Function*. Molecular Biology of the Cell, 2008. **19**(5): p. 1932-1941.
57. Percudani, R., A. Pavesi, and S. Ottonello, *Transfer RNA gene redundancy and translational selection in Saccharomyces cerevisiae*. Journal of Molecular Biology, 1997. **268**(2).
58. Hani, J. and H. Feldmann, *tRNA genes and retroelements in the yeast genome*. Nucleic Acids Research, 1998. **26**(3): p. 689-696.
59. Kuhn, R.M., L. Clarke, and J. Carbon, *Clustered tRNA genes in Schizosaccharomyces pombe centromeric DNA sequence repeats*. Proceedings of the National Academy of Sciences, 1991. **88**(4): p. 1306-1310.
60. Annaluru, N., et al., *Total Synthesis of a Functional Designer Eukaryotic Chromosome*. Science, 2014. **344**(6179): p. 55-58.
61. Ellis, T., T. Adie, and G.S. Baldwin, *DNA assembly for synthetic biology: from parts to pathways and beyond*. Integrative Biology, 2011. **3**(2).
62. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases*. Nature Methods, 2009. **6**(5): p. 343-U41.
63. Engler, C. and S. Marillonnet, *Combinatorial DNA assembly using Golden Gate cloning*. Methods Mol Biol, 2013. **1073**: p. 141-56.
64. Sleight, S.C., et al., *In-Fusion BioBrick assembly and re-engineering*. Nucleic Acids Research, 2010. **38**(8): p. 2624-2636.
65. Li, M.Z. and S.J. Elledge, *Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC*. Nat Methods, 2007. **4**.
66. Cooper, E.M., et al., *The Build-a-Genome course*. Methods in molecular biology (Clifton, N.J.), 2012. **852**.
67. Dymond, J.S., et al., *Teaching synthetic biology, bioinformatics and engineering to undergraduates: the interdisciplinary Build-a-Genome course*. Genetics, 2009. **181**(1): p. 13-21.

68. Stemmer, W.P.C., et al., *Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides*. Gene, 1995. **164**(1): p. 49-53.
69. Annaluru, N., et al., *Assembling DNA Fragments by USER Fusion*, in *Gene Synthesis: Methods and Protocols*, J. Peccoud, Editor. 2012, Humana Press: Totowa, NJ. p. 77-95.
70. Geu-Flores, F., et al., *USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products*. Nucleic Acids Research, 2007. **35**(7): p. e55.
71. Lynch, M., et al., *A genome-wide view of the spectrum of spontaneous mutations in yeast*. Proceedings of the National Academy of Sciences, 2008. **105**(27): p. 9272-9277.
72. Dymond, J.S., et al., *Synthetic chromosome arms function in yeast and generate phenotypic diversity by design*. Nature, 2011. **477**(7365): p. 471-476.
73. Lobocka, M.B., et al., *Genome of bacteriophage P1*. J Bacteriol, 2004. **186**(21): p. 7032-68.
74. Nagy, A., *Cre recombinase: The universal reagent for genome tailoring*. genesis, 2000. **26**(2): p. 99-109.
75. Guo, F., D.N. Gopaul, and G.D. Van Duyne, *Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse*. Nature, 1997. **389**(6646): p. 40-46.
76. Sauer, B., *Functional expression of the cre-lox site-specific recombination system in the yeast Saccharomyces cerevisiae*. Mol Cell Biol, 1987. **7**(6): p. 2087-96.
77. Schwenk, F., U. Baron, and K. Rajewsky, *A cre-transgenic mouse strain for the ubiquitous deletion of loxP-flanked gene segments including deletion in germ cells*. Nucleic Acids Research, 1995. **23**(24): p. 5080-5081.
78. Hoess, R.H., A. Wierzbicki, and K. Abremski, *THE ROLE OF THE LOXP SPACER REGION IN P1 SITE-SPECIFIC RECOMBINATION*. Nucleic Acids Research, 1986. **14**(5).
79. Abremski, K., et al., *PROPERTIES OF A MUTANT CRE PROTEIN THAT ALTERS THE TOPOLOGICAL LINKAGE OF RECOMBINATION PRODUCTS*. Journal of Molecular Biology, 1988. **202**(1).
80. Feil, R., et al., *Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains*. Biochem Biophys Res Commun, 1997. **237**(3): p. 752-7.
81. Lindstrom, D.L. and D.E. Gottschling, *The Mother Enrichment Program: A Genetic System for Facile Replicative Life Span Analysis in Saccharomyces cerevisiae*. Genetics, 2009. **183**(2): p. 413-422.
82. Shen, Y., et al., *SCRaMbLE generates designed combinatorial stochastic diversity in synthetic chromosomes*. Genome Research, 2015.
83. Giaever, G., et al., *Functional profiling of the Saccharomyces cerevisiae genome*. Nature, 2002. **418**(6896): p. 387-91.
84. Costanzo, M., et al., *The genetic landscape of a cell*. Science, 2010. **327**(5964): p. 425-31.
85. Dymond, J. and J. Boeke, *The Saccharomyces cerevisiae SCRaMbLE system and genome minimization*. Bioengineered bugs, 2012. **3**(3).
86. Schuller, D. and M. Casal, *The use of genetically modified Saccharomyces cerevisiae strains in the wine industry*. Applied Microbiology and Biotechnology, 2005. **68**(3): p. 292-304.

87. Nevoigt, E., *Progress in metabolic engineering of Saccharomyces cerevisiae*. Microbiol Mol Biol Rev, 2008. **72**.
88. Siewers, V., et al., *Heterologous production of non-ribosomal peptide LLD-ACV in Saccharomyces cerevisiae*. Metabolic Engineering, 2009. **11**(6).
89. Siewers, V., R. San-Bento, and J. Nielsen, *Implementation of Communication-Mediating Domains for Non-Ribosomal Peptide Production in Saccharomyces cerevisiae*. Biotechnology and Bioengineering, 2010. **106**(5).
90. Kometani, T., et al., *Practical production of (S)-1,2-propanediol and its derivative through baker's yeast-mediated reduction*. J Biosci Bioeng, 2001. **91**(5): p. 525-7.
91. DeLoache, W.C., et al., *An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose*. Nat Chem Biol, 2015. **11**(7): p. 465-471.
92. Khalil, A.S. and J.J. Collins, *Synthetic biology: applications come of age*. Nat Rev Genet, 2010. **11**(5): p. 367-79.
93. Wei, N., et al., *Enhanced biofuel production through coupled acetic acid and xylose consumption by engineered yeast*. Nat Commun, 2013. **4**.
94. Li, X., et al., *Expanding xylose metabolism in yeast for plant cell wall conversion to biofuels*. eLife, 2015. **4**.
95. Lee, S.K., et al., *Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels*. Current Opinion in Biotechnology, 2008. **19**(6): p. 556-563.
96. Liew, W.H., M.H. Hassim, and D.K.S. Ng, *Review of evolution, technology and sustainability assessments of biofuel production*. Journal of Cleaner Production, 2014. **71**: p. 11-29.
97. Nikolić, S., et al., *Production of bioethanol from corn meal hydrolyzates by free and immobilized cells of Saccharomyces cerevisiae var. ellipsoideus*. Biomass and Bioenergy, 2010. **34**(10): p. 1449-1456.
98. Nielsen, J., et al., *Metabolic engineering of yeast for production of fuels and chemicals*. Current Opinion in Biotechnology, 2013. **24**(3): p. 398-404.
99. Valle-Rodríguez, J.O., et al., *Metabolic engineering of Saccharomyces cerevisiae for production of fatty acid ethyl esters, an advanced biofuel, by eliminating non-essential fatty acid utilization pathways*. Applied Energy, 2014. **115**: p. 226-232.
100. Wei, N., et al., *Simultaneous Utilization of Cellobiose, Xylose, and Acetic Acid from Lignocellulosic Biomass for Biofuel Production by an Engineered Yeast Platform*. ACS Synthetic Biology, 2015. **4**(6): p. 707-713.
101. Ha, S.-J., et al., *Engineered Saccharomyces cerevisiae capable of simultaneous cellobiose and xylose fermentation*. Proceedings of the National Academy of Sciences, 2011. **108**(2): p. 504-509.
102. Ostergaard, S., L. Olsson, and J. Nielsen, *Metabolic Engineering of Saccharomyces cerevisiae*. Microbiology and Molecular Biology Reviews, 2000. **64**(1): p. 34-50.
103. Desai, T.A. and C.V. Rao, *Regulation of Arabinose and Xylose Metabolism in Escherichia coli*. Applied and Environmental Microbiology, 2010. **76**(5): p. 1524-1532.
104. Harhangi, H.R., et al., *Xylose metabolism in the anaerobic fungus Piromyces sp. strain E2 follows the bacterial pathway*. Archives of Microbiology, 2003. **180**(2): p. 134-141.

105. Johansson, B. and B. Hahn-Hagerdal, *The non-oxidative pentose phosphate pathway controls the fermentation rate of xylulose but not of xylose in Saccharomyces cerevisiae TMB3001*. FEMS Yeast Res, 2002. **2**.
106. Bettiga, M., B. Hahn-Hagerdal, and M.F. Gorwa-Grauslund, *Comparing the xylose reductase/xylitol dehydrogenase and xylose isomerase pathways in arabinose and xylose fermenting Saccharomyces cerevisiae strains*. Biotechnol Biofuels, 2008. **1**.
107. Brat, D., E. Boles, and B. Wiedemann, *Functional expression of a bacterial xylose isomerase in Saccharomyces cerevisiae*. Appl Environ Microbiol, 2009. **75**.
108. Karhumaa, K., et al., *Comparison of the xylose reductase-xylitol dehydrogenase and the xylose isomerase pathways for xylose fermentation by recombinant*, in *Saccharomyces cerevisiae*. *Microb Cell Fact*. 2007.
109. Kuyper, M., et al., *Metabolic engineering of a xylose-isomerase-expressing Saccharomyces cerevisiae strain for rapid anaerobic xylose fermentation*. FEMS Yeast Res, 2005. **5**.
110. Rodriguez-Pena, J.M., et al., *The YGR194c (XKS1) gene encodes the xylulokinase from the budding yeast Saccharomyces cerevisiae*. FEMS Microbiol Lett, 1998. **162**(1): p. 155-60.
111. Lee, S.-M., T. Jellison, and H.S. Alper, *Directed Evolution of Xylose Isomerase for Improved Xylose Catabolism and Fermentation in the Yeast Saccharomyces cerevisiae*. Applied and Environmental Microbiology, 2012. **78**(16): p. 5708-5716.
112. Nieves, L.M., L.A. Panyon, and X. Wang, *Engineering Sugar Utilization and Microbial Tolerance toward Lignocellulose Conversion*. Front Bioeng Biotechnol, 2015. **3**: p. 17.
113. Krahulec, S., M. Klimacek, and B. Nidetzky, *Engineering of a matched pair of xylose reductase and xylitol dehydrogenase for xylose fermentation by Saccharomyces cerevisiae*. Biotechnology Journal, 2009. **4**(5): p. 684-694.
114. Matsushika, A., et al., *Ethanol production from xylose in engineered Saccharomyces cerevisiae strains: current state and perspectives*. Applied Microbiology and Biotechnology, 2009. **84**(1): p. 37-53.
115. Jin, Y.-S., et al., *Optimal Growth and Ethanol Production from Xylose by Recombinant Saccharomyces cerevisiae Require Moderate d-Xylulokinase Activity*. Applied and Environmental Microbiology, 2003. **69**(1): p. 495-503.
116. Toivari, M.H., et al., *Conversion of Xylose to Ethanol by Recombinant Saccharomyces cerevisiae: Importance of Xylulokinase (XKS1) and Oxygen Availability*. Metabolic Engineering, 2001. **3**(3): p. 236-249.
117. Blount, B.A., T. Weenink, and T. Ellis, *Construction of synthetic regulatory networks in yeast*. FEBS Letters, 2012. **586**(15): p. 2112-2121.
118. Winston, F., C. Dollard, and S.L. Ricupero-Hovasse, *Construction of a set of convenient saccharomyces cerevisiae strains that are isogenic to S288C*. Yeast, 1995. **11**(1): p. 53-55.
119. Sikorski, R.S. and P. Hieter, *A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae*. Genetics, 1989. **122**(1): p. 19-27.
120. Yanisch-Perron, C., J. Vieira, and J. Messing, *Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors*. Gene, 1985. **33**(1): p. 103-119.

121. Lee, M.E., et al., *A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly*. ACS Synthetic Biology, 2015. **4**(9): p. 975-986.
122. Daniel Gietz, R. and R.A. Woods, *Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method*, in *Methods in Enzymology*, G. Christine and R.F. Gerald, Editors. 2002, Academic Press. p. 87-96.
123. Blount, B.A., A.J.M. Driessen, and T. Ellis, *GC preps: fast and easy extraction of stable yeast genomic DNA*. Manuscript in reparation, 2016.
124. Von Dollen, C., et al., "*Build a Genome*" *Designing and Synthesising SC2.0*. 2009, John Hopkins University: JHU. p. 97.
125. Jovicevic, D., B.A. Blount, and T. Ellis, *Total synthesis of a eukaryotic chromosome: Redesigning and SCRaMbLE-ing yeast*. BioEssays, 2014. **36**(9): p. 855-860.
126. Cherry, J.M., et al., *Saccharomyces Genome Database: the genomics resource of budding yeast*. Nucleic Acids Research, 2012. **40**(D1): p. D700-D705.
127. Richardson, S., et al., *Automated Design of Assemblable, Modular, Synthetic Chromosomes*, in *Parallel Processing and Applied Mathematics*, R. Wyrzykowski, et al., Editors. 2010, Springer Berlin Heidelberg. p. 280-289.
128. Abremski, K. and R. Hoess, *Phage P1 Cre-loxP site-specific recombination*. Journal of Molecular Biology, 1985. **184**(2): p. 211-220.
129. Eilbeck, K., et al., *The Sequence Ontology: a tool for the unification of genome annotations*. Genome Biol, 2005. **6**(5): p. R44.
130. Ji, H., et al., *Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes and LTR sequences*. Cell, 1993. **73**(5): p. 1007-1018.
131. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases*. Nat Methods, 2009. **6**.
132. Casini, A., et al., *Bricks and blueprints: methods and standards for DNA assembly*. Nat Rev Mol Cell Biol, 2015. **16**(9): p. 568-576.
133. Gibson, D.G., *Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides*. Nucleic Acids Research, 2009. **37**(20): p. 6984-6990.
134. Ray, A. and K.W. Runge, *The Yeast Telomere Length Counting Machinery Is Sensitive to Sequences at the Telomere-Nontelomere Junction*. Molecular and Cellular Biology, 1999. **19**(1): p. 31-45.
135. Kupiec, M., *Biology of telomeres: lessons from budding yeast*. FEMS Microbiology Reviews, 2014. **38**(2): p. 144-171.
136. Mirabella, A. and M.R. Gartenberg, *Yeast telomeric sequences function as chromosomal anchorage points in vivo*. The EMBO Journal, 1997. **16**(3): p. 523-533.
137. Mitchell, L.A. and J.D. Boeke, *Circular permutation of a synthetic eukaryotic chromosome with the telomerator*. Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(48): p. 17003-17010.
138. Chakshusmathi, G., et al., *A La protein requirement for efficient pre‐tRNA folding*. The EMBO Journal, 2003. **22**(24): p. 6562-6572.
139. Lin, Q., et al., *RADOM, an Efficient In Vivo Method for Assembling Designed DNA Fragments up to 10 kb Long in Saccharomyces cerevisiae*. ACS Synthetic Biology, 2015. **4**(3): p. 213-220.
140. Gibson, D.G., et al., *One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome*.

Proceedings of the National Academy of Sciences, 2008. **105**(51): p. 20404-20409.
141. Kuijpers, N.G., et al., *A versatile, efficient strategy for assembly of multi-fragment expression vectors in Saccharomyces cerevisiae using 60 bp synthetic recombination sequences*. Microbial Cell Factories, 2013. **12**(1): p. 1-13.
142. Kawai, S., W. Hashimoto, and K. Murata, *Transformation of Saccharomyces cerevisiae and other fungi: Methods and possible underlying mechanism*. Bioengineered Bugs, 2010. **1**(6): p. 395-403.
143. Bentley, D.R., et al., *Accurate whole human genome sequencing using reversible terminator chemistry*. Nature, 2008. **456**(7218): p. 53-59.
144. Franke, S., et al., *One-step assay for the quantification of T4 DNA ligase*. Analytical and Bioanalytical Chemistry, 2015. **407**(4): p. 1267-1271.
145. Makuc, J., et al., *The putative monocarboxylate permeases of the yeast Saccharomyces cerevisiae do not transport monocarboxylic acids across the plasma membrane*. Yeast, 2001. **18**(12): p. 1131-1143.
146. Lambert, J.M., R.S. Bongers, and M. Kleerebezem, *Cre-lox-Based System for Multiple Gene Deletions and Selectable-Marker Removal in Lactobacillus plantarum*. Applied and Environmental Microbiology, 2007. **73**(4): p. 1126-1135.
147. Partow, S., et al., *Characterization of different promoters for designing a new expression vector in Saccharomyces cerevisiae*. Yeast, 2010. **27**(11): p. 955-64.
148. Mingeot-Leclercq, M.-P., Y. Glupczynski, and P.M. Tulkens, *Aminoglycosides: Activity and Resistance*. Antimicrobial Agents and Chemotherapy, 1999. **43**(4): p. 727-737.
149. Wach, A., et al., *New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae*. Yeast, 1994. **10**(13): p. 1793-1808.
150. Palmeros, B.z., et al., *A family of removable cassettes designed to obtain antibiotic-resistance-free genomic modifications of Escherichia coli and other bacteria*. Gene, 2000. **247**(1–2): p. 255-264.
151. Andrews, J.M., *Determination of minimum inhibitory concentrations*. Journal of Antimicrobial Chemotherapy, 2001. **48**(suppl 1): p. 5-16.
152. Slater, M.L., S.O. Sharrow, and J.J. Gart, *Cell cycle of Saccharomycescerevisiae in populations growing at different rates*. Proceedings of the National Academy of Sciences, 1977. **74**(9): p. 3850-3854.
153. Gietz, R.D. and R.H. Schiestl, *High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method*. Nat Protoc, 2007. **2**.
154. Ude, S., et al., *Biofilm formation and cellulose expression among diverse environmental Pseudomonas isolates*. Environmental Microbiology, 2006. **8**(11): p. 1997-2011.
155. Kim, S.R., et al., *Strain engineering of Saccharomyces cerevisiae for enhanced xylose metabolism*. Biotechnology Advances, 2013. **31**(6): p. 851-861.
156. Sanderson, K., *Lignocellulose: A chewy problem*. Nature, 2011. **474**(7352): p. S12-S14.
157. D'Amore, T. and G.G. Stewart, *Ethanol tolerance of yeast*. Enzyme and Microbial Technology, 1987. **9**(6): p. 322-330.

158. Sarthy, A.V., et al., *Expression of the Escherichia coli xylose isomerase gene in Saccharomyces cerevisiae*. Applied and Environmental Microbiology, 1987. **53**(9): p. 1996-2000.
159. Moes, C.J., I.S. Pretorius, and W.H. Zyl, *Cloning and expression of the Clostridium thermosulfurogenes D-xylose isomerase gene (xyLA) in Saccharomyces cerevisiae*. Biotechnology Letters. **18**(3): p. 269-274.
160. Amore, R., M. Wilhelm, and C.P. Hollenberg, *The fermentation of xylose —an analysis of the expression of Bacillus and Actinoplanes xylose isomerase genes in yeast*. Applied Microbiology and Biotechnology. **30**(4): p. 351-357.
161. Jeffries, T.W., *Engineering yeasts for xylose metabolism*. Current Opinion in Biotechnology, 2006. **17**(3): p. 320-326.
162. Sitepu, I.R., et al., *Oleaginous yeasts for biodiesel: Current and future trends in biology and production*. Biotechnology Advances, 2014. **32**(7): p. 1336-1360.
163. Van Vleet, J.H. and T.W. Jeffries, *Yeast metabolic engineering for hemicellulosic ethanol production*. Current Opinion in Biotechnology, 2009. **20**(3): p. 300-306.
164. Zhang, W. and A. Geng, *Improved ethanol production by a xylose-fermenting recombinant yeast strain constructed through a modified genome shuffling method*. Biotechnology for Biofuels, 2012. **5**(1): p. 1-11.
165. Kuyper, M., et al., *High-level functional expression of a fungal xylose isomerase: the key to efficient ethanolic fermentation of xylose by Saccharomyces cerevisiae?* FEMS Yeast Research, 2003. **4**(1): p. 69-78.
166. Gong, C.-S., et al., *Production of Ethanol from d-Xylose by Using d-Xylose Isomerase and Yeasts*. Applied and Environmental Microbiology, 1981. **41**(2): p. 430-436.
167. Young, E., S.-M. Lee, and H. Alper, *Optimizing pentose utilization in yeast: the need for novel tools and approaches*. Biotechnology for Biofuels, 2010. **3**.
168. Traff, K.L., L.J. Jonsson, and B. Hahn-Hagerdal, *Putative xylose and arabinose reductases in Saccharomyces cerevisiae*. Yeast, 2002. **19**(14): p. 1233-41.
169. Richard, P., M.H. Toivari, and M. Penttila, *Evidence that the gene YLR070c of Saccharomyces cerevisiae encodes a xylitol dehydrogenase*. FEBS Lett, 1999. **457**(1): p. 135-8.
170. Latimer, L.N., et al., *Employing a combinatorial expression approach to characterize xylose utilization in Saccharomyces cerevisiae*. Metabolic Engineering, 2014. **25**: p. 20-29.
171. Amore, R., et al., *Cloning and expression in Saccharomyces cerevisiae of the NAD(P)H-dependent xylose reductase-encoding gene (XYL1) from the xylose-assimilating yeast Pichia stipitis*. Gene, 1991. **109**(1): p. 89-97.
172. Altschul, S.F., et al., *Basic local alignment search tool*. J Mol Biol, 1990. **215**(3): p. 403-10.
173. *Benchling*. 2014 [cited 2016; Biology Software]. Available from: http://www.benchling.com.
174. Fang, F., et al., *A vector set for systematic metabolic engineering in Saccharomyces cerevisiae*. Yeast, 2011. **28**(2): p. 123-36.
175. Stearns, T., H. Ma, and D. Botstein, *[23] Manipulating yeast genome using plasmid vectors*, in *Methods in Enzymology*. 1990, Academic Press. p. 280-297.

176. Karhumaa, K., B. Hahn-Hägerdal, and M.-F. Gorwa-Grauslund, *Investigation of limiting metabolic steps in the utilization of xylose by recombinant Saccharomyces cerevisiae using metabolic engineering*. Yeast, 2005. **22**(5): p. 359-368.
177. Thomsson, E., et al., *Carbon Starvation Can Induce Energy Deprivation and Loss of Fermentative Capacity in Saccharomyces cerevisiae*. Applied and Environmental Microbiology, 2003. **69**(6): p. 3251-3257.
178. Albers, E., et al., *Effect of nutrient starvation on the cellular composition and metabolic capacity of Saccharomyces cerevisiae*. Appl Environ Microbiol, 2007. **73**(15): p. 4839-48.
179. Demeke, M.M., et al., *Rapid Evolution of Recombinant <italic>Saccharomyces cerevisiae</italic> for Xylose Fermentation through Formation of Extra-chromosomal Circular DNA*. PLoS Genet, 2015. **11**(3): p. e1005010.
180. Bengtsson, O., B. Hahn-Hagerdal, and M.F. Gorwa-Grauslund, *Xylose reductase from Pichia stipitis with altered coenzyme preference improves ethanolic xylose fermentation by recombinant Saccharomyces cerevisiae*. Biotechnol Biofuels, 2009. **2**.
181. Demeke, M.M., et al., *Development of a D-xylose fermenting and inhibitor tolerant industrial Saccharomyces cerevisiae strain with high performance in lignocellulose hydrolysates using metabolic and evolutionary engineering*. Biotechnology for Biofuels, 2013. **6**(1): p. 1-24.
182. Hamacher, T., et al., *Characterization of the xylose-transporting properties of yeast hexose transporters and their influence on xylose utilization*. Microbiology, 2002. **148**.
183. Xu, H., et al., *PHO13 deletion-induced transcriptional activation prevents sedoheptulose accumulation during xylose metabolism in engineered Saccharomyces cerevisiae*. Metabolic Engineering, 2016. **34**: p. 88-96.
184. Hector, R.E., et al., *Engineering industrial Saccharomyces cerevisiae strains for xylose fermentation and comparison for switchgrass conversion*. J Ind Microbiol Biotechnol, 2011. **38**.
185. Mitchell, L.A., et al., *qPCRTag Analysis - A High Throughput, Real Time PCR Assay for Sc2.0 Genotyping*. 2015(99): p. e52941.
186. Villanueva, H., *Is Gene Expression Affected Due to Terminator Swapping in Synthetic Yeast after SCRaMbLE is Induced?*, MRes Thesis in *Systems and Synthetic Biology*. 2015, Imperial College London.
187. Becker, D.M. and L. Guarente, *[12] High-efficiency transformation of yeast by electroporation*, in *Methods in Enzymology*. 1991, Academic Press. p. 182-187.
188. Jeppsson, M., et al., *The expression of a Pichia stipitis xylose reductase mutant with higher K(M) for NADPH increases ethanol production from xylose in recombinant Saccharomyces cerevisiae*. Biotechnol Bioeng, 2006. **93**.